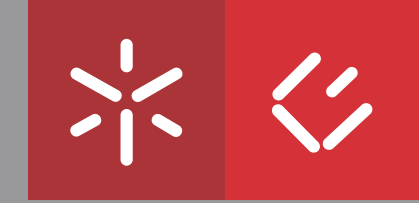


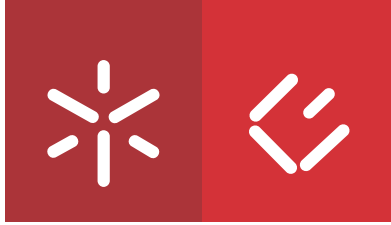


**THE ADOPTION OF BIG DATA BY NATIONAL
STATISTICS OFFICES – AN EXPLORATORY
RESEARCH**

Fabio dos Santos Cardoso

Universidade do Minho
Escola de Economia e Gestão





Universidade do Minho
Escola de Economia e Gestão

Fabio dos Santos Cardoso

THE ADOPTION OF BIG DATA BY NATIONAL STATISTICS OFFICES – AN EXPLORATORY RESEARCH

Doctoral Thesis
Ph.D. Thesis in Business Administration

Work developed under supervision of
Professor Ana Cristina Almeida Carvalho
University of Minho - Portugal
Professor João Eduardo Quintela Alves Sousa Varajão
University of Minho - Portugal
Professor Dimária Silva e Meirelles
Presbyterian University Mackenzie - Brazil

January 2024

DIREITOS DE AUTOR E CONDIÇÕES DE UTILIZAÇÃO DO TRABALHO POR TERCEIROS

Este é um trabalho académico que pode ser utilizado por terceiros desde que respeitadas as regras e boas práticas internacionalmente aceites, no que concerne aos direitos de autor e direitos conexos.

Assim, o presente trabalho pode ser utilizado nos termos previstos na licença abaixo indicada.

Caso o utilizador necessite de permissão para poder fazer um uso do trabalho em condições não previstas no licenciamento indicado, deverá contactar o autor, através do RepositóriUM da Universidade do Minho.

Licença concedida aos utilizadores deste trabalho



Atribuição-NãoComercial-SemDerivações

CC BY-NC-ND

<https://creativecommons.org/licenses/by-nc-nd/4.0/>

Acknowledgements

My first acknowledgement is for God to the permanent presence in my whole life, including in the darkest moments of this PhD journey.

In sequence, I offer my acknowledgment to my wife, Analice Araújo, for her permanent and incontestable support to me during those years. She had to abnegate from some projects, and besides those personal renouncing, she stayed though by my side. Analice this achievement is ours.

My mother Claudete Santos, my daughter Ana Clara Cardoso, my brother Luíz Cardoso and my sister Simone Rodrigues, thank you so much for your supports. During the journey, in each time one of you stayed by my side, bringing solutions or motivating me to go beyond.

Still in the family, I offer my acknowledgment to my aunts Zenaide and Zélia, my cousins Iris, Tânia, Sérgio and Júlio. When it was necessary, they were there and did not disappoint me.

Regarding the closer friends, my acknowledgment has been offered to Ricardo Oliveira, Bartira Rignel, and Robson Alves. After decades of friendship, I am happy and proud to keep calling all of you as my brothers and sister.

A great acknowledgment to my supervisors Ana Carvalho, João Varajão, and Dimária Meirelles. This marathon is in the final stage because I received more than supervision from you. Actually, you prepared an academic professional. Thank you.

My friends from University of Minho, you were permanent fellows during this stony trip. Bassem, Núbio, Talita, João, Mahamoud, Heshem, Loaloe, Yasmin, Mona, Nadine, Nada, Tilmar, Maria, Luís, Samuel, Sandrine, Alene, Enrikson, Dilson, Ziad, Tarek, Zakarias, Hussein, Maher, Samer, Kevian, thank you so much for your company.

Still about friendship, my acknowledgment to Rafael Lopes, Solange Oliveira, and Lara Baldin.

My grateful to the IBGE's employees who supported me with this research during several times. It is impossible to cite all of them. So, I say my thanks to all of you, specially to Marcelo Holanda, Georgia Assumpção, Ana Cristina von Calabach, Luíz Arbex, Bianca Walsh, Paulo Tostes, Hugo Baptista, Fabiane Lucena, Pamela Calazans, Aline Bezerra, Eduardo Derbli, Flávia Pinto, and Diana Baptista.

Throughout the development of this research, I have received financial support from the Instituto Brasileiro de Geografia e Estatística (IBGE), who sponsored partially this study.

STATEMENT OF INTEGRITY

I hereby declare having conducted my Thesis with integrity. I confirm that I have not used plagiarism or any form of falsification of results in the process of the Thesis elaboration. I further declare that I have fully acknowledged the Code of Ethical Conduct of the University of Minho.

Abstract

Official Statistics play a relevant role in modern societies. Their indices, rates, values, and time series are used to support decisions at the individual and collective levels. Behind this information are the data generators that provide Official Statistics to society. These organizations are called National Statistical Offices (NSOs).

As part of this public function, NSOs work through processes of collecting, storing, and analysing data from society to update Official Statistics. For more than two centuries, NSOs have performed the function of generating official statistics in a linear fashion, reporting results, and informing society. However, in the last two decades, the role of NSOs has been shaken by disruptive innovations related to digital transformation. One impact is the advent of Big Data. Innovative techniques and methodologies have enabled companies to manage large volumes and distinct types of data in a timely and permanent manner. This change has affected NSOs because the risk of obsolescence has impacted not only internal processes but also the response time required by society.

In this challenging context, NSOs are faced with a dilemma: either adopting Big Data technologies and methodologies to maintain their social relevance or face fatal obsolescence. The current research therefore intends to study how and why are NSOs adopting or avoiding Big Data technologies and methodologies.

To understand the adoption of Big Data (ABD), three NSOs in different realities were investigated. In South America, the selected case was the Brazilian NSO, IBGE. In Europe, Portugal's INE and the UK's ONS were chosen. These NSOs represent different geographical contexts, occupy a different the position in Global Statistical System and present different degrees of autonomy relative to the Central National Government.

The Technology-Organization-Environment (TOE) Framework was used as the analytical model. A rigorous and broad review of the literature allowed us to extract and consolidate a set forty-three factors previously used to study the adoption of Big Data. Data collection methods included a documentary collection and application of twenty semi-structured interviews with members of the three NSOs: ONS, INE, and IBGE.

Data analysis led us to identify ten additional new factors related to ABD in NSOs, including six technological, three organisational and one environmental. During the cases comparison, findings showed that the adoption of Big Data technologies and methodologies have similarities and common characteristics in the Official Statistics sector, despite each institution following a distinct trajectory. Based on this, we propose a matrix that depicts different positions in the evolution from avoidance to adoption of Big Data by NSOs. Moreover, ABD in NSOs advances on a cumulative trajectory where associations of specific factors define the path of adoption or avoidance. In conclusion, NSOs must not see ABD as a matter of choice, but as inevitable if they are to survive.

Key-words: adoption, Big Data, National Statistics Offices, Official Statistics, TOE framework.

Resumo

As Estatísticas Oficiais desempenham um papel importante nas sociedades modernas. Seus índices, taxas, valores e séries temporais são usados para apoiar decisões em nível individual e coletivo. Por trás dessas informações estão os geradores de dados que fornecem estatísticas oficiais à sociedade. Essas organizações são chamadas de Institutos Oficiais de Estatísticas (INEs).

Como parte dessa função pública, os INEs trabalham por meio de processos de coleta, armazenamento e análise de dados da sociedade para atualizar as estatísticas oficiais. Por mais de dois séculos, os INEs desempenharam a função de gerar estatísticas oficiais de forma linear, relatando os resultados e informando a sociedade. Entretanto, nas últimas duas décadas, o papel das INEs foi abalado por inovações disruptivas relacionadas à transformação digital. Um dos impactos é o advento do Big Data. Técnicas e metodologias inovadoras permitiram que as organizações gerenciassem grandes volumes e tipos distintos de dados de maneira oportuna e permanente. Essa mudança afetou os INEs porque o risco de obsolescência atingiu não apenas os processos internos, mas também o tempo de resposta exigido pela sociedade.

Nesse contexto desafiador, os INEs se deparam com um dilema: adotar tecnologias e metodologias de Big Data para manter sua relevância social ou enfrentar a obsolescência fatal. Esta investigação pretende então estudar como e porque é que os INEs adotam ou evitam tecnologias e metodologias de Big Data.

Para compreender a adoção de Big Data (ABD), foram estudados três INEs em diferentes realidades. Na América do Sul, o caso selecionado foi o IBGE, o INE brasileiro. Na Europa, foram escolhidos o INE de Portugal e o ONS do Reino Unido. Estes INEs representam diferentes contextos geográficos, diferentes posições no Sistema Estatístico Global e apresentam diferentes graus de autonomia em relação ao governo nacional central.

O modelo Tecnologia-Organização-Ambiente (TOE) foi definido como modelo analítico. Uma revisão rigorosa e ampla da literatura permitiu extrair e consolidar um conjunto de quarenta e três fatores usados anteriormente para estudar a ABD. Os métodos de recolha de dados incluíram uma coleta documental e a aplicação de vinte entrevistas semiestruturadas com membros dos três INEs: ONS, INE e IBGE.

A análise dos dados levou-nos a identificar dez novos fatores relacionados com a ABD nos INEs, incluindo seis tecnológicos, três organizacionais e um ambiental. Durante a comparação dos casos, os resultados mostraram que a adoção de tecnologias e metodologias de Big Data possuem semelhanças e características comuns no setor de Estatísticas Oficiais, apesar de cada instituição seguir uma trajetória distinta nesta adoção. Com base nisso, propomos uma matriz que representa diferentes as posições na evolução dos INEs na adopção de Big Data desde o evitamento à adopção. Além disso, a ABD nos INEs progride em uma trajetória cumulativa em que associações de fatores específicos definem o caminho da adoção ou da evitamento. Em conclusão, os INEs não podem encarar a ABD como uma questão de escolha, mas sim como uma inevitabilidade para garantir a sobrevivência.

Palavras-chaves: adoção, Big Data, Estatísticas Oficiais, Institutos Nacionais de Estatísticas, modelo TOE.

Table of Contents

Chapter 1 – Context and research roadmap	1
1.1. Official Statistics – The history and features of a business base on knowledge and technology. 4	
1.1.1. A summary of the Official Statistics History	5
1.1.2. Attributions and Uses of National Statistics Offices	7
1.1.3. The challenge of the adoption of Big Data	13
1.2. Thesis structure	14
Chapter 2 – Adoption models for Big Data - a literature review	16
2.1. Literature Review approach	17
2.2. Adoption of Big Data – definitions and frameworks	22
2.2.1. Conceptualizing Big Data	23
2.2.2. Theoretical frameworks for the adoption of Big Data	36
2.3. The Technology-Organization-Environment model (TOE)	42
Chapter 3 – TOE as a framework for the adoption of Big Data	47
3.1. Factors of the adoption of Big Data	49
3.1.1. Technology factors	56
3.1.2. Organisation factors	58
3.1.3. Environment factors	60
Chapter 4 – Methodology: how the research has been built	63
4.1. Addressing the Research problem and the research question	65
4.2. Why the Case studies method?	68
4.3. Designing the research	70
4.3.1. Selecting the cases	70
4.4. Data collection	73
4.4.1. Documental analysis	73
4.4.2. Semi-structured Interviews	75
4.5. Data Analysis Techniques	80
Chapter 5 – Understanding the National Statistics Offices	85
5.1. Features and Summarised trajectory	86
5.2. Outputs and deliveries	88
5.3. Innovation cycles	89
5.4. The Census operation	91
5.4.1. The Census in the United Kingdom	92
5.4.2. The Census in Portugal	95

5.4.3. The Census in Brazil	98
Chapter 6 - Data analysis	101
6.1. How National Statistics Offices adopt Big Data - TOE factor analysis	102
6.1.1 Technology factors	104
6.1.2 Organization factors	151
6.1.3. Environment factors	197
6.2. Summarizing Factor analysis	221
Chapter 7 - Findings and discussions	223
7.1. Dimensional precedence	224
7.2. Most relevant factors for the adoption of Big Data in National Statistics Offices	226
7.3. The idiosyncrasies of adoption of Big Data processes in NSOs	229
7.4. Interdimensional association of factors for the adoption of Big Data	230
7.5. Big Data Adoption or Big Data Avoidance? A cumulative process	236
7.6. Big Data for Official Statistics: beyond disruption	242
Chapter 8 - Contributions, Limitations and Propositions	244
8.1. Summarizing the journey	245
8.2. Contributions	246
8.3. Limitations	249
8.4. Future studies	250
8.5. Final Consideration: NSOs, an organisational species at risk of extinction?	250
9. References	252
Appendix 1 - Interview script to NSO members	272

List of Tables

Table 1 – Examples of National Statistics Offices (NSOs) Missions Statements.....	page 07
Table 2 – Fundamental Principles of Official Statistics.....	page 09
Table 3 – Results from the Strings Search.....	page 19
Table 4 – References for developing the Big Data adjudicative concept.....	page 25
Table 5 – Big Data: the current conceptualizations.....	page 26
Table 6 – Concept Centric Matrix.....	page 30
Table 7 – Selected Attributes Matrix.....	page 35
Table 8 – TOE factors used to study the Adoption Big Data.....	page 49
Table 9 – Factors per dimensions in the Adoption of Big Data studies.....	page 51
Table 10 – Polysemy in TOE factors nomenclature.....	page 52
Table 11 – Low convergence factor: the example of <i>Security and privacy regulations</i>	page 54
Table 12 – Technological factors description: the example of <i>Compatibility</i>	page 55
Table 13 – Technological factors.....	page 56
Table 14 – Organisational factors.....	page 58
Table 15 – Environmental factors.....	page 60
Table 16 – Main figures.....	page 72
Table 17 – Interviewees profiles.....	page 78
Table 18 – Text of the Privacy Commitment Protocol.....	page 79
Table 19 – Design of the code/factors and elements Matrix.....	page 83
Table 20 – Summarized Methodological Design.....	page 84
Table 21 – Technological factors: Summarized findings.....	page 106
Table 22 – Organisational factors: Summarized findings.....	page 153
Table 23 – Environmental factors: Summarized findings.....	page 199

List of Figures

Figure 1 – Searches in Google, from 2008-Jan to 2022-Dec, using the themes "Big Data", "Web Scraping", "IoT", "Artificial Intelligence", and "Cloud computing"	page 02
Figure 2 – Relevant terms mentioned in the NSOs strategic missions statements.....	page 08
Figure 3 – Life cycle of the Official Statistics.....	page 11
Figure 4 – Business architecture of the Official Statistics.....	page 12
Figure 5 – The systematic search process.....	page 21
Figure 6 – Timeline of the innovation adoption frameworks evolution.....	page 37
Figure 7 – The relevant frameworks present in this literature review.....	page 40
Figure 8 – TOE Framework: original graphic structure.....	page 43
Figure 9 – Aggregating and simplifying the TOE factors.....	page 55
Figure 10 – A revised TOE model in the adoption of Big Data context.....	page 62
Figure 11 – Case study research flow.....	page 65
Figure 12 – Flow of multiple-case study designs.....	page 69
Figure 13 – Coding process by Saldaña (1993)	page 82
Figure 14 – Official Statistics by topics – an exemplificative scheme	page 88
Figure 15 – Census Invitation letter for British citizens	page 94
Figure 16 – Submission proof of the Census questionnaire.....	page 95
Figure 17 – Census invitation letter for Portuguese citizens.....	page 96
Figure 18 – Receipt of participation in the Census 2021.....	page 97
Figure 19 – Total of Deaths by Coronavirus (COVID-19) in the World (2020 –2023)	page 98
Figure 20 – Dissemination posters for the Brazilian Census 2022.....	page 99
Figure 21 – Data collectors' work kit	page 100
Figure 22 – TOE factors for the Interviews analysis.....	page 103
Figure 23 – The adoption of Big Data in NSOs under the TOE Framework perspective...	page 222
Figure 24 – TOE framework for adoption of Big Data in National Statistics Offices: most relevant factors.....	page 227
Figure 25 – The factors chain for the Technology Competence.....	page 231

Figure 26 – The factors chain for the Technology Resources.....	page 232
Figure 27 – The factors chain for the Trialability.....	page 233
Figure 28 – The factors chain for the adoption of Big Data: types, design, technologies, and methodologies.....	page 235
Figure 29 – Avoidance vs. Adoption Matrix for Big Data in the National Statistics Offices.....	page 238
Figure 30 – How factors can block the ABD in NSOs.....	page 239

Acronyms

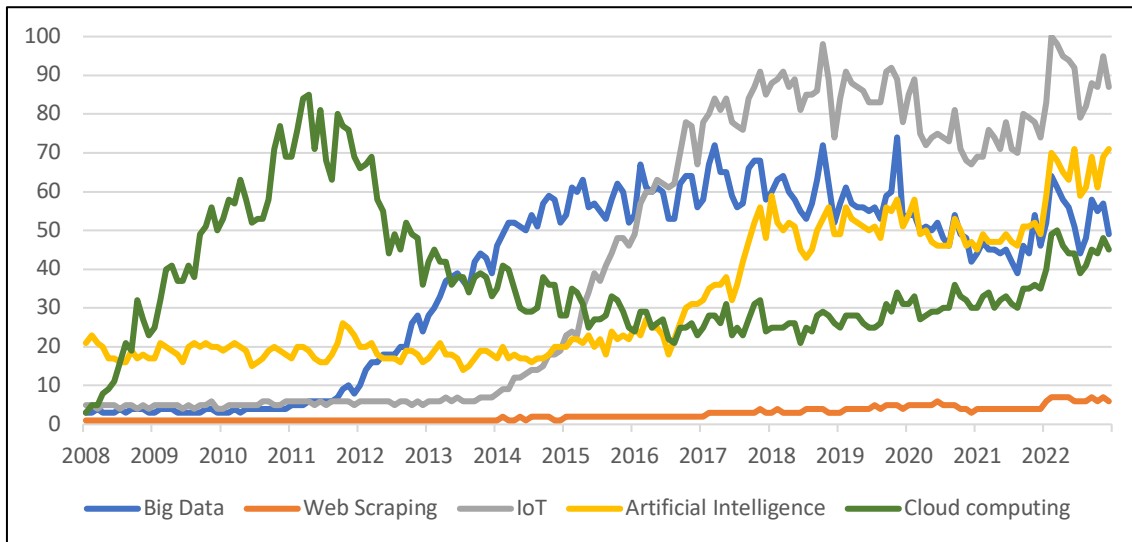
ABD	Adoption of Big Data
AI	Artificial Intelligence
BA	Business Analytics
BD	Big Data
BDS	Big Data Solutions
BDTM	Big Data Technologies and Methodologies
BR	Brasil
CBS	Netherlands National Statistics Office
CC	Cloud Computing
CEPAL	Economic Commission for Latin America
CS	Cloud Storage
GSS	Global Statistical System
ESCoE	Economic Statistics Centre of Excellence
IBGE	Instituto Brasileiro de Geografia e Estatísticas
INE	Instituto Nacional de Estatística
IoT	Internet of Things
ML	Machine Learning
NIESR	National Institute of Economic and Social Research
NSO	National Statistics Office
NSOs	National Statistics Offices
OECD	Organisation for Economic Co-operation and Development
ONS	Office for National Statistics
OS	Official Statistics
PT	Portugal
RSS	Royal Statistics Society
UK	United Kingdom
UN	United Nations
UNSD	United Nations Statistics Division
WS	Web Scraping

Chapter 1 – Context and research roadmap

The last decade has seen an unprecedented rise in data as a source of value creation for public and private organisations. In the United Kingdom, the Data Manifesto published by the Royal Statistical Society (RSS) in 2014 highlighted the importance of data to global society. The document stated: "*What steam was to the 19th Century, and oil had been to the 20th, data is to the 21st .*" (Royal Statistical Society, 2014, p.1).

In this research, data is considered part of the study object, including the term "Big Data" and its applications for National Statistics Offices. There are several definitions of Big Data in the literature, including one proposed by Chen *et al.* (2014, p.2): "*Big Data refers to the datasets that could not be perceived, acquired, managed, and processed by traditional Information Technology and software/hardware tools within a tolerable time.*" Côté-Real *et al.* (2017, p. 380) describe the term as "*a new generation of technologies and architectures designed to extract large-scale economic value from a wide variety of data involving high-speed capture, discovery, and/or analysis.*" With different definitions, the standard Big Data conceptualisation remains challenging for academia.

In the last fifteen years, from 2008 to 2022, the term "Big Data" had the peak of searches on Google webpage in September 2019, as presented in Figure 1. After the COVID-19 pandemic, Big Data technologies, such as Cloud computing, Artificial Intelligence, and the Internet of Things (IoT), received more searches from users worldwide. Only Web Scraping kept a low level of search. The main message Figure 1 brings is that Big Data has remained a hype theme for half a decade. Despite the reduction of searches about "Big Data", there is an increase in searches for its technologies. In this way, the overview for the themes of "Cloud computing", "Artificial Intelligence", and "IoT" shall be considered as additional searches about Big Data, likewise an aggregate result. So, the adoption of Big Data (ABD) is not a mere IT fashion topic but a reality with impacts on organisations beyond their technological department.



Source: Google trends. Retrieved 2023, January 30th.

Figure 1 – Searches in Google, from 2008-Jan to 2022-Dec, using the themes "Big Data", "Web Scraping", "IoT", "Artificial Intelligence", and "Cloud computing".

Beyond the definitions, Big Data is recognised by Schwab (2016) as a disruptive technology with revolutionary capacity when applied in organisational contexts, especially in decision-making processes. This author lists the benefits obtained by the adoption of Big Data (ABD): (i) "*faster and better decision-making process*", (ii) "*decision-making process in real-time*", (iii) "*open data to support and foster innovation*", (iv) "*more efficient public services*", (v) "*new job opportunities creation*", (vi) "*new types of public and private services.*" All of them, by his predictions, might result in more quality services deployed to citizens and customers, profound change in legal and commercial norms, and improvement of global integration in areas such as public services, as the Official Statistics. The impacts can positively improve citizens, companies, and States.

Despite those possible opportunities, Big Data is also considered a disruptive technology. This disruptive feature may bring risks when adopted by governments and companies. Schwab (2016) draws attention to the lack of governmental accountability, loss of individual privacy, and individual profile featuring, "profiling", to use for civil manipulation by private groups, such as in the novel "*1984*" written by Orwell (2013). As a matter of fact, Harari (2017, 2019) also reinforces the risks associated with the use of Big Data, drawing attention to the civil control made by private groups. Exemplifying, the author cites two examples: the Brexit referendum and the United States presidential elections, both in 2016. Harari emphasised the risk that a "digital dictatorship" will guide humanity. In the same way, O'Neil (2016) goes beyond and calls the erroneous use of Big Data a tool to control citizens' life.

This dystopia is approached in Amer and Noujaim's (2019) documentary "The Great Hack." The movie showed how Big Data had been used by Cambridge Analytica company, associated with Facebook, to influence electors in Argentina, Trinidad and Tobago, the United Kingdom, the USA, and Brazil. In the same film, the statement done by the Royal Statistical Society about data is confirmed, and the sentence is clear: today, data has more value than oil.

Notwithstanding the internal risks embedded in Big Data technologies and methodologies, this technological-organisational asset calls the attention of companies, States, and academia. One motive for this interest is the association between Big Data and decision-making processes. Sheng et al.'s (2017) research shows the link between those two themes. In their article, the authors showed the link between decision-making and Big Data: "(...) *also point out that data-driven decision-making is a promising trend and the primary factor attributed to successful Big Data utilisation is the organisational alignment in every aspect in the companies. (...) From the information management perspective, Big Data research primarily concerns data acquisition and process effectiveness. The availability and feasibility of information are critical to organisational success in terms of strategic decision-making.*" (p. 102).

Some social actors are responsible for the relevance of data. One of them is a group of public data generators named National Statistical Offices (NSOs). They are in charge of delivering Official Statistics to the respective societies. Regarding the relevance of the NSOs, this research focuses on their adoption of Big Data technologies and methodologies. As official data generators, the NSOs collect, store, and analyse data to deliver index prices, rates of inflation and unemployment, among others (Radermacher, 2020).

In fact, NSOs are critically relevant in this period of History because they are public institutes, accountable, subject to laws, and transparent in their data production (United Nations, 2014). Different from private companies, the Official Statistics Offices, as a business sector, has to work to support transparency, accuracy, and social trust (United Nations, 2014).

In this context, this research presents a strong motivation that can be summarised as the demand for innovation by NSOs in their processes to produce Official Statistics, considering the adoption of Big Data (ABD henceforth) solutions (technologies and methodologies). So, this research seeks to identify and analyse factors that inhibit or boost ABD. If the NSOs suffer obsolescence, this circumstance can represent risks to societies and decision-makers. The reason is that using outdated data can guide decision-makers and governments to fail.

Furthermore, studies about the ABD in NSOs are scarce, despite the relevance of those institutes. This gap provides an opportunity to address this research, contributing to the body of knowledge on ABD, management, and Official Statistics as a business sector. Indeed, this business sector presents a plethora of specifications, and these details and elements are explained in the next section.

1.1. Official Statistics – The history and features of a business base on knowledge and technology.

It is necessary to start by defining and differentiating the terms related to Official Statistics. Radermacher (2020) approaches Official Statistics as a business sector in his research. This author brings definitions for three main concepts to the Official Statistics field: “(i) statistics”, “(ii) statistical results”, and “(iii) statistical institutions.” According to this author, statistics is *“the science of learning from data.” 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.

The production and application of statistics can be routinely used by a commercial company, a non-governmental organisation and a soccer team. When the statistics are generated and broadcast by a governmental authority, it becomes official information. This official information refers to Official Statistics. Official Statistics means *“any set of statistics produced by an organisation named under secondary legislation and described by that organisation as an official statistic or part of a set of official statistics”* (UK Statistics Authority, 2022).

Statistical statistics offices are the producers of Official Statistics. 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.”* Related to 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.

Statistical Offices are, therefore, public organisations working at three different geographic levels: local, national and international. In this research, we consider the organisations which work at the

national and international levels. This distinction is relevant because of the comparative approach related to the case study method adopted in this study.

Official Statistics and their operators are not a 21st Century phenomenon. All those statistics offices, their outputs and outcomes, the methodologies and techniques result from a long historical trajectory. This trajectory of knowledge adjudication and constant innovation is described in the following subsection.

1.1.1. A summary of the Official Statistics History

Kotz (2005) and Stigler (1986) highlight the evolution of statistics as a science and Official Statistics as a subfield of this science, both part of the European Modern History. However, Kotz and Stigler diverge about the timeline. Kotz (2005) has adopted a longer trajectory since the 16th Century.

The use of counts and numbers to measure military capabilities and taxation dates remotes from Antiquity (Kotz 2005), with registers in the Mediterranean zone. In the 16th Century, seminal studies about Official Statistics started to be published in three European countries: Germany, the United Kingdom, and France.

As remarked by Kotz (2005), during the 17th Century, in Germany, the publication from Johann Peter Süssmilch (1707-1767) represented the initial event for the Official Statistics in that European Zone. As a Protestant Pastor, Süssmilch recorded religious events, such as marriages, baptisms, and funerals. Indeed, these 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.

The United Kingdom is another country where statistics and Official Statistics methods arose early. Kotz (2005) remarks that the adoption of the word "statistics" is an appropriation of the similar German word "statistik." John Gaunt (1620-1674) was one of the first involved in social and demographic measurements in the United Kingdom (UK). As a London citizen, he published a book on the city's mortality metrics based on official registers.

Another relevant British author for statistics and Official Statistics was John Sinclair (1754-1835). Sinclair's twenty-one volumes about statistical accounts in Scotland adopted an innovative method of data collection: the questionnaire. The compilation from questionnaires applied in 938 Scottish Parishes and responded to by the respective church ministers generated the data explored by

Sinclair (Kotz (2005)). Thus, Sinclair's effort represents a social statistical survey developed over two hundred years ago!

Still in the UK, during the first half of the 19th Century, the Royal Statistical Society was founded in London (Royal Statistical Society, 2022). With 188 years of history, this institution is the active think tank for Official Statistics in the British islands.

The third country is France, as remarked by Kotz (2005). The oldest register of Official Statistics in France is a public financial publication from 1581. Surveys followed this historical register to collect administrative data in the second half of the 16th Century. In fact, the French focus on Official Statistics crossed the centuries and kept steady progress until the French Revolution. Chaptal, Neufchateau, Laplace and Duvillare are some names of intellectuals who contributed to developing French Official Statistics in the 18th and 19th Centuries.

In the same period, specifically in 1843, a Belgian scientist, Adolphe Quetelet, presented his research to the French academia. His studies concentrated on demographic themes. The target of his studies was addressing 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).

Attempting to answer those questions, Quetelet developed a theoretical construct still used nowadays in Official Statistics: the average man. The proposal was to achieve the possibility of understanding a great set of data, such as a nation's society. So, Quetelet understood citizens as units of this society. In this way, he tried to reduce the extremes in matters as behaviours, physical features and natural actions. The reduction proposed by him adopted a statistic tool: normalisation of the society features in a average, the average man. This concept was created to concentrate in a symbolic-model-citizen, predominant elements from each society. He pointed out that his intention was not to create a flat view of mankind. France should have its own average man, such as England, Germany and Scotland should have their own (Quetelet, 2013).

Until the 21st Century, Quetelet's (2013) legacy still holds. In 2010, the British Office for National Statistics did a research to describe the average man and the average woman in England (BBC, 2019).

The relevance of Official Statistics is also apparent in other countries, including Portugal and Brazil, which feature in the current study along with the UK. In 1864, Portugal ran its first demographic Census, adopting modern statistical methodologies (Instituto Nacional de Estatística, 2022). Before

this year, other governmental accountings with military proposals took place in the 13th, 15th, 16th, 17th, and 18th centuries. Since the end of the Middle Ages, national accountings were present in the routines of the Portuguese government (Instituto Nacional de Estatística, 2021a).

In the case of Brazil, the reference year is 1870, when the “Diretoria Geral de Estatísticas” (General Statistics Directory) was founded. This organisation was the first entity to centralise the Official Statistics in this South American country. Two years later, in 1872, that organisation planned, organised and ran the first Brazilian demographic Census (Diretoria Geral de Estatística, 1874).

Considering the extensive history of Official Statistics, a reflection emerges in this context: in the current days, the Official Statistics shall be up to date with the current data technology applications. However, this is a huge challenge to be addressed by the Official Statistics operators in the execution of their attributions. Those operators are named in this study as National Statistics Offices (NSOs), considering the geographic approach: national and international levels.

1.1.2. Attributions and Uses of National Statistics Offices

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 citizens the capacity to scrutinise the performance of governments.

In this way, NSOs play a central role in the business of Official Statistics. Radermacher (2020) highlights those offices as keepers of the uniform production of Official Statistics, following effective and efficient patterns. As shown in Table 1, this perspective can be observed in the mission statements of several National Statistics Offices (NSOs), that claim a clear commitment with society and its players.

The selection of those three examples is defined by the choice of the three case studies to be addressed in this research: IBGE, from Brazil; INE, the Portuguese NSO; and ONS, from United Kingdom. Chapter 5 describes each case in detail.

Table 1 – Examples of National Statistics Offices' Mission Statements

<i>“High quality data and analysis to inform the UK, improve lives and build the future.”</i> ONS - Office for National Statistics
<i>“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,</i>

the analysis, the innovation and the dissemination of the national statistical activity and ensuring integrated data storage.”

INE - Instituto Nacional de Estatística

“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

It is possible to verify the relevance of some concepts presented in the mission statements using a tag cloud chart (Figure 2). The relevance of the concepts: “information”, “quality” and “analysis” indicates a common basis for these NSOs to define their missions. Those concepts can be translated to values and targets to be achieved by those organisations. 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.



Legend: Tag cloud for the NSOs strategic mission statements. The prevalence of three words (information, quality, and analysis) shows the commitment of those organisations with the duty to deliver Official Statistics to the society.

Figure 2 – Relevant terms mentioned in the NSOs strategic mission statements

In fact, this “constitutional bill” works as a list of limits and protocols that NSOs must observe. Everything out of this frame is considered misconduct by NSOs’ peers. The United Nations defines this document as the “Fundamental Principles of Official Statistics” (United Nations, 2014). Ten principles guide the operations of Official Statistics by NSOs. Table 2 shows the principles as they are presented by United Nations Statistics Division (United Nations, 2014).

Table 2 - Fundamental Principles of Official Statistics

Principle 1 <i>Official statistics provide an indispensable element in the information system of a democratic society, serving the Government, the economy, and the public with data about the economic, demographic, social and environmental situation. To this end, official statistics that meet the test of practical utility are to be compiled and made available on an impartial basis by official statistical agencies to honour citizens' entitlement to public information.</i>
Principle 2 <i>To retain trust in official statistics, the statistical agencies need to decide according to strictly professional considerations, including scientific principles and professional ethics, on the methods and procedures for the collection, processing, storage, and presentation of statistical data.</i>
Principle 3 <i>To facilitate a correct interpretation of the data, the statistical agencies are to present information according to scientific standards on the sources, methods and procedures of the statistics.</i>
Principle 4 <i>The statistical agencies are entitled to comment on erroneous interpretation and misuse of statistics.</i>
Principle 5 <i>Data for statistical purposes may be drawn from all types of sources, be they statistical surveys or administrative records. Statistical agencies are to choose the source with regard to quality, timeliness, costs and the burden on respondents.</i>
Principle 6 <i>Individual data collected by statistical agencies for statistical compilation, whether they refer to natural or legal persons, are to be strictly confidential and used exclusively for statistical purposes.</i>
Principle 7 <i>The laws, regulations and measures under which the statistical systems operate are to be made public.</i>
Principle 8 <i>Coordination among statistical agencies within countries is essential to achieve consistency and efficiency in the statistical system.</i>
Principle 9 <i>The use by statistical agencies in each country of international concepts, classifications and methods promotes the consistency and efficiency of statistical systems at all official levels.</i>
Principle 10 <i>Bilateral and multilateral cooperation in statistics contributes to the improvement of systems of official statistics in all countries.</i>

Source: United Nations. Retrieved from "<https://unstats.un.org/unsd/dnss/gp/FP-Rev2013-E.pdf>". Accessed in 01/06/2022.

These ten principles cover ten specific themes about the execution of Official Statistics. The first principle announces the relevance and the use of Official Statistics. Next, the second principle approaches the preservation of trust from Official Statistics based on scientific methods and professional ethics. Data interpretation and reinterpretation are the themes of principles three and four. Principle five covers the data sources used, including surveys and administrative records.

Individual privacy, regulations and laws are principles six and seven themes. Regarding cooperation at the international level, principles eight, nine and ten not only propose but also encourage partnerships among NSOs.

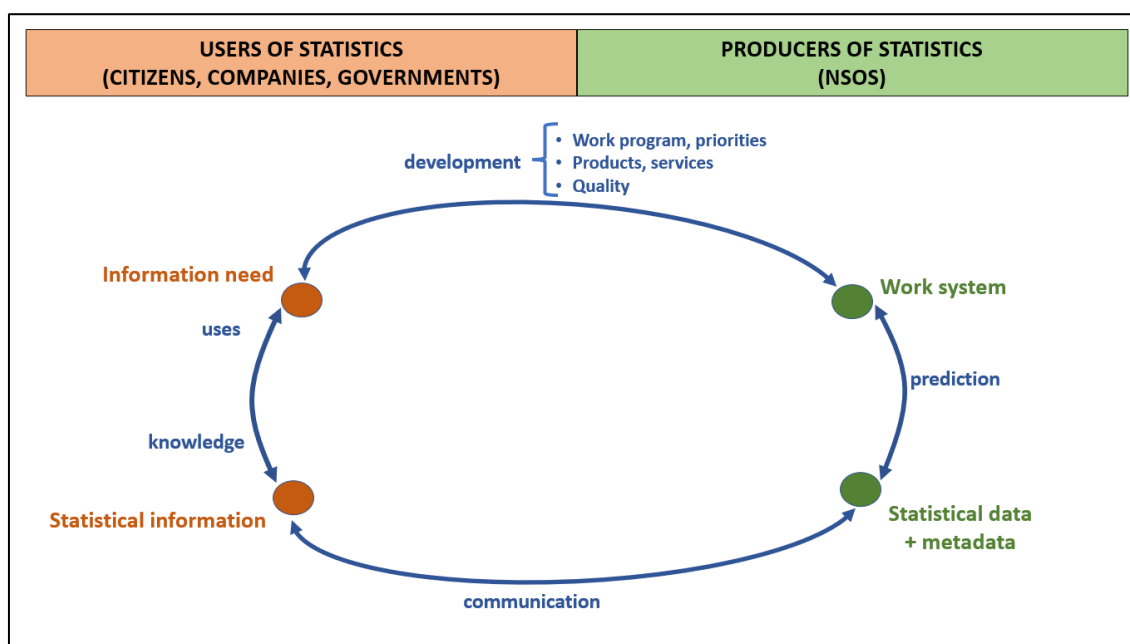
It is necessary to highlight that those principles result from a consensus between United Nations Statistics Division (UNSD) and the NSOs. This set of organisations integrates the Global Statistical Division. The exchange of information and methodologies among NSOs, and from NSOs to UNSD characterises the Global Statistical System, a global organisational network, as remarked by Cardoso et al. (2021, p.3):

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

In order to guide and advance the use of Official Statistics, those principles work as a starting point for NSOs to develop their research and targets. Indeed, the principles are not a self-centred set of rules. Radermacher (2020) remarks that Official Statistics are also being used to inform about the United Nations’ Millennium Development Goals (MDGs). In this case, surveys and administrative data support measuring the countries’ progress towards achieving the goals.

So, it is possible to define Official Statistics as a type of business (Radermacher, 2020). As a business, it has its production cycle; in this case, a data production cycle. There is also a business architecture to support the Official Statistics Business (Radermacher, 2018). Both are structured and aligned with the Fundamental Principles of Official Statistics (Radermacher, 2020).

Figure 3 presents the data generation life cycle of Official Statistics. It is possible to observe statistics users as the origin of the demands for Official Statistics outcomes. Development, production, and communication are stages in the Official Statistics life cycle.

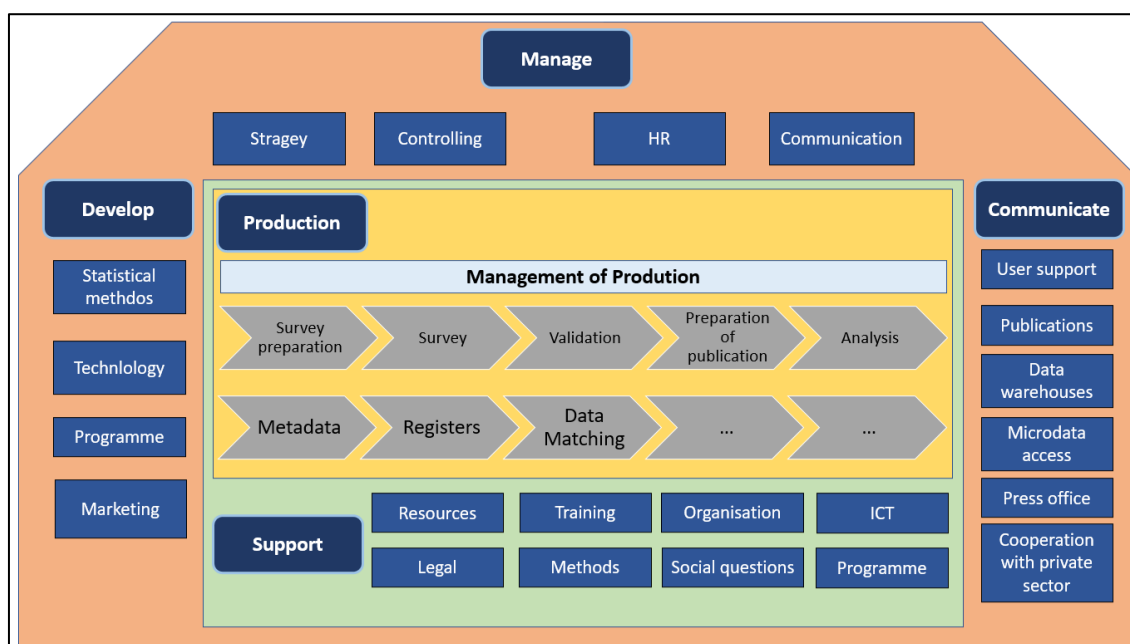


Legend: There are two groups involved during the life cycle of the Official Statistics: producers and users. The first group develops the work system, manage the statistical data and its metadata, generate predictions, and communicate the outcomes by the statistical information system. Users share information needs and their data.
Source: Based on Radermacher (2020).

Figure 3 – Life cycle of the Official Statistics

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

So, the Official Statistical life cycle, together with the Ten Fundamental Principles, allows the development of the business architecture followed by the National Statistics Offices (Radermacher, 2018). In Figure 4, this extensive scheme shows the value chain of the Official Statistics in the rectangle named “Production”. Surveys and metadata are at the core of this value chain. The methodology and the technology are highlighted in the domain “Develop”. Training, legal, organisation, and ICT in all perspectives (capabilities, complexities, readiness, infrastructure) are among the elements that constitute the support line for NSOs to run official statistics businesses. Communication is the final stage of this architecture. It is understood to cover a wide range of dissemination channels, from technological capacity to offer user access (data warehouses and microdata access), to publishing (publications and press documents).



Legend: The business architecture of the National Statistics Offices according to Radermacher (2018). Five core areas (Manage, Production, Support, Develop, and Communicate) manage the processes to regularly deliver Official Statistics to the society.

Source: Base on Radermacher, 2018

Figure 4 – Business Architecture of the Official Statistics

All these attributions and uses related to Official Statistics provide a descriptive contextualisation that introduces the reader to the business where ABD is studied in this PhD research. To address, more specifically, ABD by NSOs, it is mandatory to consider the core elements of Official Statistics and National Statistics Offices.

Not only surveys and metadata integrate the data collection executed by NSOs. With a periodicity of a decade (Instituto Nacional de Estatística, 2022), several NSOs deploy a vast operation, the Demographic Census. This national accounting of the population takes place simultaneously with other surveys. Census data are the base of future public policies. There is another Census, the agricultural Census, focused on countryside activities (IBGE, 2020a), deployed in countries where agriculture and cattle industries are significant. In this context, Brazil (IBGE, 2020a) and Portugal (Instituto Nacional de Estatística, 2021b) are examples.

It is relevant to cite that the NSOs' portfolios are similar. Despite the singularity of one or two specific national surveys, they conform to a pattern of research types followed internationally. The resulting data is sent to the UNSD by NSOs. The common patterns and the quality of the surveys make it possible for the UNSD to join and process data from different organisations. The UNSD presents on its webpage (United Nations 2021) a set of data aggregated and compiled by them.

1.1.3. The challenge of the adoption of Big Data

In the midst of the extensive routine of surveys and Census, NSOs and the GSS are facing a contemporary dilemma: innovation or obsolescence. Regarding the Official Statistics that are used by society, in the last two decades, citizens, companies, and other stakeholders demanded faster and faster data from the GSS and NSOs. Let's suppose that NSOs do not deploy the data in real-time. In that case, those stakeholders will seek alternative sources of information from competitor actors, such as Tech Giants like Google and Facebook. Cardoso et al. (2021, p.3) highlighted this scenario of threats to the role of NSOs:

“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 (...). Nowadays, those organisations deal with the challenge of losing relevance to societies in the face of new data providers (...). 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 (...) and digital advertisement (...).”

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

In fact, technological obsolescence represents a challenge to NSOs. Failing to innovate puts the role of NSOs under threat. Allin (2021) considers that NSOs must redefine their role in society in this Real-time age when data users do not wait more than seconds to receive information. So, the threat of obsolescence comes with the opportunity to renew society's confidence and attention. In this way, Allin (2021, p.66) remarks:

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

This redefinition of the NSOs' organisational role includes the innovation cycle. Many NSOs plan and work with a ten-year innovation plan timeline based on the Census operation. However,

currently, the demand for innovation occurs almost daily. Consequently, the innovation cycle of NSOs must be reduced, maybe to yearly (or even shorter) periods. It is not simple to promote this change. Considering the history of Official Statistics, this is a scientific field where innovation has taken place gradually, not in a disruptive way.

Additionally, the incapacity to work based on short-time projects is another relevant threat to NSOs. Because NSOs work in long-term projects, they may be less well adapted to short-term projects. Bakici et al. (2022) concluded that ABD demands project readiness from the adopter. This capability is also for operational and strategic projects. Following the Bakici approach, Bilal et al. (2016) indicate that project readiness is critical for the success of ABD in the construction industry.

In summary, those two risks: (i) innovation cycle and (ii) short-term project readiness, are a sample of the risks related to NSOs and ABD. In fact, those two risks are very relevant but isolated.

Concluding, the following years are critical to the survival of the Global Statistical System and National Statistics Offices. If those organisations lose public relevance and are not considered necessary by their stakeholders, the main asset of this business, trust, will be missed. So, in this challenging context, the current research advances to answer a central question: How and why do National Statistical Offices adopt or avoid Big Data technologies and methodologies?

1.2. Thesis structure

Official Statistics constitute a crucial set of information that supports decision-making in modern societies. The significance of this relevance can be attributed to a historical process, which has been summarized in this Chapter 1. Building upon this historical trajectory, the study progresses to explore the contemporary characteristics of Official Statistics as a distinct field. It delves into the information life cycle and the business architecture, with a particular focus on their implications in the context of ABD.

Subsequently, in Chapters 2 and 3, the literature review delves into the concept of Big Data and its application in Official Statistics. Chapter two also presents various models of technology adoption, following a chronological perspective. Taking into consideration the multidimensional nature of ABD, these models are thoroughly discussed and evaluated to identify the most suitable approach. Ultimately, Chapter 3 brings the chosen model, TOE. It is introduced, explained, and detailed, emphasizing the essential factors necessary to comprehend ABD.

The approach employed in this research is then elucidated in Chapter 4. Multiple case studies form the core of this investigation, along with an explanation of the criteria used for their selection. Three National Statistics Offices from distinct countries—IBGE (Brazil), INE (Portugal), and ONS (United Kingdom)—constitute the chosen cases. Furthermore, the data analysis methodology is described, which combines literature review with documentary research to develop an interview script. Over the course of one semester, a total of twenty interviews were conducted across the three National Statistics Offices: IBGE, ONS, and INE.

Following the methodology explanation, in Chapter 5, the cases are described and comprehensively analysed, particularly with regard to the major aspects of ABD within the National Statistics Offices. The innovation cycles are discussed in relation to the technological advancements experienced by IBGE, INE, and ONS. Additionally, a relevant operation conducted by these organizations—the Census—is compared among the three cases. This comparative analysis contributes to the overall research, providing insights into how three different Censuses have evolved in distinct countries and underscoring the significance of Big Data technologies in supporting such operations.

The Chapter 6 brings the analysis of the interviews, considering that the statements and speeches are interpreted through the lens of the selected analytical model, TOE. A comprehensive set of fifty-two factors was examined in the testimonials from members of IBGE, INE, and ONS. This rich material enables a robust interpretation of the reality of ABD in Official Statistics production and underscores the importance of this adoption for the survival of these organizations.

Finally, Chapter 7 brings the analysis of the outcomes. It is structured in terms of findings and contributions, which includes categorizing the three organizations in relation to their adoption of Big Data. From the contributions drawn and the conclusions derived (Chapter 8), the thesis concludes by proposing a research agenda and sending a clear message to National Statistics Offices: ABD is not a matter of choice but rather a necessity for their survival.

Chapter 2 – Adoption models for Big Data - a literature review

Studies about the adoption of Big Data (ABD) should combine technological aspects with organizational factors, and environmental variables (Baker, 2012). In the current decade, Big Data arises as a mandatory organizational and technological asset to support companies dealing with contemporary challenges, such as the COVID-19 pandemic (Soto-Acosta, 2020). In this scenario, organizations are obligated to deal with digital transformation, which is “*an evolutionary process that leverages digital capabilities and technologies to enable business models, operational processes and customer experiences to create value*” (Morakanyane et al. 2017, p. 437). Vial (2019) defines digital transformation as an improvement process that changes organizational structures based on a mix of information, computing, communication, and connectivity technologies. Under this perspective of process change and digital capabilities empowerment, this research studies ABD as a critical asset to overcome the obsolescence in public, private, and non-governmental institutions.

There is substantial noise around the definition of Big Data and the adoption of Big Data (ABD). In this thesis, Big Data is considered an organizational and technological asset, which merges different types of technologies and methodologies. Cloud computing, web scraping, the Internet of Things (IoT), and Artificial Intelligence are examples of the technological companions of Big Data, as remarked by The United Nations Statistical Commission (2020). Zhang (2012, p. 1) highlighted that Big Data “*needs more computing power and storage provided by cloud computing platforms*”. In turn, Pramana et al. (2021) indicate web scraping as a source of Big Data applied to the Official Statistics surveys. Sun and Huo (2021) reinforce that IoT, Cloud, and Artificial Intelligence are Big Data technologies.

Although earlier studies focused mostly on technological issues, there have been calls for research into the organizational dynamics that affect ABD (Picoto *et al.*, 2021; Olszak and Zurada, 2020; Félix *et al.*, 2018; Côte-Real *et al.*, 2017; Phillips-Wren *et al.*, 2015). The ABD phenomenon has been examined as a process, and there are several frameworks used to explain an individual perspective or organizational scope, such as the examples presented in the next pages.

Hence, this chapter presents the literature review that outlines the state of the art about ABD. There is a detailed description about the systematic procedure to compose the references dataset. In

sequence, ABD is explained and discussed under the lens of the main frameworks and definitions used. The third section focuses on the selected framework, TOE, to bind this research.

2.1. Literature Review approach

Knowledge review is a process to establish the current bases of scientific knowledge and spread its boundaries (Cronin and George, 2020) across different communities of practice and research. As a process, knowledge review has many positive outputs, such as overcoming the isolation of research communities nestled in academic silos. Knowledge-synthesis vehicles (KSV) are the drivers to the delivery of results from the review processes. One of those KSV is the literature review, that can be operated in two different orientations: adjudicative or redirecting. A literature review with an adjudicative orientation promotes “*the integration from different findings, compiles and characterizes existing knowledge from across communities of practice*” (Cronin and George 2020, p.2). When a literature review redirects the current knowledge (redirecting orientation), the output is the proposal of a new narrative and approach to a specific scientific issue (Cronin and George, 2020).

In this research, the adjudicative approach was adopted. The objective was to seek well-established practices to the ABD phenomenon, as well as identify systematic standards of this adoption process. According to Cronin and George (2020), a knowledge-synthesis vehicle, under the adjudicative orientation, enables the comparison between qualitative constructs. The operational drive for this comparison is the systematic literature review.

In fact, the literature review (LR) improves academic research development by aggregating different contributions (Ferfolja and Burnett, 2002; Webster and Watson, 2002; Aguinis and Ramani, 2020). It builds support to advance the current knowledge, highlights the gaps in the current state of the art in a specific area, justifies a selected research theme, and identifies an expected contribution to spread the research. Moreover, Kunisch et al. (2018) explain that the LR has two main functions: to announce the current knowledge set and guide future research. Additionally, Briner *et al.* (2009, p. 24) consider the LR as a process, enabling the study to reach a “*wider insight into how to make decisions using evidence, critical thinking, and a mindful approach to practice.*” This insight supports the motivation to develop the research question in a novel, critical, and unparalleled way Köhler (2016, p.402). Finally, about attributes, Hiebl (2021) remarks a LR shall be: (i) structured, (ii) comprehensive, and (iii) transparent.

There are different ways to structure the LR. Ferfolja and Burnett (2002, p. 3) list four types of LR organization: (i) chronological – research sources organized by historical context; (ii) “classic” studies – the focus of the LR is on the prime publications and authors from a specific topic; (iii) topical or thematic – the LR is structured on conceptual categories, also named sections and subsections; and (iv) inverted pyramid – from a large thematic perspective to a well-focused issue, the LR advances towards the core issue of the research question. Subjects organized by thematic sections can be used as the foundation of an LR’s framework (Webster and Watson, 2002). Additionally, an LR can be structured in concept-centric and author-centric logics (Webster and Watson 2002). The first is the research that adopts contributions based on a central concept, such as “Adoption of Big Data”. An author-centric LR is organized in function of the specific publications of an author or a group of authors.

Our LR adopted a concept-centric framework (Webster and Watson, 2002), organized in topical or thematic sections (Ferfolja and Burnett, 2002). The concept-centric framework enabled us to deploy a systematic LR on a specific subject, in this case, the adoption of Big Data (ABD). Concerning topical or thematic structures, Ferfolja and Burnett (1998) argue that they organize the discussion. As parts of those structures, sections and subsections develop a “research space” (Ferfolja and Burnett, 2002) to further results, including those from empirical studies, such as ABD technologies and methodologies by the National Statistics Offices. Our LR, therefore, operates with the use of key concepts. Webster and Watson (2002) consider that a robust LR must be based on a set of key concepts. Additionally, the authors reinforce the argument that concepts confer organization and boundaries to the LR.

About the selected concepts for this research, they are presented in the research question: “*How and why do National Statistical Offices adopt or avoid Big Data technologies and methodologies?*” The primary concepts are National Statistics Offices, Adoption, and Big Data. Derived from these primary concepts are the terms: acceptance, and the names of each National Statistics Office (IBGE, INE, ONS) under study. Additionally, the database search used those terms, organized in sets of keywords or strings.

The principal sources used are two databases: *Scopus* (SCP) and *Web of Science* (WOS). Despite the existence of other databases, SCP and WOS gathered the relevant journals related to business, management and innovation adoption. Harzing and Alakangas (2016) reinforce the reliability of SCP and WOS, and additionally highlight that both databases return more accurate outputs than

Google Scholar. The search included three types of published materials: journal articles, conference papers, and book chapters. Summaries, websites, and magazine articles were not considered in this LR. Ten relevant scientific fields were identified as relevant for this study: Computer Science; Engineering; Mathematics; Decision Sciences; Social Sciences; Business, Management, and Accounting; Multidisciplinary; Economics, Econometrics, and Finance. The keywords used for the searches were:

1. National Statistics Office (or NSO);
2. National Statistics Offices (or NSOs);
3. Big Data;
4. Adoption;
5. Acceptance;
6. Innovation.

Those six keywords were used in different combinations (strings) during twenty-seven search rounds in SCP and WOS. The Boolean connector “and” linked the keywords to compose the strings. During the current search string process, English as the primary language was used. Portuguese, as the mother language from Portugal (INE) and Brazil (IBGE) was applied in this string search. However, the results in Portuguese were null. Table 3 summarizes the adopted set of strings and the search results.

Table 3 - Results from the Strings Search

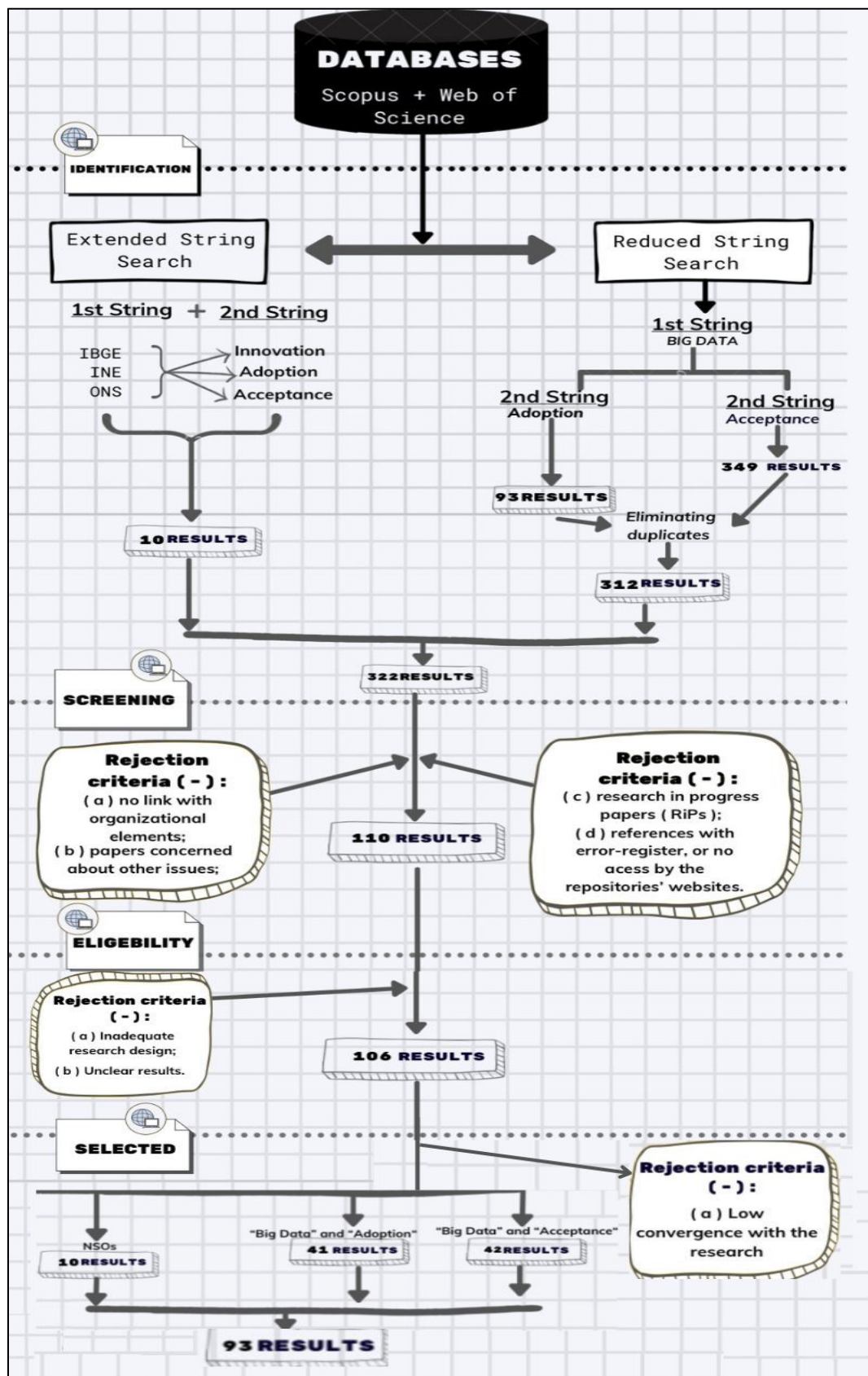
Results from the Strings Search		References obtained
<i>Strings</i>	<i>No. of rounds</i>	<i>Results</i>
“National Statistics Office” and “adoption” and “[country name]”	3	0
“National Statistics Office” and “acceptance” and “[country name]”	3	0
“National Statistics Office” and “innovation” and “[country name]”	3	0
“National Statistics Office” and “Big Data” and “[country name]”	3	10
“[NSO name]” and “adoption”	3	4
“[NSO name]” and “innovation”	3	5
“[NSO name]” and “Big Data”	3	1
“[NSO name]” and “acceptance”	3	0
“Big Data” and “adoption”	1	93
“Big Data” and “acceptance”	2	219
TOTAL	27	322

Considering the low number of results (only ten) in the initial rounds, other string searches were executed. In these additional searches, the reference to NSOs was dropped, and strings combining

“Big Data” with “adoption” or “acceptance” were used. This procedure followed the example of Ram et al. (2019, p. 566). Firstly, ninety-three (93) results from “Big Data” and “adoption” were found. Secondly, the string “Big Data” and “acceptance” generated three hundred forty-nine (349) references. After eliminating 130 duplicated or failed registers, 219 documents were considered for screening with the string “Big Data” and “acceptance.” Finally, three hundred and twenty-two (322) references were obtained after twenty-seven string searches, considering adoption (1 round), acceptance (2 rounds), and NSOs (24 rounds). Summarizing, the aggregated results (WOS+SCP) are presented also in Table 1.

Figure 5 shows the steps taken in this process that followed the stages proposed by Santos et al. (2020): (i) identification, (ii) screening, (iii) eligibility, (iv) selection. Identification is the initial stage, described in the previous paragraph and Table 1. In the screening stage, the title, abstract, keywords, and graphic elements were read. The exclusion criteria followed these patterns: (a) studies about technology *per se*, without link to organizational elements (e.g., coding, mathematical model to algorithm development); (b) papers concerned about other issues (e.g. medicine, nursing, astronomy, etc.); (c) research in progress papers (RiPs) without academic findings, or just proposing an investigation agenda; (d) references with register-errors in the databases, or no access by the principal repository websites (ScienceDirect, Springer Link, JStor, EBSCO, etc.). Two hundred and fifteen (215) references were thus rejected, and the number of references was reduced from 322 to 110. In the eligibility stage, four studies were removed because the study design and the results were unclear. So, the total references reviewed was 106, which were read in full.

Finally, at the selection stage, thirteen references were rejected because of low convergence with the research theme. The ten articles about NSOs remained in the selected studies. Our LR has a set of 93 references.



Legend: The four stages of the literature review process: (i) Identification, (ii) Screening; (iii) Eligibility; (iv) Selection. The initial 322 references were reduced to a final 93.

Figure 5 – The systematic search process

Considering the results of the Literature Review, the following section proposes to establish standard concepts about the central terms of this research. So, the strings “Big Data”, “Adoption”, and “Acceptance” are discussed based on the 93 references. This set of references constitutes the basis for achieving four objectives:

- (i) develop consolidated definitions used in this thesis, such as Big Data and Big Data applied in Official Statistics;
- (ii) identify the theoretical frameworks used in the literature to study the adoption of Big Data;
- (iii) focus specifically on studies using the Technology-Organization-Environment framework (TOE), and
- (iv) devise a consolidated list of factors affecting ABD in the 3 dimensions of the TOE framework: technological, organizational, and environmental.

2.2. Adoption of Big Data – definitions and frameworks

As remarked at the beginning of the current chapter, the initial discussion about Big Data positioned it as an asset within Information Technology (IT) with impacts to the organization. In this perspective, IT adoption constitutes a fundamental pillar on which to base the study related to ABD.

IT adoption has been a relevant research topic in Information Systems (IS) and Management for the last 30 years (Iacovou *et al.*, 1995; Chau and Tam, 1997; Karahanna *et al.*, 1999; Brown and Venkatesh, 2005; Ahuja and Thatcher, 2005; Komiak and Benbasat, 2006; Pavlou and Fygenson, 2006; Sarker and Valacich, 2010; Trigo *et al.*, 2013; Angst and Agarwal, 2016; Bayerl *et al.*, 2016; Duan *et al.*, 2017; Kim and Son, 2017; Retana *et al.*, 2018; Wunderlich and Veit, 2019; Bose and Leung, 2019; Jung *et al.*, 2019; Ho *et al.*, 2020). Most studies are concentrated in the last ten years (e.g., Al-Rahmi *et al.* 2019; Angwar 2018; Baig *et al.* 2019; Bilal *et al.* 2016, 2016, 2018; Bonner 2013; Cabrera-Sánchez and Villarejo-Ramos 2019; Cabrera-Sanchez and Villarejo-Ramos 2019; Félix *et al.* 2018; Kwon *et al.* 2014; Lai *et al.* 2018; Martinez-Mosquera and Luján-Mora 2019; McMahon *et al.* 2020; Memon *et al.* 2016; Park and Kim 2019; Ramadoss and Elango 2015; Shahbaz *et al.* 2020; Silva *et al.* 2019; Sun *et al.* 2018; Verma and Bhattacharyya 2017; Yadegaridehkordi *et al.* 2018).

To establish the state of the art related to ABD we therefore review models applied to IT adoption, and then more specifically to ABD. The TOE model demanded particular attention as it was followed in this research. So, we start by a discussion of the concepts and definitions of Big Data and ABD.

2.2.1. Conceptualizing Big Data

Big Data is a concept with high diversity in its definition. It is possible to affirm this situation based on the several different definitions presented in academic articles (Ylijoki and Porras 2016; Chen *et al.* 2017; Brock and Khan 2017; Urbinati *et al.* 2019; McMahon *et al.* 2020; Bakici *et al.* 2022). Looking just at the articles included in this literature review (see section 2.1), we find many different conceptualizations of Big Data.

Big Data is therefore a polysemic concept. Polysemy is the “*multiplicity of meanings*” (Ravin and Leacock, 2000) from different words which express a concept. Considering this polysemous feature of Big Data, the current study proposes to build a conceptualization of this term in an adjudicative (Cronin and George, 2020) sequence. A clear conceptualization of the term Big Data will improve the analytical capacity of this research, from the theoretical review to the interview analysis.

In this way, conceptualization or concept formation is a complementary process of theory formation, as remarked by Kaplan (1964). In the same perspective, Welch *et al.* (2016) define concepts as elements of a theory or, in their own words, “building blocks”. Additionally, Podsakoff *et al.* (2016) stated that a concept is “*a cognitive symbol that has meaning*”, adopted and use by researchers. In addition to their definition, Podsakoff *et al.* (2016) highlighted that concept formation is a core part of social science research.

Based on the relevance to social research, Welch *et al.* (2016) assert that the concept has a dynamic nature. Those authors understand that concepts must be “*tested, challenged and revised*” to verify their validity in a specific study. These procedures check if a concept present high intensity and low extension, or the opposite status. Looking at the plethora of Big Data definitions, there is an indication that this concept is very extensive and poor in intensity.

Welch *et al.* (2016) also indicates that parts of a concept – its attributes – must be compared as part of the conceptual test. This comparison process is relevant to identify if the attribute presents more coverage or accuracy. In the following pages and tables, the concept of Big Data is analysed and rebuilt with a specific purpose: to develop an updated conceptualization to be used in the context of National Statistical Offices and Official Statistics.

There are distinct methodologies to test, challenge and revise a concept. Morakanyane *et al.* (2017) adopt a concept-centred matrix to evaluate a concept, checking its core fundamentals present in the literature state of the art. Another method of conceptualization is proposed by Podsakoff *et al.* (2016) with the use of the literature review. This procedure defines ten steps/questions to check if the concept is reliable and robust. Although the previous two methods use literature review as the background, there are other possibilities to conceptualize a research term. For example, Welch *et al.* (2016) evaluated a concept with the use of a research case. The case therefore works as the basis of concept building.

In this doctoral investigation, Morakanyane *et al.*'s (2017) method is the selected one because the applied technique – a concept-centric matrix – is objective with a clear orientation of each step, and presents the stages in sequential tables. Indeed, those features are compatible with the structure of the LR adopted in this investigation (see section 2.1). The next paragraphs present the steps of building the concept of Big Data in the National Statistics Offices contexts following this method.

Firstly, Morakanyane *et al.*'s (2017) method starts with a structured literature review. The references extracted from the literature review define the landscape and coverage of the analysed concept. In the case of Big Data, our systematic literature review confirmed a polysemic status of its definition.

Considering the ninety-three references, our LR found that most of the articles, conference papers, and book chapters mention, cite but do not define the term “Big Data”. This is an indicator that the polysemy is present and acts as an inhibitor to the adjudicative research. Probably, many authors avoid conceptualizing the term Big Data considering the limited time and space available in a published manuscript. Also, the lack of consensus about the concept is an additional reason to avoid defining it.

The next stage in the Morakanyane *et al.*'s (2017) method is selection. In the sample of ninety-three references, just twenty presented a clear, objective, and specified conceptualization. Nineteen references are extracted from journals and reviews in Management, Business, and Information Systems. Those publications are top ranked by the academic portal SCIMAGOJR (<https://www.scimagojr.com/>) with high scores Q1 and Q2. Additionally, an international peer-reviewed conference paper was included in this set. In total, twenty references were applied to develop a Big Data adjudicative concept.

Table 4 shows the authors, year of publication, article title and journal. The term Big Data is part of the titles in nineteen references. Only one does not mention the term in the title; notwithstanding, it defines it. Another relevant find is that reputable reviews and journals are more receptive to publish manuscripts with Big Data conceptualization. This is not a subject frequent in conference papers and periodicals Q3 or Q4, as identified in the ninety-three references of this literature review. Additionally, the journal titles bring an indicator of the multidisciplinary related to Big Data. Nine of the nineteen journals have the term “management” in their titles, showing Big Data is a concept present not only in “information systems” studies but also in “management” research.

Table 4 – References for developing the Big Data adjudicative concept

Authors	Title	Journals and Conference Titles
Gangwar, 2018	Understanding the Determinants of Big Data Adoption in India: An Analysis of the Manufacturing and Services Sectors	Information Resources Management Journal
Sun et al., 2018b	Understanding the Factors Affecting the Organizational Adoption of Big Data	Journal of Computer Information Systems
Baig et al., 2019	Big Data adoption: State of the art and research challenges	Information Processing and Management
Yadegaridehkordi et al., 2018b	Influence of big data adoption on manufacturing companies' performance: An integrated DEMATEL-ANFIS approach	Technological Forecasting and Social Change
Salleh et al., 2015	SEC-TOE Framework: Exploring Security Determinants in Big Data Solutions Adoption	PACIS 2015 Proceedings - Pacific Asia Conference on Information Systems (PACIS)
Sun et al., 2020	Organizational intention to adopt big data in the B2B context: An integrated view	Industrial Marketing Management
Ram et al., 2019	Adoption of Big Data analytics in construction: development of a conceptual model.	Built Environment Project and Asset Management
Bakici et al., 2022	Big Data Adoption in Project Management: Insights from French Organizations	IEEE Transactions on Engineering Management
Paradza and Daramola, 2021	Business Intelligence and Business Value in Organizations: A Systematic Literature Review	Sustainability
Maroufkhani et al., 2022	Determinants of big data analytics adoption in small and medium-sized enterprises (SMEs)	Industrial Management & Data Systems,
Baig et al., 2021	A Model for Decision-Makers' Adoption of Big Data in the Education Sector	Sustainability
Lai et al., 2018b	Understanding the determinants of big data analytics (BDA) adoption in logistics and supply chain management: An empirical investigation	International Journal of Logistics and Management
Al-Rahmi et al., 2019	Big Data Adoption and Knowledge Management Sharing: An Empirical Investigation on Their Adoption and Sustainability as a Purpose of Education	IEEE ACCESS

Mcmahon et al., 2020	Requirements for Big Data Adoption for Railway Asset Management	IEEE ACCESS
Müller et al., 2018	The Effect of Big Data and Analytics on Firm Performance: An Econometric Analysis Considering Industry Characteristics	Journal of Management Information Systems
Urbinati et al., 2019	Creating and capturing value from Big Data: A multiple-case study analysis of provider companies	Technovation
Günther et al., 2017	Debating Big Data: A literature review on realizing value from big data.	The Journal of Strategic Information Systems
Brock and Khan, 2017.	Big data analytics: does organizational factor matters impact technology acceptance?	Journal of Big Data
Ylijoki and Porras, 2016	Conceptualizing Big Data: Analysis of case studies.	Intelligent systems in accounting, finance and management
Chen et al., 2017	How Lufthansa Capitalized on Big Data for Business Model Renovation.	MIS Quarterly Executive

In sequence, an in-depth analysis of the references (Morakanyane *et al.*, 2017) took place. During this stage, it was possible to identify how researchers understand and define the concept of Big Data. The analysis was based on the definition, characteristics, and the attributes of the concept. Attributes are a central lens of this analysis, considering their relevance to conceptualization (Welch *et al.*, 2016). So, the development of a concept centric matrix was deployed, showing the different concepts and sources, presented in Table 5.

It is necessary to remark that the twenty different definitions represent a polysemic phenomenon. As an example, there is a distinct understanding of Big Data in the same scientific periodical, such as the “IEEE Access” journal, where Al-Rahmi *et al.* (2019) and McMahon *et al.* (2020) conceptualized Big Data with different meanings and approaches.

Table 5 – Big Data: the current conceptualizations

Authors	Big Data Definitions
Gangwar, 2018	<i>“It is a collection of huge and complex amalgamation of data sets which make it difficult to process using traditional data processing platforms.”</i>
Sun <i>et al.</i> , 2018b	<i>“In this research, from a business solution perspective, we define big data as a new technology that is primarily characterized by its advanced business intelligence and analytics (BI&A) function.”</i>

Authors	Big Data Definitions
Baig <i>et al.</i> , 2019	<i>"Big Data is defined as a large dataset that could not be controlled or examined through conventional databases."</i>
Yadegaridehkordi <i>et al.</i> , 2018b	<i>"Big Data is currently considered as a fast-emergence amount of data from different processes that gradually poses a challenge to industrial organizations all around the world."</i>
Salleh <i>et al.</i> , 2015	<i>"The term (Big Data) is often associated to three unique characteristics of data: Volume, Variety, and Velocity or more widely known as the 3Vs."</i>
Sun <i>et al.</i> , 2020	<i>(...) "describe the data sets and analytical techniques in applications that are so large (from terabytes to exabytes) and complex (from sensor to social media data) that they require advanced and unique data storage, management, analysis, and visualization technologies."</i>
Ram <i>et al.</i> , 2019	For the purpose of clarity, BD (Big Data) is defined here 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."
Paradza and Daramola, 2021	"The latest term that evolved from BI and BA is Big Data or Big Data Analytics which represents larger volumes and complex sets of data requiring dedicated tools to synthesize information while still upholding the same emphasis on reporting and predictive analytics."
Bakici <i>et al.</i> , 2022	"In this article, Big Data are defined as datasets in such a large scale that typical tools and software cannot store, manage, and analyse it as before. The Big Data come in various forms (structured and unstructured) from various sources and requires a fast-real-time analysis. Broadly, Big Data technologies include distributed file systems (Hadoop MapReduce framework), NoSQL, MPP systems, cloud computing platforms (storage and computing resources), in-memory database processing, and data mining tools."
Maroufkhani <i>et al.</i> , 2022	<i>"The term "Big Data" refers to the massive amounts of data, including both unstructured and structured, which is accessible in real-time."</i>
Baig <i>et al.</i> , 2021	<i>"Big Data refers to huge and multifaceted data sets that need potent storage systems and tools [2]. The characteristics of Big Data make it distinguishable from conventional data. Most of the research characterized Big Data into seven V's [3], namely: Volume (amount of data), Velocity (speed of data), Variety (structure of data), Veracity (accuracy of data), Variability (meaning of data), Volatility (validity of data), and Value (use of data)."</i>

Authors	Big Data Definitions
Lai <i>et al.</i> , 2018b	<i>"BD is put forward to describe the massiveness and complexity of data which usually require organizations to improve processing abilities and analysing capabilities (...)."</i>
McMahon <i>et al.</i> , 2020	<i>"Big Data refers to the explosion in the quantity (and sometimes, quality) of available and potentially relevant data, largely due to the result of recent and unprecedented advancements in data recording and storage technology."</i>
Al-Rahmi <i>et al.</i> , 2019	<i>"Big Data are Information asset categorized by high Velocity, Volume and Variety to have Technology or Analytical techniques for its conversion towards Value."</i>
Müller <i>et al.</i> , 2018	<i>"Unprecedented growth in data volume, variety, and velocity has emerged over the course of the past decade, a phenomenon often referred to as 'Big Data'."</i>
Urbinati <i>et al.</i> , 2019	<i>"Big Data refers to datasets that are large in volume, diverse in data sources and types, and created quickly, resulting in greater challenges to harvest, manage, and process them through traditional systems and capabilities."</i>
Günther <i>et al.</i> , 2017	<i>"Large volumes of extensively varied data that are generated, captured, and processed at high velocity."</i>
Brock and Khan, 2017.	<i>"'Big Data' is considered as the datasets that challenge the ability of typical applications and technologies in managing and analysing the data. Big data also challenges the human imagination."</i>
Ylijoki and Porras, 2016	<i>"Big Data is a multidisciplinary phenomenon. Big Data is an emerging, multidisciplinary research area."</i>
Chen <i>et al.</i> , 2017	<i>"'Big data' is the term for data sets that are so large or complex that traditional data processing applications are inadequate to deal with them."</i>

Morakanyane *et al.* (2017) remark the concept centric matrix is the core of a conceptualization process. Presented as a table, this construct has the function of identifying patterns in the structure of the definitions, the adopted keywords, and the variables used to develop the definitions. Welch *et al.* (2016) understands variables as attributes, parts of a concept. In this section, concepts' variables are considered attributes. The attributes, structures and patterns of Big Data concepts, are presented in Table 6 – Concept-Centric Matrix.

After an attentive reading of the concepts, it is possible to split them into attributes, as proposed by Welch *et al.* (2016). Each concept of Big Data present, at a minimum, one of three attributes: (i) essence - what Big Data is; (ii) special features – what makes Big Data different from other current technologies and methodologies; (iii) uses/purpose – what Big Data is deployed for. Matrix

1 shows these three attributes, in this order, and using font format to distinguish them: *italic* – essence, **bold** – special feature, underline – uses/purpose. Some of the concepts are incomplete and just adopt one or two attributes. This structural divergence is another symptom of the polysemic in Big Data conceptualization.

Those combined attributes develop a focus of each concept. Based on the most relevant features, it is possible to remark the concepts are clustering in four different groups or categories: (a) coverage, (b) data features, (c) strategy, and (d) uses. Coverage involves the domain of the Big Data concept: “technological application”, “multidisciplinary topic”, “information asset”... Data features relates to properties of the data, since “huge amounts of data”, “unprecedented amounts of data”, “structured and unstructured data”, “large datasets”. Strategy indicates how Big Data influences the decision-making processes in an organization concerned about “analysis” and “management” based on data. Finally, uses categorizes how Big Data is used in organizations and which are its purposes, such as “enable enhanced insight”, “process automation”, or “conversion towards value”. This categorization is presented in Table 6, the Concept-Centric Matrix.

The result from the concept-centric matrix is clear. There is a prevalence of “data features” over the other three categories. It is a prevalence in temporal and numeric perspectives. Since 2015 until 2021, this category has been adopted by eleven references, ten journals and the one conference paper. However, this majority indicates much more a path dependence than a robust convergence about the concept. Conceptualizations related to “data features” presented a peculiar aspect. They use previous references to propose the concept of Big Data, lacking in originality and propositions. Maybe, this scenario can be result of an attempt to reduce the polysemic by the authors, basing their definition in previous articles.

The category “uses” of the Big Data was addressed by Sun *et al.* (2018), Ram *et al.* (2019), Al-Rahmi *et al.* (2019), Paradza and Daramola (2021). This is a more recent approach in the literature. “Coverage” has also emerged in the last five years and is used three times (Ylijoki and Porras (2016), Sun *et al.* (2020), Bakici *et al.* 2022). “Strategy” appeared in just two citations, but the last reference was from four years ago: Brock and Khan (2017) and Lai *et al.* (2018b). Beyond the quantitative results, the concept centric matrix expressed an evolution of the Big Data concept. In the earlier references the data features category was the only one used. After improvements from the academic debate, and the spread of the concept, it turns its focus on to coverage and uses.

Table 6 – Concept-Centric Matrix

Authors	Definition	Attributes			Categories
		<i>Essence</i>	Special feature	<u>Uses/purpose</u>	
Ylijoki, O.; Porras, J. (2016)	Big data is a multidisciplinary phenomenon. (p. 296) Big data is an emerging, multidisciplinary research area. (p. 297)	<i>Big data is a multidisciplinary phenomenon.</i>	[NONE]	<u>Big data is an emerging multidisciplinary research area.</u>	Coverage
Sun, S.; Hall, D. J.; and Cegielski, C. G. (2020)	Big data and big data analytics are used to “describe the data sets and analytical techniques in applications that are so large (from terabytes to exabytes) and complex (from sensor to social media data) that they require advanced and unique data storage, management, analysis, and visualization technologies. (p. 110)	<i>Big data and big data analytics are used to “describe the data sets and analytical techniques in applications</i>	that are so large (from terabytes to exabytes) and complex (from sensor to social media data)	<u>that they require advanced and unique data storage, management, analysis, and visualization technologies.”</u>	Coverage
Bakici, T.; Nemeh, A.; and Hazir, Ö. (2021)	In this article, big data are defined as datasets in such a large scale that typical tools and software cannot store, manage, and analyze it as before. The big data come in various forms (structured and unstructured) from various sources and requires a fast-real-time analysis. Broadly, big data technologies include distributed file systems (Hadoop MapReduce framework), NoSQL, MPP systems, cloud computing platforms (storage and computing resources), in-memory database processing, and data mining tools (p. 1).	<i>In this article, big data are defined as datasets in such a large scale</i>	that typical tools and software cannot store, manage, and analyze it as before. The big data come in various forms (structured and unstructured) from various sources and requires a fast-real-time analysis	<u>Broadly, big data technologies include distributed file systems (Hadoop MapReduce framework), NoSQL, MPP systems, cloud computing platforms (storage and computing resources), in-memory database processing, and data mining tools.</u>	Coverage
Salleh, K. A.; Janczewski, L.; and Beltran, F. (2015)	term ‘Big Data’. The term is often associated to three unique characteristics of data: Volume, Variety, and Velocity or more widely known as the 3Vs (p. 1)	<i>term ‘Big Data’. The term is often associated to three unique characteristics of data: Volume, Variety, and Velocity or more widely known as the 3Vs (P. 1)</i>	[NONE]	[NONE]	Data features
Günther, W. A.; Mehrizi, M. H. R.; Huysman, M.; and Feldberg, F. (2017)	Large volumes of extensively varied data that are generated, captured, and processed at high velocity. (p. 191)	<i>Large volumes of extensively varied data</i>	that are generated, captured, and processed at high velocity	[NONE]	Data features
Chen, H. M.; Schütz, R.; Kazman, R.; and Matthes, F. (2017)	“Big data” is the term for data sets that are so large or complex that traditional data processing applications are inadequate to deal with them. (p. 19)	<i>Big data” is the term for data sets</i>	that are so large or complex	<u>that traditional data processing applications are inadequate to deal with them.</u>	Data features

Table 6 – Concept-Centric Matrix (Continuation)

Authors	Definition	Attributes			Categories
		Essence	Special feature	Uses/purpose	
Gangwar, H. (2018)	it is a collection of huge and complex amalgamation of data sets which make it difficult to process using traditional data processing platforms (P. 1)	<i>it is a collection of huge and complex amalgamation of data sets</i>	which make it difficult to process using traditional data processing platforms (P. 1)	[NONE]	Data features
Yadegaridehkordi, E.; Hourmand, M.; Nilashi, M.; Shuib, L.; Ahani, A.; and Ibrahim, O. (2018)	Big data is currently considered as a fast-emergence amount of data from different processes that gradually poses a challenge to industrial organizations all around the world (p. 201)	<i>Big data is currently considered as a fast-emergence amount of data from different processes</i>	that gradually poses a challenge	<i>to industrial organizations all around the world (p. 201)</i>	Data features
Müller, O.; Fay, M.; and vom Brocke, J. (2018)	Unprecedented growth in data volume, variety, and velocity has emerged over the course of the past decade, a phenomenon often referred to as "big data." (p. 489)	<i>Unprecedented growth in data volume, variety, and velocity has emerged over the course of the past decade, a phenomenon often referred to as "big data."</i>	[NONE]	[NONE]	Data features
Baig, M.I.; Shuib, L.; and Yadegaridehkordi, E. (2019)	Big data is defined as a large dataset that could not be controlled or examined through conventional databases (p. 3)	<i>Big data is defined as a large dataset</i>	that could not be controlled or examined through conventional databases	[NONE]	Data features
Urbinati, A.; Bogersb, M.; Chiesaa, V.; and Frattinia, F. (2019)	Big Data refers to datasets that are large in volume, diverse in data sources and types, and created quickly, resulting in greater challenges to harvest, manage, and process them through traditional systems and capabilities. (p. 21)	<i>Big Data refers to datasets</i>	that are large in volume, diverse in data sources and types, and created quickly, resulting in greater challenges to harvest, manage, and process them through traditional systems and capabilities	[NONE]	Data features
McMahon, P.; Zhang, T.; and Dwight, R. (2020)	'Big Data' refers to 'the explosion in the quantity (and sometimes, quality) of available and potentially relevant data, largely due to the result of recent and unprecedented advancements in data recording and storage technology' (p. 6).	<i>'Big Data' refers to 'the explosion in the quantity (and sometimes, quality) of available and potentially relevant data, largely due to the result of recent and unprecedented advancements in data recording and storage technology'.</i>	[NONE]	[NONE]	Data features

Table 6 – Concept-Centric Matrix (Continuation)

Authors	Definition	Attributes			Categories
		<i>Essence</i>	Special feature	<u>Uses/purpose</u>	
Baig, M.I.; Shuib, L.; and Yadegaridehkordi, E. (2021)	Big data refers to huge and multifaceted data sets that need potent storage systems and tools. The characteristics of big data make it distinguishable from conventional data. Most of the research characterized big data into seven V's, namely: Volume (amount of data), Velocity (speed of data), Variety (structure of data), Veracity (accuracy of data), Variability (meaning of data), Volatility (validity of data), and Value (use of data). (p. 1)	<i>Big data refers to huge and multifaceted data sets</i>	that need potent storage systems and tools. The characteristics of big data make it distinguishable from conventional data. Most of the research characterized big data into seven V's, namely: Volume (amount of data), Velocity (speed of data), Variety (structure of data), Veracity (accuracy of data), Variability (meaning of data), Volatility (validity of data), and Value (use of data).	[NONE]	Data features
Maroufkhani, P.; Iranmanesh, M.; and Ghobakhloo, M. (2022)	The term "big data" refers to the massive amounts of data, including both unstructured and structured, which is accessible in real-time (p. 3).	<i>The term "big data" refers to the massive amounts of data,</i>	including both unstructured and structured,	<u>which is accessible in real-time</u>	Data features
Brock, V.; Khan, H. U. (2017)	"Big Data" is considered as the datasets that challenge the ability of typical applications and technologies in managing and analysing the data. Big data also challenges the human imagination. (p. 1)	<i>"Big Data" is considered as the datasets</i>	that challenge the ability of typical applications and technologies (...) Big data also challenges the human imagination.	<u>in managing and analysing the data</u>	Strategy
Lai, Y.Y.; Sun, H.F.; and Ren, J.F. (2018)	BD is put forward to describe the massiveness and complexity of data which usually require organizations to improve processing abilities and analyzing capabilities (Chen et al., 2013). (p. 678)	[NONE]	BD is put forward to describe the massiveness and complexity of data	<u>which usually require organizations to improve processing abilities and analyzing capabilities</u>	Strategy
Sun S.; Cegielski C.G.; Jia, L.; and Hall, D.J. (2018)	"In this research, from a business solution perspective, we define big data as a new technology that is primarily characterized by its advanced business intelligence and analytics (BI&A) function." (p. 194)	<i>big data as a new technology</i>	[NONE]	<u>that is primarily characterized by its advanced business intelligence and analytics (BI&A) function "</u>	Uses
Ram, J.; Afridi, N. K.; and Khan, K. A. (2019)	For the purpose of clarity, BD is defined here 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"(p. 566)	<i>For the purpose of clarity, BD is defined here as "high-volume, high-velocity and/or high-variety information assets</i>	.that demand cost-effective, innovative forms of information processing	<u>that enable enhanced insight, decision making and process automation"</u>	Uses

Table 6 – Concept-Centric Matrix (Continuation)

Authors	Definition	Attributes			Categories
		<i>Essence</i>	Special feature	<u>Uses/purpose</u>	
Sun S.; Cegielski C.G.; Jia, L.; and Hall, D.J. (2018)	"In this research, from a business solution perspective, we define big data as a new technology that is primarily characterized by its advanced business intelligence and analytics (BI&A) function." (p. 194)	<i>big data as a new technology</i>	[NONE]	<u>that is primarily characterized by its advanced business intelligence and analytics (BI&A) function "</u>	Uses
Ram, J.; Afridi, N. K.; and Khan, K. A. (2019)	For the purpose of clarity, BD is defined here 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"(p. 566)	<i>For the purpose of clarity, BD is defined here as "high-volume, high-velocity and/or high-variety information assets</i>	that demand cost-effective, innovative forms of information processing	<u>that enable enhanced insight, decision making and process automation"</u>	Uses
Al-Rahmi, W.M.; Yahaya, N.; Aldraiweesh, A.A.; Alturki, U.; Alamri, M.; Bin Saud, M.S.; Kamin, Y.B.; Aljeraiwi, A.A.; and Alhamed, O.A. (2019)	Big Data are Information asset categorized by high Velocity, Volume and Variety to have Technology or Analytical techniques for its conversion towards Value (p. 47247)	<i>Big Data are Information asset</i>	categorized by high Velocity, Volume and Variety to have Technology or Analytical techniques	<u>for its conversion towards Value</u>	Uses
Paradza, D.; Daramola, O. (2021)	The latest term that evolved from BI and BA is Big Data or Big Data Analytics which represents larger volumes and complex sets of data requiring dedicated tools to synthesise information while still upholding the same emphasis on reporting and predictive analytics (p. 28).	<i>The latest term that evolved from BI and BA is Big Data or Big Data Analytics which represents larger volumes and complex sets of data</i>	requiring dedicated tools to synthesise information	<u>while still upholding the same emphasis on reporting and predictive analytics</u>	Uses

This trend indicates that Big Data conceptualization is suffering a refinement owing to its multidisciplinary nature, not only circumscribed to the information systems and information technology domains.

Like the categories, the attributes support the conceptualization analysis, indicating a possibility to develop a phrase-template or a “pattern that suggests” (Morakanyane et al., 2017) a concept for Big Data:

It is something... (+) with/that [a special feature] added to a [complexity level] (+) to deliver/deploy solutions form a specific/general proposal or exemplification.

Based on this phrase-template, it is possible to evolve to an adjudicated conceptualization of Big Data. To build this concept it is necessary to select which attributes can be adopted from the references in the concept centric matrix. Table 7 presents the selected references and the extracted attributes to build a new Big Data concept.

All four categories (use, data features, strategy and coverage) were selected to conceptualize the Big Data. There were the uses of five attributes from the first part, “essence”, three from “special feature”, and six from “uses/purpose”. In the end, the developed conceptualization of Big Data balances the trade-off between large coverage and strong accuracy, as remarked by Welch et al. (2016).

Considering the selected fourteen attributes, it is possible to remark that Big Data, as a concept, presents the prevalence of two categories: data features and uses. The conceptualization to this term must approach the types of data that this technology/methodology are based on. In the same context, the uses and applications for Big Data have to be considered as part of its meaning. If Big Data is a disruptive innovation (Ylijoki and Porras 2018; Maroufkhani et al. 2020), this disruption demands an applicability to take place.

Then, the adjudicative concept of Big Data can be present in two sentences, as shown below:

*As a multidisciplinary phenomenon, Big Data are used to define the huge data sets and complex analytical techniques in applications **with various data formats (structured and unstructured) from various data sources, requiring a fast-real-time capability to generate, capture, store, process, analyse and visualize data.** Big data adopters are able to enhance insight, decision-making and process automation, by extracting value from data.*

Table 7 - Selected Attributes Matrix

Authors and Year	Categories	Essence	Special feature	Uses/purposes
Ylijoki, O., & Porras, J. (2016)	Converge	<i>Big data is a multidisciplinary phenomenon.</i>		
Sun, S., Hall, D. J., & Cegielski, C. G (2020)		<i>Big data and big data analytics are used to "describe the data sets and analytical techniques in applications"</i>	that are so large (from terabytes to exabytes) and complex (from sensor to social media data)	<u>that they require advanced and unique data storage, management, analysis, and visualization technologies."</u>
BAKICI, Tuba; NEMEH, André; HAZIR, Öncü. (2021)			that typical tools and software cannot store, manage, and analyze it as before. The big data come in various forms (structured and unstructured) from various sources and requires a fast-real-time analysis	<u>Broadly, big data technologies include distributed file systems (Hadoop, MapReduce framework), NoSQL, MPP systems, cloud computing platforms (storage and computing resources), in-memory database processing, and data mining tools .</u>
Günther, W. A., Mehrizi, M. H. R., Huysman, M., & Feldberg, F. (2017)	Sources		that are generated, captured, and processed at high velocity	
Baig M.L, Shuib L., Yadegaridehkordi E. (2021)		<i>Big data refers to huge and multifaceted data sets</i>		
Maroufkhani, P., Iranmanesh, M. and Ghobakhloo, M. (2022)				<u>which is accessible in real-time.</u>
Lai, YY; Sun, HF; Ren, JF (2018)	Strategy			<u>which usually require organizations to improve processing abilities and analyzing capabilities</u>
Sun S., Cegielski C.G., Jia L., Hall D.J. (2018)	Uses			<u>that is primarily characterized by its advanced business intelligence and analytics (BI&A) function."</u>
Ram, J., Afridi, N. K., & Khan, K. A. (2019)				<u>that enable enhanced insight, decision making and process automation"</u>
Al-Rahmi W.M., Yahaya N., Aldraiweesh A.A., Alturki U., Alamri M., Bin Saud M.S., Kamin Y.B., Aljeraiwi A.A., Alhamed O.A. (2019)		<i>Big Data are Information asset</i>		
Paradza, D., & Daramola, O. (2021)		<i>The latest term that evolved from BI and BA is Big Data or Big Data Analytics which represents larger volumes and complex sets of data</i>		

This Big Data conceptualization is a reference that can be used in following studies about the theme. It is composed of a wide range of attributes, requirements and applications which makes this as a feasible and updated concept.

Focusing on the Official Statistics domain, this concept can also be adopted. However, this context presents singularities described in the first chapter, in the section “Global Statistical System”. It is possible to summarize those singularities in the ten principles of the Official Statistics.

Then, the developed conceptualization of Big Data, presented above, can be applied to Official Statistics with some adjustments. In view of this demand, the Big Data concept for Official Statistics can be defined as:

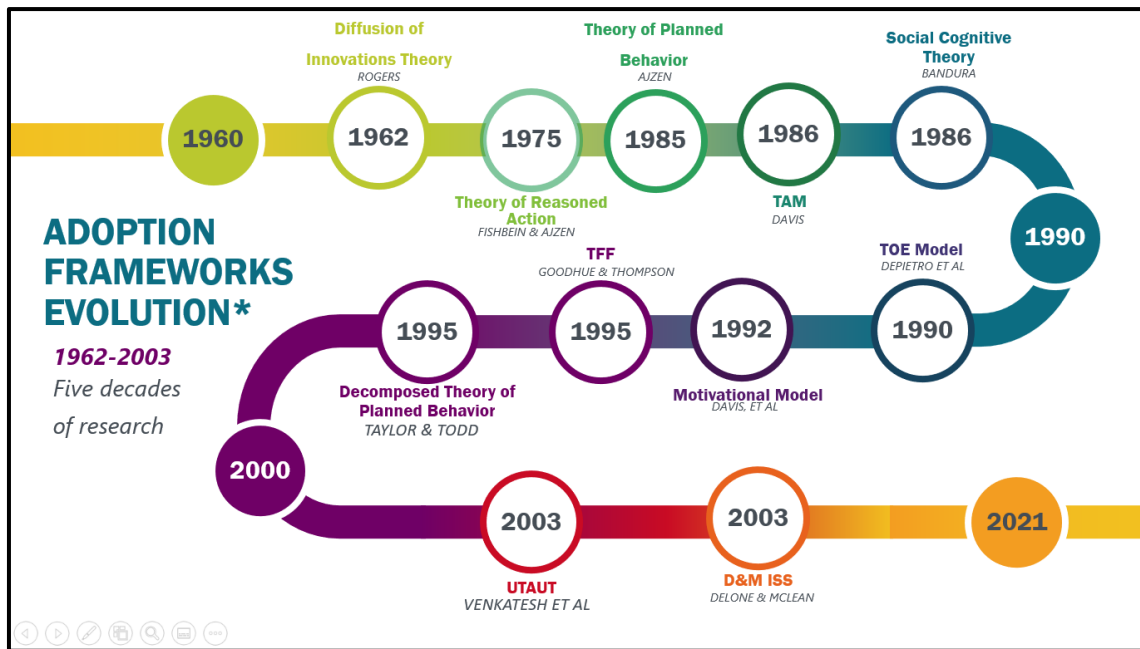
Big data are used to define the huge data sets and complex analytical techniques in applications with various data formats (structured and unstructured), from various data sources, requiring a fast-real-time capability to generate, capture, store, process, analyse and visualize data. Following the principles of Official Statistics, those datasets and techniques can be adopted to produce reliable, accurate and timely public data.

Concluding this section, the conceptualization of Big Data for Official Statistics is the current meaning to be understood when the term Big Data will be cited in the following pages. This adjusted concept fills a remaining gap in the literature involving Big Data and Official Statistics.

The adoption models used to study Big Data adoption are reviewed in the next subsection. Considering the number of frameworks used in Big Data studies, the models are analysed under the applicability to this research focus.

2.2.2. Theoretical frameworks for the adoption of Big Data

The diversity of frameworks can be an indicator of the complexity in studying and understanding the technological adoption process. Figure 6 presents those frameworks organized by the year of the original publication. This timeline represents the evolution of frameworks applied in Technology adoption studies. The more representative are briefly described next.



Legend: Innovation adoption frameworks organized by the year of the original publication.
Source: adapted from Varajão *et al.* (2021).

Figure 6 – Timeline of the innovation adoption frameworks evolution

Varajão *et al.* (2021) listed eleven models and analytical structures as the most used in the literature to understand technological adoption processes or the adoption project to deploy a new information technology asset. This will be used to analyse the 93 articles selected in the systematic literature review described in section 2.1 in order to identify which IT adoption models have been applied to the study of ABD. We start by reviewing eight of those models.

Diffusion Of Innovation (DOI) theory. Developed by Rogers (1962, 1983, 1995, 2004), this framework was originally adopted in rural sociology studies. The analytical level of this model is the individual, and the influence of adopter/user's decision in the organizations (Gallivan, 2001). About the analytical dimensions, the DOI framework considers the adopter/user, the organization, and the technology. Its set of factors are composed by innovation, communication, social system, time, individual education and cultural level, knowledge transfer, barriers to knowledge transfer, individual learning capacity, and assimilation gap (Fichman and Kemerer, 1999; Gilbert and Cordey-hayes, 1996; Lundblad, 2003; Miranda *et al.*, 2016).

Theory of Reasoned Action (TRA). This framework was applied to study behavioural processes, and emerged from psychological research (Fishbein and Ajzen, 1977). Although its initial application did not approach information system management and organizational behaviour, the TRA is relevant as a framework-base. It supported the development of other

frameworks, like TAM and UTAT. At the individual level, TRA applies two factors (attitude towards behaviour and subjective norms) to identify individuals' behaviour intention (Varajão *et al.*, 2021). Additionally, TRA proposes that user/adopter behaviour can be measured by beliefs and expectations related to the outcomes and attributes from specific actions (Mital *et al.*, 2018). Currently, TRA is applied in distinct research fields studies, from the Adoption of Internet of Things (Mital *et al.*, 2018) to “*the consumer attitude towards green consumption*” (Paul *et al.* 2016, p. 124).

Theory of Planed Behaviour (TPB). Another individual-level framework, the Theory of Planed Behaviour (or TPB), is an evolution of TRA (Varajão *et al.*, 2021). It was built by one of the TRA's authors, Ajzen (1991). This framework proposes that an individual's behaviour results from three factors: subjective norms, attitude towards behaviour, and the individual-self-perception of the attitudes control (Paul *et al.*, 2016). Like TRA, the TPB also supported the development of the Technology Acceptance Model (TAM). Mital *et al.* (2018, p. 341) classify TPB and TAM in function of the dimensions of analysis. TPB analyses reduced parts of the user perceptions. In its turn, TAM has a wider analytical focus, considering the mix of perception and behaviour.

Technology of Acceptance Model (TAM). As an evolution of TRA and TPB, the Technology of Acceptance Model (or TAM) is one of the most powerful frameworks to study how users decide to adopt or avoid vanguard technologies (Al-Rahmi *et al.*, 2019). Davis (1986) and Davis *et al.* (1989) developed TAM considering that the user's attitude toward using technologies receives influence from two specific factors, perceived usefulness and perceived ease of use (Varajão *et al.*, 2021). Following Davis's original proposal, other studies added new factors to this framework, reinforcing its adaptability and feasibility (Mital *et al.*, 2018).

Technology-Organization-Environment Model (TOE). Technology-Organization-Environment model (or TOE) is an adaptative framework to study technology adoption (Varajão *et al.*, 2021). DePietro *et al.*, (1990) developed it as an evolution of the DOI model (Angwar 2018, p.4). Based on its three dimensions, the technological, organizational, and environmental, TOE's unit of analysis is the organization, and not the individual. This feature supports the model to be used in studies about ABD by companies, as remarked by Baig *et al.* (2019).

Task-Technology Fit (TTF) - This framework complements a gap in the Technology of Acceptance Model (TAM) concerned about IT users task motivation (Varajão *et al.*, 2021). The authors of TTF, Goodhue and Thompson (1995), developed Technology-to-Performance Chain

structure, composed by three factors: (i) IT characteristics, (ii) tasks, and (iii) individual users. Those three factors can support the understanding of why a person uses an information system and improves or reduces his/her performance.

DeLone and McLean IS Success Model (DMISM). Distinct from some previous frameworks, DeLone and McLean IS Success Model (or DMISM), has been developed specifically to Information System studies (Varajão *et al.*, 2021). The authors, DeLone and McLean (2003), proposed the use of this model in e-commerce systems success adoption. One important feature of DMISM is the link between the individual and the organizational dimensions, offering a bottom-up perspective of the adoption process inside an organization. Wang and Liao (2008) tested DMISM to measure the success of e-Government systems. Petter *et al.* (2013) operated an extensive literature review and identified twenty success factors to be achieved in a well-deployed adoption process, with the use of DMISM.

Unified Theory of Acceptance and Use of Technology (UTAUT). Venkatesh et al. (2003, p. 428-432) developed this framework with the aggregation of features from other previous models: (i) TRA, (ii) TAM, (iii) MM, (iv) TPB, (v) DOI, (vi) SCT; (vii) model of Personal Computer Utilization (Thompson et al., 1991); (viii) hybrid model using TAM and TPB (Taylor and Todd 1995). In this model, a set of individual factors associated with environmental elements influence the behavioural intention to adopt technology. Respectively, the individual components are age, gender, experience, and voluntariness of use, and the environmental elements are facilitating conditions, performance expectancy, effort expectancy, and social influence (Venkatesh *et al.* 2003, p. 447).

However, the technological adoption field of study is not restricted to those frameworks. There are theoretical constructs adapted from the other areas, such as project management.

2.2.2.1. Frameworks used to study the adoption of Big Data.

We now focus specifically on the models used in the 93 references included in the literature review. In total, twenty-two frameworks or theoretical constructs are mentioned. Twenty-six of the 93 references do not apply any specific framework, especially literature review articles. In some papers, two or more frameworks are adopted (e.g., Pillai and Sivathanu, 2020; Queiroz and Farias Pereira, 2019). Figure 7 presents the frequency of frameworks in this literature review.

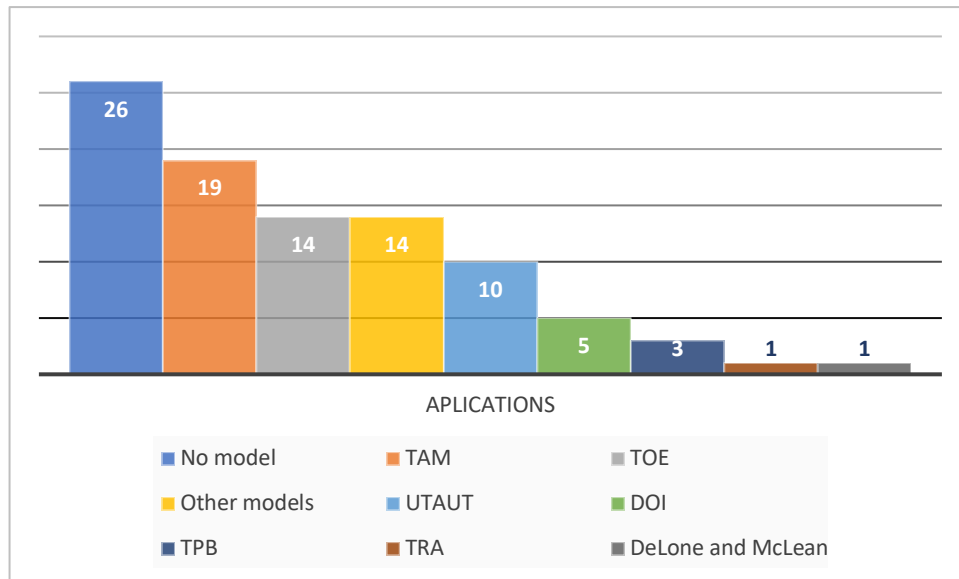


Figure 7 – The relevant frameworks present in this literature review

Specifically, about ABD, five frameworks were identified: Diffusion of Innovation (DOI), Technology-Organization-Environment model (TOE), Technology Acceptance Model (TAM), Task-Technology Fit (TTF), Unified Theory of Acceptance and Use of Technology (UTAUT). The frameworks' features considered in this discussion are those relevant to study ABD processes. The review will focus mainly on the following elements: levels of analysis (individual or organizational), dimensions (internal or external), and influencing factors (antecedent variables).

Diffusion of Innovation theory (DOI)

Since its development by Rogers (1983, 1995), the DOI model is widespread in technological adoption studies (Daradkeh, 2019). ABD is also studied with the support of DOI. Baig *et al.* (2019) applied a combination of DOI with the TOE framework to understand the adoption decision. The authors remarked that five factors influence ABD: relative advantage, complexity, compatibility, trialability, and observation. Daradkeh (2019) also associated DOI with TOE to understand which factors are relevant to the adoption of Big Data Analytics and identified thirty-four. In both studies (Baig *et al.* 2019 and Daradkeh 2019), the researchers applied the combination of frameworks to understand ABD from the individual to the organizational level. In a different approach, Saheb (2020) used DOI to explain the impact of digital services of social connectivity and personalization on the adoption behaviour.

Technology of Acceptance Model (TAM)

Al-Rahmi et al. (2019) used TAM to understand how users adopt or avoid Big Data technologies in the educational sector, at the individual level. In this study, the authors added eight factors to the TAM framework: (i) perceived risk; (ii) users age diversity; (iii) cultural diversity; (iv) motivators; (v) behaviour intention; (vi) knowledge management sharing; (vii) ABD readiness; (viii) sustainability for education. Other research about the adoption of Big Data analytics in healthcare organizations (Shahbaz et al., 2020) included two factors – gender and resistance to change – to the original framework, considering the internal dimension of the organization.

Task-Technology Fit (TTF)

Sinha et al. (2019) applied the TTF framework to verify the effectiveness of the Internet of Things (IoT) and Big Data in disaster management. Gangwar (2020) studied the integration of TAM and TTF to understand how the usage of Big Data Analytics affects business performance. Considering the research developed by Gangwar (2020), TIF has the potential to evolve from the individual to the organizational analytical level.

Unified Theory of Acceptance and Use of Technology (UTAUT)

The UTAUT model is used in ABD studies, considering both the individual and the organizational levels. Cabrera-Sanchez and Villarejo-Ramos (2019), based on the UTAUT model, investigated the adoption of Big Data Analytics in companies from different sectors: human resources, finance, marketing, and retail. Like other frameworks, UTAUT is adaptable, and in this study, it worked with two added new factors: perceived risk, and resistance to use. Expanding UTAUT's applicability, Cabrera-Sánchez and Villarejo-Ramos (2020) added a new adoption factor to the original framework: opportunity cost. In that study, the investigation core was the adoption process of Big Data in 199 Spanish service companies. Silva et al. (2019), applying UTAUT, concluded that the intention to adopt Big Data in Peruvian companies is influenced by Performance Expectancy, Social Influence, and Facilitating Conditions. This framework represents a resilient and adaptable model, and it is well-explored to understand the ABD process at the individual level.

TRA, TPB, and DeLone and McLean IS Success Model (DMISM)

These three frameworks are present in two articles included in the LR. Firstly, Kohli and Tan (2016) studied the adoption of electronic health records by healthcare companies. In this

research they verified that the knowledge developed with the use of DeLone and McLean IS Success Model supported the innovation process. Secondly, Theory of Reasonable Action and Theory of Planned Behaviour are two frameworks relevant to be used in studies involving data applications such as blockchain adoption (Wong et al., 2020). There is no strict link between those three frameworks and ABD.

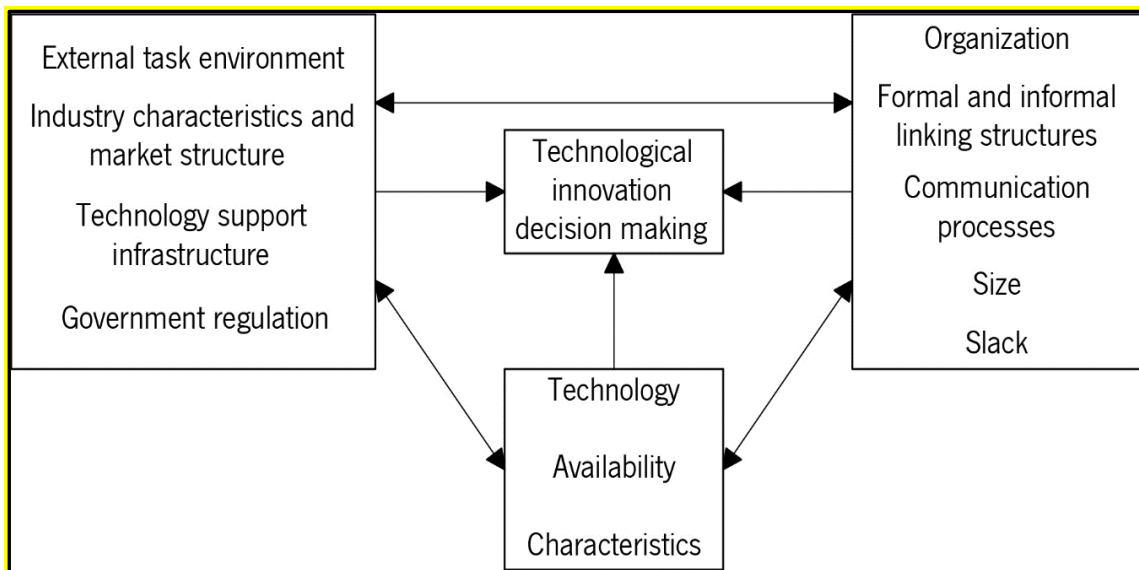
Technology-Organization-Environment Model (TOE)

In our literature review (LR), 14 references (16%) use TOE as the principal framework to study ABD. Salleh et al. (2015) remarked the power of TOE is clustered in its factor analysis using three analytical dimensions: technological, organizational, and environmental. Additionally, Salleh and Janczewski (2016) highlighted TOE's capacity to explain an adoption element, such as IT security, under internal and external perspectives, including factors such as top management support, organisational culture, and IT complexity.

Considering the status of the National Statistics Offices, and their organizational network structure, it is critical to adopt a model that considers the organizational analytical level. Following this logic, and considering the characteristics of the reviewed frameworks, the TOE model is the selected framework to be used in this doctoral research. In the following subsection, TOE model's origins, features, and factors are discussed with the focus on ABD.

2.3. The Technology-Organization-Environment model (TOE)

The TOE model encompasses three dimensions: technology, organization and environment (DePietro et al., 1990), as presented in Figure 8. This triple analytical dimensions represent an advantage when TOE is compared with other frameworks, such as DOI, with only two analytical dimensions (Park and Kim 2019; Lai et al. 2018b). Organizational elements, such as culture, skills, company size, or internal resources (slack) are considered in this framework. In this perspective, TOE analyses also the economic and business environment and evaluates how those external factors affect the adoption process. About the Environment dimension, industry characteristics, the external stakeholders, national economic scenario, national business environment, and public regulation are considered relevant factors to influence the adoption process (Baker, 2012).



Legend: Original version of TOE framework, developed by DePietro, et al (1990).
Source: Tornatzky and Fleischer (1990); Baker (2012).

Figure 8 – TOE Framework: original graphic structure

The seminal publication about the TOE framework took place thirty-two years ago, by DePietro et al. (1990). Previously, Rogers (1962), Fishbein and Ajzen (1977), Ajzen (1991), Davis (1986), and Bandura (1985) approached the theme of technological innovation by considering the individual and the technological dimensions. In order to bring the external influence to technological adoption, DePietro et al. (1990) included the environment as the external dimension of the technological innovation process, thus completing its present configuration of the three delimiting contexts of innovation decisions: technology dimension, organization dimension, environment dimension. Baker (2012) remarks the three dimensions offer a real picture of the organization during the adoption process.

Firstly, there is the Technology dimension, which comprises an external and an internal component. The external component is where the IT adoption is analysed under the perspective of technology availability, demanded methods, risks and costs. The internal component looks at decision-making capacities, and current assets (DePietro *et al.*, 1990). This availability means the technological assets that need to be acquired by the organization from external suppliers. It includes licenses and copywriting, number of suppliers able to offer IT solutions, and the costs of the acquisition. Inside the firm, there is another set of adoption factors related to the internal technological conjuncture (the central box in Figure 8). In other words, an organization must consider the current state of its technological equipment. Additionally, the complexity of the IT adoption process must be examined. Complexity refers to the incremental or discontinuous

changes involved (Tushman and Nadler, 1986). Indeed, the changes associated with technology adoption can provoke competence-destroying and unclear scenarios (Tushman and Anderson, 1986). Another factor of discontinuity is the methods necessary to adopt the IT solution. As an example, a National Statistics Office can be obligated to adopt new data analysis methodologies if it adopts a Big Data solution, which can impact the way it carries out Official Statistics surveys.

Lai, *et al.* (2018b, p. 680) signal that the “*technological context focuses on the characteristics related to the technology itself that may have a positive or negative impact on the decision-making process.*” This ambivalence about positive or negative technological impacts can create uncertainty in the organizations. Complexity arises also with the demand for workflow changes. Those changes represent risks by disrupting processes and systems provoked by the IT adoption. Regarding risk and uncertainty, DePietro *et al.* (1990) indicate that incremental adoption is more feasible in scenarios where massive productions are the business core.

Thus, the Technology dimension arranges internal and external factors related to the organization's daily-base operations. Inside the firm, factors such as technological equipment, decision-making processes, destroying competences, and process features (massive or non-massive) must be considered before the IT adoption. From the external context, technological availability, required methods, cost, and suppliers are the set of factors to be evaluated by the adopters.

Secondly, the Organization dimension represents a portrayal of the firm's condition during the adoption process. DePietro *et al.* (1990, p. 154) define this dimension as the output of planned and unplanned interactions between the organizational members. Those interactions work in formal and informal structures. The set of processes, the centralization level and the hierarchical design are elements of the formal structure. On the other hand, the informal structure is a summary of the informal links between co-workers, culture, and collective behaviours. In this way, formal and informal structures can be boosters or blockers of the IT adoption (DePietro *et al.*, 1990). The same adoption is reinforced by the complexity of the organization and limited by the formality and centralization. However, not only structures play a determinant role in the organizational dimension. Subscribing this assertion, Tushman and Nadler (1986) remark that top management support is a key factor to advance with the adoption into an organization. Additionally, these two authors consider also project readiness, communication networks and

problem-solving practices as organizational factors. The set of formal and informal factors are complementary in the successful IT adoption process (DePietro *et al.*, 1990).

Therefore, the Organization dimension highlights the internal conjuncture of the company (DePietro *et al.* 1990). This conjuncture is summarized in organizational factors. In this way, some of them present descriptive measurements and others qualitative insights. Size (number of employees), slack (budget or revenues) and scale (amount of work dedicated to achieving an output) are examples of quantitative organizational factors. In turn, top management support, project readiness, communication networks, problems-solving practices, culture, hierarchical structure, principal stakeholders, and nature (public or private) integrate the set of qualitative organizational factors.

It is necessary to highlighted the role of education and training (Tushman and Nadler, 1986) as a factor capable of changing culture, behaviours, and develop skills. It can be evaluated in a hybrid way. Quantitatively, the number of trained employees can be considered. Under a qualitative approach, training plans together with organizational coverage (areas and departments) can identify if the adoption process is widespread or focused. Tushman and Nadler (1986) consider this large coverage as cross-section trainings.

Thirdly, there is the Environment dimension, the main innovation of the TOE model (Lai *et al.*, 2018a). DePietro *et al.* (1990) consider the environment as the arena where firms search for IT assets, identify the adoption possibilities, and acquire the adopted resources. The authors point out the core factors that compose this dimension: industry features, competitors, government, and regulatory system. The Environment dimension has the function of delimiting the boundaries between the organization and its surrounding area. However, sometimes this delimitation is part of a grey zone, where internal and external actors work in both spaces. DePietro *et al.* (1990) explain that those actors bring and leave knowledge, skills, and experiences from and to the environment, such as project teams, or consulting companies. In this exchange, the quality and availability of external assets play a fundamental role in the adoption of new technologies.

Other environmental items integrate a plethora of factors. Market and Industry Structure, added by uncertainty and volatility can be a booster or a constraint to adoption (DePietro *et al.*, 1990). Not only structure, but also the competition level of the industry, reinforces the adoption of new technologies. In his turn, Baker (2012) added the value chain where the organization is inserted as a relevant environmental factor. Beyond those examples, DePietro *et al.* (1990) defined a

subset of the environmental factors, labelled as technology support infrastructure (TSI), representing the technology availability in the environment.

Labour costs, skills available in the labour market, and the existing technology suppliers are the components of the Environment dimension. The study of McGuire (2017) reinforces the relevance of the skills available in the labour market. She researched employees in England, Scotland, and Wales, considering their capacity to deal with statistical and Big Data tasks. The result remarked that the labour market is determinant to ABD when professionals with statistical skills are available to be hired in this same market.

In addition to the three-dimensional consideration of the factors influencing IT adoption, the TOE framework presents a suitable design to explain innovation adoption at the organizational level (Baker, 2012). Its graphic structure (Figure 7) explains the linkages between internal and external adoption factors.

By including the influence of environmental factors, the TOE model is considered a derivation form of Institutional Theory (Sun et al., 2018). Its overarching analytical capacity, with the three-dimensional factor-sets, includes a strong feature of a knowledge-synthesis vehicle (Cronin and George, 2020). This robustness makes Lai *et al.* (2018b, p. 680) attribute TOE the status of "*the most widely used and prominent method to conduct research on the technology adoption*".

In summary, TOE is a robust framework to study new technology adoption. In the next chapter, ABD is discussed under the lens of TOE, based on the references obtained in the LR.

Chapter 3 – TOE as a framework for the adoption of Big Data

The selection of TOE as the analytical framework to study the adoption of Big Data (ABD) is based on the systematic literature review discussed in section 2.1. Fourteen studies use TOE to understand ABD processes. This model presents a convergence of different authors about the feasibility of studying ABD at the organisational level, not at the individual one.

A framework has to be feasible and related to the studied issue/research problem. In this perspective, the TOE framework receives the adjudication of the authors. Baig *et al.* (2019) remark that TOE is the most feasible framework to study ABD, considering its three dimensions: Organisation, Technology, and Environment. From the standpoint of Ram *et al.* (2019): “*Given its multi-context coverage and inclusiveness, TOE provided a robust theoretical underpinning to conceptualise various factors that drive the adoption of Big Data....*” Lai *et al.* (2018b) understand TOE is valuable and appropriate to study ABD because it examines “*the multiple-level determinants in deciding the use of new information technology*”. In turn, Park and Kim (2019) emphasise that TOE has the advantage of focusing on internal and external contexts during the ABD processes launched by organisations.

Another strength of TOE is the possibility of associating its analytical capacity with other frameworks. Verma and Bhattacharyya (2017) applied TOE in association with the Perceived Strategic Value (PSV) construct to study ABD in Indian firms. Those authors listed three positive features of TOE in technological solutions adoption, such as Big Data. Firstly, TOE differentiates the adoption motivations from the solution qualities to the organisational capabilities. Second, the evolutionary path from the DOI framework (Rogers, 1983, 2004; Sun *et al.*, 2018) to TOE improves this last model with added resilience and adaptability. Thirdly, the TOE framework assesses the adoption of technological innovations in distinct economic sectors and companies. As examples of this last feature, studies are using this framework to explain the ABD processes in different types of contexts: services and manufacturing (Angwar, 2018; Park and Kim, 2019), B2B (Sun *et al.*, 2020), and construction (Ram *et al.*, 2019).

In addition, Bremser *et al.* (2017) verified that TOE supports the study of ABD in ten cases. This set of organisations verified that the first group of TOE factors looks at strategic drivers of ABD inside companies. The other group of factors examines ABD at the operational level, not at the

strategic one. Besides its three dimensions, TOE can also be applied from a temporal perspective, from the strategic decision until the deploying effort of adoption. Another example of the temporal application is the research developed by Bremser (2018), where the ABD agenda building is studied using TOE factors. *Yadegaridehkordi et al.* (2018b) based their research on TOE's three dimensions to extract sets of factors that influence ABD. The authors used those factors in a Multi-Criteria Decision Model (MCDM) to identify their influence on ABD in manufacturing companies from the top managers' perspective. Indeed, the coverage of strategic and functional drivers, the temporal perspective (from strategic planning to adoption), and the multilevel analytical capability (C-Level, Management-Level,...) underscore the suitability of TOE to analyse ABD processes.

Concerning the framework's flexibility, TOE is expandable. Angwar (2018) studied ABD from the organisational perspective and added a fourth dimension to the construct: security. The author added the Security dimension for this reason *"(...) the integrity of Big Data is always in question, and with that, security"* (p. 8). She called this variant TOES, a technology-organisation-environment-security framework. The study by Angwar (2018) shows how TOE is flexible enough to be used in ABD research, considering this can take place in the manufacturing and services sectors.

Regarding technological innovation, the TOE model proposes three different types: incremental, radical, and synthetic (Baker, 2012). All of them may result in the reinforcement or destruction of competencies, and they can impact the culture, processes, and the workers inside the organisation, in a hard or soft way. This possibility of understanding the adoption process under three different types provokes significant organisational changes to represent a model's strength. Considering the organisational perspective of this research, this strength is a criterion for using the TOE model.

In our research, the TOE model is considered relevant in its three dimensions: Technology, Organisation, and Environment. The first dimension refers to the features of Big Data as a technological innovation. Nowadays, ABD represents a technological innovation to NSOs, especially in transitioning traditional survey techniques to new alternatives, such as using web scraping in the consumer index price survey (Statistics Canada, 2020). Hence, TOE is the robust framework for studying ABD in technological change environments, such as NSOs' operations.

Internal resources, culture and skills are relevant variables during the ABD process in the organisational dimension. In addition, the environmental dimension investigates the impacts of the industry features vis-à-vis the new technology. In this thesis, the Global Statistical System represents the industry.

Thus, it is necessary to clarify how TOE is deployed in this research. Regarding the framework, it uses factors to analyse the ABD processes in the three dimensions (Sun et al., 2020). Following this analytical perspective, this thesis analysed, in a structured procedure, how many factors the fourteen studies used to understand ABD and the applicability of those factors in ABD in the National Statistics Offices (NSOs) environment.

3.1. Factors of the adoption of Big Data

The application of the TOE model involves using its technological, organisational, and environmental factors to understand the research object: technology adoption by an organisation. Since the first publication on TOE (DePietro et al., 1990), the number of factors varied in function of the research objects.

Considering this is an adjudicative research (Cronin and George, 2020), it is necessary to check the definitions of all factors. In sequence, the organisation of similar names with similar definitions must be done. Finally, a convergent label to identify each factor is used. This procedure reduces the noise related to the factors set and allows the researcher to develop an entire code book (Saldaña, 2013) for the qualitative analysis.

So, as an outcome of the literature review, the twenty references in Table 8 presented different TOE factors. Many of those factors use different names to express the same meaning. In this section, we review and consolidate all the factors influencing ABD covered by those twenty references.

Table 8 – TOE factors used to study the Adoption of Big Data

Authors	Title	Journal /Conference Title
Salleh <i>et al.</i> , 2015	SEC-TOE Framework: Exploring Security Determinants in Big Data Solutions Adoption	PACIS 2015 Proceedings - Pacific Asia Conference on Information Systems (PACIS)
Salleh and Janczewski, 2016	Technological, organisational, and environmental security and privacy issues of Big Data: A literature review.	Procedia Computer Science

Authors	Title	Journal /Conference Title
Verma and Bhattacharyya, 2017b	Perceived strategic value-based adoption of Big Data Analytics in emerging economy A qualitative approach for Indian firms	Journal of Enterprising Information Management
Gangwar, 2018	Understanding the Determinants of Big Data Adoption in India: An Analysis of the Manufacturing and Services Sectors	Information Resources Management Journal
Sun et al., 2018b	Understanding the Factors Affecting the Organizational Adoption of Big Data	Journal of Computer Information Systems
Yadegaridehkordi et al., 2018	Influence of Big Data adoption on manufacturing companies' performance: An integrated DEMATEL-ANFIS approach	Technological Forecasting and Social Change
Lai et al., 2018a	Understanding the determinants of Big Data analytics (BDA) adoption in logistics and supply chain management: An empirical investigation	International Journal of Logistics Management
Ram et al., 2019	Adoption of Big Data analytics in construction: development of a conceptual model.	Built Environment Project and Asset Management
Park and Kim, 2019	Factors Activating Big Data Adoption by Korean	Journal of Computer Information Systems
Baig et al., 2019	Big Data adoption: State of the art and research challenges	Information Processing and Management
Sun et al., 2020	Organisational intention to adopt big data in the B2B context: An integrated view	Industrial Marketing Management
Picoto et al., 2021	The influence of the technology-organisation-environment framework and strategic orientation on cloud computing use, enterprise mobility, and performance.	Revista Brasileira de Gestão de Negócios
Bakici et al., 2022	Big Data Adoption in Project Management: Insights from French Organisations	IEEE Transactions on Engineering Management
Al Hadwer et al., 2021	A systematic review of organisational factors impacting cloud-based technology adoption	Internet of Things

Authors	Title	Journal /Conference Title
	using Technology-organization-environment framework	
Paradza and Daramola, 2021	Business Intelligence and Business Value in Organisations: A Systematic Literature Review	Sustainability
Baig et al., 2021	A Model for Decision-Makers' Adoption of Big Data in the Education Sector	Sustainability
El-Haddadeh et al., 2021	Value creation for realising the sustainable development goals: Fostering organisational adoption of big data analytics.	Journal of Business Research
Ghaleb et al., 2021	The Assessment of Big Data Adoption Readiness with a Technology–Organisation–Environment Framework: A Perspective towards Healthcare Employees	Sustainability
Maroufkhani et al., 2022	Determinants of Big Data analytics adoption in small and medium-sized enterprises (SMEs)	Industrial Management & Data Systems

As presented in Table 9, the number of factors in the references reaches a total of 234 items! This tremendous and diverse number of factors constitutes a barrier to research advance in this area. Considering the epistemological proposal of Cronin and George (2020), 234 factors were submitted to check for similarity.

Table 9 – Factors per dimensions in Adoption of Big Data studies

Dimensions	Number of Factors
Technology	99
Organisation	77
Environment	58
Total	234

In an attempt to reduce the semantic noise, the factor description of each factor was checked and compared using a convergence criterion. If the factor's descriptions present similarities and the text approaches the same theme, this is considered a factor with high convergence. In the opposite direction, a factor with discrepant descriptions is considered a low convergence factor. Following this thought, the research proposed rearranging the factors based on the respective

descriptions. In the end, high convergence factors kept the original name. Low convergence factors were joined under a wide common name to allow the reader a better understanding of the meaning.

There are factors with close nomenclature. An example is the organisational factor, named “top management support”, which presented similar names but similar descriptions, as shown in Table 10. This factor counted fourteen definitions and three non-definitions. Much more than an accounting issue, different names for the same factor indicate the necessity to check all descriptions and verify whether they are part of a thematic group.

Table 10 – Polysemy in TOE factors nomenclature

Authors	Authors' Factor Description	Same factor, different labels
Sun <i>et al.</i> (2018, p. 198)	Managers are willing to allocate sufficient resources and encourage the initiative adoption of Big Data (e.g., top executives responsible for data management, CIOs' willingness to adopt Big Data).	Management Support
Yadegaridehkordi <i>et al.</i> (2018, p. 201)	The authors gave no definition.	Management Support
Park and Kim (2019, p. 3)	Management support refers to the degree to which management recognises the importance and relevance of Big Data.	Management Support for Big Data
El-Haddadeh <i>et al.</i> (2021)	Top management support explains the extent to which top management understands the importance of innovation and the extent to which it is involved in new innovative activities, such as ABD	Top Management Support
Gangwar (2018, p. 7)	Top management support refers to the degree to which top management understands the strategic importance of IS innovation and the extent to which it is involved in IS activities.	Top Management Support
Verma and Bhattacharya (2017, p. 361)	Referred to devoting time to the (IS) program in proportion to its cost and potential, reviewing plans, following up on results and facilitating the management problems involved	Top Management Support

Authors	Authors' Factor Description	Same factor, different labels
	with integrating ICT with the management process of the business (p. 361)	
Baig <i>et al.</i> (2019)	Management support is a degree to which it perceives the significance and applicability of Big Data adoption (Tornatzky et al., 1990). Usually, top management does not support the 'change' factor. In big data, adoption process changes are required when top management becomes reluctant to change for improvement, then the whole organisation starts following higher management decisions, which will linger the adoption process. Management support is necessary to encompass the organisational policies, rules, handling of data storage issues, technological and financial capability. Therefore, management support is a significant factor that contributes to the adoption of Big Data.	Management Support for Big Data
Bakici <i>et al.</i> (2021, p. 12)	Commitment, availability of data, skilled human Resources, training.	Top Management and Strategy
Lai <i>et al.</i> (2018, p. 361)	The degree to which top management understands the importance of the IS function and the extent to which it is involved in IS activities.	Technology Management Support
Ghaleb et al. (2021, p.12)	Refers to the extent to which managers grasp and embrace a new technology system's technological potential.	Top Management Support
Paradza and Daramola (2021)	No definition by the authors	Top Management Support
Al Hadwer <i>et al.</i> (2021)	No definition by the authors	Top Management Support

On the opposite side, an example of a low convergence name is the environmental factor "Security and privacy regulations", as presented in Table 11. In this case, four different definitions, two non-definitions and six different names integrate the factor's conceptualisation. "Security and privacy regulations" is a factor related to the plethora of laws and acts to regulate

one specific issue: security and privacy. Clients' data, collected and processed by a company's Big Data application, must have privacy protection from any leak or irregular dissemination.

Table 11 – Low convergence factor: the example of “Security and privacy regulations”

Authors	Authors' Factor Description	The same factor, different names
Sun; Cegielski; Jia; and Hall. (2018, p. 198)	Data collection from individuals causes individuals' security, privacy concerns. (e.g., legal implications of collecting customers' private information).	Security, Privacy and ethical concerns in collecting data.
Salleh; Janczewski; and Beltran. (2015)	Security and privacy regulatory concerns refer to organisational concerns in ensuring compliance to security and data privacy regulations	Security and Privacy Regulatory Concerns
Baig; Shuib; and Yadegaridehkordi. (2019, p. 10)	Security, privacy and risk factors will not only affect the big data adoption process but will also destroy the company reputations.	Security, Privacy and Risk
Baig; Shuib; and Yadegaridehkordi. (2021, p.7)	Security and privacy concerns are associated with individual safety and moral values. Security and privacy concerns were the obstructing factors of ABD.	Security and Privacy Concerns
Salleh and Janczewski. (2016)	No definition by the authors	Privacy Regulatory Concerns
Al Hadwer; Tavana; Gillis; and Rezania. (2021)	No definition by the authors	Cloud Locality

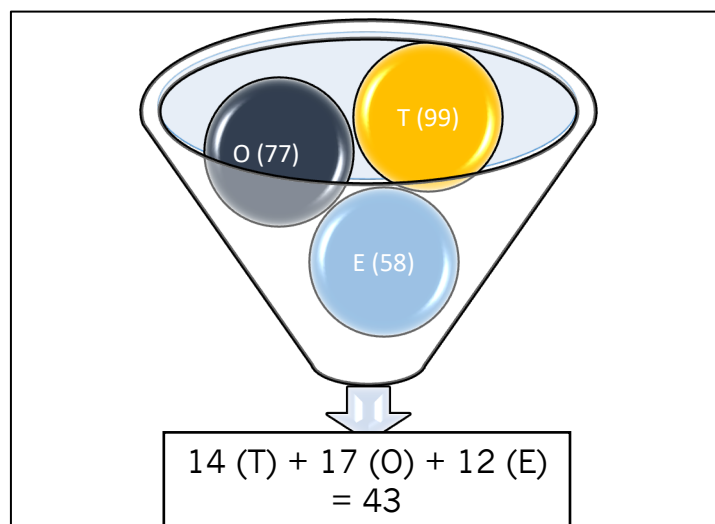
The factors' polysemy is also present in the Technology Dimension but to a lower extent. For example, the factors “Complexity” and “Compatibility” have nineteen and fourteen references. Besides the vast number of citations, it is possible to assert that both names define specific concepts. Table 12 brings three definitions of “Compatibility”. They all consider how a Big Data solution is compatible or incompatible with the current organisational processes and systems.

In this factors analysis, an association to a familiar name took place regarding the redundancies and the convergences of each factor's descriptions. The reduction of factors was drastic, and it was possible to gather several factors in a shorter-term/definition group. Beyond the polysemic reduction, the other outcome for future TOE studies is the possibility of using a set of references proposed by this literature review.

Table 12 – Technological factors description: the example of “Compatibility.”

<u>Compatibility means...</u>
<i>"The characteristics of big data are perceived as being consistent with the existing IT architecture in an organisation (e.g., scalability, integration into the existing information systems). (p. 198)"</i> Sun, Cegielski, Jia & Hall (2018)
<i>"The concept of compatibility refers to the level at which innovation is perceived as consistent with internal organisational processes and information systems. (p. 283)"</i> Picoto, Crespo & Carvalho (2021)
<i>"Compatibility examines "the degree to which a new system is consistent with the current system within the company."</i> Maroufkhani, Iranmanesh, & Ghobakhloo (2020)

Summarising, Figure 9 shows the procedure of the assembling factors. From the original number, 234 (99 Technological, 77 Organisational, and 58 Environmental), to the final, the reduced set presented 43 factors. This adjudication in the factors promoted the capability to analyse ABD with a consistent set of factors. This consistency comes from (i) the sources, as relevant articles, (ii) the description associated with the noise reduction/elimination, and (iii) the evaluation of which factor describe ABD in organisations.



Legend: The reduction of the TOE factors presented in the literature about Adoption of Big Data. The original 99 technological, 77 organisational, and 58 environmental factors were refined to 14 technological, 17 organisational, and 12 environmental factors, after eliminating redundant meanings.

Figure 9 – Aggregating and simplifying the TOE factors.

This section presents an extensive list of the factors and their descriptions. Those descriptions and the factors' names are detailed in Tables 11, 12 and 13. It is necessary to highlight that those tables result from a synthesis, as described before, including names and definitions.

Hence, the final factors set is the base to analyse ABD under the TOE's perspective. Following this interpretation, the research's analytical lens refers to the 43 factors. In this way, the three dimensions are considered subsets of factors, ordering groups of Technological, Organisational, and Environmental factors. Regarding those selection criteria, the result is detailed in the following three subsections regarding the TOE dimensions.

3.1.1. Technology factors

The Technology dimension kept the totality of the fourteen reviewed factors. There is a reason to support all factors: Big Data is not a single technological application. As cited at the beginning of this chapter, it presents a combination of technologies, such as AI, IoT, Cloud, and Web Scraping. In this perspective, an extensive list of factors allows the content analysis to capture details of ABD in NSOs. So, the technological factors used in this research are those presented and defined in Table 13.

Table 13 – Technology factors

Name	Authors	Factor Description
Compatibility	Picoto et al. (2021, p. 283)	"The concept of compatibility refers to the level at which innovation is perceived as consistent with internal organisational processes and information systems."
Data quality	Lai et al. (2018a, p. 682)	Data quality means the degree to which the data needed for Big Data are accessible, consistent, and complete.
Data integration	Park and Kim (2021)	Data integration refers to the degree to which the data is analysed by the Big Data systems, considering the relevance of Big Data applications and the integration of those with the data collected.
Complexity	Sun et al. (2018, p. 198)	Complexity refers to "the characteristics of Big Data that are perceived as being difficult to understand and use (e.g., the difficulty of learning related knowledge for employees who will use Big Data applications),"
Perceived benefits	Park and Kim (2019, p.2)	Perceived benefits refers to "the degree to which firms perceive benefits (such as cost reduction, operation improvement, and marketing performance) deriving from Big Data."

Name	Authors	Factor Description
Technology resources	El-Haddadeh et al. (2021)	Technology resources represent tangible and intangible resources such as hardware, software, human resources, skills, and experience that an organisation acquires for implementing Big Data.
Security and privacy	Park and Kim (2021, p. 287)	Security and privacy are associated with privacy invasions and security risks posed by Big Data.
Relative advantage	Sun et al. (2018, p.198)	Relative advantage means "the characteristics of Big Data are perceived as being better than those of the idea it supersedes (e.g., its unique role for innovation, competition, productivity, customer value creation, and good business problem solution)."
Trialability	Baig et al. (2019, p. 10)	"Trialability is the degree to which companies can experiment with Big Data before fully implementing or committing to Big Data adoption."
Technology competence	Sun et al. (2018, p.198)	Technology competence means the "organization has sufficient internal IT expertise and technological infrastructure to adopt Big Data (e.g., IT knowledge and skills within the organisation)."
Cost of adoption	Baig et al. (2019), Sun et al. (2018b), Bakici et al. (2022)	Cost of adoption refers to the financial resources demanded and executed expenses related to Big Data adoption (e.g., the costs of using Big Data technology, the significant initial investment required to embrace ABD).
Accuracy	Based on Baig et al. (2019), Yadegaridehkordi et al. (2018a), and Bakici et al. (2022).	Accuracy refers to ability to overcome the data collection and evaluation limitations due to sampling, allowing the organisation to deploy faster, predictable, accountable and truthful outcomes.
Internal versus external technologies	Baig et al. (2019, p. 9)	Internal versus external technologies refers to "the hardware and software technologies used for Big Data adoption. Technologies revealed by retailers are internal technologies, whereas those provided by vendors are called external technologies."
Vendor support	Baig et al. (2019)	Vendor support can be defined as the services and activities to support the ABD in NSOs when the solutions are bought with commercial suppliers.

3.1.2. Organisation factors

Orlikowski (2000) highlighted that technology influences the organisation and vice-versa, limiting or spreading the adoption into the departments and areas. In this way, Organisational the selection of factors was less straightforward than technological ones. Firstly, two reviewed factors were discarded. "System perspective" did not present a description. "Optimism" presented a description indicating this is not a factor but an output from ABD.

Secondly, the most cited factor, "top management support," indicated a divergence point. Some authors indicate that management support is widespread, from line management to the top level. Others consider this factor exclusively from the strategic layer of a company. Despite this double approach, after examining the descriptions, it was concluded that this is the same factor. However, this indication will be checked during the interview analysis and define if this is indeed a single factor, or two conflated under one name.

Thirdly, some factor labels are misleading and could lead to a misinterpretation of their real meaning. "IS strategy orientation" and "Technological capabilities" are organisational factors. However, they could be understood as technological factors. As a strategic orientation, "IS strategy orientation" integrates the strategic planning of each company. "Technological capabilities" is closer to employees than computers. It means the skills of the whole organisation and the employees to deal with the adoption of Big Data. This challenge includes the organisational impacts, such as processes and methodologies. Another issue of the Organisational factors is the word "readiness". To avoid noise, this word, in this study, means "*willingness or a state of being prepared for something*". For that matter, the list of organisational factors is thus presented in Table 14.

Table 14 – Organisation factors

Name	Authors	Factor Description
Top Management support	Park and Kim (2019)	Top Management support is the degree to which high-level management recognises the importance of Big Data, sponsoring and supporting the adoption process.
Firm Size	Sun et al. (2018, p.198)	Firm size refers to "the firm's annual revenue and number of employees that could support the adoption of Big Data (e.g., leading companies with more revenue)."

¹ Cambridge Dictionary. Retrieved from <https://dictionary.cambridge.org/dictionary/english/readiness>. Accessed in 01/02/2021.

Name	Authors	Factor Description
Financial capacity	Baig et al. (2019)	Financial capacity means the skills of the staff to manage budget and financial resources in order to make available the monetary conditions for ABD.
Change efficiency and efficacy	Baig et al. (2019)	Change efficiency and efficacy refers to the idea that the employees of NSOs should be capable of handling changes related to ABD and obtaining better results from this adoption.
Business strategy orientation	Baig et al. (2019, p.9)	Business strategy orientation "refers to a strategy that is used to learn about business analytics and making strategic decisions for organisations in adopting Big Data."
Human resources and Skills	Baig et al. (2019, p.9)	"Human resources play a significant role in employing and maintaining technological innovation. Sufficient skilled human resources are required for Big Data adoption. Strong programming, statistical and analytical skills help in implementing Big Data in retail organisations."
Decision-making culture	Baig et al. (2019, p.9)	Decision-making culture "refers to as a belief that the adoption of Big Data is a key to success in enhancing organisational productivity. At the company level, executive or managerial staffs make a decision."
Technological capabilities	Park and Kim (2019, p.3)	Technological capabilities refers to "the degree to which firms have the ability to operate Big Data systems in terms of technological readiness."
Organisational structure	(2018, p. 198)	Organisational structure implies "a well-organised structure is well-suited to the adoption of Big Data (e.g., cross-organisational collaboration structure, IT departments, and staff configurations)."
Organisational learning culture	Salleh et al. (2015)	Organisational learning culture is "the learning characteristics and orientation of an organization, especially in the process of complex technology adoption"
Organisational data environment	Verma and Bhattacharyya (2017, p.361)	Organisational data environment "refers to the extent to which data resources are managed in an organisation"
Organisational readiness	Gangwar (2018, p.6)	Organisational readiness is "the extent to which financial, technological, and skilled human resources are accessible to an organisation wishing to embrace a technology."
IS strategy orientation	Sun et al. (2018, p. 198)	IS strategy orientation occurs when "the firm's IS strategy prioritises Big Data usage (e.g., information strategy, information governance policy)."

Name	Authors	Factor Description
Information security culture	Salleh et al. (2015)	Information security culture is the totality of patterns of behaviour in an organisation that contributes to the protection of information of all kinds
Project management readiness	Bakici et al. (2021)	Project management readiness refers to the organisational expertise for working with project management, portfolio of projects, project frameworks and different sizes of the projects in a project-orientated organisational structure.
Organisation Performance	Picoto et al. (2021); Yadegaridehkordi et al. (2019)	Organisation performance refers to the organisations that transform their business processes and systems to adopt Big Data applications, or in the following stage, the positive results derived from this adoption.
Improved design and execution efficiencies	Ram et al. (2019, p. 573)	Improved design and execution efficiencies means "the adoption of BD enables augmented design and execution capabilities."

3.1.3. Environment factors

The environment dimension presented thirteen factors initially. Although, one was redistributed: "Trust in Supplier". This factor description evokes the "Vendor support" concept, a Technological factor. In this way, "Trust in Supplier" is excluded from Environmental Dimension and absorbed by Organisational.

It is essential to highlight that the Environmental dimension kept features from the original TOE model (DePietro *et al.*, 1990). Those features are present in the factors' names and descriptions. "Market turbulence", "Industry features", and "Competitive pressure" are factors present in the original reference from DePietro *et al.* (1990) and remain in articles thirty-one years newer (e.g., El-Haddadeh *et al.* 2021; Baig *et al.* 2021). Table 15 presents the Environmental factors that influence ABD.

Table 15 – Environment factors

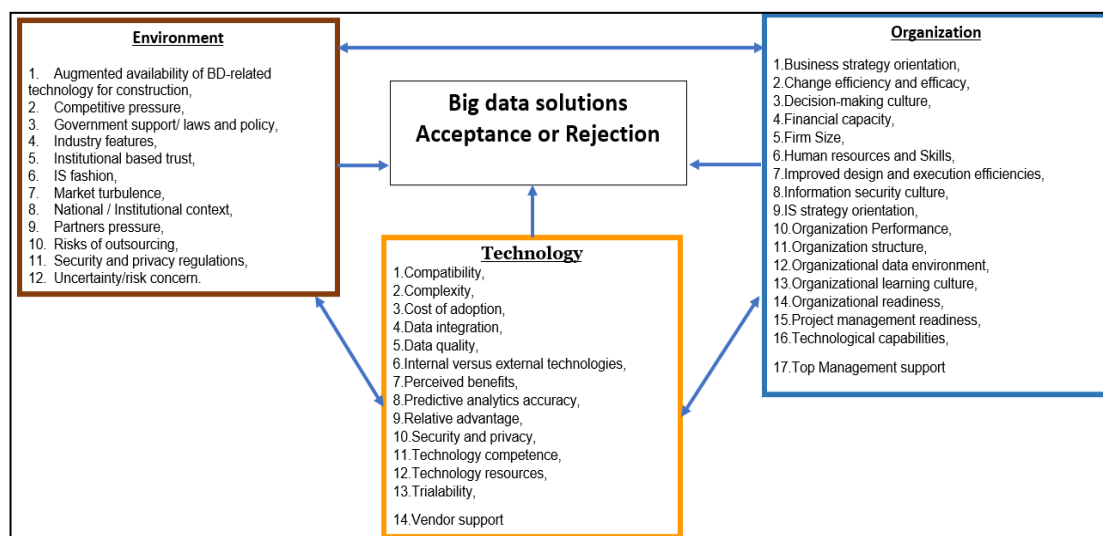
Name	Authors	Factor Description
Competitive pressure	El-Haddadeh et al. (2021, p.404)	Competitive pressure are "the actions and movements from the competitors that influence the organisation's adoption of Big Data."

Name	Authors	Factor Description
Government support, laws, and policy	Ghaleb et al. (2021)	Government support, laws and policy can be defined as institutional support or restriction for organizations adopting Big Data.
IS fashion	Sun et al. (2018, p.198)	IS fashion is relevant when "Information is obtained through external communication channels by focusing on an Organisation's peers and perceived experts such as vendors and customers (e.g., managerial interactions, social channels and trends)."
Risks of outsourcing	Baig et al. (2019, p.10)	Risks of outsourcing occurs when "hiring from outside firms to implement Big Data adoption-related services. Privacy, security, reputation, transaction, compliance, and contractual risks are associated with outsourcing."
Industry features	Verma and Bhattacharyya (2017, p.361)	Industry features refers to "the sector to which the business belongs."
Uncertainty/risk concern	Sun et al. (2018, p. 198)	Uncertainty/risk concern regards the degree to which unexpected environmental events and unpredictable occurrences can occur, that constrain Big Data adoption in organisations.
National / Institutional context	Bakici et al. (2021)	National/institutional context can be understood as the relevance of the interactions, cooperation, and institutional ties in the domestic environment, which promotes or inhibits Big Data adoption by an organisation.
Augmented availability of BD-related technology (AABDRT)	Ram et al. (2019)	AABDRT means the availability of off-the-shelf or open-source data management applications, and how this availability is important to improve the use and analyses of Big Data.
Institutional-based trust	Sun et al. (2018) and Yadegaridehkordi (2018)	Institutional-based trust means the trust that an organization has in the face of the society, considering the innovation and adoption process it uses to deliver outcomes.
Market turbulence	Sun et al. (2018), Baig (2019), and Yadegaridehkordi (2019)	Market turbulence refers to changes in citizen demands, the operational environment, and government actions with impact firms' business architecture.
Partner Pressures	Sun et al. (2018, p. 198),	Partner pressures refer to the pressure to follow partners and maintain the firm's internal balance with them (e.g., the readiness of suppliers in external collaborations).

Name	Authors	Factor Description
Security and privacy regulations	Ahmad et al. (2015)	Security and privacy regulations relates to "organisational concerns in ensuring compliance to security and data privacy regulations."

With these Environment factors, the final set of TOE factors is complete. Figure 10 shows the TOE reviewed under this literature review. Those factors were the base for the development of the qualitative methodology. From the development of the interview script until the analysis, those factors were the lens to understand how ABD runs in NSOs.

Regarding de adjudicative perspective proposed by Cronin and George (2020), this version of the TOE framework represents an evolution from the original version presented in Chapter 2. Following the data analysis, this framework may suffer a further update with the possibility of discarding factors without relevance in the context of NSOs. Alternatively, there is the possibility to include new emergent factors related to this specific context.



Legend: The TOE framework version based on the literature review about Adoption of Big Data.

Figure 10 – A revised TOE model in the context of the Adoption of Big Data

The next chapter describes how this systematic literature review supported the methodology building. Chapter 4 also explains how the cases and the interview guide were constructed.

Chapter 4 – Methodology: how the research has been built

Research without a well-defined methodology is not research. For this reason, the methodology is not only a set of data collection techniques. It is a road map to address the research question and guide the student to reach the academic objectives. Regarding this relevance, the methodology has to be adequate for the research approach (Hatch and Cunliffe, 2013). This approach also includes the philosophical perspective of the study.

Saunders et al. (2009) understand the approach as the research philosophy along four different categories: (i) Positivism, (ii) Realism, (iii) Interpretivism and (iv) Pragmatism. Those philosophies of research present specific attributes to view and understand reality (ontology), define and accept what can be considered fundamental knowledge (epistemology), the role of values in research and to the researcher (axiology) and specify how to obtain data (data collection techniques). Four philosophies composed of four attributes allow the researcher to plot them and define the best philosophical perspective to address the research question.

Following the same understanding, Fisher (2007) considers a more significant number of research philosophical approaches. Fisher's list includes Critical Realism, Standpoint Research, Action Research, Postmodernism and Hermeticism. With additional four types, this methodological construct considers that the research philosophies operate in a double-axe matrix. Ontological Realism, associated with extreme objectivity, is the set for Positivism. Other philosophies vary from Ontological Realism to Ontological Nominalism but consider subjectivity as part of the research. The main difference is the portion of subjectivity for each philosophical approach.

Previously, Burrell and Morgan (1979) adopted a similar classification to organise scientific assumptions: (i) ontological, (ii) epistemological, (iii) human nature, and (iv) methodological. Ontology is related to the meaning of the social phenomenon to be studied. Epistemology refers to the knowledge to be transmitted. In research about Human nature, the focus relays on the human perspective of social phenomenon. Methodological assumption receives influences from (i), (ii), and (iii).

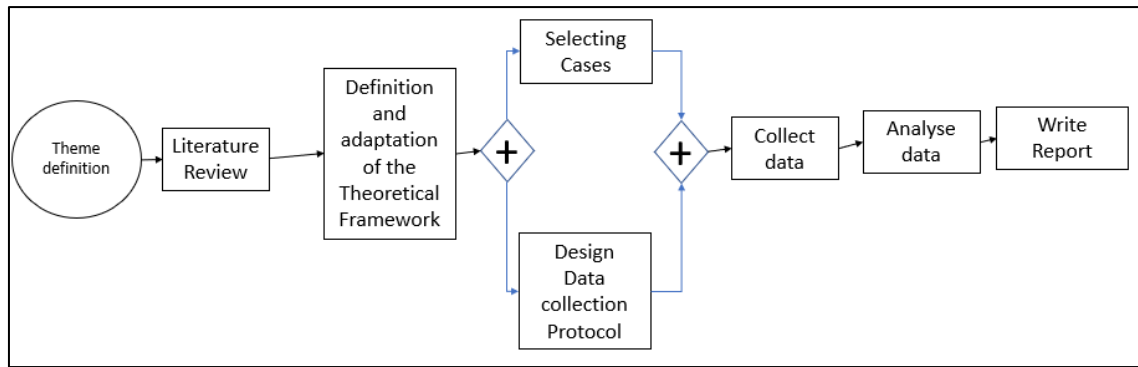
Indeed, academic research demands a transparent methodology. In the current study, the adopted philosophy is Realism. According to Fisher (2007), this philosophy is objective, with a

straightforward research question, problem, and objectives/targets. However, unlike Positivism, it considers subjectivism in the investigation process, including hidden reality layers. Another feature of Realism is the capacity to interpret the findings into categories. It is necessary to remark other Fisher's considerations about Realism. According to the author, this research philosophy allows the social scientist to verify the theory's validation in contact with a new event/phenomenon. In summary, Realism offers the possibility to generalise the study findings to similar contexts or circumstances. However, unlike Positivism, it is impossible to predict contexts or scenarios in a rigorous way.

An exploratory journey, such as this doctoral thesis, demands methods to deal with empirical scenarios. In fact, the adoption of Big Data (ABD) technologies and methodologies by National Statistics Offices (NSOs) is an almost unexplored topic. In this context, Eisenhardt's (1989) road-map for case studies can be adopted. This selection is due to the low number of academic sources about the selected topic. It is necessary to use triangulation and multiple techniques to understand the organisational phenomenon that is ABD by NSOs.

Case studies are regularly used in investigations about technology adoption and its impacts on organisational processes and routines. Moura *et al.* (2021) applied a case study strategy to understand the performance of project team members in the Information Systems domain. Trigo *et al.* (2022) used a case study to analyse how a new approach in software delivery, DevOps, arose in a large European company. Chua *et al.* (2012) focus on a specific company with the same method to verify how Social Capital could boost the performance of complex IT-Projects. Francisco and Meirelles (2016) investigated the introduction of new technologies for digital learning processes with the case study method.

More than a picture, the case study method captures the context where the researched phenomenon occurs. The method can be understood as a process with eight tasks in this perspective. From the research theme definition to the design of data collection, moving to the presentation of the results (case report), the case study can be depicted as a flow of several stages (Yin, 2014; Trigo, 2022). Figure 11 presents eight stages of the case study flow. This same sequence was used to develop the research methodology of this doctoral research. It is associated with the Eisenhardt's (1989) road map, detailed in this chapter.



Legend: The case study research flow adopted by Trigo, *et. al* (2022). Based on Trigo, Varajão e Sousa (2022).

Figure 11 – Case study Research Flow

The case study is one of the several qualitative methods, as remarked by Cresswell and Cresswell (2018, p. 292): “(a) *narrative research*, (b) *phenomenology*, (c) *grounded theory*, (d) *ethnography*, and (e) *case studies*.” Additionally, the authors highlighted that a case study is appropriate to interpret, especially “a *program, event, activity, or process*”. This interpretation is restricted to a period. About this delimitation, the current research studies a technological adoption process, during a specific period, including the demographic Census operations deployment. It is essential to highlight that the three selected NSOs launched their Census within a period of two years: INE Portugal in 2021, ONS UK in 2021, and IBGE Brazil in 2022.

4.1. Addressing the Research problem and the research question

Cresswell and Cresswell (2018) define a research problem as “*an issue or concern that needs to be addressed*”. Alves-Mazzotti and Gewandsznajder (2002) present a more specific vision, stating that the research problem must “*identify points of consensus, as well as controversy, areas of shade and gaps that should be clarified*.” The authors reinforce the importance of researching problems on which little or no research has been carried out, in the following statement: “*The production of knowledge (...) is a collective construction by the scientific community, a continuous process of searching, in which each new piece of research is inserted, complementing or contesting contributions previously made to the study of the subject.*” (p. 180).

In the same way, Noor (2008) mentions that the research problem is a determinant of the choice of methodology. From the standing point of Kothari (2004, p. 25-26), the research problem shall be built based on a group of criteria. Some of them are “(i) a *subject which is overdone should not be normally chosen, for it will be a difficult task to throw any new light in such a case*; (ii) *controversial subject should not become the choice of an average researcher*; (iii) *too narrow or*

too vague problems should be avoided. (...)”. In the face of those criteria, it is necessary to define a research problem with technique. One of this technique’s elements is the possibility to develop ideas through discussions with peers and individuals, searching for *“a position to enlighten the researcher on different aspects of the proposed project (p. 28)”*.

Additionally, Kothari (2004) proposed a validation test to check whether a research problem proposal is reliable. This validation is composed of 5-checking-elements: (a) problem owner – an individual, a group, or an organisation; (b) objective(s) to be attained; (c) alternative means/channels to attend the objective(s); (d) a portion of uncertainty about the access of the sources; (e) an environmental conjuncture where the problem takes place/occurs.

Firstly, the problem owner has been identified as the National Statistics Offices (NSOs). Next, the objectives to be attained are understanding why and how NSOs adopt or avoid Big Data solutions. Thirdly, alternative information sources will be official documents, structured interviews, and direct observations. The information access policy from each NSO represents the fourth item. One NSO, e.g., INE (Portugal), may be closer to external researchers compared to a similar IBGE (Brazil). Lastly, the environment where the problem takes place is the environment of disruptive changes, where NSOs have to adopt or avoid Big Data solutions.

Based on the analysis of Kothari's (2004) elements, the research problem has to demonstrate consistency and resilience. Besides this validation, the research problem shall be verified in a double-check review under another approach. So, for Cresswell and Cresswell (2018, p. 57), *“(...) if a concept or phenomenon needs to be explored and understood because little research has been done on it or because it involves an understudied sample, then it merits a qualitative approach. Qualitative research is especially useful when the researcher does not know the important variables to examine. This type of approach may be needed because the topic is new, the subject has never been addressed with a certain sample or group of people, and existing theories do not apply with the particular sample or group under study.”*

In this case, the research problem is examined relative to those four questions: (I) Has some research been done? (II) Is the topic new? (III) Has the subject never been addressed with a specific sample? (IV) Have no existing theories been applied to the particular sample or group under study?

The checklist proposed by Cresswell and Cresswell (2018) indicated that the current research problem is correctly addressed. The answer to question (I) is objective: Yes, but on a reduced

scale. There is no expressive number of academic works in Management and Information Systems to understand how NSOs are adopting Big Data solutions. Next, the answer to question (II) reports that the adoption of Big Data applications by NSOs is a research issue quite innovative and not widely explored. In sequence, the answer to question (III) is: no, not yet. Specifically, the TOE framework has not been commonly applied in adoption studies related to organisational networks, such as Global Statistical System. Finally, the answer to question (IV) is negative. The adoption theory has not been applied to ABD in NSOs. The theories of adoption models and pervasiveness may, but have not been applied, in an innovative perspective, to cases like NSOs.

Thus, after the double validation checking, the research problem is able to be explored and detailed into a research question. This central question is addressed to understand the actual dilemma lived by NSOs. As described in Chapter 1, until the last decade, those organisations were the primary official data generators for societies. After the information technology revolution and the spread of social media, mobile phones, ubiquitous internet, 5G speed and others, NSOs are losing prominence, as remarked by Allin (2021).

Eisenhardt (1989) places the research question definition in the first stage of a case study journey. According to the author, this choice allows the student to focus her/his efforts and develop solid constructs and measures. So, the research question has to summarise this scenario and the challenges to NSOs. Additionally, considering the case study method, the research question must be composed of “how” and “why” questions (Yin, 2014). In this way, the final version of the central question is this:

How and why do National Statistical Offices adopt or avoid Big Data technologies and methodologies?

Following this perspective, the central objective of this research is to answer the research question and explain how and why NSOs adopt or avoid Big Data solutions in their processes and business cycle.

Then, if Big Data is an innovation for NSOs and it might be adopted to offer better processes and outcomes, this is probably a context of innovation adoption and acceptance. After this initial step, it is essential, also, to verify if, when adopted, Big Data solutions become pervasive inside the NSO or remain localised as an asset used only by one or a couple of departments. In this way, multiple case studies arise as an adequate research strategy (Yin, 2014).

4.2. Why the Case studies method?

Eisenhardt (1989) remarks that the case study method is suitable for unexplored research topics. Regarding multiple case studies, initially, each of the three NSOs will be evaluated as a specific case. This specification was followed by a search for patterns in the adoption or avoidance of Big Data. The proposal is to identify convergences or divergences in the adoption processes for each NSO.

Considering the classification proposed by Yin (2014), there are six different possibilities to describe multiple case studies: (i) linear-analytic structure; (ii) comparative structure; (iii) chronological structures; (iv) theory-building structures; (v) suspense structure; (vi) un-sequence structure. It is possible to propose that linear-analytic and comparative structures are more pertinent to this study. The reason for this selection considers the exploratory nature of this research. As reported in Chapter 2, the issue of ABD by NSOs has not been explored in-depth and large scale. So, the three case studies, and the conclusions extracted from them, allow the development of comparisons and standards. Specifically, in this research, two types of analyses can be adopted related to multiple case studies: *pattern matching* and *cross-case synthesis* (Yin, 2014).

Furthermore, the literature review explains ABD as a non-standard process in different business sectors. The current research attempts to identify if ABD is similar to organisations in the same business sector, the Official Statistics. In this context, the multiple case studies with three different NSOs (IBGE, INE, ONS) can offer precise and objective understandings.

Not only the analytical possibilities but also the features sustain the selection of the method. Yin (2014) remarks case study is a method to be applied in empirical inquiries investigating a real-world phenomenon. This phenomenon occurs in a grey zone where the limits are unclear between it and the context.

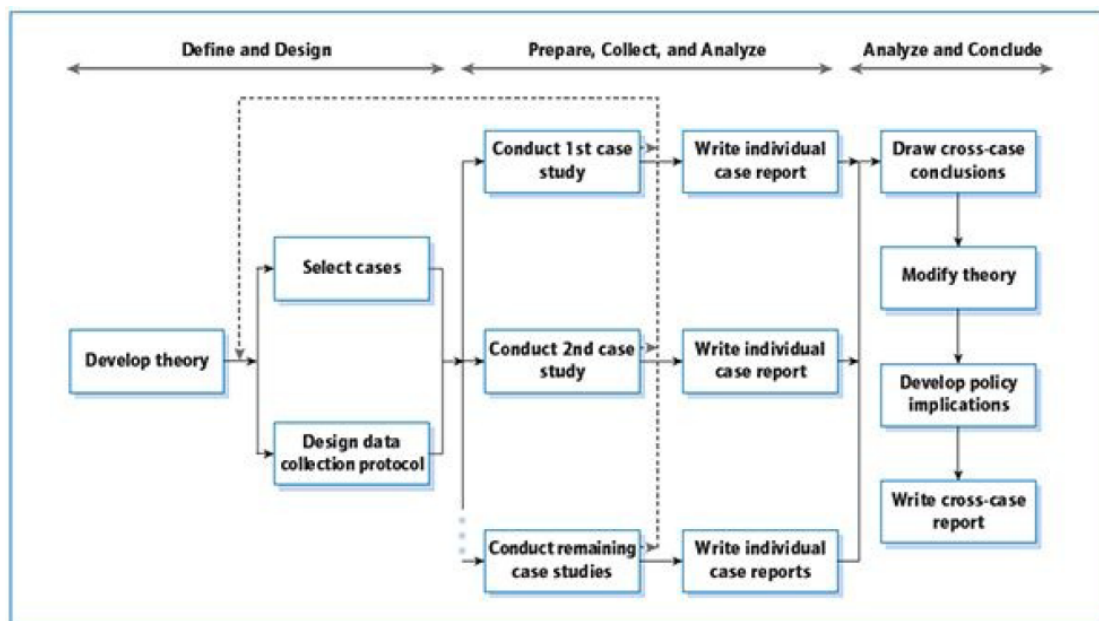
It is possible to infer that ABD is part of the context. At the same time, it is also a phenomenon, considering the dimensions of the TOE framework. In the same way, there is another strength of case study methods. It has the capacity to offer insights related to a (i) new or (ii) not-well-explored topic. ABD in NSOs fits in the topic (ii).

Regarding the method's typology, it is necessary to highlight the selected case study type. Yin (2014) defined four types of case study design. Types 1 and 2 are exclusively dedicated to single-

case designs. This research adopts a multiple-case design, precisely Yin's fourth type, the embedded multiple-case design. Explaining the selection, this type looks at a phenomenon replication, such as ABD.

The proposal of type 4 is to analyse an initial proposition and verify how it takes place in each case. In this way, the current research investigates how and why Big Data is adopted or avoided in the three NSOs (IBGE, INE, ONS). Another feature of the adopted type in this research is delimitation. Regarding the focus on ABD, the study did not spill over to all routines from the NSOs. When supporting processes and systems were studied, they were restricted to the adoption of Big Data process interface.

In summary, this research follows the same sequence proposed by Yin (2014, p. 60), as presented in Figure 12. Related to the multiple case studies, the flow of activities contains eight stages: (1) Develop theory, (2) Select cases, (3) Design data collection protocol, (4) Conduct the case studies, (5) Writing of the three individual case reports, (6) Draw cross-case conclusions, (7) Modify theory, (8) Develop policy implications.



Legend: The eight stages of multiple case studies activities flow: (1) Developing the theory, (2) Selecting the cases, (3) Designing the data collection protocol, (4) Carrying out the case studies, (5) Writing the reports on the three individual cases, (6) Drawing cross-cutting conclusions, (7) Modifying the theory, (8) Developing policy implications. Source: Yin (2014).

Figure 12 – Flow of multiple-case study designs

Those stages converge with the roadmap proposed by Eisenhardt (1989). Stake (1995) exemplifies this generalisation, citing a case study about a child's behaviour. In this example, a particular child had been studied, and the research conclusions could be extended to other

children in similar contexts. In a similar way, Yin (2014) considers the case study able to generalise findings. No statistical findings were to be tested, but theoretical and conceptual conclusions were.

4.3. Designing the research

As defined by Yin (2014), the research design includes an explanation of the selection criteria for each case, how the data collection will run, and the logical link between the research question and the data collection instrument. Each stage is detailed in the following subsections.

4.3.1. Selecting the cases

The selection of the cases follows the logic attributed by Yin (2014, p. 57): “*The logic underlying the use of multiple-case studies is the same. Each case must be carefully selected so that it either (a) predicts similar results (a literal replication) or (b) predicts contrasting results but for anticipatable reasons (a theoretical replication).*” This research attempts to achieve similar or contrasting results, looking for possible generalisations.

Stake (1995) highlights that the diversity of the attributes shall not be the sole criterion for selecting a case. Variety and balance are other features to be realised in a case study sample/selected group. As the author remarks, case selection has to offer the researcher the possibility to learn about the differences. In the same way, Eisenhardt (1989) remarks that case selection is not random but based on theory and literature review, considering a specific population. In this study, the population is the members of the Global Statistical System (GSS), the NSOs.

Under this perspective, the selection of the cases followed three criteria: organisational nature, geographical location, and position in the Global Statistical System. Those three criteria are analysed under the TOE model perspective. In fact, the differentiation from these criteria provides information that allows us to examine TOE factors in the Technology, Organisational and Environment dimensions.

1st criterion: Governmental Nature – Autonomous or Independent

This criterion is analysed to identify differences among NSOs with autonomous and independent natures. The first means a strictly formal hierarchy from a country's central political power. In sequence, the second refers to administrative independence, free from ministerial control. Additionally, the difference between organisational perspectives can support the analysis of each

case. In this way, the TOE framework remarks that, within Organisational factors, the governmental nature can be classified as centralised/formal (Baker, 2012). On the other hand, a non-governmental statute can be read as a decentralised/organic one. Those variations in the organisational nature allow the research related to factors in the three TOE dimensions.

In the Environment dimension, factors such as *national/institutional context*, *partners pressures*, and *Security and privacy regulations* can be differentiated when an NSO works under an independent orientation or in a limited one. Under the Organisational dimension, *business strategy orientation*, *decision-making culture*, and *financial capacity* can be better supported or constrained under a specific type of governmental nature. Related to the Technology dimension, factors such as *technology competence* and *technology resources* may advance or restrict the ABD. Considering that an independent NSO will probably have a more robust financial capacity to improve its competences and infrastructures, those two technological factors can influence the adoption.

2nd criterion: Geographical location

It is relevant to consider that ABD can be influenced by the business sector (Gangwar, 2018; McMahon et al., 2020) or gender (Shahbaz *et al.*, 2019). Perhaps, the national and international context can interfere with the ABD process too.

Then, the selected NSOs are organisations from three nations with asymmetrical realities. This allows us to explore the environmental factors in TOE, including the *national/institutional context*. The national context may explain or inform patterns to adopt or avoid Big Data. Competitive pressure and partner pressure are other two factors related to national asymmetries.

A country's regulation system, or the barriers to importing technology, determine the access opportunities for companies to adopt national or foreign solutions. This criterion can influence factors such as *Government support/ laws and policy*, *Human resources and Skills*, and *Vendor support*. Laws and policies vary from one country to each other. This environmental factor may support or constrain the development of the labour force and its skill. Additionally, if a country is attractive to technological companies, *vendor support* will probably be superior to a country with a closed economy.

In this way, an open economy, such as the Portuguese, can offer access to buy national or international Information and Systems Technologies (IST). In contrast, in Brazil, the import tariffs are high, and, in many cases, it is easier and cheaper to develop internal solutions.

Complementarily, when a national economy is in recession, NSOs may suffer from budget restrictions, impacting their *financial capacity*. In fact, asymmetric geographic contrasts are potential sources for this comparative research. Illustrating those differences, Table 16 shows basic figures to highlight the geographic contrast.

Table 16 – Main figures

Indexes / Numbers	Brazil	Portugal	United Kingdom
Human Development Index (2017)	0,76	0,86	0,92
Gross Domestic Product (2020-US\$ trillion)	1,75	0,20	2,89
Population in 2018 (millions)	211	10	67

*Source: Our World in Data. Retrieved from <https://ourworldindata.org/>, 2022 November 1st.

3rd criterion – Position in the Global Statistical System

NSOs respond to their countries' respective governments and the United Nations (UN) through the United Nations Statistics Division - UNSD (United Nations Statistics Division, 2019). As remarked in Chapter 1, the joint operation of NSOs with UNSD forms the Global Statistical System (GSS). In other words, the position in the GSS offers the possibility to compare the three NSOs and understand if a member of this network accesses the technology and resources more simply than others.

Even so, an NSO in a central position in GSS can participate in knowledge exchanges more intensely than a peripheral partner. This exchange flow can improve an NSO, making it able to modernise its outcomes and reflecting them in technological factors such as *compatibility*, *complexity*, *cost of adoption*, *data integration*, and *data quality*. Indeed, those five technological factors present significative levels of deployed knowledge.

On the other hand, the influence of the GSS can differ in each NSO. This potential singularity can allow comparisons in the environmental factor *industry features*. If an NSO receives more support from the GSS and, in this way, adopts Big Data faster, the *industry features* can represent a critical factor.

These organisations work as official data providers, and the results of their work processes - surveys, studies and projections - guide and support the stakeholders' decision-making processes from public and private organisations across the globe. Each selected NSO for this research plays a significant role in local societies.

Then, it is possible to use TOE to analyse innovation in organisations embedded in an organisational network. The three NSOs, INE-PT, IBGE-BR, and ONS-UK, are connected by a central node, UNSD. This is the core/brain of the Global Statistical System. On the regional scale, European NSOs work integrated into Eurostat, the statistical office of the European Union.

In summary, these three organisations, IBGE, INE and ONS, serve around 276 million citizens. The results of their work supply supranational institutions, UNSD and EUROSTAT, and the governments of 6 nations: Brazil, Portugal, England, Scotland, Wales, and Northern Ireland.

4.4. Data collection

As a stage of the research project, data collection includes deciding the collection instrument(s) regarding the selected method. Quantitative research uses specific methods, like questionnaires, while qualitative studies apply another type, such as structured interviews (Cresswell and Cresswell, 2018). Besides the type of research, data collection seeks to answer the research question (Yin, 2014). To achieve this target, the data collection methods shall be *“flexible and opportunistic”* (Eisenhardt, 1989).

The design of the data collection instrument follows the literature review and the case selection under Yin's (2014) perspective. Following this understanding, Cresswell and Cresswell (2018, p. 262) remark, *“the data collection steps include setting the boundaries for the study through sampling and recruitment, collecting information through unstructured or semi-structured observations and interviews, documents, and visual materials (...).”*

In fact, not only the interview but also the selection and collection of documents are included in this critical research stage. Yin (2014) and Cresswell and Cresswell (2018) identify the documental collection as a rich source of research information. This understanding is discussed in the following subsection.

4.4.1. Documental analysis

How and why do National Statistical Offices adopt or avoid Big Data technologies and methodologies? To achieve the objective of answering this question, data from different sources need to be collected using triangulation procedures (Yin, 2014; Eisenhardt, 1989).

In the first place, it is worth explaining why to use documental analysis to collect data about a dynamic issue such as ABD in NSOs. Bowen (2009) identifies positive and negative points about

using this collection data technique. The positives are (i) efficient method, (ii) availability, (iii) cost-effectiveness; (iv) lack of obtrusiveness and reactivity; (v) stability; (vi) exactness; (vii) coverage. The negatives are (i) insufficient detail; (ii) low retrievability; (iii) biased selectivity. The balance is positive and proves this is a well-disseminated research technique. As an initial tool for this research, the documental analysis will be used to understand and explore the three case studies and obtain support to elaborate the case study analysis.

Following this perspective, after the selection of the cases, documents were collected, exclusively in digital format, considering this research was advanced during the COVID-19 pandemic. So, the documental search is the first data collection in this methodology. Bardin (2016) considers document collection as an autonomous research source. Her content analysis technique attributes a similar relevance to documents and interviews as data sources. Documents extracted from the three organisations, including website publications, digital repositories and social media channels, were obtained. INE, IBGE, and ONS offer an amount of content in their virtual addresses. Additionally, the site from UNSD was also consulted. This source considers the directives and proposals for adopting Big Data by NSOs.

The documents were classified into four types: reports, descriptions, publications and outcomes. In the business architecture of Official Statistics (Radermacher, 2020), those four research data sources are related to the phase called “communication”, specifically in “user support”, “press office”, and “publications” segments.

Related to the first type, reports inform the state of ABD in NSOs, such as the Big Data report from The United Nations (The United Nations Statistical Commission, 2020). Descriptions involve information about the features, operations, organisational structure, history and achievements of the NSOs. For example, the history of the first Brazilian Census is available on IBGE’s website (IBGE, 2022b). The third documental source, publications, refers to the contents in the NSOs’ websites about Big Data for Official Statistics and the operations deployed by ONS, IBGE, and INE. Finally, outcomes are the descriptive material resulting from the NSOs operations. In this context of operations, the Census is the most relevant one to the studied NSOs. Considering the magnitude of the Census and its impact on NSOs, this was the more relevant operation/outcome to be observed by this researcher in the three cases. Between 2021 and 2022, those three institutions launched nationwide census operations to portray the three populations. This common period was a unique window of opportunity to realise how Big Data technologies were

adopted and deployed or, on the other hand, avoided by the NSOs. Another feature of the Census operations is the fact that it is a well-standardised operation for GSS members. This feature allows comparisons between IBGE, ONS and INE.

In studies about new technologies, Yin (2014) proposed to observe this new technology at work. So, observations can be done by taking photographs during fieldwork or collecting physical artefacts, such as an electronic device used in a technological journey. Van Campenhoudt *et al.* (2017) go beyond and consider the electronic register in the field as a variation of this data collection procedure. So, photos, videos, informal conversations with NSOs members, and screenshots of the three Censuses were elements of this documental collection. Van Campenhoudt *et al.* (2017) remark that field findings shall be organised in a grid of evidence.

It is necessary to remark that this thesis does not focus only on ABD during Census operations. Other processes and surveys are also considered, such as price index surveys, unemployment rate surveys, inflation rate, and industrial production indexes.

Thus, the documental analysis associated with the literature review supported the development of the interview script in two complementary ways. Using the documents from the NSOs, it was possible to elaborate the questions with the proposal to photograph the stages where each organisation was placed in the ABD journey: avoidance, initial adopter, medium adopter, and advanced adopter. Another relevant contribution from the documental analysis was understanding the relevant organisational roles to be interviewed. This diagnostic represented time-saving and focus gain to the research during the interviews stage.

4.4.2. Semi-structured Interviews

The semi-structured interview is a main evidence source in social research (Van Campenhoudt *et al.*, 2017). The reason for this heavy use is the instrument's flexibility, which attributes some degree of freedom to the researcher in interacting with the respondent (Van Campenhoudt *et al.*, 2017). Based on Kvale (1996), those interviews tried to take a photograph of the world of these organisations and to understand the researched phenomena, ABD.

Yin (2014) asserts that multiple case studies shall do a cross-check between each case conclusion to offer a generalisable result. In this study, beyond this, another triangulation took place. The cross-check between documents, and interviews was performed.

The interview script was less standardised than a quantitative questionnaire, taking into account the possibility of adapting it in the evolution of the research. So, the interview script, based on Kvale (1996) and Cresswell and Cresswell (2018), was flexible but still structured. The target was to understand how the interviewees realised and dealt with ABD in NSOs.

Related to the typology, the interview process proceeded according to Yin's (2014, p. 110) prescriptions: (i) following a line of inquiry, (ii) adopting an unbiased perspective, (iii) attending a friendly and non-threatening approach. Regarding the structure, also using Yin's types, the script fits in a "shorter case study interview" model, with a one-hour duration in general. This script is also open-ended, using a conversational manner, ever following the case study background (ABD, TOE framework, NSOs reality).

Regarding the questions, the structure of the interview script joined five blocks, as presented in Appendix 1. The first block asked the interviewee about her/his professional journey in the NSO and the respective achievements. In block two, the questions asked this professional about the innovations and changes she/he witnessed during this period as an NSO employee. In the sequence, the third block approached queries related to Big Data technologies and methodologies, representing a set of questions about the Technology dimension. Fourthly, block four approached topics about the state of the NSO related to factors from the Organisation dimension. Finally, the Environment block used questions about the national and international contexts, following the factors presented in the TOE framework.

It is essential to highlight that the last block of questions, the Environment one, was adapted for each national reality. Those specifications influenced factors such as *risk concern*, *risk of outsourcing*, *government support/laws and policies*, and *institutional-based trust*. In this way, questions about partnerships and external risks were adapted to each case.

In describing the interview approach, it is necessary to cite the key decision-makers to be interviewed. Each NSO was contacted with the proposal to interview five professionals from specific organisational areas. Two organisational units were defined: Presidential Office (PR) and the Census Operations Office. The area circumscribes other roles: technology, economic indices, and regional offices. In this last set, the criteria were not so restricted, considering the singularities from each NSO's structure. Those five positions in the three NSOs (IBGE-BR, INE-PT, and ONS-UK) totalised fifteen planned interviews. Alternative positions would be interviewed in cases of unavailability from one of those professionals. Summarising the interview plan, Table 17 presents

the position of the representatives from the three NSOs who participated in this research, considering the position in each NSO, background and working time.

Regarding the case names, it was necessary to reinforce them and their respective acronyms (Table 17, second column). The acronym is used to compose the “*number_NSOName*”, presented in column 3. Each interviewee code refers to a single respondent. This set of codes will be used in the data analysis, as presented in Chapter 6.

It is necessary to remark that the initially planned interviews were fifteen. However, in the Brazilian case, this quantity had to be doubled. The reasons for this extension in the IBGE’s interviews are explained still in this chapter.

In order to describe the process of accessing the NSOs, it is necessary to explain the specifications and procedures to contact, schedule and interview the professionals. Each Statistical Office had its singularities.

In the Brazilian case, IBGE, to reach the professionals, considering the researcher is an employee of this institution, he used his professional network and established contact with secretaries and staff members to schedule the interviews. In this context, the difficulty was the agenda of the interviewees. During the first semester of 2022, IBGE prepared the demographic Census operation. Professionals in top positions were very busy; sometimes, interviews were held outside office hours, such as on Sunday nights or holidays, and before office hours.

Intense contradictions emerged from the initial five interviews, showing the necessity to keep the data collection in that NSO. A new round of five interviews was executed to shed light on the contradictions between the replies and understandings of ABD in IBGE. At the end of ten interviews in different areas, it was possible to assert that there is no clear and effective redundancy in the IBGE staff about ABD or what it means for that NSO. It is important to remark that the Brazilian interviewees were very motivated to participate in the research. This behaviour was essential to the successful approach.

With support from some IBGE members, the researcher accessed the professionals from ONS-UK. This support is an example of the links between different NSOs staffs resulting from the GSS network. This was the first contact effort. It was successful, and the researcher advanced with the five planned interviews. As Van Campenhoudt *et al.* (2017) and Bardin (2016) remark, the interview looks for differences and particularities in the narratives. Like in the Brazilian case, ONS-UK was in the final stage of running a Census operation. Again, the interviewees were busy but

available to cooperate. Different from the Brazilian case, the British interviewees demonstrated redundancy in their speeches and presented the same understanding about the ABD in the ONS-UK.

Table 17 – Interviewee profiles

NSO	Acronym	Interviewee position in the NSO	Background	Tenure (years)
Office for National Statistics	ONS	Presidential Office	PhD in Finance	6
		Directorate of Methodology and Quality	MBA	5
		Data Science Campus	PhD in Artificial Intelligence and Neuroscience	5
		Directorate of Macroeconomic Statistics and Analysis	Master in Economics	7
		Training and Capabilities	PhD in Statistics	7
Instituto Nacional de Estatística	INE	President	Ph.D. in Economics	4
		Director of Methodology at Statistics	Ph.D. in Business Administration	31
		Regional Division - Oporto	Master's degree in services management	33
		Census Office	Geographics and Regional Planning	23
		Diretoria Regional da Madeira	M.Sc. In Software Engineering	16
Instituto Brasileiro de Geografia e Estatísticas	IBGE	President	PhD in Demographics	3
		Former-President	PhD in Industrial Engineering	39,5
		Directorate of Surveys	Undergraduate in Statistics	41
		Vice-Director of Technology	PhD in Innovation	38
		Management Advisor	Master in Public Administration	20
		Census Coordinator	Master in Public Administration	46
		Operational Data Manager	Master in Information System	34
		Cartography and Geodesy Coordinator	Master in Geomatics	20
		Big Data HUB Coordinator	PhD in Population, Territory and Public Statistics	20
		Corporate Education Analyst	Pedagogue/Educator	6

Finally, the third case, INE, from Portugal, was accessed in a different way. Considering the University of Minho is also in Portugal, the researcher visited INE and presented his research proposal to the members of that NSO. This action supported an initial openness to advance the

access of top-level professionals. Another critical support came from the Professors from the University of Minho. Some of them had partnerships with INE in the past. These professional ties reinforced the request for access to INE by a PhD student from that University.

Like the British case, INE members demonstrated a strong understanding of ABD for Official Statistics and how this adoption improved their internal processes. During the Portuguese journey of interviews, it was possible to realise how enthusiastic the members were with the results of the Census 2021 and the applications of Big Data technologies and methodologies.

Beyond the cases particularities, the interview followed a protocol to interact with the interviewees. After the acceptance by the interviewee, the researcher sent an e-mail or an inbox message on LinkedIn to confirm the invitation acceptance and asked for the authorisation of the interviewee to participate in this research. Just after a formalised agreement from the interviewee was the interview was scheduled. If the interview had taken place via electronic platforms (ZOOM or MS TEAMS), the researcher read the commitment protocol for research privacy at the beginning of the meeting. On the other format, in a face-to-face interview, this commitment protocol was read at the beginning, ahead of the audio recording. The privacy commitment protocol is presented in Table 18. After the protocol reading, the researcher asked the interviewee, and all of them authorised the interview records.

Table 18 – Text of the Privacy Commitment Protocol

<i>“Firstly, I inform you this interview is part of my PhD research process. It is an investigation anchored at the University of Minho in Portugal, and Presbyterian University Mackenzie, in Brazil.</i> <i>Secondly, the National Statistics Offices’ adoption of Big Data technologies is the focus, and your perspective concerned about Office for National Statistics is fundamental to this research.</i> <i>Additionally, I highlight that interview privacy is ensured. In this context, the content’s confidentiality is guided by those three legal registers: the British Data Protection Act 2018, the European Data Protection Law, and the Brazilian Data Protection Law.</i> <i>Thus, I ask your permission to record our interview during the Zoom meeting. The registers will be transcribed in an anonymised format, and the recorders will be destroyed after the PhD research conclusion. “</i>

At the period when the interviews took place, the COVID-19 pandemic was still lashing the world. In the first semester of 2022, the three NSOs adopted sanitary protocols from the home-office labour regime, including restricting access to their physical installations. So, most interviews were conducted on ZOOM or MS TEAMS platforms. When the interview was not on digital platforms, it was run physically in NSO headquarters.

In the end, the result was fruitful: 20 interviews during the first semester, with respondents from different organisational departments, respecting an equivalence of business domain (Radermacher, 2020). The result was around two-hundred transcript pages and seventeen hours of records.

Summarising, the interview journey was the most crucial data collection stage. It started on January 3rd, 2022 and finished on June 3rd, 2022. Twenty respondents, ten from IBGE and five from ONS and INE each, participated in this batch of interviews. The adopted criteria were to select respondents in top organisational roles or strategic positions related to the ABD process.

4.5. Data Analysis Techniques

This research adopted different data sources and data analysis methods. According to Van Campenhoudt et al. (2017), data analysis shall include comparing the expected and the achieved results and investigating the differences between both. Eisenhardt (1989) goes in the same way. The author remarks that using divergent techniques allows the researcher to watch for hidden evidence, unclear at a first impression. Following this orientation, the data analysis could be organised in three stages.

Firstly, the documental analysis adopted content analysis (Bardin, 2013). Beyond the frequency counting, the relevance of each key term was considered. In this technique, the direction/way refers to a positive, a negative, or a neutral influence. If a key term, such as “*skills*”, was cited in a context or omitted, there was a specific understanding to be attributed. Adverse contexts meant negative perspectives. Positive influences refer to an adoption context. The neutral influence was considered when the critical term was omitted or without any qualification, just cited. Associated with the content analysis, text mining was also used to identify words and terms repeatedly used in the documents.

Secondly, the interviews were analysed on two levels (Van Campenhoudt et al., 2017). The initial level considered the interviewee’s organisational role and interaction with other NSO areas. This

role was also associated with the dimensions of the TOE framework and its factors. Next, the second level analyses the comparison between the answers. However, this was not a simple comparison. The same question had the answer compared with other representatives of the same NSO and the other NSOs. Considering this is not a quantitative study, this approach proposed to create categories of answers, similar to a clustering arrangement. In this way, some answers represented a context of adoption and others a context of avoidance.

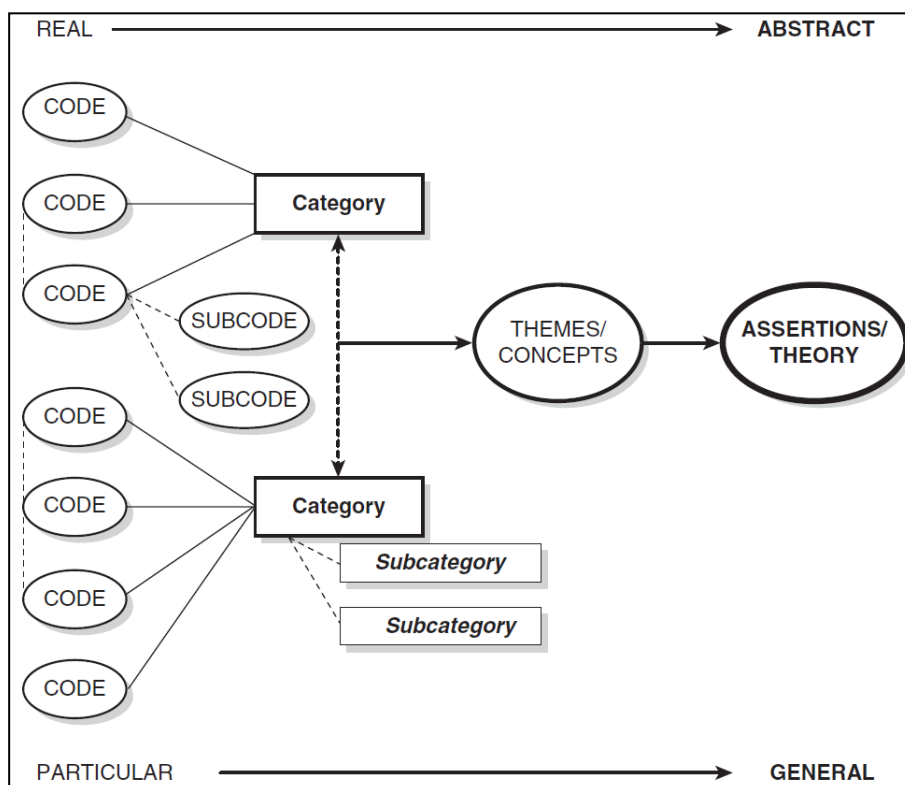
Associated with this technique, the code book analysis from Saldaña (2013) was adopted. Developing a code book to analyse the interviews, the procedure to categorise answers was clear and organised. Additionally, content analysis was used to understand the adoption process, reported from the interviews and their reactions during these interview meetings.

Thirdly, the cases reports were developed during the analysis. Yin (2014) asserts that the multiple case study method has to generate a report for each case. The set of individual case reports was joined and defined, following Yin's perspective, the cross-case conclusion. Those documents were built with the use of the NVivo software in version 12. Demanding a set of categories and codes for analysing the interviews, NVivo worked harmonically with the adopted research design. The TOE dimensions were inserted in the software as the categories, and the factors were used as codes.

The coding process is a rich analytical tool because it makes it possible to link the code with the applied theory, as presented in Figure 13. Saldaña (2013) shows how to generalise findings from the particular to the universal perspective. Regarding the reality of NSOs, the coding process allowed the application of the findings to the improvement of IBGE, INE and ONS, and other NSOs.

As an illustration, one organisational factor, *financial capacity*, was understood as a code to verify how the Big Data was adopted or avoided in an NSO, related to the factor's definition obtained in the literature review. In the example of *financial capacity*, the definition was "the skills of the staff to manage *budget and financial resources in order to make available the monetary conditions for the Big Data adoption*". During the interview analysis, the answers or comments about budget management, financial resources acquisitions, and the development of monetary conditions were identified. In sequence, the analysis advanced to identifying this code/factor in the ABD process. So, if an NSO presented a capacity to access financial resources associated with positive budgetary management, this factor contributed to ABD. On the contrary, if an NSO lost financial

resources, suffered budget cut-offs, and was procrastinating the use of Big Data, this code/factor was verified as reinforcing the avoidance of Big Data.



Legend: Saldaña's coding process. From the real/particular to the abstract/general perspective, promoting affirmations and theoretical advances based on codes and categories.
Source: Saldaña (2013).

Figure 13 – Coding Process by Saldaña (1993)

Despite the extensive effort to analyse the interviews under the lens of each factor, the result was rewarding. A robust outcome was elaborated from those analyses with the evidence of elements strictly linked with ABD for each code/factor. In fact, those analyses brought summaries from each code/factor. Again, in the example of *financial capacity*, the result of analysis understood this code/factor impacts ABD in NSO because it is (i) a requirement for the adoption of Big Data; demands (ii) budget, cost, and stakeholder management; suffer pressures for (iii) considerable investments in technology and training; and forces the NSO to keep (iv) proactively seeking funding. The adoption effort probably fails if those elements are not achieved/executed.

At the end of the analysis, a relevant result was a matrix developed based on the three TOE dimension, their respective factors and elements. This matrix allowed the research to verify if codes/factors were more critical to ABD than others. A code/factor might be irrelevant, relevant, or critically relevant. Just one factor/code presented irrelevant contributions to ABD in NSOs and was discarded, the *IS fashion*. Others were relevant but not critical, like "*improved design and*

execution efficiencies” and “*vendor support*”. Those codes/factors were identified in the answers of the interviewees, but their impacts were not significant for ABD. On the other side, if the NSO members remarked that a code/factor was massively relevant, as the *financial capacity*, it was considered critical.

Following this analytical path, the matrix added an expressive value for this research: a summary of the findings related to ABD elements in the context of NSOs, as presented in Table 19.

Table 19 – Design of the code/factors and elements Matrix

Categories	Factors/codes (FC)	Elements (EL)
Technology (T)	T_001; T_002; T_003; ...	T_FC_01 (EL01, EL02; EL03,); T_FC_02 (EL01, EL02; EL03,); T_FC_03 (EL01, EL02; EL03,); ...
Organisation (O)	O_101; O_102; O_103; O_104 Financial capacity; ...	O_FC_01 (EL01, EL02; EL03,); O_FC_02 (EL01, EL02; EL03,); O_FC_03 (EL01, EL02; EL03,); <i>(Requirement for ABD; budget, cost, and stakeholder management; huge investments in technology and training; proactively seeking funding.)</i> ...
Environment (E)	E_FC_01; E_FC_02; E_FC_03; ...	E_FC_01 (EL01, EL02; EL03,); E_FC_02 (EL01, EL02; EL03,); E_FC_03 (EL01, EL02; EL03,); ...

This matrix of findings in the case study was based on the Eisenhardt (1989) proposal. According to the author, the researcher has to shape hypotheses and enfold the literature after the data analysis. Internal validity and measurability are necessary to sharpen the construction definition, also named the shaping hypothesis (Eisenhardt, 1989). About enfolding literature, comparing the findings with the literature reinforced the internal validity of the construct, the TOE framework. An expected outcome for this thesis is the resultant TOE version for ABD in NSOs. At this moment, the last step of the case study journey (Eisenhardt, 1989) occurs: “*reaching closure*”, when theoretical saturation is realisable. After the TOE’s adapted version for ABD was built, there were improvements in the theory. Those theoretical improvements are discussed in Chapter 7, “*Results*”.

Another finding from the analysis was the influences and linkages a code/factor deploys on or receives from each other. This result showed that TOE's factors work under a web of interactions, where each relevant factor acts like a node of influence and trend. One example is the technological factor named "*technology resources*". It is influenced by factors from the three dimensions and represents the results of added actions and practices executed by the respective NSO.

Finally, it is necessary to remark that the multiple-case study method demands a generation of a repository with the research's records, files, prints, and videos (Yin, 2014). In this study, the adopted software, NVivo, followed the techniques proposed by Jackson and Bazeley (2019). Documental, interviews and Direct Observation analysis are processed by NVivo, version 12 for Microsoft Windows.

Considering the described methodology, Table 20 summarises the features of the research design adopted in this thesis. It is a proposal to figure out how ABD takes place in NSOs under an exploratory perspective, with openness for future studies.

Table 20 – Summarized Methodological Design

Approaches	Definitions
Philosophical	Realistic
Research features	Empirical and Exploratory
Method	Multiple Case Studies
Data collections	Documental research, structured interviews, direct observation
Sampling	INE-PT; IBGE-BR; ONS-UK.
Target Group	Presidents, Census managers, IT managers, and Regional Local Offices.
Data analysis	Content analysis, speech analysis and text mining
Geographic horizon	Transnational

After the detailed description of the methodology, the next chapter presents the cases descriptions based on the documental analysis. The cases presentations supported the following interview analysis.

Chapter 5 – Understanding the National Statistics Offices

This research studied the adoption of Big Data (ABD) in three National Statistics Offices (NSOs): the Office for National Statistics (ONS) in the United Kingdom, the *Instituto Nacional de Estatística* (INE) – Statistics Portugal - in Portugal, and Instituto Brasileiro de Geografia e Estatística (IBGE) – the Brazilian Institute of Geography and Statistics - in Brazil. Each organisation advanced in a singular path for adopting technologies and innovations. In the case of Big Data technologies and methodologies, the same happens. This chapter briefly presents the three NSOs' trajectories and features.

After this, these organisations' primary statistical operation, the demographic Census, is discussed under the comparison of the three NSOs. All of them had launched Census operations in the last three years, achieving different results.

Next, about the operations deployed by the NSOs, the data generated by them is presented under the structure proposed by the United Nations Statistics Division (UNSD). Official Statistics include different topics beyond economics indices. Those data, surveys, indices, and rates represent the outcomes and deliveries that NSOs present to society.

The last subsection approaches the studies about the association between Big Data and Official Statistics. Additional methods are necessary to advance ABD in NSOs.

The following sections explain the description of ONS, INE, and IBGE based on a mix of documental references and testimonials drawn from interviews with members of the NSOs. A couple of questions about professional trajectory were applied in the interviews' scripts. The result is a rich view of the NSOs by their employees, with clarifying notes.

Innovation cycle is one of the mentioned notes. It includes the perceptions from the NSOs' employees about the tipping points where technological changes took place. Another important aspect of the cases is the Census operations management. In this section, the British, the Portuguese, and the Brazilian NSOs had their demographic Census scrutinized. The justification for this detailed examination is the relevance Census has to NSOs, as their major survey. Finally, the outputs and deliveries addressed by NSOs are presented in this chapter. The relevance of the outputs and their standardised structure signals how the Global Statistical System influences the NSOs routines.

5.1. Features and Summarised trajectory

The NSOs as the current organisations we know were created in different periods of the last one hundred years. In chronological order, INE is the oldest institute, followed by IBGE, and ONS is the most recent. However, the activity of Official Statistics in the three countries had previous operators before those three.

The “*Instituto Nacional de Estatística*” of Portugal (INE) was created in 1935 (Valério and Bastien, 2010). It collects and disseminates Official Statistics to the Portuguese society, which totalises approximately 10 million inhabitants. This organisation generates economic and demographic data collected in the 18 districts of the mainland and the autonomous regions of Madeira and the Azores.

Regarding the regions of Madeira and the Azores, both have dedicated structures that work with INE. Those governmental bodies are not independent NSOs, but members of the Portuguese Statistical System. “*Diretoria Regional de Estatística da Madeira*” (Diretoria Regional de Estatística da Madeira, 2023) and “*Serviço Regional Estatístico dos Açores*” (SREA, 2023) work as partners of INE in the Portuguese archipelagos of Madeira and Azores.

INE is a public agency subordinate to the Minister of the Presidency (Instituto Nacional de Estatística, 2023). In the Portuguese case, this link means that the NSO has limited budget autonomy.

Portugal is a member of the European Union. The 27 countries of the European Union, through their respective NSOs, are part of the EUROSTAT group (Eurostat, 2023). This set of NSOs composes a regional body tasked with maintaining and improving the quality of the Official Statistics of the member-states of the European Union.

In 1938, IBGE was created by the fusion of two previous organisations: the Brazilian National Statistics Office and the National Geographic Council (Penha, 1993; Gonçalves, 1995). Since that genesis, IBGE has presented a singularity compared to other NSOs. Its operations included not only Official Statistics but also Official Geographic data (IBGE, 2023a). During those eighty-five years, IBGE has achieved a double challenge: deploying Official Statistics and managing all geographic information for the citizens of the fifth-biggest country in the world.

Related to the link with the Brazilian Federal Government, IBGE is a national agency subordinated to the Ministry of Planning (Ministério do Planejamento, 2023). This feature is remarkable because it justifies the limited autonomy that NSO has.

Nowadays, IBGE provides information for approximately 210 million Brazilian citizens. It operates in 26 states and the Federal District, collecting data from 5570 Brazilian cities. In addition, this organisation disseminates demographic, geographic, social and economic data from Official Statistics.

One remarkable feature in the IBGE trajectory is the capacity to develop outcomes about the complex Brazilian reality. This competence is proven by constructing experimental statistics to count the indigenous populations (IBGE, 2023b) and verify their conditions. Another example of the IBGE response capacity was a survey created to measure the impacts of the COVID-19 pandemic on Brazilians, named PNAD COVID-19 (Penna *et al.*, 2020).

In the case of the British Office for National Statistics (ONS), it delivers Official Statistics for a group of countries integrated by England, Scotland, Wales and Northern Ireland (Office for National Statistics, 2023a). This organisation presents a singularity in comparison with its counterparts in Brazil and Portugal: it is a public-independent research institute with the status of a non-ministerial organisation (Office for National Statistics, 2023b).

ONS has another distinctive feature, its shorter lifetime. Unlike INE and IBGE, the British NSO was created only twenty-seven years ago, in 1996. Pullinger (1997, p. 291) remarked that the "*merge of the Central Statistical Office (CSO) and the Office of Population*" created the ONS. Despite its relative youth, ONS absorbed the two hundred years of heritage on Official Statistics in the United Kingdom. Beyond the historical register, ONS works for different nations and cultures. This diversity of nationalities represents a context different from INE and IBGE, which work only for people in a respective country.

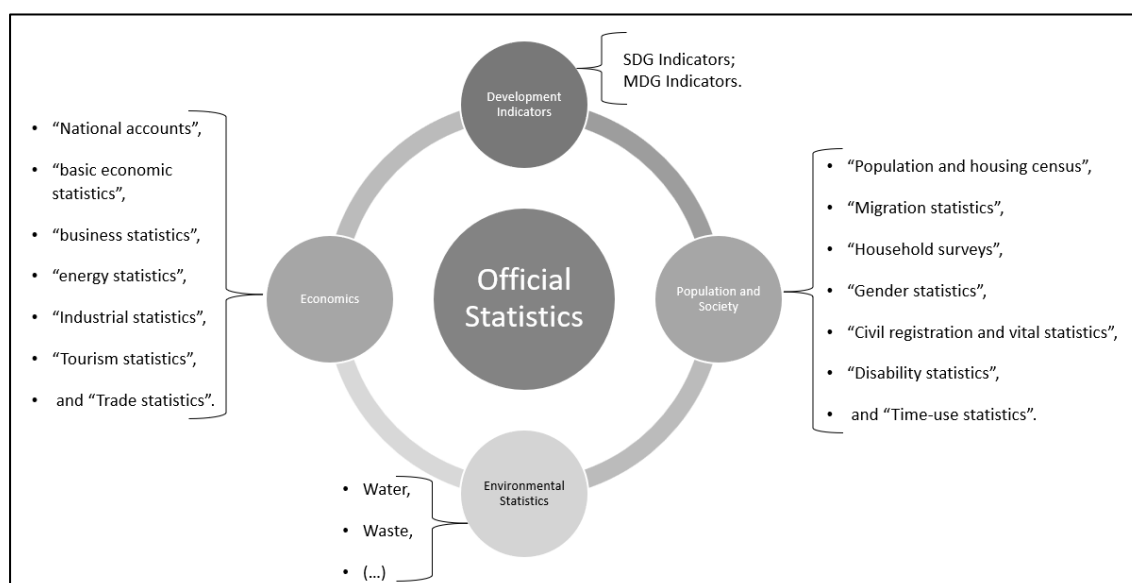
As a non-ministerial public body, ONS reports only to the British Parliament and the UK Statistic Authority (UK Statistics Authority, 2023). This last organisation acts like a regulatory agency with the "*objective of promoting and safeguarding the production and publication of official statistics that 'serve the public good'*" (UK Statistics Authority, 2023). In this way, it is relevant to verify in further analysis (Chapter 6) if this institutional design has some influence on the ABD process.

Speaking of institutional structures, it is possible to inquire if an NSO, with more autonomy than a similar one, can adopt Big Data faster and more broadly. An unfolding result of the linking

structures may be the other TOE's element, *government regulation*. In a formal and centralised relationship, the regulations are based on laws and governmental acts. The Brazilian case can confirm this, and a quick view of IBGE's regulatory bases shows more than ten pages referencing official rules (IBGE, 2021). On the other hand, ONS presents fewer formal norms, as shown on the National Census webpage (Office for National Statistics, 2022a). In the same path, data regulation laws affect those organisations, and they are under distinct legal norms. INE-PT and ONS-UK are covered by The General Data Protection Regulation – GDPR (European Commission, 2022), and IBGE-BR is working under a domestic governmental norm, Personal Data Protection General Law - PDPGL (República Federativa do Brasil, 2018).

5.2. Outputs and deliveries

During a search on the UNSD webpage (United Nations, 2023), it is possible to verify, in the menu "Topics", five Official Statistics main categories. In Figure 14 there is an illustrative scheme of those categories. The proposal presented an illustrative list of Official Statistics, not an extensive one.



Legend: A summarized portfolio of the Official Statistics standards outputs/deliveries following United Nations Statistics Division guidance. Based on United Nations Statistics Division – Department of Economics and Social Affairs. Available at: <https://unstats.un.org/UNSDWebsite/>

Figure 14 – Official Statistics by topics – an exemplificative scheme.

As presented above, four topics summarise the main Official Statistics developed by NSOs. Economics statistics support decision-making with data about business sectors, national accounts and trade statistics. In the opposite circle, Population and Society include demographic statistics such as Census data and migration. Environmental statistics is an emergent topic,

reporting data from water, waste, forests, and associated issues. Finally, the Development Indicators represent two sets of Official Statistics, including the "Sustainable Development Goals" (United Nations, 2022) and the "Millennium Development Goals" (United Nations Statistics Division, 2023d). Those two sets work as monitors to check how countries cooperate to promote a sustainable feature.

All those surveys, indices and rates are generated by NSOs, including ONS, INE, and IBGE, as presented on their websites. A simple comparison between the NSOs' websites makes it possible to realise how standardised the Official Statistics portfolio is. This feature allows the exchange of practices and proceedings in the GSS network. One possibility is to spread the ABD among the NSOs.

Those outcomes are part of the business architecture presented in Figure 4 (Chapter 1). Big Data technologies and methodologies can be adopted or avoided in this business architecture element of surveys.

5.3. Innovation cycles

During the interviews, the members of NSOs were asked two questions about their professional trajectory. The first inquired about their trajectory in the respective institute, regarding their achievements and professional experience. Secondly, the members answered about their perceptions of the relevant innovations they had witnessed during their journey in the NSO.

After collecting those questions, they are compared with the documental sources. The proposal was to verify if the described facts are also relevant to the respective NSO. In fact, in the three cases, the convergence was strong. Accordingly, the interviews and documents showed that ONS, INE, and IBGE work on innovation cycles.

IBGE and INE presented similar innovation cycles to the demographic Census operation, occurring every ten years. This information was relevant to this research because it shows when those NSOs receive from the Governments the window of opportunity to update their technologies and processes. Members from both institutes subscribe to this fact. (In this research, the citations are presented with the code of the "interviewees" between brackets after the quotation marks.)

"Another point that I would also like to draw attention to is the modernisation of data collection and verification. Related to the collection, the introduction of the CAPI was

fundamental because it made the collection quicker. CAPI gave more time for the interviewer to work on the approach process." [06_IBGE]

In 2010, during the demographic Census, IBGE innovated by adopting Computer Assisted Personal Interviewing - CAPI (Instituto Brasileiro de Geografia e Estatísticas, 2023a). Those devices substituted the sheet-form questionnaires for Census data collection. For IBGE members, this innovation represented an icon of the institute adopting new technologies. As a remarkable event, the members cite the prize won by IBGE from UNESCO for adopting CAPIs in Census 2010 (UNESCO, 2023). Notwithstanding, the interviews did not mention breakthrough innovations after the CAPIs adoption.

"IBGE won the technology adoption prize (by UNESCO). So, I would say that this was the greatest technological advance that I have witnessed. Production of technologies in the collection of the Demographic Census. In collection in all senses: in collection management and in the collection itself, with the CAPIs." [09_IBG_HUB]

Regarding the advances of INE, there is no mention of international prizes. However, the members highlighted the use of electronic forms since the first decade of this century. In sequence, the Portuguese professionals pointed out a continuation of technological adoption. The data collected in digital form through INE's website (Instituto Nacional de Estatística, 2011) started in the CENSUS 2011.

"Here in Portugal, Census operations are always marked by some component of innovation. In 2001, the great innovation in the Census was optical reading and character recognition (...), which was a very big leap in what was done until then. (...) Then, after ten years, in the 2011 Census, one of the very innovative components that were done in Portugal quite successfully - and let me say that it is a pride to work in an Institute like this - refers to the introduction of the internet response in the Census. In 2011 we already introduced the internet response, and I can tell you that we achieved at the time a 50% response rate." [04_INE_COC]

The realised difference between IBGE and INE refers to the innovation cycle continuation. While IBGE's cycle appears to be interrupted, the Portuguese one maintained the effort to transform processes after the Census operations in 2011 (Instituto Nacional de Estatística, 2011).

A particular innovation cycle was verified in the ONS case. Since its foundation in 1996, this institute deployed two Census, 2001 and 2011 (Office for National Statistics, 2022b). In 2016,

with twenty years of executed services, ONS suffered a structural reform. An act from the British Parliament determined a review of the ONS from a data supplier to a changing driver for other governmental bodies. This ambitious proposition was registered in the study developed by Bean (2016).

This initiative resulted in relevant innovations. The significant output was the creation of the "Data Science Campus", also launched in 2017 (Data Science Campus, 2023). Interviewees reported this as a tipping point in the ONS trajectory.

"They decided to establish a so-called Data Science Campus. So that not only their employees could be trained to understand the (data science) concepts, basic techniques, but also they want to attract some kind of external collaboration within the institution." [01_ONS_PRO]

In summary, the data collection revealed two types of innovation cycles for NSOs. The cycle associated with the Census operations is verified in IBGE and INE. Conversely, the innovation cycle provoked by external agents refers to ONS. In this case, the external agent is the major stakeholder, the British Parliament.

5.4. The Census operation

The previous section identified the Census operation as a window for innovation for IBGE and INE. This attributed relevance is consensual among NSOs and supranational agencies, such as the United Nations Statistics Division (UNSD). Best practices and challenges are discussed in forums organised by the UNSD (United Nations Statistics Division, 2023b). The International Committee on Census Coordination (ICCC) coordinates those forums and the exchange of information.

According to the UNSD (2008), the demographic/population Census means:

"(...) the total process of collecting, compiling, evaluating, analysing and publishing or otherwise disseminating demographic, economic and social data pertaining, at a specified time, to all persons in a country or in a well-delimited part of a country."

ICCC runs recommendations and best practices for NSOs to deploy the demographic Census. Those prescriptions are necessary, as this is the major Official Statistics operation performed by NSOs. UNSD (2008) presents best practices for Census operations. Some examples of those guidelines are (i) definition of the topics to be investigated, as immigration and fertility; (ii)

recommended tabulations for the outcomes; (iii) products to be delivered, as percentage population change during the intercensal period(s), average annual growth rate, or populational density (persons per square kilometre).

Those standards and information exchanges between the UNSD and NSOs are examples of the Global Statistical System (GSS) performance. As cited in Chapter 1, the GSS is the global network that integrates the NSOs among them and with UNSD. Beyond those integrations, each NSO presents specifications to deploy the Census operations. The following subsections present and describe the United Kingdom, Portugal, and Brazil census. This description includes documents and registers obtained with citizens in the three countries through informal talks. Those citizens offered those testimonials freely, requiring only anonymity. All of them were contacted by the researcher through the professional network LinkedIn.

5.4.1. The Census in the United Kingdom

According to the ONS website, the initial results from CENSUS 2021 were published on June 28th 2022 (Office for National Statistics, 2023c). One year before those broadcasts, the COVID-19 pandemic was slashing the world with new variants, such as Omicron (World Health Organization, 2022). In this way, the CENSUS 2021 and the pandemic were part of the British people's life two years ago (Office for National Statistics, 2022b).

Indeed, the pandemic impacted Census operations around the world. Considering the extension of this disruptive, unpredictable event, the UNSD ran surveys to measure the pandemic's damages on Census operations in each country (United Nations Statistics Division, 2023c). Despite this dramatic context, ONS planned, deployed and concluded its Census.

In fact, the mitigation of the pandemic by ONS was possible because this Census was entirely digital, using administrative records and internet questionnaires filled up by the citizens. As a member of ONS cited, this Census combined different data sources.

"We have run a Census this year. I don't know if this Census count comes under your definition of Big Data. But I expect it's quite big. It's going to give such an amazing opportunity to use as a basis for modelling and for a spine core infrastructure people and households. And, so, this is a great program to combine administrative data, census data, and all of this definition of Big Data." [03_ONS_DIT]

The use of self-answers by citizens was a successful strategy considering the reply to ONS is mandatory by British law. Firstly, ONS sent an invitation letter, as presented in Figure 15. The citizen, or the household group, had a period to access the weblink, fill out the digital form, and send the reply. In the sequence, the ONS website presented a kind of receipt, shown in Figure 16.

After the results, the UK Census 2021 can be classified as remarkable regarding two facts. Firstly, the capacity of ONS to "keep the pace" and deliver a Census, despite the worst sanitary crises in the last one hundred years, represents an achievement. Secondly, online data collection allowed citizens to reveal sensitive information (like sexual orientation) that might be more difficult to obtain in face-to-face interviews (Moss, 2023). In fact, this Census outcome revealed a previously neglected face of British society.




The Resident
Room
Iguana, Houses
Road
Canterbury
CT2
SR-RSE2-ZC-07 D_ICA_ICLR1

794807



2130 0004 3904

Contact us
www.census.gov.uk
 Phone us free
 0800 141 2021
 Text relay
 (18001) 0800 141 2021
 Reference number:
 5232 2590 92

Dear Resident

We need your help with the census, which gathers vital information to help plan services such as transport, education and healthcare. Please complete the census on **Sunday 21 March 2021** or as soon as possible after.

You must complete the census by law or you could be fined up to £1,000.

How to complete your census

- 1 Read this letter containing your individual access code. 
- 2 Go to www.census.gov.uk, select "Start census" and enter your individual access code:

K3

9D

7

M

Q

4

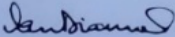
G


- 3 Answer the questions and submit when complete. 

 Your answers are protected by law and will be kept confidential.

The census should take around 10 minutes per person to fill in.
 Help is available for everyone. Speak to the manager of this accommodation or contact us. You can also request a paper form, including large print.

Thank you



Professor Sir Ian Diamond – National Statistician



 Office for National Statistics
 www.census.gov.uk


ICLR1/794807 / 1 of 1

Legend: Invitation letter, sent by post service, inviting citizens to answer the census questionnaire on the ONS-UK website. Source: document from the researcher's files.

Figure 15 – Census Invitation letter for British citizens

Finally, the Census 2021 represented a gain for British society because the results were available to support better decision-making in less than six months. ONS did not announce that this Census was a result of the exclusive use of Big Data applications. Nevertheless, Big Data methodologies and technologies were critical to advance this colossal survey, as reported by the ONS member.



Legend: Receipt of participation of citizens who answered the digital questionnaire of the British Census. Source: document from the researcher's files.

Figure 16 – Submission proof of the Census questionnaire

5.4.2. The Census in Portugal

In the Portuguese Census, INE followed a similar trajectory to ONS, despite some singularities. During 2021, INE advanced its Demographic Census, working on a hybrid approach. Part of the data collection was performed by digital internet forms filled up by citizens, just like, in ONS. Simultaneously, a workforce of eleven thousand data collectors were hired on temporal contracts (INE, 2021). Their mission was to deliver the invitation letters to each house. An example of this letter is presented in Figure 17. Additionally, they supported the elderly population in filling out the forms. In cases where the residents did not have electronic devices, the data collectors interviewed the residents and collected data by CAPIs.

INSTITUTO NACIONAL DE ESTATÍSTICA
STATISTICS PORTUGAL

CENSOS 2021

RESPONDA PELA INTERNET

a/c do Residente no Alojamento

7GMGVX

A PARTIR DE 19 DE ABRIL
EM CENSOS2021.INE.PT

LINHA DE APOIO
210 54 20 21

DEPOIS, DOBRE E DESTAQUE ESTAS BANDAS PELOS PICOTES

O Instituto Nacional de Estatística, IP (INE), solicita e agradece a resposta dos residentes deste alojamento aos Censos 2021, **a partir de 19 de abril**.

RESPONDA PELA INTERNET

1 FÁCIL - Aceda a **censos2021.ine.pt**, ou use o QR CODE

2 SEGURO - Digite o código e a password abaixo indicados.

CÓDIGO: 7GMi

PASSWORD: ub7Aqk

3 RÁPIDO - Responda às perguntas e quando terminar seleccione "ENTREGAR".
Guarde o comprovativo de resposta.

O INE AGRADECE A SUA COLABORAÇÃO

Se não tiver condições para responder pela Internet nas próximas duas semanas:

- ➔ Solicite o apoio de familiares ou amigos
- ➔ Dirija-se à sua Junta de Freguesia e peça apoio (leve consigo esta carta)
- ➔ Aguarde pela visita do Recenseador*
- ➔ Telefone para a linha de apoio

* Os contactos presenciais seguem o protocolo de segurança de saúde pública no âmbito da situação epidemiológica

Nos termos da Lei n.º 22/2008 de 13 de maio, do Decreto-lei n.º 136/2012 de 2 de julho, e do Decreto-lei n.º 54/2019 de 18 de abril a resposta aos Censos 2021 é obrigatória, sendo garantida a confidencialidade e a salvaguarda do segredo estatístico de toda a informação.

COLABORE COM O INE
RESPONDA COM BREVIDADE
EVITE A DESLOCAÇÃO DO RECENSEADOR

GUARDE ESTA CARTA E TENHA-A CONSIGO SE PRECISAR DE CONTACTAR COM O INE

Legend: Invitation letter, sent by post service, inviting citizens to answer the census questionnaire on the INE-PT website. Source: document from the researcher's files.

Figure 17 – Census Invitation letter for Portuguese citizens

After the form filling, the citizen received from INE's website a receipt of participation in the CENSUS 2021, as demonstrated in Figure 18. The citizens must keep this document because CENSUS participation is mandatory in Portugal.

In fact, the hybrid approach in the Portuguese Census was successful. Like the ONS, half a year after the data collection, on 2022 November 23rd, INE published the results of the CENSUS (Instituto Nacional de Estatística, 2022). During that journey, the institute realised the Portuguese

population had a large segment of elderly people. This group might have some restrictions to accessing the internet.



INSTITUTO NACIONAL DE ESTATÍSTICA
 STATISTICS PORTUGAL



Comprovativo de entrega do questionário dos Censos 2021

Este documento valida a entrega do questionário dos Censos 2021 relativo ao alojamento

CÓDIGO	DATA/HORA DA ENTREGA (PORTUGAL CONTINENTAL)
A58PC	28-04-2021 23:46:53
NOME DO REPRESENTANTE	
[REDACTED]	
MORADA	
RUA ALFONSO CHAVEZ, 120 LISBOA 1000-043 LISBOA	
PESSOAS	
HOMENS: 0 MULHERES: 0	
NOTAS:	
Vago para venda ou arrendamento	

Instituto Nacional de Estatística
 Av. de António José de Almeida
 1000-043 Lisboa
 Portugal

Linha de apoio
 21 054 20 21

Legend: Receipt of participation of citizens who answered the digital questionnaire of the Portuguese Census. Source: document from the researcher's files.

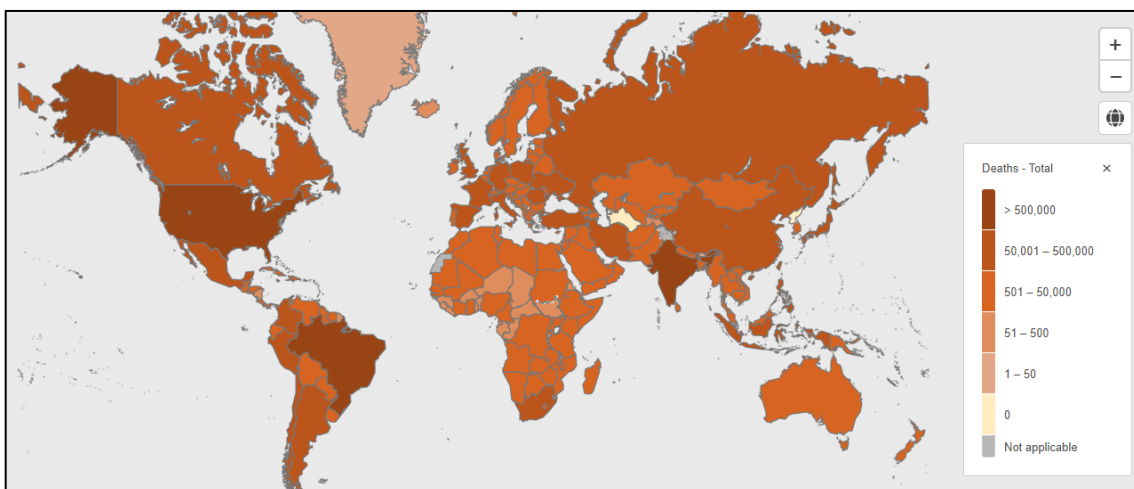
Figure 18 – Receipt of participation in the CENSOS 2021

Related to Big Data technologies and methodologies, INE adopts Artificial Intelligence with the use of machine learning algorithms to support citizens in filling out forms. As the INE representative cited, this adoption represented an innovation for the Portuguese institute.

"Now, in 2021, there was a reinforcement of this response component. There was the introduction of autocomplete, which, as the person writes, we suggest entries for the text in the open questions. We had the use of machine learning in the coding of the descriptives, and one of the other innovations was the incorporation and the possibility of getting information from the administrative records." [04_INE_COC]

5.4.3. The Census in Brazil

Since 2020, Brazil has been one of the countries most affected by the COVID-19 pandemic. Behind the United States, with 1,111,342 deaths, Brazil is the second country in total lost lives (WHO, 2023), with 699.276 deaths, during the period from 2020 to 2022, as presented in Figure 19. This externality has impacted different Brazilian business sectors, including the Official Statistics.



Legend: Global map categorizing, by colors, the total of deaths by Coronavirus (COVID-19) in the World (2020-2023). Source: World Health Organisation (WHO). In: WHO Coronavirus (COVID-19) Dashboard. Extracted from: <https://covid19.who.int/>

Figure 19 - Total of Deaths by Coronavirus (COVID-19) in the World (2020-2023)

Initially, the Census had the original plan to begin in 2020. Because of COVID-19, it was postponed to 2021 (IBGE, 2020b). A new delay occurred in 2021 (IBGE, 2022a). The justification was based on the Brazilian Government's allegation that there were insufficient funds to deploy the Census. Finally, IBGE launched the Census on 2022 August 1st (Gandra, 2022).

Unlike ONS and INE, the Census deployed by IBGE focuses mainly on direct data collection based on the work of hired data collectors. To advance this strategy, IBGE disseminated visual content to promote the engagement of Brazilians. Two examples are the posters presented in Figure 20.



Legend: Informative posters to disseminate the Census to the Brazilian citizens.

Source: IBGE (2022). Dissemination material. Extracted from: <https://censo2022.ibge.gov.br/pecas-de-divulgacao/pecas-de-divulgacao.html>, on 09/01/2023.

Figure 20 – Dissemination posters for the Brazilian Census 2022

Related to technological adoptions, no reference was mentioned on the IBGE website. Just the use of the CAPIs is presented in the dissemination materials. Figure 21 shows the set of materials used for data collection by Census professionals. This set includes one cap, one waistcoat, and a CAPI.

The data collection suffered delays because of the reduced number of Census professionals (Almeida, 2022). Originally, IBGE had planned to hire 183,000 professionals. During the peak of the collection, 120,000 data collectors were working. This deficit impacted the conclusion of the Census. In this scenario, the dissemination of the final results, scheduled for April 2023 (Silveira, 2023), was yet to occur.

On 2023, June 28th, finally, IBGE broadcast the initial results from the last demographic Census. This new populational counting presented around 203 million Brazilian citizens. For this doctoral

research two numbers are relevant: the percentage of non-responses and the number of surveys deployed by Internet, in an experimental mode. Five percent of the Brazilians refused to answer the Census survey. Although the proportion of 95% of responses can be understood as a strong value, the non-responses are significative. On the other side, the number of responses by Internet was reduced when compared with the door-to-door interviews. Just 0,56% of the Census questionnaires were answered using the internet. The majority of the data collection, 98,88%, took place using face-to-face interviews. Another residual percentage, 0,56%, was executed by telephone call interviews.



Legend: Data collectors' uniform on the left. CAPI model used in the Brazilian Census on the right.
Source: photos kindly provided by IBGE staff. Rio de Janeiro, 2022 August 1st.

Figure 21 – Data collectors' work kit.

In summary, the delays suffered by the Brazilian Census are a strong warrant for others NSOs. When the three cases are compared, INE and ONS advanced with their Census, and IBGE just concluded it in 2023, after three years of attempts. The focus on data collection by the internet associated with ABD allowed INE and ONS to overcome the impacts of the pandemic. On the opposite side, Brazil insisted on the same formula of the Census 2011, despite the adverse context. Maybe if IBGE had adopted a hybrid strategy as INE, the Census would be concluded on time, consuming less resources from the taxpayers.

Chapter 6 - Data analysis

The adoption of Big Data (ABD) by NSOs can be verified in the operational process changes to deploy Official Statistics. Additionally, ABD can also be verified in initiatives to promote the use of Big Data technologies and methodologies in and out of the NSO.

One example of the effort to associate Big Data with Official Statistics is the creation of Hubs in the Global Statistical System (United Nations Statistics Division, 2023e). Those entities have the mission to promote and spread the use of Big Data for NSOs in specific regions. Currently, there are four Big Data Hubs in the world, located in four regions on four continents: Brazil (America), Rwanda (Africa), China (Asia), United Arab Emirates (Middle East). Each Hub is supported and managed by the NSO in the respective country. In this way, the Brazilian Hub receives support from IBGE (Instituto Brasileiro de Geografia e Estatísticas, 2023b).

The opened catalogue of courses (United Nations Statistics Division, 2023a) to teach Big Data techniques or the creation of different Hubs are examples of adoption acts. In fact, those acts impact the TOE model on its Organisation and Environment dimensions. Adding a new business unit into the IBGE structure, as the Hub is, represents a change in the organisational processes and interactions between departments. On the other hand, the environmental dimensions are affected by new interactions between the Hub host and the other regional NSOs.

Related to the statistics and ABD, this interaction figures in the Technology dimension. According to Allende-Alonso *et al.* (2019), a Big Data set delivers unnormalised data. This feature demands that Statisticians and Data Scientists use additional statistical methods to interpret and extract value from massive datasets.

Indeed, this journey to develop Big Data methodologies for Official Statistics is complex and represents an ongoing challenge. Bollineni-Balabay *et al.* (2015), van den Brakel and Kreig (2016), Van Den Brakel et al. (2017), and Schiavoni et al. (2021) guided researches to develop or adapt statistics tools and methods to incorporate Big Data in the Official Statistics productions. In this way, for NSOs, it is necessary to use the expression Big Data technologies and methodologies. A mandatory link between the statistical methodology and the technology allows to collect, store, process and analyse the data.

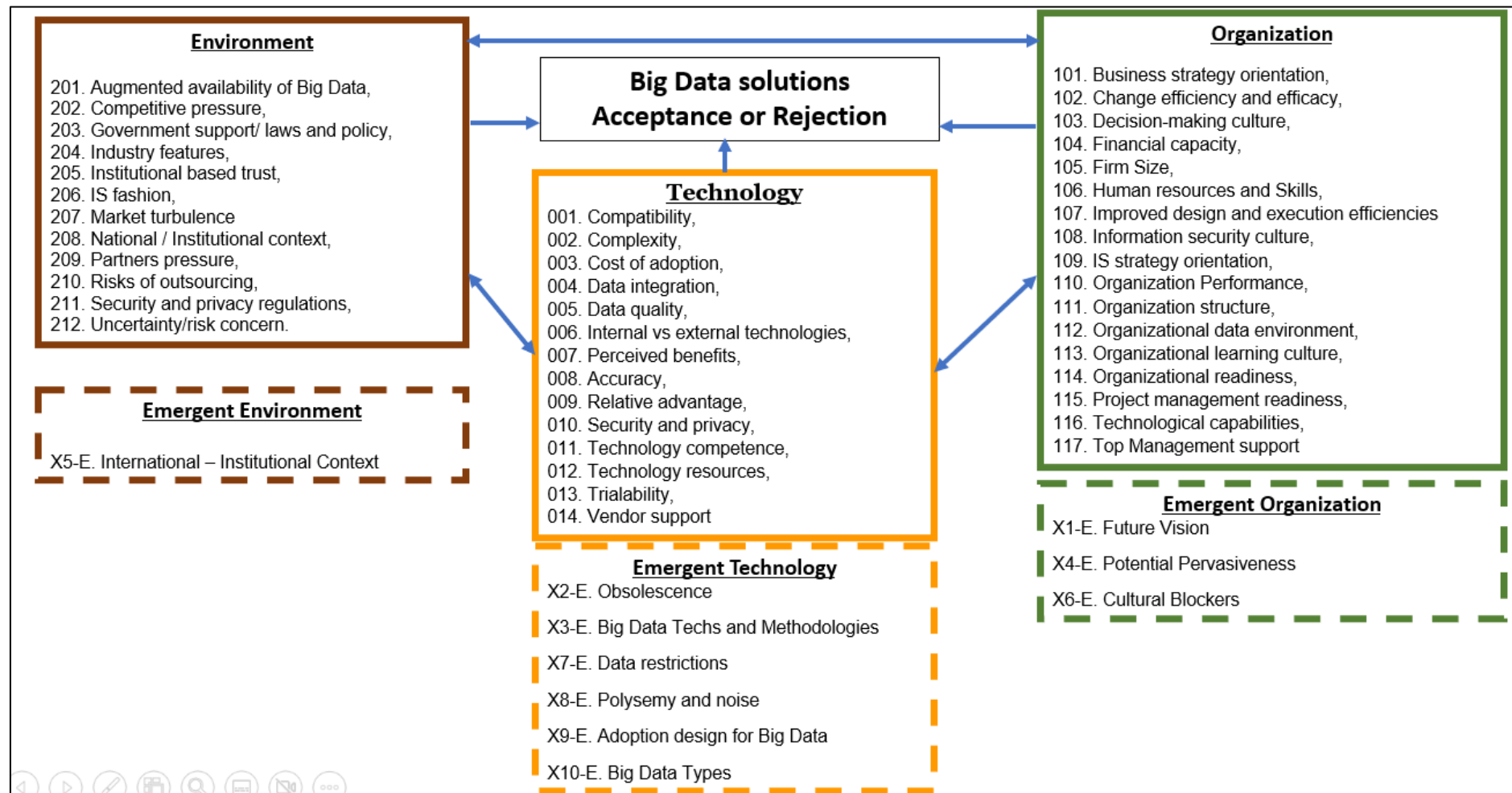
Next, the sections and subsections present the analyses of the interviews with ABD in perspective. This adoption refers to technologies and methodologies for Big Data applied in the Official Statistics processes.

6.1. How National Statistics Offices adopt Big Data - TOE factor analysis

During one semester, the interviews with ONS, IBGE, and INE members ran in a systematic way. All interviewees demonstrated a thorough understanding of Official Statistics as well as a genuine perception from the adoption of Big Data (ABD). Aside from the fact that the interviews took around forty minutes, each piece of work used by them was precise and carefully selected. The result was relevant raw data to be analysed as proposed in the Chapter 4.

Interviewees did not assign the same importance and attention to the different TOE factors. The most cited factors presented more quotations and descriptions than the others, with fewer mentions. In fact, the factors that drive ABD in NSOs do not run in a flat and balanced way. So, the analysis advances in a systematic mode, following the lists of factors per dimensions, with the proposal to bring a basic order in the sets of factors and simplify the localization of each factor in this chapter, like a catalogue, as presented in Figure 22.

According to the original proposal of DePietro *et al.* (1990), TOE should be, initially, approached from the Technology dimension. Hence, the Technology dimension is the first to be analysed, followed by the Organisation and Environment dimensions.



Legend: Continuous Rectangles depict TOE factors found in the Literature review. Dashed Rectangles depict emergent TOE factors resulting from this research. The colours represent the three TOE dimensions: Environmental factors are brown; Organisational factors are green; Technological factors are orange. Arrows represent influences from one factor to each other.

Figure 22 – TOE factors for the Interview analysis.

6.1.1 Technology factors

The literature review presented fourteen TOE factors in the Technology dimension related to the adoption of Big Data (ABD). As presented in Chapter 4, this set of factors was used as a code book to analyse the content of the interviews (Saldaña, 2013).

During the interview analysis, those fourteen factors were validated by members of the three NSOs (ONS, INE, IBGE). However, some factors presented validation and high relevance, while others were validated, but showed less importance. Additionally, six original factors emerged from the content analysis specifically related to ABD in NSOs: "*obsolescence*," "*Big Data technologies and methodologies*," "*data restrictions*," "*polysemy and noise*," "*adoption design for Big Data*," and "*Big Data types*". Clarifying, Figure 22 presented those emergent factors in the yellow dashed rectangle, and in the continuous one, the factors from the literature review.

Each new factor was defined to support the respective content analysis. The first technological emergent factor, "*Obsolescence*", means the incapacity of the traditional methodologies and technologies to address current issues in Official Statistics (OS), such as large volumes of data, randomness, and imputation. Second, "*Big Data Technologies and Methodologies*" include descriptive characteristics of Big Data technologies and methodologies used by NSOs. Thirdly, "*Data restrictions*" refers to every condition, norm, economic feature, or technological issue that restricts the access, processing, use, and broadcast of a data source or data set required by an organisation engaged in BD adoption. In regard to "*Polysemy and noise*", it is related to the different interpretations about what Big Data for Official Statistics means. The emergent factor named "*Adoption design for Big Data*" refers to the plans and actions designed for a successful adoption of Big Data. Last, but not least, "*Big Data types*" define the data sources and types used in ABD processes in NSOs.

During the analysis, each Technology factor presented relevant elements about ABD in the three NSOs – ONS, INE, and IBGE.

Further, the analysis of each factor, related to ABD in NSOs, includes the respective definition, the elements/items that integrate the factor, illustrating quotes from the interviews and respective comments. As the anonymity of the interviewees was guaranteed, each quote is identified only with the institute of the interviewee, between brackets. Finally, in the end of each factor analysis, there is a summarized list of integrating elements/items. This list informs the findings of this research

related to the respective factor. Like with the quotes, each item has the names of the respective NSO between brackets. In this way, it is possible to compare and verify in which case each item was identified, allowing the reader to compare the three cases.

For this purpose, Table 21 has the objective to show where the respective item was more salient. In this context, the first column informs the factor name. In sequence, column two shows the consolidated findings. Inside this column, each element has an order number and brackets with the respective NSO where it is manifest. Some elements were observed in the three NSOs. Others just in one or two cases. Finally, the last three columns inform the presence of each element in the respective case, following the order number from column two. So, there is the synthesis of the results from the Technology dimension in Table 21.

After Table 21, the detailed analysis of each Technology factor is presented.

Table 21 – Technology factors: Summarised findings

Technology Factors	Consolidated Findings	Findings per case		
		ONS-UK	INE-PT	IBGE-BR
001.Compatibility	1. Why is compatibility important for ABD in NSOs? Because the adoption must integrate Big Data to the internal organisational processes and information systems. [ONS, INE, IBGE] 2. How does compatibility influences ABD? By meeting specific requirements: a. Managing long-term projects. [INE] b. Alignment of technological and methodological procedures to deploy new outcomes for Official Statistics. [ONS, INE, IBGE] c. First milestone: compatibility of traditional and new data sources. [ONS] d. Following NSOs calendar: Big Data outcomes inserted in the Official Statistics (OS) schedule. [ONS] 3. Benefit of the compatibility to ABD: the comparability between NSOs, improving the benchmark. [IBGE]	1; 2b; 2c; 2d.	1; 2a; 2b.	1; 2b; 3.
002.Complexity	1. Why complexity for ABD: avoidance of a potential barrier for the adoption. [ONS; INE; IBGE] 2. How does complexity influence ABD? It affects ABD initiatives when: a. The replacement between Big Data methods and survey methods is too complex. [ONS] b. Coverage issues arise: “Just because it’s big, it doesn’t mean that its coverage is right.” [ONS] c. Methodological challenges include the imputation and missingness. [ONS] 3. Benefit of reducing complexity: Merging of the Census operation and administrative Census. [INE]	1; 2a; 2b; 2c.	1; 3.	1.
003.Cost of adoption	1. Why and how does cost of adoption impact ABD? Because it includes cost elements necessary to ABD, such as: a. ABD is expensive, complex, and risky. [IBGE] b. Cost drivers: technology and digital tools, infrastructure, training, and datasets access. [ONS, INE, IBGE] c. Specific budget for ABD. [ONS, INE, IBGE] d. Training costs: a critical item associated with the lack of skills. [ONS; IBGE] 2. Requirement: dealing with high costs in earlier stage of adoption. [ONS, INE, IBGE]	1b, 1c, 1d, 2.	1b, 1c, 2.	1a, 1b, 1c, 1d, 2.
004.Data integration	1. Why is data integration relevant for ABD? Because it analyses the complementarity between Big Data and traditional surveys. [ONS, INE, IBGE] 2. How does data integration influence ABD? Big Data and surveys work in a complementary way. Exemplifying: administrative records used as complements to the OS. [ONS; INE; IBGE] 3.Benefits: a. More precise measurements with real-time data collection. [ONS] b. Ability to use different types of data, including semi-structured and unstructured. [ONS; IBGE] 4. Outcome: a nationwide administrative database. [INE]	1; 2; 3a; 3b.	1; 2; 4.	1; 2; 3b.

Technology Factors	Consolidated Findings	Findings per case		
		ONS-UK	INE-PT	IBGE-BR
005.Data quality	1. Why is data quality relevant for ABD in NSOs? Because OS demand reliable and consistent data to deliver the outcomes. [ONS, INE, IBGE] 2. How does data quality influence ABD? The influence advances by the quality requirements demanded to adopt Big Data as inputs to the OS. [ONS; INE; IBGE] 3. Requirements: a. Rigorous quality for ABD projects. [ONS; INE; IBGE] b. Achievement of quality standards, as harmonisation, before the integrations with OS. [ONS; INE] c. Validation of administrative records before integration with traditional data sets. [ONS; INE] d. Quality tests to verify the quality of big datasets. [ONS; INE; IBGE]	1; 2; 3a; 3b; 3c; 3d.	1; 2; 3a; 3b; 3c; 3d.	1; 2; 3a; 3d.
006.Internal versus external technologies	1. Why is the internal vs external technologies factor relevant for ABD in NSOs? Because the origin of the technology can promote cost and time savings. [ONS, INE, IBGE] 2. How does the internal vs external technologies factor influence ABD? The technological strategy defines the origin of the technology and the type of adoption. [ONS; INE; IBGE] 3. Results: combination amongst internal, external, and hybrid solutions via consortia. [ONS] 4.Requirements: a. First stage: attempt to use internal solutions. [ONS; INE; IBGE] b. Second stage: call for outsourcing. [ONS; INE; IBGE] c. Suppliers: Universities, Research Centres, private companies. [ONS; INE; IBGE]	1; 2; 3; 4a, 4b, 4c.	1; 2; 4a, 4b, 4c.	1; 2; 4a, 4b, 4c.
007.Perceived benefits	1. Why is the perceived benefits factor relevant for ABD in NSOs? Because NSOs' top managers base their decisions on the perceived benefits of the Big Data for OS. [ONS, INE, IBGE] 2. How does the perceived benefits factor influence ABD? This factor works as a promoter for ABD through: a. The potential capacity to manage sorts of data storage as a service (cloud). [ONS] b. The possibility to model and deploy innovative data collection processes, as using mobile devices. [ONS] c. The option for transitioning to use on time data streaming (data sources) from retail chains. [ONS] 3. Requirements: a. Plan and execute the adequate Big Data project for the NSO reality. [ONS] b. Dedicated budget as a support for ABD projects. [ONS] c. Organisational redesign with a dedicated Big Data unit. [ONS; INE] 4. Benefits: a. The improvement in the reporting time. [ONS; INE] b. Better, faster, and more precise OS outcomes. [ONS; INE] c. Data quality improvement: accuracy and granularity gains. [ONS]	1; 2a; 2b; 2c; 3a; 3b; 3c; 4a; 4b; 4c.	1; 3c; 4a; 4b.	1

Technology Factors	Consolidated Findings	Findings per case		
		ONS-UK	INE-PT	IBGE-BR
008.Accuracy	1. Why is the accuracy factor relevant for ABD in NSOs? Because it refers to the improvements for the data collection process, such as high-frequency and granularity. [ONS; INE] 2. How does the accuracy factor influence ABD? Accuracy represent bonus for the NSO daily-basis operations, including prediction and estimation capacities for the OS outcomes. [ONS; INE] 3. Requirement for accuracy: a. Training the staff in Big Data skills. [ONS] 4. Benefits: a. Data collection with high-frequency, high-speed and granularity. [ONS; IBGE] b. More accurate insights and predictions based on Big Data technologies and methodologies. [ONS; INE] c. Precise estimations for economic indices, such as the British Gross Domestic Product estimation. [ONS]	1; 2; 3; 4a; 4b; 4c.	1; 2; 4a; 4b.	4a.
009.Relative advantage	1. Why is the relative advantage factor relevant for ABD in NSOs? Because it shows to the NSO staff the advantage of adopting Big Data technologies and methodologies instead of avoiding them. [ONS; INE] 2. How does the relative advantage factor influence ABD? This influence occurs through the potential results of ABD for OS, such as innovative sorts of scale, types of data and methods. [ONS; INE] 3. Benefits of relative advantage: on-time outcomes over traditional and time-consuming methods. [ONS] 4. Negative influence of the relative advantage factor: the lack of relative advantage perception for ABD resulting from cultural resistance by NSO staff. [IBGE]	1; 2; 3.	1;2.	4.
010.Security and privacy	1. Why is the security and privacy factor relevant for ABD in NSOs? Because it can be considered a hindrance for ABD. [IBGE] 2. How does security and privacy influence ABD? This factor is perceived as a daily, expensive, and inglorious struggle, considering the risk of cyberattacks. [IBGE] 3. Requirement for security and privacy: corporate policy and programs focusing on prevention. [ONS; INE; IBGE]	3.	3.	1; 2; 3
011.Technology competence	1. Why is technology competence essential for ABD in NSOs? Because it enables the staff to require/develop BD technologies and methodologies, and to design the adoption process. [ONS; INE; IBGE] 2. How does technology competence allow ABD? It allows ABD through three pillars: a. Mixed strategy: hiring policy and skills development. [ONS] b. Internal skills development: long-term training projects associated with experimental statistics construction. [ONS] c. Moving the staff to a data-driven culture. [ONS] 3. Requirement for technology competence: a. Continuous investment in training. [ONS] b. Execution of core actions: to propose, adopt, plan, deploy, and establish Big Data technologies and methodologies for experimental statistics. [ONS]. 4. Benefit of technology competence: individual productivity gain with “R” and “Python” applications. [ONS]	1; 2a; 2b; 2c; 3a; 3b; 4.	1.	1.

Technology Factors	Consolidated Findings	Findings per case		
		ONS-UK	INE-PT	IBGE-BR
012. Technology resources	1. Why is the technology resources factor determinant for ABD in NSOs? Because technology resources allow the NSO to run Big Data, that implies the use of uncommon and untraditional technologies, such as coding-controlled applications. [ONS; INE] 2. How does the technology resources factor run ABD in NSOs? Technology resources run ABD through: a. Increasing computing, storage, and processing power, and data integration. [ONS] b. Accessing data sources. [ONS] c. Managing technological infrastructure by skilled professionals. [ONS] 3. Examples of technology resources for Big Data in NSOs: cloud storage, free per-review version-controlled (Git Hub and Git Lab), Python and “R” applications. [ONS]	1; 2a; 2b; 2c; 3.	1.	
013. Trialability	1. Why is trialability necessary for ABD in NSOs? Because trialability involves experimental statistics, regarding tests with low-sensitive datasets. Testing and verifications are mandatory to verify if the proposed BD solution is adequate for the NSO need. [ONS; INE] 2. How does trialability support ABD in NSOs? Trialability supports ABD through: a. Experimental statistics operating a pipeline of trial projects. [ONS; INE; IBGE] b. Adoption of orderable and peer-reviewable Big Data applications. [ONS] 3. Requirements for trialability: a. Development of a trailable mentality before the trialability. [ONS] b. Dedicated budget, professionals, IT infrastructure, and time (schedule). [ONS; INE; IBGE] c. Capacity to take risks through innovative methodological approaches. [ONS; INE] d. Use of data for trials: low-security, low-sensitivity datasets. [ONS; INE; IBGE]	1; 2a; 2b; 3a; 3b; 3c; 3d.	1; 2a; 3b; 3c; 3d.	2a; 3b; 3d.
014. Vendor support	1. Why is vendor support necessary for ABD in NSOs? Because vendor support avoids the lock-in dependence from a specific supplier. [ONS; INE] 2. How does vendor support help ABD in NSOs? Vendor supports works as services to support the adoption of demand programming hours and Business Intelligence services. [INE]	1.	1; 2.	
OX2. Obsolescence	1. Why is the obsolescence factor critical for ABD in NSOs? Because obsolescence presents the limited capacity to launch and manage Big Data long-term project by some NSOs. [IBGE] 2. How does the obsolescence factor support ABD in NSOs? Obsolescence constrains ABD through: a. An outdated organisational structure: IT over-centralisation. [IBGE] b. Maintenance of traditional methodologies that are insufficient for the ABD Adoption challenge. [ONS; INE; IBGE] 3. As a consequence of obsolescence, there are delays in the extended response time to achieve outcomes. [IBGE] 4. NSOs can react to obsolescence by: a. Talking about ABD more than actually doing anything about it - “Much theorization, less action.” [IBGE] b. Trying to overcome it by an escape route: mapping the flow of the data and all the processes being applied. [ONS]	2b; 4b.	2b.	1; 2a; 2b; 3; 4a.

Technology Factors	Consolidated Findings	Findings per case		
		ONS-UK	INE-PT	IBGE-BR
OX3. Big Data Technologies and Methodologies (BDTM)	1. Why is the Big Data technologies and methodologies (BDTM) factor determinant for the ABD? Because the definition of technologies and methodologies drives the costs, project-time, and training demand. [ONS; INE; IBGE]. 2. How does the BDTM factor work in ABD? BDTM works through the adoption of: a. Technologies (Cloud, AI/machine learning, web scraping, and Big Data analytics) [ONS; INE; IBGE]; b. and Methodologies (clustering-based alternative method for the AI to detect outliers (TCBAN), natural language processing (NLP), sentiment-related information, real-time measurement) [ONS]. 3. Requirements for BDTM adoption: a. AI, cloud storage and cloud analytics - high investments, safety environment. [ONS] b. Web Scraping: reduced investment and time. [ONS; INE; IBGE] c. Big Data analytics as an auxiliary technology [ONS; INE; IBGE] 4. Benefits of the BDTM for ABD: a. Cloud computing and cloud analytics – security, scalability, and insights for experimental statistics. [ONS]. b. Web Scraping - the simplest, cheapest, and fastest Big Data technology to be adopted by NSOs in common application as prices indices calculus [ONS; INE; IBGE]. c. Artificial intelligence / machine learning applications – quality gains in Census operations, and satellite imagery recognition, and joined applications with web scraping [ONS; INE; IBGE].	1; 2a; 2b; 3a; 3b; 3c; 4a; 4b; 4c.	1; 2a; 3b; 3c; 4b; 4c.	1; 2a; 3b; 3c; 4b; 4c.
OX10. Big Data Types (BDT)	1. Why are Big Data types (BDT) critical to NSOs when adopting Big Data? Because the set of BDT handle different types of data, and those types of data are the raw material of the adoption. [ONS; INE; IBGE] 2. How do NSOs adopt BDT? NSOs adopt specific types of Big Data, such as this non-exhaustive list of nine types: (1) road traffic data; (2) shipping data; (3) satellite imagery data; (4) retail price collection; (5) electronic invoice; (6) mobile phone data; (7) administrative records; (8) synthetic data; (9) on-time transports and commercial transactions [ONS; INE; IBGE]. 3. Requirement for considering data as a BDT: a. According to the expected outcome, the type varies [ONS; INE]. b. Common criteria - huge and constant volumes of data, available in real-time [ONS; INE]. c. Demanded process - extraction, transformation, and loading in real-time [ONS; INE]. d. Complements the traditional data [ONS; INE; IBGE]. 4. Benefit of BDT: incremental changes, not disruptions [ONS; INE; IBGE].	1; 2; 3a; 3b; 3c; 4.	1; 2; 3a; 3b; 3c; 4.	1; 2; 3d; 4.
OX9. Adoption design for Big Data	1. Why is adoption design for Big Data mandatory for ABD in NSOs? Because it can reduce the resistance of the staff against BDTM, and the risk of adoption failure. [ONS; INE] 2. How does adoption design for Big Data influence the adoption in NSOs? It influences by the actions of mentors and coaches who run the adoption projects in the departments [ONS; INE]. 3. Requirements: a. Motivation from the Board to the staff (top-down) [ONS; INE]. b. Dealing with methodological challenge: starting with the question; find the question using Big Data [ONS; INE].	1; 2; 3a; 3b.	1; 2; 3a; 3b.	

Technology Factors	Consolidated Findings	Findings per case		
		ONS-UK	INE-PT	IBGE-BR
OX7. Data restrictions	1. Why are data restriction relevant for ABD in NSOs? Because they represent a limitation related to data access, a fundamental asset for ABD in NSOs. [ONS; INE; IBGE]. 2. How do data restrictions affect ABD in NSOs? Data restrictions manifests in: a. Obstacles: administrative records and datasets with reduced access [ONS; INE; IBGE]. b. Legal acts constraining the advance of ABD projects by NSOs [ONS; INE; IBGE]. c. Strong resistance from National Administrations and their public bodies [INE; IBGE]. d. Tabu: mobile data access [ONS; INE; IBGE].	1; 2a; 2b; 2d.	1; 2a; 2b; 2c; 2d.	1; 2a; 2b; 2c; 2d.
OX8. Polysemy and noise	1. Why is the polysemy and noise factor a constrain for NSOs when adopting Big Data? Because the concept of “Big Data” remains unclear when applied to OS. [ONS; INE; IBGE] 2. How does the polysemy and noise factor constrain ABD in NSOs? The lack of a clear Big Data definition provokes resistance from the staff to adopt BDTM. [ONS; INE; IBGE] 3. Damages of the polysemy and noise factor for ABD: a. Lack of consensus among the NSOs members. [ONS; INE; IBGE] b. Entropy increasing in task groups and team projects. [ONS] c. Supporting resistance groups to avoid ABD. [IBGE] 4. Alternative to solve the polysemy and noise issue: demand for a standardized definition of “Big Data” for OS. [ONS; INE; IBGE]	1; 2; 3a; 3b; 4.	1; 2; 3a; 4.	1; 2; 3a; 3c; 4.

001. Compatibility

Compatibility is defined by Picoto *et al.* (2021, p. 283) as *"the level at which innovation is perceived as consistent with internal organisational processes and information systems"*. Considering this definition, compatibility manifests its importance for ABD, demanding (1) organized deployment, (2) meeting specific requirements, and (3) improving the benchmark between the NSOs.

In this context, why is compatibility important for ABD in NSOs? Because the adoption must integrate Big Data in internal organisational processes and information systems. If compatibility is neglected by NSO employees, ABD can become a problem instead of a solution.

Additionally, how does compatibility influence ABD? By meeting specific requirements such as (a) managing long-term projects; (b) alignment of technological and methodological procedures to deploy new outcomes for Official Statistics; (c) first milestone: compatibility of traditional and new data sources; (d) adjustments for the dissemination calendar: Big Data outcomes inserted in the Official Statistics schedule.

In fact, compatibility for ABD demands long-term project management. When NSOs run long-term projects for Big Data, compatibility is improved following a gradual adoption. This gradualism allows comparability gains between NSOs, as IBGE members remarked.

In this adoption process, ABD must be compatible with current technological and methodological procedures for Official Statistics and its outcomes, like was explained by members from the three NSOs. The first milestone refers to the compatibility between traditional and new data sources.

A first milestone for compatibility refers to the methodological association between traditional datasets, as National accounts, and innovative types of administrative records, as tax-fiscal records. Compatibility for ABD in NSOs refers to this association, related to datasets, methodologies and technologies.

"So, working with Big Data, a lot of my activity is around the use of these datasets, bringing them into National Accounts where we can do this. The technical methodologies issues within around this. So, value added tax gives us a measure of output which is hugely useful (...) If there's a big firm, they reply monthly. Smaller firms can reply quarterly or annually. (...) So, you got to find some methods to allocate

out the annual data to get a monthly or quarterly path to be able to put it with the rest of the data and to produce something.” [04_ONS]

As cited before, *compatibility* is managed by long-term projects such as the Census internet replays, which started in 2011 and were established in Portugal ten years later. INE is confident about advancing the demographic Census to a Census based on administrative registers because of this long path of innovation.

“(…) In 2011, we had already introduced the internet response and I can tell you that at that time we got 50% of the responses. In comparison, for example, our friends from the United Kingdom had 8% or something like that. We were the top three states in the European Union with the highest percentage of responses via the Internet. Ten years ago, there was also a very important innovation.” [04_INE]

Regarding compatibility improvement, Big Data brings the benefit of comparability. Official Statistics, developed using Big Data technologies and methodologies, allow the comparability of indexes and outcomes from one country to another.

“Comparability. Big Data favours international comparability. What do I mean by that? Other countries are already migrating to this way of collecting prices to make the indices. (...) the benefits of Big Data are present in these processes.” [09_IBG_HUB]

Finally, the results from ABD must be adequate for the current Official Statistics dissemination schedule, as ONS employees pointed. This calendar adequacy is part of the process to insert the innovative Big Data outcomes into the NSO portfolio.

For ABD in NSOs, the compatibility factor is summarised as

1. Why is compatibility important for ABD in NSOs? Because the adoption must integrate Big Data to the internal organisational processes and information systems. [ONS, INE, IBGE]
2. How does compatibility influences ABD? By meeting specific requirements:
 - a. Managing long-term projects. [INE]
 - b. Alignment of technological and methodological procedures to deploy new outcomes for Official Statistics. [ONS, INE, IBGE]
 - c. First milestone: compatibility of traditional and new data sources. [ONS]

d. Following NSOs calendar: Big Data outcomes inserted in the Official Statistics (OS) schedule. [ONS]

3. Benefit of the compatibility to ABD: the comparability between NSOs, improving the benchmark. [IBGE]

002. Complexity

Complexity is *“the characteristics of Big Data that are perceived as being difficult to understand and use (e.g., the difficulty of learning related knowledge for employees who will use Big Data applications),”* as defined by Sun et al. (2018a, p. 198).

Employees from an NSO must deal with complexity because this factor, if not addressed, can be a barrier to ABD. If complexity becomes too great, ABD can be abandoned by the NSO. This dropout includes incompatibility between Big Data methods and survey methods, or methodological challenges, as imputation and missingness.

In this way, the factors of *compatibility* and *complexity* work together, as presented in citations, to ensure that ABD in NSOs is not a simple phenomenon. In the interviewees' words, ABD is associated with technological and methodological procedures to deploy new outcomes from NSOs.

The difficult mentioned in the definition of complexity is part of ABD in NSOs context. In fact, there is a demand to make a replacement between Big Data methods (innovative) and survey methods (traditional). Another challenge for ABD in NSOs includes dealing with imputation and missingness during the data collection. It represents a methodological frontier to be expanded when an NSO is working with large and non-normalized amounts of data.

As highlighted by a member of ONS, Big Data is important and necessary. However, it does not have a complete coverage of the country and society.

“Just because it's big (data), it doesn't mean that its coverage is right. (...) Nobody yet, including Stats Nederland, has got to a position where the perfect replacement between Big Data methods and the survey methods is achieved. We have not gotten in place at this moment.” [02_ONS]

The adoption of data lakes and data streaming to generate Official Statistics demands the convergence between administrative records with surveys, in a mandatory way. An example is the Portuguese Census operation, where the use of administrative records filled up gaps in the traditional Census data collection procedures.

So, the *complexity* of ABD is an obstacle to be overcome. It is complex because some results from Big Data are not originally adequate to Official Statistics, requiring imputation and missingness treatment. Below, the interviewee cites this methodological challenge.

“It is the big methodological challenge. We’ve also started thinking now about imputation. So, one of the big challenges that we go with Big Data is that the target population from which we got the data that isn’t the same as the target population we’re interested. Our challenge is to understand the missingness, and how to address it..” [02_ONS]

By the way, the adoption of Big Data technologies and methodologies involves distinct *complexity* levels. From a wider perspective, *complexity* appears in the Portuguese Census operations regarding the association between Census data collection and the administrative records. This association focuses on the creation of a Census based on the administrative data, an administrative Census, with regular updates in the demographic database, such as in emigration and permanent resident population counting.

“In addition, these administrative bases, which is not only for the administrative Census, but it's the backbone to expand our production of statistics related to households, social statistics...” [01_INE]

“We have confirmation that it is feasible to adopt an administrative Census methodology.” [04_INE]

For ABD in NSOs, the *complexity* factor is summarised as:

1. Why complexity for ABD: avoidance of a potential barrier for the adoption. [ONS; INE; IBGE]
2. How does complexity influence ABD? It affects ABD initiatives when:
 - a. The replacement between Big Data methods and survey methods is too complex. [ONS]
 - b. Coverage issues arise: “Just because it's big, it doesn't mean that its coverage is right.” [ONS]
 - c. Methodological challenges include the imputation and missingness. [ONS]
3. Benefit of reducing complexity: Merging of the Census operation and administrative Census. [INE]

003. Cost of adoption

Considering Baig et al. (2019), Sun et al. (2018b), Bakici et al. (2022), the *cost of adoption* refers to the financial resources demanded and executed expenses related to ABD (e.g., the costs of using Big Data technologies, the large initial investment required to embrace Big Data). According to Bakici et al. (2021, p. 11), "*when Big Data tools are expensive, they would be more difficult to adopt by companies that do not have sufficient resources.*" So, cost of adoption impacts ABD, imposing high costs for earlier stages of adoption.

In fact, cost of adoption is critical to ABD because this factor includes the cost drivers of this innovation and remarks the necessity to a dedicated budget and training costs management. Those elements are mandatory for an NSO to adopt Big Data technologies and methodologies. Without those cost elements, the NSO does not access knowledge and technological resources for ABD.

ABD is expensive, complex, and risky. It is expensive because many technological applications and infrastructures are not free, such as cloud storage or streaming. Another cost driver is the access to external datasets developed by other companies different from the NSOs. In this example, the cost is a restrictor because innovative methods for delivering experimental statistics can demand those external data sources. As a restriction, the limitation of datasets access means complexity and risk.

"And just to conclude, Big Data is awfully expensive. You know, it's not something that is going to be cheaper in the next few years. It will cheapen in the medium term." [02_IBG]

The cost of Big Data is so significant that ONS is considering allocating a specific budget for its implementation. So, the cost of adoption includes the budget dedicated for acquiring Big Data technologies. The progress of the adoption process is determined by budgetary capacity, which is linked to the financial capacity of the respective NSO.

"I think it is important to think the importance of budgetary spending. Big Data sources are hard to find. The technologies, if we want to develop from scratch, will take a lot of effort and time, and money as well. So, at the current stage, I think ONS wants a budget solution for Big Data. If we opt for a small budget, quality will sometimes be sacrificed, which will cause some problems. As a governmental body, ONS deals with a background of tight budgets. I think it is always all over the world. I think this kind of question probably is universal." [01_ONS]

“So, ONS wanted us to do the research so that they can improve the quality of their statistics because they are different layers of quality based on their budget.” [01_ONS]

“Budget is always an onerous. We never have enough money to do everything that we would like to do and that is a good thing.” [03_ONS]

Another relevant cost driver for ABD in NSO is training. The training cost is a requirement to overcome the lack of Big Data skills in the staff of an NSO. This item is a cost-intensive and critical part of the adoption process. It is necessary to train a large number of employees to initiate the adoption process inside NSO departments. Cost management is an obligation to advance in ABD. This type of adoption demands financial resources on a large scale. Indeed, ONS is an example of costs of training reverted in Big Data skills.

“The training effort deployed by ONS was something quite strong. The ONS launched and funded the master’s in data Analytics of Government. I was the manager of that training program. There was also a trainee program. Everything was funded. This ONS’ effort is something I consider brilliant. (...) I contracted various universities according to the strategic goals that we wanted to achieve.” [05_ONS_PRO]

For ABD in NSOs, the *cost of adoption* factor is summarised as:

1. Why and how does cost of adoption impact ABD? Because it includes cost elements necessary to ABD, such as:

- a. ABD is expensive, complex, and risky. [IBGE]
- b. Cost drivers: technology and digital tools, infrastructure, training, and datasets access. [ONS, INE, IBGE]
- c. Specific budget for ABD. [ONS, INE, IBGE]
- d. Training costs: a critical item associated with the lack of skills. [ONS; IBGE]

2. Requirement: dealing with high costs in earlier stage of adoption. [ONS, INE, IBGE]

004. Data integration

Based on Park and Kim (2021, p.287), *data integration* refers to the degree to which the data is analysed by the Big Data systems, considering the relevance of Big Data applications and the integration of those with the data collected. Under this concept, *data integration* for ABD in NSOs refers to accessing and processing administrative records and aggregating them for

complementing Official Statistics outcomes. A successful data integration allows the NSO to delivery more precise measurements, and deal with different types of data.

Data integration is critical to ABD because this factor analyses the complementarity between Big Data and traditional surveys. Indeed, ABD becomes a reality when it is integrated into the NSO's current processes and deliveries.

"It is a harmonization. Actually, it is a complementation of registers. Like administrative records that you can use in association with the survey, and you can join other data to update the Official Statistics."[07_IBG_COC]

Data integration refers to the complementarity between Big Data and surveys, associating methodologies and technologies. In the initial stages, Big Data was understood as a disruptive hype. However, when the advancement of ABD took place, it became clear that this innovation came to improve *data quality* for the current Official Statistics, not substitute them.

"(...) When I joined ONS, their "mantra" was that surveys were in the last year. That only we wouldn't need surveys anymore. That all we would need would be Big Data, transactional data, administrative data... and that we would be able to produce even better statistics than we were producing before. Since then, I think we've accepted that it was a very naïve approach...kind of a data strategy. Now, our understanding is that we need integrated data assets and administrative records. Basically, to pull together administrative records, transactional data, survey data...or, also called Big Data whatever."[02_ONS]

Data integration is critical to ABD because a new developed dataset has to be compatible with the traditional data structure in the NSO's operations. In this process, measurements become more precise using real-time data collection. This approach represents an improved quality for NSOs' data collection. Therefore, a more precise and faster data collection process means an improvement in the service level of NSOs. This improvement positively impacts organisational performance.

"I think that they are starting to use alternative data sources to see whether they can measure the economic activity more precisely, more in a real-time sense."

[01_ONS_PRO]

The changes embraced by NSOs have delivered recognised benefits, including the ability to use different types of data and data sources. Related to the types, unstructured and semi structured data can be added to complement the Official Statistics outcomes.

“The ability to use any type of data whether it is survey data or large volume, administrative records, or unstructured text data, such as documents, or unstructured data, such as images and video. ” [03_ONS]

“Now, I can see another part of the Big Data, the real-time processing, referring to unstructured statistics, using the famous NoSQL. ” [01_IBG]

“IBGE works essentially with structured data. So, there is a very interesting and very important movement in IBGE to interact with government agencies and use the administrative records made by other government agencies” [04_IBG]

Referring to data sources, administrative records, videos and images represent the target source to be explored with focus on the Official Statistics improvement. *Data integration* and the organisational data environment are benefited from the new internalised knowledge. One example is the gain of data quality when INE adopted administrative records in the Portuguese Census. Actually, a Portuguese administrative database has been developed with a nationwide coverage.

“We have the big project to change from the traditional Census to the Census based on an administrative base.” [01_INE]

For ABD in NSOs, the *data integration* factor is summarised as:

1. Why is data integration relevant for ABD? Because it analyses the complementarity between Big Data and traditional surveys. [ONS, INE, IBGE]
2. How does data integration influence ABD? Big Data and surveys work in a complementary way. Exemplifying: administrative records used as complements to the OS. [ONS; INE; IBGE]
3. Benefits:
 - a. More precise measurements with real-time data collection. [ONS]
 - b. Ability to use different types of data, including semi-structured and unstructured. [ONS; IBGE]
4. Outcome: a nationwide administrative database. [INE]

005. Data quality

Based on Lai et al. (2018a, p. 682), *data quality* means the degree to which the data needed for Big Data applications are accessible, consistent, and complete. This factor is critical for ABD because Official Statistics demand reliable and consistent data to deliver their outcomes. So, data quality includes the requirements for the adoption of datasets related to Big Data.

In fact, the quality of data is a fundamental element in keeping the Official Statistics trustworthy and reliable. In the case of NSOs, ABD follows this assumption.

All the three cases studied – ONS, INE, and IBGE – presented a solid commitment with data quality for their current Official Statistics outcomes. When ABD is the focus, firstly ONS, secondly INE and IBGE, demonstrated the same rigour in the data quality criteria as used in Official Statistics.

Regarding the process to verify the quality of data derived from Big Data applications, the NSOs deploy a quality test applied in a current Official Statistics outcome. If the Big Data outcome is approved in this control test, it is integrated into the Official Statistics portfolio of the respective NSO.

An example is the administrative records and the harmonisation methods they are submitted to in order to achieve the quality patterns of Official Statistics. This is the reason why there is a segmented run by NSOs named “experimental statistics”, where outcomes (data) derived from ABD are tested and validated. Those results can be evolved to Official Statistics just after the quality criteria are achieved by the Big Data results.

“(...) I'm talking about quality of data already. But it is very common the statistical challenge around understanding these data sources. And the last challenge is how do we ensure, publish, disseminate, explain why we use Big Data. How do we create a high-quality result from a source that can be organic? Maybe collecting for administrative purposes. Maybe from sources that we have no control over like satellite imagery. It is a particularly important part of our work to ensure that the user trusts our work until it is concluded. So, it is just thinking about public trust and what we need to do to keep people informed.” [ONS_02]

In this context, the new methods used by NSOs have to maintain the data quality of the Official Statistics. Consequently, by advancing ABD in NSOs, it is mandatory to achieve established quality standards. When the Big Data outcome is working on those standards, NSOs understand Official

Statistics acquired an improvement in their accuracy and updating capacity. This is one of the more realised benefits of ABD in NSOs. This benefit comes with “*the use of better-quality data sources*” [ONS_02] associated with other elements like:

- (i) “*the ability to bring administrative data into our production technologies for routine statistical production;*” [ONS_02]
- (ii) “*to incorporate these sorts of administrative data into our systems;*” [ONS_02]
- (iii) “*to collect and use to being a clean data set that is nationally representative*” [ONS_02]
- (iv) “*(...) to validate the entire methodology of that data, how that data was collected and worked on, and sometimes this is not very clear. So, I need to validate that these data are really reliable and useful to complement the Official Statistics from IBGE.*” [04_IBG]

Those *data quality* advances are associated with the validation of administrative records. Just when the data quality requirements are achieved, are the administrative records able to be integrated into datasets for Official Statistics generation.

“It is a mode. You can call it “collecting”. But the source is not a survey. The source is an administrative record. (...) Once again, the example of the Tax Authority: it extracts information and sends this information to us, many times without meta-information. There is no metadata; there is nothing. We have to create that. We have to clean that up, put those bases in a way that is workable.” [01_INE]

An example of Big Data technology submitted to the data quality rigour, is the Web scraping used for prices index calculation. This technology adoption involves complexity, considering the level of noise in the extracted data. It is relevant but, from the interviewee’s perspective, not very useful yet. This technology is in an ongoing process to be better applied with significant *data quality* patterns for the Official Statistics.

“(...) alternative price data is quite messy. We all know Big Data is good, but the messiness is very difficult to tackle, and this is another area that we try to build the capacity. So, two areas (AI and web scraping), as the technological front, that I am personally evolved into the technology sets.” [01_ONS]

For ABD in NSOs, the *data quality* factor is summarised as:

1. Why is data quality relevant for ABD in NSOs? Because OS demand reliable and consistent data to deliver the outcomes. [ONS, INE, IBGE]

2. How does data quality influence ABD? The influence advances by the quality requirements demanded to adopt Big Data as inputs to the OS. [ONS; INE; IBGE]
3. Requirements:
 - a. Rigorous quality for ABD projects. [ONS; INE; IBGE]
 - b. Achievement of quality standards, as harmonisation, before the integrations with OS. [ONS; INE]
 - c. Validation of administrative records before integration with traditional data sets. [ONS; INE]
 - d. Quality tests to verify the quality of big datasets. [ONS; INE; IBGE]

006. Internal vs External technologies

Baig et al. (2019, p. 9) define this factor as *“the hardware and software technologies used for Big Data adoption. Technologies revealed by retailers are internal technologies, whereas those provided by vendors are called external technologies.”* The origin of technology for ABD is relevant because it can represent cost and time savings to the respective NSO. Internal, external, and hybrid sources are selected by the NSOs based on their technological strategy and the suppliers' availability.

Moreover, there are stages to be followed. The initial work is to attempt an internal solution. If it is not possible, calls for outsourcing projects take place.

During the research, it was possible to verify that ONS demands more external and hybrid solutions than INE and IBGE. The justification is that British NSO has advanced in the initiatives and projects related to ABD more than the others.

“Yeah, Both! So, I think you must have a mixture, a hybrid. The cloud providers ... you know... Amazon, Google, Microsoft, and others... they are the sorts of groups that we would be talking about for deploying and supporting the work of specific applications for us. Many of those applications are Statistical Office depending specialised. So, they tend to be developed in-house. And lot of other things, like survey tools, that we develop in partnership with all the other Statistical Offices. So, a mixture of them all.”

[03_ONS]

Additionally, other interviewees reinforced this mix of internal, external, and hybrid solutions. The IT strategy emerges as a driver of the ADB process, indicating what needs to be internal and what

can be external or hybrid. Regarding this strategic guidance, it is noteworthy to highlight the partners and suppliers, such as universities, research centres and private companies.

Differently from the ONS and IBGE, INE presented a degree of openness for external solutions mainly when the supplier is a Portuguese University. In fact, this strategy signals an effort to promote the development of Portuguese technology.

"As for machine learning in specific, it had an in-house part but with collaboration from the University of Porto. In the development of methodologies, the academia has been requested. But it doesn't mean that there isn't also a collaboration with companies." [04_INE]

One example of the hybrid effort with partners is the ESCoE, advanced by ONS. The Economic Statistics Centre of Excellence (ESCoE) is a consortium led by ONS, where universities and research centres not only pursue technology but methodology topics as well.

"So, part of the solution might be developed from the internal research team, especially the capacity built by the Big Data campus. (...) But from my understanding, the innovative methodology is not developed by them. Basically, they try to borrow the technology, try to learn the language developed by other experts and areas to see whether they can use the algorithm, and test the algorithm do themselves to see the impact. So, that's one way of doing things. It's very, very expensive to develop something new. And secondly, for the external collaboration with their internal projects that to be supported by the so-called ESCOE. They want to see the external researchers' output as they want external researchers to devise algorithms to develop the methodology for them to use their data. So, these are two ways for the technology and for the algorithm, and for the methodology." [01_ONS]

For ABD in NSOs, the Internal vs External technologies factor is summarised as:

1. Why is the internal vs external technologies factor relevant for ABD in NSOs? Because the origin of the technology can promote cost and time savings. [ONS, INE, IBGE]
2. How does the internal vs external technologies factor influence ABD? The technological strategy defines the origin of the technology and the type of adoption. [ONS; INE; IBGE]
3. Results: combination amongst internal, external, and hybrid solutions via consortia. [ONS]
4. Requirements:
 - a. First stage: attempt to use internal solutions. [ONS; INE; IBGE]

- b. Second stage: call for outsourcing. [ONS; INE; IBGE]
- c. Suppliers: Universities, Research Centres, private companies. [ONS; INE; IBGE]

007. Perceived benefits

Park and Kim (2019, p. 2) define *perceived benefits* as “the degree to which firms perceive benefits (such as cost reduction, operation improvement, and marketing performance) deriving from Big Data.” This factor affects ABD because the NSO’s top managers base their decisions on the perceived benefits of Big Data for Official Statistics. If the technology or methodology to be adopted does not show clear benefits, the adoption proposal can be abandoned.

Thus, the benefits work as promoters for ABD, allowing NSOs to manage sorts of data storage as a service, modelling and deploying innovative data collection processes, such as using mobile devices, and transitioning to use on time data streaming (data sources) from retail chains.

ONS and INE members remarked that ABD brings benefits to the statistical processes. The major gain is an improvement in the reporting time of the outcomes. There is also an advance in data quality and accuracy of Official Statistics regarding the granularity in their analysis when Big Data is adopted.

Moreover, the diversification of data sources, data storage, data tools, outcomes, and professional skills are other benefits highlighted by interviewees from ONS. Associated with those gains, there is the benefit related to organisational redesign, with the creation of dedicated structures for ABD. In fact, the more an NSO is advanced in ABD, the more benefits are verified by its staff.

Interviewees from ONS cited four top benefits achieved when the adoption of Big Data technologies and methodologies run into an NSO. Those benefits are listed below:

- (i) Manage sorts of data storage as a service (cloud);
- (ii) Model and deploy innovative data collection processes, as using mobile devices;
- (iii) Transitioning to use on time data streaming (data sources) from retail chains;
- (iv) Plan and execute the adequate Big Data project for the NSO reality.

This innovation involves new data sources, data types, data collection devices, analytical data models, and the technological capacity to move from medium data organisation to a Big Data organisation.

In the same way, the improvement of the Official Statistics involves focus on *the data quality* improvement, such as the granularity gains in the analysis when Big Data is adopted. Those gains also support improvements in analysis accuracy of the outcomes, such as the adoption of data tools, mainly “R” and “Python” applications. So, better, faster, and more precise Official Statistics outcomes mean perceived benefits for the ONS and INE members.

“So, the benefit is that we can produce better analysis. We can explain the trends of our statistics better. Additionally, we’re able to address granularity in our reporting.” [02_ONS]

“Here, there is a great benefit, (...) transformations are so fast that if we continue to do Censuses every ten years, we won't be able to portray society in the most correct way because this period (ten years) means a very big delay. And so, there is this possibility that, administratively, we can have the Census every year or every six months or even (...) on a monthly basis.” [04_INE]

The *Perceived benefits* factor also means having a dedicated budget to support the ABD process through the NSO departments. Without budget support, the respective NSO does not have financial capacity to run ABD initiatives.

“So, ONS wanted us to do the research. So, they can improve the quality of their statistics because they adopt different layers of quality based on their budget”

[01_ONS]

Finally, there is an organisational perceived benefit derived from ABD. It refers to the creation and sponsorship of dedicated units of Big Data into the NSO structure. This kind of unit has the mission to provoke the adoption spill over into different directories and departments. The Portuguese case presented a department to organise Big Data and to call internal partners to its Big Data projects. In the same way, ONS created the Data Science Campus. On the other side, IBGE, despite the creation of the Big Data Hub, did not demonstrate the use of this unit to spread Big Data into the organisation. It showed that the Hub works more as an external body, not part of that NSO.

“What is done with this unit in terms of the technicians? Who are specialists? Their subject is how to organise this data. Then, they must work as a unit that communicates correctly with the others, to know their demands, supports... We have a big gain there...” [01_INE]

For ABD in NSOs, the *perceived benefit* factor is summarised as:

1. Why is the perceived benefits factor relevant for ABD in NSOs? Because NSOs' top managers base their decisions on the perceived benefits of the Big Data for OS. [ONS, INE, IBGE]

2. How does the perceived benefits factor influence ABD? This factor works as a promoter for ABD through:

- a. The potential capacity to manage sorts of data storage as a service (cloud). [ONS]
- b. The possibility to model and deploy innovative data collection processes, as using mobile devices. [ONS]
- c. The option for transitioning to use on time data streaming (data sources) from retail chains. [ONS]

3. Requirements:

- a. Plan and execute the adequate Big Data project for the NSO reality. [ONS]
- b. Dedicated budget as a support for ABD projects. [ONS]
- c. Organisational redesign with a dedicated Big Data unit. [ONS; INE]

008. Accuracy

Based on Baig et al. (2019), Yadegaridehkordi et al. (2018a), and Bakici et al. (2022), accuracy refers to overcoming the data collection and evaluation limitations, due to sampling, and capacity gains to deploy faster, more predictable, accountable and trustworthy outcomes.

For NSO members, accuracy is relevant for ABD because it refers to the improvements to the data collection process, such as high-frequency and granularity. Those gains represent bonus for the NSO's daily-base operations, including prediction and estimation capacities for the Official Statistics outcomes.

In the same way, the NSO members understand that accuracy is achieved by training the staff and developing their Big Data skills. This understanding is identified by ONS and INE staffs. Moreover, the Portuguese NSO ensured that ABD improved the accuracy in the Census operations. Following this perspective, IBGE remarks accuracy gains are also verified in price index surveys when Big Data is adopted.

Hence, accuracy improvement positively impacts data collection and, consequently, the outcomes delivered by the NSOs. In fact, accuracy in Big Data collection refers to high-frequency, high-speed, and granularity. This accuracy improvement is a front to be explored during ABD projects.

“Yeah, accuracy, high frequency, and alternative data sources are knowledge that I definitely see exploring.” [01_ONS]

“But I should also say that technological advances have also helped us to create better surveys, and actually, some of our surveys are the timeliest and the most accurate now.” [02_ONS]

“So, in the IBGE, we currently have the production of the price index using web scrapping. Benefits: quality in several ways. There is quality because the collection is more agile. So, you have a real time in the price index collection. Quality in terms of process standardisation also.” [09_HUB]

In the literature review, the factor “predictive analytics accuracy” is explained as the capacity to generate more accurate insights and predictions. INE members have the similar understanding: the *accuracy* improvement is linked to ABD because this adoption allows more precise predictions and insights.

“(…) because the interest of exploiting this Big Data is also about being able to increase the effectiveness and accuracy in our statistics and all of that at a relatively controlled cost, because that issue is important. And I think that INE and the statistical system can only gain by exploiting this information.” [04_INE]

Additionally, the effort from ONS to train the staff in data science brought accuracy in two dimensions: technological and organizational. Technology was more accurate with the gains of precision in the predictions of economics indices. The major example was the Gross Domestic Product estimation and the improvement this outcome obtained after the adoption of Big Data.

On the organisational dimension, the mindset of ONS staff shifted by the Big Data training programs. This initiative allowed ONS managers to develop an organisational reality in which traditional statisticians and Data Scientists could collaborate. In this context, trained staff promoted accuracy gains, because employees with updated skills are able to run Big Data projects. Those projects have delivered outcomes that are more predictable, accountable, and trustworthy.

For ABD in NSOs, the *accuracy* factor is summarised as:

1. Why is the accuracy factor relevant for ABD in NSOs? Because it refers to the improvements for the data collection process, such as high-frequency and granularity. [ONS; INE]
2. How does the accuracy factor influence ABD? Accuracy represent bonus for the NSO daily-basis operations, including prediction and estimation capacities for the OS outcomes. [ONS; INE]
3. Requirement for accuracy:
 - a. Training the staff in Big Data skills. [ONS]
4. Benefits:
 - a. Data collection with high-frequency, high-speed and granularity. [ONS; IBGE]
 - b. More accurate insights and predictions based on Big Data technologies and methodologies. [ONS; INE]
 - c. Precise estimations for economic indices, such as the British Gross Domestic Product estimation. [ONS]

009. Relative advantage

According to Sun et al. (2018, p.198), *relative advantage* means “*the characteristics of Big Data are perceived as being better than those of the idea it supersedes (e.g., its unique role for innovation, competition, productivity, customer value creation, and good business problem solution).*” In the context of Official Statistics, *relative advantage* is important for ABD because it shows to the NSO staff the advantage of adopting Big Data technologies and methodologies. This advantage is manifested by the potential results of ABD, such as innovative sorts of scale, types of data and methods.

One professional at ONS remarked that ABD brings a relative advantage for broadcasting outcomes. In his words “*sorts of scale, types of data, and types of methods*” associated with “*now-casting techniques give you data which is far more to the minute*”. [ONS_03] This ability to use multiple data sources to obtain on-time outcomes represents a *relative advantage* of Big Data over traditional methods for producing Official Statistics.

In the opposite side, IBGE was the case with less perception about Big Data's relative advantage over the traditional methods and technologies. As a result of a conservative culture, the resistance against ABD in the Brazilian NSO disturbs the modernization of Brazilian Official Statistics.

For ABD in NSOs, the *relative advantage* factor is summarised as:

1. Why is the relative advantage factor relevant for ABD in NSOs? Because it shows to the NSO staff the advantage of adopting Big Data technologies and methodologies instead of avoiding them. [ONS; INE]
2. How does the relative advantage factor influence ABD? This influence occurs through the potential results of ABD for OS, such as innovative sorts of scale, types of data and methods. [ONS; INE]
3. Benefits of relative advantage: on-time outcomes over traditional and time-consuming methods. [ONS]
4. Negative influence of the relative advantage factor: the lack of relative advantage perception for ABD resulting from cultural resistance by NSO staff. [IBGE]

010. Security and Privacy

Based on Park and Kim (2021, p. 287) this factor can be defined as security and privacy are associated with privacy invasions and security risks posed by Big Data. In the context of ABD for NSOs, this factor is critical because it can be considered a hindrance for the adoption of Big Data technologies and methodologies. This perceived risk influences the ABD process through the possibility of a cyberattack in sensitive databases used by Official Statistics.

In this research, IBGE demonstrated intense concern about this factor as an ABD restrictor. The terms cybersecurity and cyberattacks emerged during the interviews with IBGE members. This NSO understands that security and privacy is part of a daily-base struggle against hackers.

On the other side, the data analysed from INE and ONS did not present this concern clearly. In both cases, the care with data privacy is verified. However, different from IBGE, the other two NSOs consider this factor as a current part of the ABD process, not a hindrance to it.

The preventative measures for security and privacy in Official Statistics has been defined by corporate policies that include formal procedures. In the NSOs' procedures, there are corporate policies for security and privacy. Associated with those policies, the struggle for cybersecurity is part of the NSOs' daily routine.

" You don't mind the number of cyberattacks IBGE suffers every day. It is a very expressive quantity from everywhere: Europe, Asia, and Americas. It's a daily, expensive, and inglorious struggle." [04_IBG]

“Brazilians need to understand IBGE is the coordinator of the National Statistical System, and IBGE has a policy focused on privacy and security.” [03_IBG]

Security and privacy concerns are present in innovation initiatives, like with ABD. ONS welcomes proposals since they prioritise security and privacy.

“But we also allow people to highlight to us where their ideas are. Obviously, those ideas need to meet the security requirements. Nothing gets happened if it doesn't meet security requirements.” [04_ONS]

For ABD in NSOs, the *Security and Privacy* factor is summarised as:

1. Why is the security and privacy factor relevant for ABD in NSOs? Because it can be considered a hindrance for ABD. [IBGE]
2. How does security and privacy influence ABD? This factor is perceived as a daily, expensive, and inglorious struggle, considering the risk of cyberattacks. [IBGE]
3. Requirement for security and privacy: corporate policy and programs focusing on prevention. [ONS; INE; IBGE]

011. Technology competence

According to Sun et al. (2018a, p. 198), technology readiness or *technology competence* is verified when the “*organization has sufficient internal Information Technology expertise and technological infrastructure to adopt Big Data (e.g., IT knowledge and skills within the organization).*”

In the context of ABD in NSOs, technology competence is essential because this factor enables the NSO staff to require/develop Big Data technologies and methodologies, and at last, to design the adoption process.

This factor can be understood as an outcome of the employees' skills and process improvements. ABD demands from an NSO the hiring of top-skilled professionals or internal skills development. Those skilled employees spread their knowledge into the NSOs departments through the exchange of expertise about Big Data. As a result of this spill over, the NSO staff acquires technology competence for ABD. Not only IT professionals, but also specialists in economics, statistics, and business administration are necessary to take part in this staff.

During this research, ONS demonstrated a well-established set of technological competences, adopting practices such as hiring highly-qualified professionals in Big Data. In sequence, the British organisation provoked the cooperation between traditional statisticians and data scientists, data analysts and business analysts. An established data-driven culture was built based on those interactions.

So, this mixed strategy, hiring policy and internal skills development, brought a cultural transformation to ONS. The change allowed ONS to adopt Big Data technologies and methodologies in a gradual expansion. This managed change reduced the level of disturbance and resistance to ABD.

Related to skills development, the key for the success was in the continuous investment on long-term training programs. Instead of short trainings, ONS staff underwent a master's degree in data science to learn topics such as Big Data design, process design and standardization, digital data collection, data visualization, business models, and project management techniques.

In sequence, ONS deployed and established experimental statistics journeys to promote the engagement of the skilled staff. The rule is "learning by doing." Finally, the staff must be tasked to use Big Data technologies and methodologies for core activities, after the data quality criteria is achieved.

Thus, the raw material required to begin the ABD process is a highly qualified workforce drawn from the labour market. Big Data knowledge is delivered to NSOs through skills and top-tier human resources. Following those requirements, *technology competence* is forged to prepare the NSOs for ABD. The time and money required to hire a Big Data expert may be less than those required to prepare someone from inside the organization.

"My background: I have spent time in Industry as a start-up chief executive, and for companies and those are in academia using a background in physics (undergraduate). But then, I used what we would now describe as deep learning (PhD in artificial intelligence and neuroscience) to develop control brains for robots, and there is a link with Big Data. (...) So, I joined ONS at the same time as joining the public sector civil service in the UK government, when I applied for the role of Campus Director for the Data Science Campus." [03_ONS]

ONS members remarked that individual productivity is part of *technology competence*. Based on this perspective, the change from MS-Excel to Stata and from Stata to "R" and "Python" represents

a *technology competence* gain. This upgrade flow showed that *technology competence* emerges as an asset to deal with ABD. This asset is associated with the *human resources and skills* and the *organisational learning culture*, demonstrated the capacity to embrace change and innovation.

“Yep. So, let’s start with the technological side. So, there is a big push for the organization to move some old and heavy systems. We have quite a large and varied technology estate, and that’s been quite a journey. I think one of our biggest challenges is knowing which options to choose because technology moves so quickly that what looks sensible at one point in time can very quickly become the sort of new legacy system tomorrow.” [04_ONS]

In summary, *Technology competence* for ABD demands actions from the NSO. Indeed, those actions can be summarised in four points for the NSOs’ reality:

1. Why is technology competence essential for ABD in NSOs? Because it enables the staff to require/develop BD technologies and methodologies, and to design the adoption process. [ONS; INE; IBGE]
2. How does technology competence allow ABD? It allows ABD through three pillars:
 - a. Mixed strategy: hiring policy and skills development. [ONS]
 - b. Internal skills development: long-term training projects associated with experimental statistics construction. [ONS]
 - c. Moving the staff to a data-driven culture. [ONS]
3. Requirement for technology competence:
 - a. Continuous investment in training. [ONS]
 - b. Execution of core actions: to propose, adopt, plan, deploy, and establish Big Data technologies and methodologies for experimental statistics. [ONS].
4. Benefit of technology competence: individual productivity gain with “R” and “Python” applications. [ONS]

012. Technology resources

Based on El-Haddadeh et al. (2021), *technology resources* represent tangible and intangible resources such as hardware, software, human resources, skills, and experience that an organisation acquires for implementing Big Data.

During this research, specifically ONS presented four additional elements of *technological resources* for ABD in NSOs: computing, storage and processing power, data integration, and software licenses.

Those elements are associated with the capacity of the respective NSO staff to realize and demand uncommon and untraditional technological resources necessary to ABD. One component of those technology resources is the coding-controlled applications, as remarked by ONS employees. The adoption of those tools allows employees to manage and control the Big Data deployment in the Official Statistics processes.

However, the interviewees from ONS and INE stressed that this is a work in progress rather than a finished process. This statement brings the understanding that the technology resources for ABD in NSOs include other assets. One is the access to new data sources, allowing the development of experimental statistics. Another is the presence of top qualified professionals to manage the infrastructure dedicated for Big Data.

“On the one hand, an increase in technological capacity. A big investment not only in hardware - servers, for example, and processing power, for the other side. (...) As I said, it was a couple of decades. So, in that sense, there were many, many differences. And I can say that the biggest of all differences refers, exactly, to the data-warehouse area.” [01_INE]

“Now, in 2021, there was a reinforcement of this response component. There was the introduction of autocomplete function, which, as the person writes, we suggest entries for the text in the open questions. We had the use of machine learning in the coding of the descriptive; and one of the other innovations was the incorporation and the possibility of yielding information from the administrative registers.” [04_INE]

Moreover, the ABD in NSOs means transforming the traditional technological infrastructure and the medium data process into Big Data outcomes. This infrastructure is part of the ABD effort.

“Statistical Offices are on a very big journey to modernize our infrastructure for technology and for data used to manipulation. Every National Statistics Office is the same, and it always will be. This is never a journey that will be completed. (...) Over the period since then, their investment in infrastructure that ONS is doing really in a very large-scale. Now, we have system-wide organizational platforms and tools that support analysts, statisticians, data scientists, domain experts, such as

demographers, Model S, Health analysts, and so on. I am using a very rich range of data. “ [03_ONS]

Namely, the technologies to arrange Big Data include Cloud storage, peer-review version-controlled (Git Hub and Git Lab), and Python and “R” applications. Those solutions are always controlled by code and represent *technological resources*.

“(…) So, data infrastructure, which clouds systems, and tools with all of the data science stack that I would like to see from you know, free per review version-controlled Git Hub, Git Lab type approaches. Throw well-understood software development in documented practices. (…) the end applications (…) in Python and “R” packages and applications for text analysis, image analysis, for time series (…) right throw the publication using things like (R-)Markdown. And all of them are controlled by a code.” [03_ONS]

In summary, the technology resources factor for ABD in NSOs means:

1. Why is the technology resources factor determinant for ABD in NSOs? Because technology resources allow the NSO to run Big Data, that implies the use of uncommon and untraditional technologies, such as coding-controlled applications. [ONS; INE]
2. How does the technology resources factor run ABD in NSOs? Technology resources run ABD through:
 - a. Increasing computing, storage, and processing power, and data integration. [ONS]
 - b. Accessing data sources. [ONS]
 - c. Managing technological infrastructure by skilled professionals. [ONS]
3. Examples of technology resources for Big Data in NSOs: cloud storage, free per-review version-controlled (Git Hub and Git Lab), Python and “R” applications. [ONS]

013. Trialability

Baig et al. (2019, p. 10) define *trialability* as “*Trialability is the degree to which companies can experiment with Big Data before fully implementing or committing to Big Data adoption.*” This factor is necessary because it implies running experimental statistics, regarding tests with low-sensitive datasets. Testing and verifications are mandatory to verify if the proposed BD solution is adequate for the NSO need.

In fact, trialability is relevant for ABD through experimental statistics operating a pipeline of trial projects. Additionally, trialability works by the adoption of peer-reviewable Big Data applications.

In the context of Official Statistics, *trialability* arises as a dependent factor of security and privacy, financial capacity, human resources, skills, and time. Indeed, the three cases – ONS, INE, and IBGE – showed that *trialability* demands a professional's willingness to try and experiment innovations, combined with a risk-taking capacity. For ONS, a trailable mentality must come before *trialability*.

In the operational domain of NSOs, *trialability* is represented by experimental statistics, as remarked for the three cases. This type of statistics refers to a pipeline of trial projects related to ABD for improving or creating innovative Official Statistics outcomes.

Our finding was the fact that ONS is more advanced in projects to develop trial/experimental statistics and evaluate then to the Official Statistics standard. On the other side, INE and IBGE are early adopters of experimental statistics.

Those trial projects require customized and peer-reviewable procedures with the use of new methodological approaches, using specific types of data. In this context, low-security and low-sensitivity datasets are mandatory for experimental statistics, to avoid security and privacy damages.

Additionally, the supporting requirements for trialability/experimental statistics include budget, skilled professionals, IT infrastructure, and time (schedule). There is a consensus in the three cases that those requirements are mandatory for any Big Data project for Official Statistics.

“So, orderable, a peer reviewable, and then reusable for other applications. (...) But data scientists can practice, trying innovative approaches, new tools, application packages, etc. Not being loaded or set up on the corporate systems. (...) And we can use those for low security, low sensitivity datasets. So, we have much more control of we’ve run on those systems. Then, we have much less access to secure data.”

[03_ONS]

Trialability can also be understood as the capacity to take risks for innovation. In other words, adopters of experimental statistics are potential adopters of Big Data technologies and methodologies. Those departments, as risk takers, represent the entering gate for Big Data in an NSO.

“It depends on the dataset; it can depend on your risk appetite... So, price statistics, by legislation, we can’t revise our price statistics. Probably, Brazil is very similar. And therefore, you got to get it right the first time. There’s no go back and revise it. So, our risk appetite is a lot lower in that area than in some other statistics that where we might still be in an experimental state. And actually, we’re actively exploring how to use this data in more ways. We’re still looking to use this scan of data in prices. It just means the process. It’s probably longer and there’s more checks along the way. So, I think we’re very actively looking to deploy Big Data technologies.” [04_ONS_DEX]

In summary, the *trialability* of ABD in NSOs means:

1. Why is trialability necessary for ABD in NSOs? Because trialability involves experimental statistics, regarding tests with low-sensitive datasets. Testing and verifications are mandatory to verify if the proposed BD solution is adequate for the NSO need. [ONS; INE]
2. How does trialability support ABD in NSOs? Trialability supports ABD through:
 - a. Experimental statistics operating a pipeline of trial projects. [ONS; INE; IBGE]
 - b. Adoption of orderable and peer-reviewable Big Data applications. [ONS]
3. Requirements for trialability:
 - a. Development of a trailable mentality before the trialability. [ONS]
 - b. Dedicated budget, professionals, IT infrastructure, and time (schedule). [ONS; INE; IBGE]
 - c. Capacity to take risks through innovative methodological approaches. [ONS; INE]
 - d. Use of data for trials: low-security, low-sensitivity datasets. [ONS; INE; IBGE]

014. Vendor support

Baig et al. (2019, p. 9) consider this factor as *“helpful in implementing Big Data technologies. Vendors can fix information technology related issues more quickly and efficiently.”* So, it can be defined as the services and activities to support the ABD in NSOs when the solutions are bought from commercial suppliers.

This factor presented low relevance for this research. Despite this reduced influence, vendor support avoids the lock-in dependence from a specific supplier. If an NSO has only one vendor for the Big Data solution, this may pose a risk due to over-reliance on that vendor.

Related to the support offered by vendors, it works as services to support the adoption of demand programming hours and Business Intelligence services. However, it is possible to try to avoid becoming reliant on a single supplier.

Only INE considered this as an important factor for ABD, specifically for projects where business intelligence support and programming hours are involved. In fact, the Portuguese NSO showed a prevention attitude to avoid a lock-in dependence from a supplier/vendor.

"In other words, when INE contracts with a company-whether it's Microsoft or Amazon or the big players that are out there in the market - it is not just contracting for cloud storage. It is also contracting for intelligence." [01_INE]

"Our relationship with companies related to information systems, technologies and refers to software development. Typically, the development is in-house. We can hire companies. But our relationship with companies is almost for programming hours."
[01_INE]

In summary, the *vendor support* factor of ABD in NSOs means:

1. Why is vendor support necessary for ABD in NSOs? Because vendor support avoids the lock-in dependence from a specific supplier. [ONS; INE]
2. How does vendor support help ABD in NSOs? Vendor supports works as services to support the adoption of demand programming hours and Business Intelligence services. [INE]

OX2. Obsolescence

Obsolescence means the incapacity of the traditional methods and technologies to address current issues in Official Statistics, such as large volumes of data, randomness, and on-time data collection. It works under a timing perspective, from the past (traditional methods) to the present (Big Data methods).

Obsolescence is a concern for NSOs. The risk of Official Statistics being considered outdated information in the age of Big Data is well known by the staffs in Brazil, Portugal, and United Kingdom.

"And our old type of technology and old type of methods weren't really suitable for that. But that also means that what you call Big Data is a moving target. So, what we

called Big Data five years ago probably isn't big anymore, given the technology that we've got at the moment. So, you know? Years ago, the Census of the population in the UK, which was rounding about sixty-five million, would be considered Big Data. That is not Big Data anymore. Now, you can analyse it on your PC." [02_ONS]

In the context of ABD, obsolescence is critical because this factor presents the limited capacity to launch and manage Big Data long-term project by some NSOs. So, obsolescence is manifested as over-centralised technological and departmental structures, and the maintenance of traditional methodologies against the Big Data alternatives.

In the case of IBGE, the over-centralization in an organisational unit to manage technological subjects can be understood as obsolete and a constrain for the ABD. Another effect of obsolescence is the response time or outcomes delay derived from that over-centralization.

"Precisely, because of the survey that we did. We can't respond at the speed that citizens need. We have a lot of difficulty, because our time for data collection and processing is very long in relation to citizens necessities." [10_IBG]

The Brazilian NSO is outdated in its deliveries, as its own members remarked. One example of delay, as presented in Chapter 5, is the extension of the Brazilian demographic Census with estimated conclusion in 2023, May, after several postponements.

In this scenario, a feature that sustains obsolescence was found. It refers to theorizing more than executing proposals, specially about ABD projects. This statement was remarked by IBGE members.

" So, I said, 'what is the use of theorizing and not doing anything?'. But this is the IBGE culture (...) There was not a strong conception of Big Data in any directorate." [01_IBG].

Different from UK and Portugal, Brazil demonstrated a limited capacity to launch and manage Big Data long-term projects. Regarding the previous citation, obsolescence dominates where there are more discussions than actions.

In fact, ABD is a hard challenge to be addressed, and it becomes critical in cases where the NSO is suffering obsolescence. An ONS member remarked that "*Big Data is a moving target*" with

changing requirements constantly. So, the insistence to focus on traditional methodologies while Big Data projects are avoided represents limited capacity to improve and update Official Statistics.

Despite the damages and risks of obsolescence, there is an escape route for this failing. An ONS member pointed the way to overcome obsolescence and adopt Big Data. First, it is mandatory to map the flow of the data and all the processes involved. In sequence, the potential improvements must be listed. Finally, the technology resources must be defined to be deployed in a future ABD project.

Obsolescence in NSOs for ABD means:

1. Why is the obsolescence factor critical for ABD in NSOs? Because obsolescence presents the limited capacity to launch and manage Big Data long-term project by some NSOs. [IBGE]
2. How does the obsolescence factor support ABD in NSOs? Obsolescence constrains ABD through:
 - a. An outdated organisational structure: IT over-centralisation. [IBGE]
 - b. Maintenance of traditional methodologies that are insufficient for the ABD Adoption challenge. [ONS; INE; IBGE]
3. As a consequence of obsolescence, there are delays in the extended response time to achieve outcomes. [IBGE]
4. NSOs can react to obsolescence by:
 - a. Talking about ABD more than actually doing anything about it - “Much theorization, less action.” [IBGE]
 - b. Trying to overcome it by an escape route: mapping the flow of the data and all the processes being applied. [ONS]

OX3. Big Data Technologies and Methodologies (BDTM)

Big Data technologies and methodologies (BDTM) can be described as the features of the Big Data technologies and methodologies adopted by NSOs. In this context, the case studies indicate that NSOs work mainly with four Big Data technologies: Cloud, Artificial Intelligence (AI), web scraping, and Big Data analytics (BDA). Related to the methodologies, the cited examples are clustering-based alternative method for AI to detect outliers (TCBAN), natural language processing (NLP), sentiment-related information, and real-time measurement.

Considering this exemplification, the BDTM factor is determinant because the definition of technologies and methodologies drives the costs, project-time, and training demand. In the same way, BDTM works in ABD through the access and use of the listed technologies and methodologies presented before.

Those technologies are known by members of the three cases. However, each NSO is in a different adoption stage relative to the others. The British case presented an intense level of adoption, using the four technologies during its processes to develop experimental statistics and improve Official Statistics. In an intermediate stage, INE use partially specific applications of some technologies. The later adopter is IBGE, with focus on one application, and just testing others in isolated environments. An explanation for those distinct stages can be found in the specifications of each technology to be adopted. As example, web scraping demands less resources than Cloud, demonstrating that financial capacity matters for ABD in NSOs.

" The cloud, web scraping, AI, (...) All of them. On those three, we are developing applications or efforts in that direction. Web scraping, we have been experimenting in the area of price index."[02_INE]

Considering the Cloud technology, it demands significant investments, because there is a necessity to hire different Cloud service providers. ONS is the advanced adopter of Cloud applications. The British members remarked that an NSO must use more than one Cloud service provider to avoid supplier over-dependence. More contracts imply more costs.

The Cloud services works in two categories for the NSOs: storage and analytics. Related to the process of keeping Big Data, Cloud storage works not only as a data warehouse, but also as an environment where it is possible to organize huge amounts of data safely. As remarked by ONS members, the adoption of Cloud services allows NSO staff to advance in ABD projects, because it is safe to run Big Data in the Cloud environment. Beyond security, the use of Cloud storage includes the possibility to scale and expand the storage space from small to Big Data. So, Cloud storage can be considered as an infrastructure to run Big Data applications.

Additionally, there is the Cloud analytics services. In this category, the providers offer services of analytics to support the ONS team to get insights. Those insights are used to develop experimental statistics.

"So, we use pretty much all three main Cloud providers. Because we've got the cloud computing capability essentially. So, we created a secure environment on it (cloud),

and we enable our analysts to access data and methods and do their analyses on the cloud.” [01_ONS]

Artificial Intelligence (AI) has been adopted by the three NSOs. In isolated initiatives, IBGE and INE use AI algorithms to recognize satellite images to measure crop harvests and estimate agriculture statistics.

INE also uses AI, via machine learning application, to support demographic Census during the data collection. An automatic-fill service helped citizens to inform their data to the Portuguese Census website. As a result, the engagement of citizens was significant, improving the accuracy of the Census.

In an advanced stage, ONS adopted AI/machine learning associated with web scraping. In this context, technologies work together with methodologies. So, methodologies, such as natural language processing (NLP), are used by AI to measure alternative data sources. Another methodology, called “TBACN”, refers to a clustering-based alternative method for detecting outliers in web scraping outcomes.

“We’re playing with artificial intelligence. But none of that has been turned into Official Statistics yet, so to speak.” [01_ONS]

“Two areas of so-called artificial intelligence or machine learning are being used. One is for the text related to the data, and we will start to adopt the so-called natural language processing – NLP - to analyse tech-related data. We can find some kind of sentiment. For example, if we want to measure the moment of economic activity, we want to see how the confidence would be, and sentiment-related information can be measure in real-time from accounted sources based on so-called NLP. It is something that we have been explored.” [01_ONS]

Web scraping is the major Big Data technology adopted by IBGE to improve the price index calculation. This is an isolated case where Big Data belongs to the current Official Statistics process in Brazil. In the same way, web scraping is also adopted by ONS to calculate retail prices. The three cases highlighted web scraping as the simplest, cheapest, and fastest Big Data technology to be adopted by NSOs.

“And secondly, they want to see how inflation is being measured. But they want to see, for example, from the web scrapped data of online supermarket trends. Whether

we can get the data ready and then find noisy information and detect them. So, the so-called outline action. So, for this area, we will adopt some kind of so-called TBACN. So, TBACN is a kind of clustering-based alternative method for the AI to detect outliers. We try to see whether these kinds of methods can improve the measurement of the alternative data, because alternative price data is quite messy. However, we all know Big Data are good. But the messiness is very difficult to tackle, and this is another area that we try to build the capacity."[01_ONS]

These statements lead to one conclusion: ABD is more than a technological process. It combines methodology and technology within the Official Statistics business architecture. The previous citation brings two examples that technologies, such as AI and Web Scraping, work associated with methods, like NLP and TBACN.

Hence, ABD in NSOs is suitable and different from other business sectors. In this context, the methodology and the technology are an embedded mix under the same concept, Big Data.

Finally, Big Data analytics is a complementary Big Data technology, working in collaboration with the other three. Not only Cloud analytics, but also Power BI, and several dashboards developed in "R" or "Python" compose the set of analytics applications in NSOs processes.

"We are now focusing on the issue of dashboards, on Power BI... We are also going to make dashboards in ArcGIS. We also have a colleague: who is an expert in "R"? Who is going to give training and "R"? Which is a language that is used for data processing? So, it is a very common and free tool..."[05_INE]

Although IoT (the Internet of Things) was cited during the interviews, this Big Data technology is in an initial stage of adoption by NSOs. However, the perspectives for IoT are exciting, mainly in agriculture statistics.

In summary, for ABD in NSOs, the factor called *Big Data technologies and methodologies* means:

1. Why is the Big Data technologies and methodologies (BDTM) factor determinant for the ABD? Because the definition of technologies and methodologies drives the costs, project-time, and training demand. [ONS; INE; IBGE].

2. How does the BDTM factor work in ABD? BDTM works through the adoption of:

- a. Technologies (Cloud, AI/machine learning, web scraping, and Big Data analytics) [ONS; INE; IBGE];

b. and Methodologies (clustering-based alternative method for the AI to detect outliers (TCBAN), natural language processing (NLP), sentiment-related information, real-time measurement) [ONS].

3. Requirements for BDTM adoption:

- a. AI, cloud storage and cloud analytics - high investments, safety environment. [ONS]
- b. Web Scraping: reduced investment and time. [ONS; INE; IBGE]
- c. Big Data analytics as an auxiliary technology [ONS; INE; IBGE]

4. Benefits of the BDTM for ABD:

- a. Cloud computing and cloud analytics – security, scalability, and insights for experimental statistics. [ONS].
- b. Web Scraping - the simplest, cheapest, and fastest Big Data technology to be adopted by NSOs in common application as prices indices calculus [ONS; INE; IBGE].
- c. Artificial intelligence / machine learning applications – quality gains in Census operations, and satellite imagery recognition, and joined applications with web scraping [ONS; INE; IBGE].

OX10. Big Data Types (BDT)

In ABD processes, data sources and types adopted by NSOs define the concept of *Big Data types*. Considering the set of technologies – Cloud, AI, web scraping, and Big Data analytics (BDA) –, they handle different types of data to adopt Big Data. This is the reason why BDT is critical to the adoption of Big Data technologies and methodologies.

As explained by several interviewees, the data types vary depending on the planned Official Statistics outcome. In fact, NSOs can be considered adopters when they use some Big Data types, such as: (1) road traffic data; (2) shipping data; (3) satellite imagery data; (4) retail price collection; (5) electronic invoice; (6) mobile phone data; (7) electronic invoice; (8) administrative registers; (9) synthetic data; (10) on-time transports and commercial transactions. This is not an exhaustive list because ABD is an ongoing journey without a conclusion about which is the most adequate *Big Data types* for Official Statistics. In this context, data is selected considering the expected outcome

for the Official Statistics. Simplifying, the statistical question comes first; the selected data comes second.

“For Official Statistics, I think I would characterize rather than using Big Data. I would use alternative or novel data sources. Things that we are not so familiar with. And you might include in this mobile phone data, Census, and satellite imagery. You might include things like GPS and movement data from shipping, airplanes, and flights. Also, subsystems, that we are including retail scanning data from supermarkets and retail shopping. So, all of those data you can talk about Big Data too” [03_ONS]

About data sources, metadata extracted from apps and social networks are able to support Official Statistics operations, always safeguarding the security and privacy of the citizens. Those metadata, associated with transactions data, become one of the raw materials of Big Data for Official Statistics. In this way, this research identified a common requirement for data be considered Big Data: the possibility to extract, transform, and load data in real-time operations. Additionally, the data extracted from on-time sources has been deposited in cloud storage applications, as in the case of ONS, or governmental data warehouses (data lakes), like INE and IBGE exemplified.

“But I think it’s not necessarily helpful to focus on the volume or the size. It’s what they bring you that is additional and gives you an added explanatory value to your existing sources. So, this is the first point. The second point is we have many good examples both in ONS or across the Global Statistical System of using this kind of data to give ourselves capacity to create or produce faster indicators. So, for example, using mobile phone data, or in our case, traffic camera data, or using credit card financial transactions, or using Facebook Mobility Movement along with Google and Apple devices movements. All of those things to give we real-time indicators that are updated on a daily basis. This fast you talk about the V’s, Velocity and etc. Coming through in to improve to vast statistics is a really powerful addition.” [03_ONS]

To summarize, Big Data types do not substitute traditional collected data. In fact, the first complements the second, incrementally. Concluding, it is possible to observe that ABD for NSOs is not a disruptive alternative; rather, the association of Big Data types with traditional sources means the reinforcement of the current business architecture of Official Statistics.

The factor called *Big Data types* for ABD can be summarized by:

1. Why are Big Data types (BDT) critical to NSOs when adopting Big Data? Because the set of BDT handle different types of data, and those types of data are the raw material of the adoption. [ONS; INE; IBGE]

2. How do NSOs adopt BDT? NSOs adopt specific types of Big Data, such as this non-exhaustive list of nine types: (1) road traffic data; (2) shipping data; (3) satellite imagery data; (4) retail price collection; (5) electronic invoice; (6) mobile phone data; (7) administrative records; (8) synthetic data; (9) on-time transports and commercial transactions [ONS; INE; IBGE].

3. Requirement for considering data as a BDT:

- a. According to the expected outcome, the type varies [ONS; INE].
- b. Common criteria - huge and constant volumes of data, available in real-time [ONS; INE].
- c. Demanded process - extraction, transformation, and loading in real-time [ONS; INE].
- d. Complements the traditional data [ONS; INE; IBGE].

4. Benefit of BDT: incremental changes, not disruptions [ONS; INE; IBGE].

OX9. Adoption design for Big Data

The plans and project design for successful ABD is referred to as *adoption design for Big Data*. This factor is mandatory for ABD because it can reduce the resistance of the staff against BDTM, and the risk of adoption failure. As a mandatory factor, it is influenced by the actions of mentors and coaches who run the adoption projects into the departments.

By the way, the design for Big Data in NSOs arose from three distinct topics: motivation for adoption, methodological challenge, and deployment. All those features are part of the ABD process.

Top managers can motivate the staff to embrace ABD projects. Resistance can arise when the motivation of the NSO's board is modest or weak. In this scenario, the lack of motivation allows the rise of resistance against ABD.

"The interest in exploiting Big Data also has to do with the possibility of increasing the efficiency and accuracy in our Statistics and all of that at a relatively controlled cost, because that's an important issue." [04_INE]

Methodological challenge for using Big Data is a critical point. Before running Big Data projects, it is necessary to determine what is the question to be answered by an experimental or Official Statistics. Just after this question is defined, is it possible to advance the selection of Big Data technologies and methodologies. This question definition includes the problem/issue to be solved.

Regarding the proposals presented by members of the three NSOs, there is a five-stage process to address a statistical issue using Big Data. The first stage refers to the research question definition, described before. In sequence, the second stage orders to select the adequate type of data to answer the question. Thirdly, there is the recruitment effort, bringing all skilled professionals able to deal with Big Data subjects. The fourth stage is the moment when in-house solutions and capabilities are checked. If they are not available, outsourcing is recommended. Finally, the fifth stage involves verifying if the solution demands a Big Data project or not.

Another methodological challenge for ABD is how an NSO team can move from the experimental to Official Statistics. This movement from experimental to Official demands methodological advances, allowing the NSO to develop new procedures to adopt Big Data technologies. However, the challenge is dependent on the robustness of the data sources and the detection of bias. Besides the possibilities, Official Statistics demand rigorous standards of *data quality*. Those patterns include the acceptance of new Big Data technologies and methodologies. So, ABD can be both an opportunity and a challenge for NSOs project teams.

“The key challenge and your ask about my critical opinion...the key challenge is to ensure that we understand that the robustness of those data sources. So, we understand the bias, we understand any issues in the process that we use to generate those data sources that we do not take them at the face value. We look how we can to a similar, a survey coverage.” [03_ONS]

Regarding deployment, Big Data projects demands top management support for departments and project teams. It is necessary to prepare mentors and coaches to spread knowledge and create a network of trained and engaged people. This is a process for determining technological competence for adoption.

“An important part of this transformation was the ONS doing a mentorship, and mentor coaching in these other departments. There was a lot of investment in taking our analysts and putting them teaching and coaching... leaving the knowledge with them to train the trainees and expanding the network of trained people.” [05_ONS]

Deployment includes plans, descriptions and actions with a mid- and long-term duration. More than just a declaration of intent, those plans indicate a committed *business strategy orientation*, which includes the use of Big Data.

In this context, Big Data projects take place in NSOs' long-term projects. One example is the British plan to deliver a huge dataset with market prices and inflation elements of the UK economy in 2024. The prevalence of national-wide datasets to be used as a data lake can be understood as a current target to be achieved by the NSO. On the same path, INE is deploying a national-wide housing database. In the Portuguese case, the target is to count the population with a yearly frequency.

“And the second front is on the inflation of the price of the measurement. ONS currently has a plan of indoor adoption or counter data measurement for CPI to be implemented next year, and if they cannot get ready next year, they want to be at least 2024. So, within two years, we should be able to see a Big Database, the CPI, to be published in a high-frequency way from their websites. So, these are the two impacts from the ONS perspective in the near term.” [01_ONS]

So, considering the context of ABD in NSOs, the factor named *adoption design for Big Data* means:

1. Why is adoption design for Big Data mandatory for ABD in NSOs? Because it can reduce the resistance of the staff against BDTM, and the risk of adoption failure. [ONS; INE]
2. How does adoption design for Big Data influence the adoption in NSOs? It influences by the actions of mentors and coaches who run the adoption projects in the departments [ONS; INE].
3. Requirements:
 - a. Motivation from the Board to the staff (top-down) [ONS; INE].
 - b. Dealing with methodological challenge: starting with the question; find the question using Big Data [ONS; INE].

0X7. Data Restrictions

Every condition, norm, economic feature, or technological issue that restricts the access, processing, use, and broadcast of a data source or data set means a *data restriction* for ABD in NSOs. In other words, *data restriction* is important because it represents a limitation to ABD in NSOs, considering data access is fundamental to the adoption of Big Data. In the context of NSOs,

this factor operates by (a) the obstacles to access administrative records, (b) legal acts to reduce the access by NSOs, (c) resistance from Public Bodies to share data, and, mainly, (d) the block of anonymized mobile data access. Thus, those types of *data restrictions* represent a barrier to overcome by NSOs in its ABD journey.

"What we need is the continuous access of Big Data. That's the hard part for me. We should make sure we have the cell phone operators providing them. Today, it still doesn't happen."[01_INE]

"Already in a more external perspective, it is, clearly, despite the law of the legal framework, (...) the exchange of data within the Public Administration, that in each body we arrive at, we always start over again. That being so, still, culturally (into the public administration), we have to chase this."[02_INE]

"Of course, even today, there are small situations where we don't have all the information through administrative channels that is needed." [04_INE]

Definitely, the restriction of data access inhibits the ABD effort. As an example, ONS uses many data sources. However, data sources from the British property market are not accessible and accessing them is a very expensive process. On the other hand, the advance of ABD can allow alternatives to overcome the price-access limitation.

If an NSO has access to government databases, it indicates that data restrictions for Big Data use in the respective national context have been reduced. Therefore, the environmental context influences the technological one.

"Both survey data from ONS has been traditionally very stringing. But also, the increasing amounts of government administrative data is shared by other Ministries departments across the Public Sector. So, what you might describe as Big Data sources which generated by commercial organization doing some cases by people and citizens and that we acquire access free typically from outside governments."
[03_ONS]

If a government and its structures limit the access to administrative records, an NSO can fail with its ABD projects. This is the example of IBGE. The Brazilian NSO is suffering with the bureaucracy and restriction to access administrative records from de Brazilian National Health System and the Tax Authority. INE is in a similar situation, suffering some restrictions from the Portuguese

Administration. The exception is ONS. After many years of lobbying, the British NSO starts to access administrative records.

The *data restrictions* factor is critical for the ABD in NSOs considering that:

1. Why are data restriction relevant for ABD in NSOs? Because they represent a limitation related to data access, a fundamental asset for ABD in NSOs. [ONS; INE; IBGE].
2. How do data restrictions affect ABD in NSOs? Data restrictions manifests in:
 - a. Obstacles: administrative records and datasets with reduced access [ONS; INE; IBGE].
 - b. Legal acts constraining the advance of ABD projects by NSOs [ONS; INE; IBGE].
 - c. Strong resistance from National Administrations and their public bodies [INE; IBGE].
 - d. Tabu: mobile data access [ONS; INE; IBGE].

OX8. Polysemy and noise

Polysemy and noise, as a factor, refers to misunderstandings about what Big Data for Official Statistics means. This factor is a constrain in ABD because the concept of “Big Data” remains unclear when applied to Official Statistics. In this context of NSOs, this factor affects ABD by the lack of a clear Big Data definition, provoking resistance from the staff to adopt BDTM.

According to one ONS member, the term “Big Data” is unnecessary because “Big Data” has existed since the first British Census, centuries ago.

“I'm never quite sure what people mean by Big Data nowadays. So, I don't like the term. I think NSOs are always into the Big Data. So, ONS has been running the Census of the entire population for 480 years! That was Big Data in 1841.” [03_ONS]

This contestation can be considered an opportunity to establish clearly what Big Data and its adoption mean for Official Statistics. However, polysemy and noise bring trouble to ABD projects. During the research, it was possible to list three problems that a lack of standardized definition provokes in NSOs: (i) the lack of consensus about the term “Big Data” brings entropy in task groups or teams projects; (ii) noise and polysemy can provoke loss of top management support; (iii) polysemy can be used as an objection by resistance groups to avoid ABD. In fact, opponents of innovation can use those polysemy issues to spoil initiatives and adoption efforts to include Big Data in the NSOs routines.

“(...) not all the Big Data researchers agree with each other. So, there are a wide group of techniques wide group of Big Data. I think the Big Data relation researchers may not join as a common ground group workers to do the research. (...) But there's one senior data scientist who simply only likes his idea without thinking about others. So, I think that as human beings, we shouldn't think that the older generation without the Big Data background is all bad, and I think they are dogfighting within the new generations as well.” [01_ONS]

In addition to those features, polysemy for ABD shows the lack of consensus about the term, including some interviewees creating their own definition of Big Data. “Big Data” can refer to sources and uses of data or be part of the data science concept.

“Actually, this concept of Big Data is a very broad concept that is not known to everyone. Everyone can talk about it, but sometimes the purest concept is not necessarily known.” [04_IBG]

“I am thinking more of Big Data as sources and topics that are used.” [09_IBG]

“Big Data is a part of the overall concept of data science.” [10_IBG]

In summary, the term “Big Data” for Official Statistics demands as standardized definition. A clear and current definition of Big Data allows (a) the simplification of projects designs, (b) loss reductions in adoption processes, and (c) gains in knowledge exchanges between NSOs. Those three benefits can be obtained by reducing polysemy and noise.

The *Polysemy and noise* factor related for ABD in NSOs is summarised by:

1. Why is the polysemy and noise factor a constrain for NSOs when adopting Big Data? Because the concept of “Big Data” remains unclear when applied to OS. [ONS; INE; IBGE]
2. How does the polysemy and noise factor constrain ABD in NSOs? The lack of a clear Big Data definition provokes resistance from the staff to adopt BDTM. [ONS; INE; IBGE]
3. Damages of the polysemy and noise factor for ABD:
 - a. Lack of consensus among the NSOs members. [ONS; INE; IBGE]
 - b. Entropy increasing in task groups and team projects. [ONS]
 - c. Supporting resistance groups to avoid ABD. [IBGE]

4. Alternative to solve the polysemy and noise issue: demand for a standardized definition of “Big Data” for OS. [ONS; INE; IBGE]

6.1.2 Organization factors

The literature review presented seventeen TOE factors in the Organisation dimension related to the adoption of Big Data (ABD). As presented in the Chapter 4, this set of factors was used as a code book to analyse the content of the interviews (Saldaña, 2013).

During the interview analysis, those seventeen factors were validated by the three NSO (ONS, INE, IBGE) members. However, some factors presented validation and high relevance, while others were validated, but showed less importance for ABD. Additionally, three original factors emerged from the content analysis: “*future vision*”, “*potential pervasiveness*”, and “*cultural blockers*”. Clarifying, Figure 22 presented those emergent factors in the green dashed rectangle, and in the continuous one, the factors from the literature review.

Each new factor was defined to support the respective content analysis. The first organisational emergent factor, “*future vision*” presents the intention for desired and proposed future scenarios for an NSO related to ABD. Secondly, *potential pervasiveness* refers to the possibility, potentiality, or demand from other areas to adopt Big Data at a later stage. Finally, “*cultural blockers*” means practices and habits related to internal avoidance to Big Data by NSOs' staff. This avoidance is a visible effort to preserve the status quo against Big Data opportunities, or internal conflicts for power and influence.

During the analysis, each organisational factor presented relevant elements about ABD in the three NSOs – ONS, INE, and IBGE.

Like before, the analysis of each factor related to ABD in NSOs, includes the respective definition, the elements/items that integrate the factor, illustrating quotes from the interviews and respective comments. Table 22 follows the same structure as Table 21, and has the synthesis of the results from the Organisational dimension. After Table 22, the detailed analysis of each Organization factor is presented.

101. Business Strategy Orientation

This factor is defined by Baig et al. (2019, p. 11) as “*a strategy that is used to learn about business analytics and make strategic decisions for organisations in adopting Big Data.*” Regarding this

definition, *business strategy orientation* is critical because this factor promotes a mid and long-term landscape for ABD. For this reason, this factor executes ABD through the targets and projects included in the corporate strategy.

Clearly, the interviewees reinforced the necessity of a *business strategy orientation* for ABD in the context of NSOs. This necessity relies on the modernization of Official Statistics. So, the plans, descriptions, and actions with a mid- and long-term duration are part of the ABD journey. More than just a declaration of intent, those strategic plans indicate a committed *business strategy orientation* where the adoption and use of Big Data is a central part.

Table 22 – Organisational factors: Summarised findings

Organisational Factors	Consolidated Findings	Findings per case		
		ONS-UK	INE-PT	IBGE-BR
101. Business Strategy Orientation	1. Why is business strategy orientation critical for ABD in NSOs? Because this factor promotes a mid and long-term landscape for the full adoption. [ONS; INE] 2. How is business strategy orientation critical for ABD in NSOs? It works on the adoption through the inclusion of Big Data projects and targets in the corporate strategy. [ONS; INE] 3. Target for NSOs in the business strategy orientation for ABD: to become a reference of Big Data for Official Statistics in the GSS. [ONS] 4. Requirements for business strategy orientation: a. Following 4 pillars – (i) team with openness to change, (ii) ambition and risk-taking, (iii) incremental innovation, and (vi) international partnerships with GSS members. [ONS] b. Demand for a strategic plan for the adoption of Big Data (ABD). [ONS; INE; IBGE]	1; 2; 3; 4a; 4b.	1;2;4b.	4b.
102. Change Efficiency and Efficacy	1. Why is change efficiency and efficacy necessary for ABD? Because this factor can promote organisational changes to support ABD, considering the collective gains in NSOs. [ONS; INE] 2. How does change efficiency and efficacy supports ABD in NSOs? It can provoke collective engagement for the ABD project, reducing staff resistance. [ONS; INE] 3. Requirement: collective collaboration as a pivot for change efficiency and efficacy. [ONS; INE] 4. Benefit associated with change efficiency and efficacy factor: reduced time for the new deliveries using Big Data. [ONS; INE]	1; 2; 3; 4.	1; 2; 3; 4.	
103. Decision-making culture	1. Why is decision-making culture critical for ABD in NSOs? Because the spark for ABD comes from a decision to adopt or avoid Big Data. [ONS; INE] 2. How does the decision-making culture work as a critical factor for ABD in NSOs? It works through a continuum from decentralised decisions (initial stages) to centralised decisions (final stages). [ONS; INE] 3. Requirement for a decision-making culture to work in ABD: bottom-up decisions in a decentralised way, while top-down decisions in a centralised way. [ONS; INE] 4. Consequences of this factor: formal culture means centralised decision-making process, while informal culture means decentralised decision-making process. [ONS; INE]	1; 2; 3; 4.	1; 2; 3; 4.	
104. Financial Capacity	1. Why is financial capacity determinant for ABD? Because ABD projects demand financial resources to advance. [ONS; INE] 2. How does financial capacity work for ABD? It works through the management of the budget, costs, and stakeholders. [ONS; INE] 3. Requirement for an NSO to acquire financial capacity: proactively seeking funding. [ONS] 4. Consequences of financial capacity for an NSO: the possibility to execute huge investments in Big Data training and technologies. [ONS]	1; 2; 3; 4.	1; 2.	

Organisational Factors	Consolidated Findings	Findings per case		
		ONS-UK	INE-PT	IBGE-BR
105. Firm Size	1. Why if firm size relevant for ABD in NSOs? Because size impacts organisational structure, and a review is necessary for ABD in NSOs. [ONS; INE; IBGE] 2. How does firm size affect ABD? It affects through the silos effect. When an NSO is large, the existence of silos blocks ABD projects. [ONS; INE; IBGE] 3. Requirement for firm size supporting ABD: organisational integration regarding security data issues and territorial coverage on a national scale. [ONS; INE] 4. Consequences of firm size: a variable number of employees in the NSOs during the Census season. [ONS; INE; IBGE]	1; 2; 3; 4.	1; 2; 3; 4.	1; 2; 4.
106. Human resources and Skills	1. Why is the human resources factor necessary for ABD in NSOs? Because there is a competition for skills in the labour market and NSOs are affected by the lack of skilled professionals in Big Data. [ONS; INE; IBGE] 2. How does the human resources factor support ABD in NSOs? It supports ABD by hiring skilled professionals in the government or the labour market and training staff in Big Data skills. [ONS; INE; IBGE] 3. Requirements for a supportive human resources factor for ABD: a. Development of new professional profiles adequate to job descriptions related to Big Data. [ONS] b. Constant skills development and training as part of the ABD's maintenance process. [ONS; INE]	1; 2; 3a; 3b.	1; 2; 3b.	1; 2.
107. Improved design and execution efficiencies	1. Why is improved design and execution efficiency part of ABD in NSOs? Because the perception of staff about the improvements from Big Data can support the adoption expansion into the NSO. [INE] 2. How does improved design and execution efficiency affect ABD in NSOs? The improved design and execution efficiencies act on ABD through the discoveries generated by this adoption and the gains from it in the NSOs routines. [INE] 3. Requirement for the improved design and execution efficiency factor in ABD - dealing with innovation as a series of challenges and discoveries, rather than as a problem or risk. [INE]		1; 2; 3.	
108. Information security culture	1. Why is information security culture a relevant factor for ABD? Because there is a constant concern about the risk of data leaks or cyber-attacks in NSOs. [ONS; INE; IBGE] 2. How does an information security culture support ABD? It prevents data leaks and cyber-attacks, generating a safety context for the ABD. [ONS; INE; IBGE] 3. Requirements of information security culture for ABD? The adoption of the security and data privacy as dogmas by NSOs' staff. [ONS; INE; IBGE]	1; 2; 3.	1; 2; 3.	1; 2; 3.
109. Information System strategy orientation	1. Why is information system strategy important for ABD in NSOs? Because ABD demands integration of the technological areas, and isolated adoption initiatives tend to fail. [ONS; INE] 2. How does information system strategy work for ABD? It works through overcoming the silo effects in the NSOs departments. [ONS; INE] 3. Requirements for the information system strategy to support ABD: a. Focus on system integration. [ONS] b. Continuous training for updating skills. [ONS] 4. Mandatory avoidances for the information system strategy: a. Focus on data process centre perspective (outdated). [IBGE] b. Bureaucratic procurement policy — 10 years for a software license acquisition. [IBGE]	1; 2; 3a; 3b.	1; 2.	4a; 4b.

Organisational Factors	Consolidated Findings	Findings per case		
		ONS-UK	INE-PT	IBGE-BR
110. Organisation Performance	1. Why is organisation performance necessary for ABD? Because ABD projects demand measures and evaluations to verify if the adoption brings positive impact for NSO deliveries. [ONS; INE] 2. How does organisation performance work on ABD? It works through the measurements of ABD performance. [ONS; INE] 3. Requirements for organisation performance to support ABD: a. Intense use of frameworks to start adoption processes. [ONS] b. Regular evaluations of the ABD effort. [ONS; INE] c. Association with data quality and the human resources and skills strategies. [ONS; INE]	1; 2; 3a; 3b; 3c.	1; 2; 3b; 3c.	
111. Organisational structure	1. Why is organisational structure important for ABD? Because ABD demands the rebalancing of power in the structure, breaking down the silo effect dominance (a constraint). [ONS; INE; IBGE] 2. How does organisational structure allow ABD? Promoting flexible structures, task teams, project teams, and achieving agility in the deliveries. [ONS; INE] 3. Requirements for the organisational structure to support ABD: a. Organisational chart review. [ONS; INE] b. Requirement for matrix or flexible structures. [ONS; INE] 4. Results from organisational structure to ABD includes the creation of dedicated structures, such as Data Science campus (ONS), Department of data management and collection (INE), and Big Data Hub (IBGE).	1; 2; 3a; 3b; 4.	1; 2; 3a; 3b; 4.	1; 4.
112. Organisational Data Environment	1. Why is organisational data environment mandatory for ABD in NSOs? Because data resources are at the core of ABD. The organisational environment is where the data will be used, defining the adoption success, or failure. [ONS; INE; IBGE] 2. How does organisational data environment enable an NSO to adopt Big Data? This factor enables the NSO staff to use any type of data under the criteria of control, security, and quality. [ONS; INE] 3. Requirements of organisational data environment to enable ABD in NSOs: a. Move from a structured data driven NSO to a non-structured data driven one. [ONS; INE] b. Address the databases/data-lakes as the operational core of the NSO. [ONS; INE] 4. Benefits of the organisational data environment for ABD: capability to adapt to new data sources, types, and formats. [ONS; INE]	1; 2; 3a; 3b; 4.	1; 2; 3a; 3b; 4.	1.
113. Organisational Learning Culture	1. Why is an organisation learning culture critical for ABD? Because this factor drives the training and learning policies. Those policies internalize Big Data skills for the NSO staff. [ONS; INE] 2. How does organisation learning culture supports ABD in NSOs? It supports ABD through prioritization of quality rather than quantity of training, with incentives for long-term programs and the acquisition of academic backgrounds (e.g.: master's in data science). [ONS; INE] 3. Benefit for an NSO when it has an organisational learning culture focused on Big Data: the capacity to embrace changes and innovation faster. [ONS] 4. Wrong focus of organisational learning culture for Big Data: quantity rather than quality. Only short-term trainings do not promote staff engagement for ABD. [IBGE]	1; 2; 3.	1; 2.	4.

Organisational Factors	Consolidated Findings	Findings per case		
		ONS-UK	INE-PT	IBGE-BR
114. Organisational readiness	1. Why is organisational readiness relevant for ABD? Because it allows the NSO staff to seek resources for a successful adoption. [ONS; INE; IBGE] 2. How does organisational readiness work for ABD? It promotes process improvements and training policies. [ONS] 3. Requirements for organisational readiness relevant for ABD: a. Collective effort to arrange resources for ABD. [ONS; INE] b. Demanded integrated ABD initiatives. [ONS; INE] c. Testing and experimentation of technologies and methodologies. [ONS; INE] 4. Benefits of organisational readiness for ABD: organisational restructure focused on ABD, allowing the NSO to offer faster deliveries. [ONS; INE]	1; 2; 3a; 3b; 3c; 4.	1; 3a; 3b; 3c; 4.	1.
115. Project Management readiness	1. Why is Project management readiness a central factor for ABD in NSOs? This factor refers to a mandatory set of skills to adopt Big Data, relating to project management techniques. [ONS; INE] 2. How is Project management readiness mandatory for ABD in NSOs? The extensive experience in long-term projects allows the NSO staff to execute ABD effectively. [ONS; INE] 3. Requirements for Project management readiness to support ABD in NSOs: a. Adoption of institutional/strategic projects for the adoption, sponsored by top management. [ONS; INE] b. The ability to simultaneously work with short- and long-term projects. [ONS; INE] 4. Avoidance of Big Data using Project management readiness: investment in not integrated Big Data projects. [IBGE]	1; 2; 3a; 3b.	1; 2; 3a; 3b.	1; 2; 3a; 3b; 4.
116. Technological capabilities	1. Why is technological capability important for ABD in NSOs? Because it allows the NSO staff to handle fast changes keeping the production of Official Statistics. The know-how to deal with changes supports ABD. [ONS; INE; IBGE] 2. How does technological capability support ABD in NSOs? It supports ABD through the development of the ability to work within standardized frameworks by the NSO staff. [ONS; INE] 3. Requirements for technological capability supporting ABD: a. Best practices - the adoption of frameworks, such as the General Statistical Business Process Model (GSBPM). [INE] b. Reinforcement of international partnerships and exchange of best practices with others NSOs. [ONS; INE; IBGE] 4. Benefit of the technological capability for ABD: the use of different Big Data sources and Big Data structures to delivery experimental and Official Statistics. [ONS]	1; 2; 3b; 4.	1; 2; 3a; 3b.	1; 3b.

Organisational Factors	Consolidated Findings	Findings per case		
		ONS-UK	INE-PT	IBGE-BR
117. Top management support	1. Why is top management support determinant for ABD in NSOs? Because top managers are the main sponsors of ABD in NSOs. [ONS; INE; IBGE] 2. How does top management support determine the adoption or avoidance of Big Data in NSOs: Top managers can promote the adoption or the avoidance of Big Data through top-down decision. They can propose policies and programs for the adoption or impose barriers to avoid Big Data in NSOs. [ONS; INE; IBGE] 3. Requirements for top management to engage in ABD: a. Development of task-groups to spread ABD among NSO staff. [ONS; INE] b. Centralized internal environment control during the adoption process. [ONS; INE] c. Freedom environment (almost anarchic) during the early stages of adoption process. [ONS; INE] d. Incentive to the development of managerial skills in NSO staff. [IBGE] 4. Benefit of top management support for ABD: emergence of a collective way to adopt and spread Big Data solutions into NSO departments. [INE] 5. ABD avoidance in NSOs: lack of top management support to run adoption projects. [IBGE]	1; 2; 3a; 3b; 3c;	1; 2; 3a; 3b; 3c; 4.	1; 2; 3d; 5.
1X1. Future Vision	1. Why is future vision important for ABD in NSOs? Because it defines the selected vision of the NSO to redesigned internal and external organizational structures for ABD. [ONS; INE] 2. How does future vision support ABD in NSOs? It defines strategic horizons and this definition formats the strategic plans for ABD. As an example, INE has a future vision for 2031. [INE] 3. Requirements for a supportive future vision for ABD: a) Evolution of the data collection: surveys + Big Data sources. [ONS; INE; IBGE] b) Creation of a regulatory agency for Official Statistics. [IBGE] 4. Benefits for the NSO of developing a future vision for ABD: a) The capacity to design and control the future. [ONS; INE; IBGE] b) Achievement of a designed future, with censuses using mostly administrative records. [ONS; INE]	3a; 4a; 4b.	1; 2; 3a; 4a; 4b.	3a; 3b; 4a.
1X4. Potential Pervasiveness	1. Why is potential pervasiveness an important factor for ABD in NSOs? It is important because it shows the potential advance of Big Data technologies and methodologies in the NOS's internal processes. [ONS; INE] 2. How does potential pervasiveness influence ABD in NSOs? It influences adoption positively showing top management the potential process improvements in distinct NSO's departments through ABD. [ONS; INE] 3. Requirements for potential pervasiveness to become actual pervasiveness: a. Mindset change, enabling the NSO staff to learn new paradigms and adopt innovations. [ONS] b. Definition of a pervasiveness driver, like ESCoE in ONS. [ONS] c. Pervasiveness out-in, where a collaboration with the external environment spreads into NSO's departments. [ONS] 4. Potential benefit of potential pervasiveness: the adoption of Big Data technologies and methodologies for "all the sections," in "on all fronts." [ONS]	1; 2; 3a; 3b; 3c; 4.	1; 2.	

Organisational Factors	Consolidated Findings	Findings per case		
		ONS-UK	INE-PT	IBGE-BR
1X6. Cultural Blockers (CB)	1. Why do cultural blockers cause avoidance of ABD in NSOs? Because this factor promotes scepticism combined with an attempt to keep the status quo, avoiding ABD in NSOs. [ONS; INE; IBGE] 2. How do cultural blockers avoid Big Data in NSOs? They promote avoidance in soft or hard levels, from isolated initiatives to hidden groups. [IBGE] 3. Cultural blockers' actions require internal struggle for power and influence, including conservative groups of the NSO staff. [IBGE] 4. Critical damage: process obsolescence and outdated Official Statistics. [IBGE] 5. Solution to block cultural blockers: workforce renewal and engagement promotion. [IBGE]	1.	1.	1; 2; 3; 4; 5.

In the British case, the ONS team is working to change the traditional way the Official Statistics have been produced. One example is the planned delivery of a huge dataset with market prices and inflation elements in 2024. The development of national-wide datasets to be used as a data lake can be interpreted as a common strategic target to be met by several NSOs. In this context, ONS is leading this advance, following its strategic plan.

“And the second front is on the inflation measurement. ONS, currently, has a plan of indoor adoption or counter data measurement for CPI to be implemented next year. If they cannot get ready next year, they want to be at least 2024. So, within two years, we should be able to see a big database, the CPI, to be published in a high-frequency way from their websites. So, these are the two impacts from the ONS perspective in the near term.” [01_ONS]

The perceived benefits of ABD with impacts on organisational performance and efficiencies appeared mainly when this adoption was part of a *business strategy orientation*. Additionally, ABD can increase the *accuracy* of the Official Statistics when there is a long-term strategic plan to support this adoption.

In this context, Big Data and surveys are complementary, not mutually exclusive methodologies and technologies. In the initial stages, Big Data was understood as a disruptive hype. However, when the advancement of ABD takes place, it becomes clear that this innovation came to improve the quality of data for Official Statistics, not substitute them. This guided and organized adoption is a result of a *business strategy orientation*, moving from the status quo to the incremental stage.

“(...) When I joined ONS, their “mantra” was that surveys were in the last year. That only we wouldn’t need surveys anymore. That all we would need would be Big Data, transactional data, administrative data... and that we would be able to produce even better statistics than we were producing before. Since then, I think we’ve accepted that it was a very naïve approach...kind of a data strategy. Now, our understanding is that we need integrated data assets and administrative records. Basically, to pull together administrative records, transactional data, survey data...or, also called Big Data whatever.” [02_ONS]

In the case of ONS, it is clear there is an openness to innovation and *trialability*, derivate from the *business strategy orientation*. This strategy has a long-term horizon, including the ambition of that NSO to be the global reference for Big Data in the Global Statistical System (GSS).

Supplementary, the collaboration with the public sector is also part of a successful *business strategy orientation*. The target of this strategy is the improvement of Official Statistics outcomes, considering the benefit of accessing large governmental datasets and administrative records, such as public health system and tax records.

“So, National Statistics is in diamonds. It is set out in its strategy. That strategy sets out pillars and the first one is radical, which may indicate that we are very open to change. Very open to trying new things. The second is to be ambitious. So, looking for opportunities for impact everywhere on how we can improve our statistical analysis. How can we work with other parts of the public sector? How can we shape decision-making in relation to Covid-19? Around changes in global supply, and so on.”

[03_ONS]

According to this interviewee, there are four pillars to run a business strategy orientation for ABD in NSOs. The first pillar is the development of a team mentality with openness to change. Second, the creation of a mindset in the staff for ambition and risk-taking when the subject is ABD. Third, the engagement of the staff to work in incremental innovation waves, avoiding drastic disruptions. Fourth, top management team should promote and expand frequent international partnerships with GSS members.

As a rule, ABD runs when the NSO develops a clear and *focused business strategy orientation*. When the opposite occurs — lack of a strategic orientation for ABD — the result is unsuccessful with reduced coverage.

“So, it must be said that, in the medium term, not being in the strategic planning can have an impact, because for the (Big Data) Hub to grow, more than it has already grown, it will need structure. If the HUB is not included in the strategic planning, it will be more complicated for it to obtain resources to expand.” [09_IBG]

In the context of NSOs, *business strategy orientation* for ABD includes:

1. Why is business strategy orientation critical for ABD in NSOs? Because this factor promotes a mid and long-term landscape for the full adoption. [ONS; INE]
2. How is business strategy orientation critical for ABD in NSOs? It works on the adoption through the inclusion of Big Data projects and targets in the corporate strategy. [ONS; INE]

3. Target for NSOs in the business strategy orientation for ABD: to become a reference of Big Data for Official Statistics in the GSS. [ONS]

4. Requirements for business strategy orientation:

- a. Following 4 pillars – (i) team with openness to change, (ii) ambition and risk-taking, (iii) incremental innovation, and (vi) international partnerships with GSS members. [ONS]
- b. Demand for a strategic plan for the adoption of Big Data (ABD). [ONS; INE; IBGE]

102. Change Efficiency and Efficacy

Based on Baig et al. (2019), *change efficiency and efficacy* factor refers to the idea that the employees of NSOs should be capable of handling changes related to ABD and obtaining better results from this adoption. Considering this definition, this factor is important for ABD because it can promote organisational changes to support the adoption, considering the collective gains for the NSOs deliveries. For this reason, this factor influences ABD by provoking collective engagement in ABD projects, while reducing staff resistance.

In the context of NSOs, gradually, employees are dealing with the changes provoked by ABD. This movement can be understood as a change to achieve more efficiency and efficacy based on collaboration. Those changes are technological, environmental, and organisational.

“I think more and more collaboration is being established. And now, at least, there shouldn't be too many misunderstandings or lack of clarity about the kind of Big Data sources that can be used and the kind of data sets that can be adopted.”

[01_ONS]

The referred collaboration implies that it works as a pivot for *change efficiency and efficacy* as a factor. In the cases of NSOs, the change includes the switch from the traditional procedures to Big Data technologies and methodologies.

“This always ends up being a mix of seeking efficiency. But it is seeking to respond to new needs that arise. And the new needs are sometimes a question of time. Of wanting things to be available more quickly, other times it's really about covering new areas of new domains.”[03_INE]

"(...) After going through the information systems, they can go directly to the Thematic Departments, Finalistic Departments, which deal, for example, with Social Statistics." [01_INE]

For ABD in NSOs, the *change efficiency and efficacy* factor refer to:

1. Why is change efficiency and efficacy necessary for ABD? Because this factor can promote organisational changes to support ABD, considering the collective gains in NSOs. [ONS; INE]
2. How does change efficiency and efficacy supports ABD in NSOs? It can provoke collective engagement for the ABD project, reducing staff resistance. [ONS; INE]
3. Requirement: collective collaboration as a pivot for change efficiency and efficacy. [ONS; INE]
4. Benefit associated with change efficiency and efficacy factor: reduced time for the new deliveries using Big Data. [ONS; INE]

103. Decision-making culture

According to Baig et al. (2019, p. 9), *decision-making culture "refers to a belief that the adoption of Big Data is a key to success in enhancing organisational productivity. At the company level, executive or managerial staff make a decision."* Regarding this definition, a *decision-making culture* is critical because the spark for the adoption comes from a decision-making related to adopt or avoid Big Data. So, this factor works through a continuum from decentralised decisions (initial stages) to centralised decisions (final stages) during the ABD process.

This decision includes to adopt or avoid Big Data technologies and which type of technology should be adopted. In this context, the *decision-making culture* is a critical factor for ABD in NSOs. In the Portuguese case, the decision sparks in a decentralised mode and, if the proposal is feasible, it advances to a centralised stage with regards to budget and security.

"The beginning of the process is decentralized. But then, there is a centralization of the decision in the following stages. Anything that involves hiring, a big investment, even of time on the part of the technicians, all of that comes to the Governing board."

[01_INE]

In the British case, the decision-making culture in the NSOs is a mix of centralised and decentralised processes, since there are centralised top-down and bottom-up decisions in methodology adoption, including technology adoption, mindset change, and training. All of those

actions work in an internally controlled environment. This control demonstrates that *top management support* is critical to a successful ABD associated with a mix of decision-making types.

“Yes, I think there was a mix (formal and informal). In the beginning, because you needed to deliver something faster, you start to get the essential things. And, over time, you start planning, defining the framework. What comes from the conception of the Data Science Campus are those two well-defined areas, research, and training (formal). And in six months, we launched the trainee and master's program, which was absurdly fast.” [05_ONS]

Regarding the formalities of the decision-making process, it is necessary to understand that NSOs are public organisations. The adoption process in this context is formal and adheres to governmental procurement frameworks. As a result, a well-established decision-making culture is required to move the NSO from the proposal stage to the adoption one.

“It is very formal. There are two government's procurement frameworks. Which cover the way that we can talk to our suppliers. The way that we can procure and commissioned work and the way that we are we ensure high-quality delivery. So, all of those pieces have a very well understanding, a very formal process which is the same for ONS and in other parts of the British government. And then, the delivery sides that we have. We describe them as major projects. A major programs and projects approach delivery have a series well understood gateways for decision-making, risk ratings, and analysis. So, from a big scale for checks as well as our programs like a Census. We offered those in a very formal way.” [03_ONS]

At the same time, there is a mixture of centralised and decentralised decision-making culture in the organisational levels. The core technologies — cloud, infrastructure — and procurement present centralised decision-making processes. On the other hand, business areas have the autonomy to commission a supplier for a particular application or methodology.

“It's a mixture of centralized in the mixture of business area. So, different directorates. It depends on what's particular bet you talk about. So, the core IT, like networks and laptops and screens and those things is in a centralized orientation. Cloud procurement, cloud infrastructure and data processing are centralized. But for example, if the business team or the demography team want to commission a supplier to do a particular piece of analysis or develop a particular mode of them, it is a test

as a proof-of-concept that they have that control. So, it does a mixture of centralized and stretcher level procurement be like." [03_ONS]

For ABD in NSOs, the *decision-making culture* factor refers to:

1. Why is decision-making culture critical for ABD in NSOs? Because the spark for ABD comes from a decision to adopt or avoid Big Data. [ONS; INE]
2. How does the decision-making culture work as a critical factor for ABD in NSOs? It works through a continuum from decentralised decisions (initial stages) to centralised decisions (final stages). [ONS; INE]
3. Requirement for a decision-making culture to work in ABD: bottom-up decisions in a decentralised way, while top-down decisions in a centralised way. [ONS; INE]
4. Consequences of this factor: formal culture means centralised decision-making process, while informal culture means decentralised decision-making process. [ONS; INE]

104. Financial Capacity

Based on Baig et al. (2019), *financial capacity* means the skills of the staff to manage budget and financial resources in order to make available the monetary conditions for ABD. Considering the definition of this factor, it is determinant for ABD because Big Data projects demands financial resources to advance.

In the context of Official Statistics, *financial capacity* works through budget, costs, and stakeholders' management, associated with funding-seeking. Those elements compose a combination that must be managed by NSOs' top administration in an attempt to make ABD achievable.

"Now, without a doubt, Big Data technology is more expensive..." [03_INE]

"It's more the question of flexibility. (...) We are a civil service company, public administration. We can't decide alone. (...) So, the cost! I would say no. It will not be impediment when things have to be done. But this cost, obviously, has to be seen within the framework of the budget limitations that public administrations live with."

[01_INE]

"It's like this: a factor that always limits these adoption issues is the available budget. (...) Technology and license issues have already led us to abandon some solutions because it was not possible to renew the licensing." [05_INE]

The NSO adopts Big Data when it adopts methodologies, technologies, and staff training to use the Big Data. This last one, training, is a cost-intensive action and represents a critical part of the adoption process. It is mandatory to train a large number of employees to spark the adoption process inside NSOs. So, *financial capacity* is an obligation to advance in the ABD process. If an NSO does not have a robust *financial capacity*, it will probably have to postpone the adoption because this type of change demands financial resources on a large scale.

“The training effort deployed by ONS was something quite strong. The ONS launched and funded the master’s in data Analytics of Government. I was the manager of that training program. There was also a trainee program. Everything was funded. This ONS’ effort is something I consider brilliant. (...) I contracted various universities according to the strategic goals that we wanted to achieve. For example, we had to represent each country in the UK: Scotland, England, Wales, and Northern Ireland. We focused on each University and negotiated one by one, discussing the tuition fees... all that. And always getting agreements, fee reductions for the British Government students because, of course, we were paying. So, it was a very big investment. ONS created the Senior Manager, which is almost a Vice Director, to lead with those training issues.” [05_ONS]

Another financial front for ABD refers to the cost of data acquisition. Many data sources or technologies are not free, such as cloud storage or analytics. The acquisition of those new datasets and technologies represents a monetary expense, with associated risk. It is also risky because this is a new way to deliver Official Statistics, and experiments are part of the effort. If there are experiments, there are risks. As a result, the cost of Big Data is so important that ONS considers allocating a specific budget for ABD.

“I think it is important to think the importance of budgetary spending. Big Data sources are hard to find. The technologies, if we want to develop from scratch, will take a lot of effort and time, and money as well. So, at the current stage, I think ONS wants a budget solution for Big Data. If we opt for a small budget, quality will sometimes be sacrificed, which will cause some problems. As a governmental body, ONS deals with a background of tight budgets. I think it is always all over the world. I think this kind of question probably is universal.” [01_ONS]

The dedicated budget is extremely important for ABD in this context of financial capacity. The advance of the adoption process depends on the budgetary capacity. So, budgeting is part of the financial capacity factor. In the case of ONS, *financial capacity* also refers to meeting budget demands and attracting investments from the UK government. The reason is that shared procurement for technology can reduce the final price acquisition to the partners. In this case, the partners are ONS and other governmental organisations. They share the technological demands and repay investment by participating in joint governmental projects.

“Budget is always onerous.” [03_ONS]

“So, ONS wanted us to do the research. They can improve the quality of their statistics because they are different layers of quality based on their budget.” [01_ONS]

“So, it depends. (...) there is obviously a constraint that we need to go to the British Treasury to get the approval.” [02_ONS]

In summary, *financial capacity* involves budget and cost management, embedded in a deployment culture, and being able to ask for resources and deliver results on time. Then, *financial capacity* is a central pillar of every NSO that plans to adopt Big Data technologies and methodologies.

ABD in NSOs depends on *financial capacity*, that means:

1. Why is financial capacity determinant for ABD? Because ABD projects demand financial resources to advance. [ONS; INE]
2. How does financial capacity work for ABD? It works through the management of the budget, costs, and stakeholders. [ONS; INE]
3. Requirement for an NSO to acquire financial capacity: proactively seeking funding. [ONS]
4. Consequences of financial capacity for an NSO: the possibility to execute huge investments in Big Data training and technologies. [ONS]

105. Firm Size

Sun et al. (2018, p. 198) define the *firm size* factor as *“the firm’s annual revenue and number of employees that could support the adoption of Big Data (e.g., leading companies with more revenue).”* So, this factor is relevant to ABD because an organisational structure review it is necessary in the NSOs, considering their size. It affects ABD through the silos effect. When an NSO has a large size, the existence of silos blocks ABD projects.

In the context of NSOs, size really matters for ABD. If an NSO must ensure a wide territorial coverage, it needs a larger number of employees, which is an additional challenge. This challenge increases significantly during the Census season.

"The size of IBGE refers to a national company, operating in all of the twenty-seven states of the (Brazilian) Federation. In each state, we have a Data Processing Centre, which answers directly to the Information Technology Directorate. So, it is a very big coordination. In IBGE, there are around 11 thousand employees on regular time, each one with at least one piece of equipment/laptop. In Census season, for example, we have the entry of practically 190 thousand temporary professionals simultaneously... all entering in a formal process, being hired in a formal process, and having an equipment with connectivity to do the work." [04_IBG]

Additionally, NSOs deal with the silo effect, and this effect becomes more intense in large-size NSOs. However, NSOs are changing this reality by reviewing their organisational structure.

"(Silos effects) Yes. Look, we're a big organization. We're covering everything, from migration and population to health, crime, economics, environment. There are silos effects, but we are looking to break those down and to bring things together as far as we can (...)" [04_ONS]

Firm size for ABD in NSOs refers to:

1. Why if firm size relevant for ABD in NSOs? Because firm size influence organisational structure, and a review is necessary for ABD in NSOs. [ONS; INE; IBGE]
2. How does firm size affect ABD? It affects through the silos effect. When an NSO is large, the existence of silos blocks ABD projects. [ONS; INE; IBGE]
3. Requirement for firm size supporting ABD: organisational integration regarding security data issues and territorial coverage on a national scale. [ONS; INE]
4. Consequences of firm size: a variable number of employees in the NSOs during the Census season. [ONS; INE; IBGE]

106. Human resources and Skills

The meaning of the factor *human resources and skills* is based on the definition developed by Baig et al. (2019, p.8). It states that *"human resources and skills play a significant role in employing and maintaining technological innovation. Sufficient human resources are required for Big Data*

adoption. Strong programming, statistical, and analytical skills help in implementing Big Data in retail organizations.”

This is a supportive factor because there is a competition for skills in the labour market and NSOs are affected by the lack of skilled professionals in Big Data. In this context, the human resources factor supports ABD by hiring skilled professionals within government or in the market labour and training the staff in Big Data skills.

In the ONS case, it was possible to verify that the public service in the UK offers a dynamic and motivational path for civil servants to achieve new positions and deal with new challenges. ONS is part of the British public service. So, it is possible that the professional has spent ten years working in three different organisations within the British public service. The time limit is four years per organization. There are two instruments for staff mobility within the British public service: “external recruitment” and “across government.” The first instrument hires professionals outside of government, and the second hires civil servants in government bodies. Hiring is a mode for ONS to improve its staff with skilled human resources to promote ABD.

“It was perfect because they needed people for this (Big Data) area that is so specific. I stayed in the position for eight months when a vacancy for internal competition opened at the ONS. This is very common here and why I explain that people don't stay long in a department. It's very common to have loans of professionals. Loans, between governmental bodies. (...) It's a possibility to develop your career and get promotions. It's not like in Brazil where you stay in a position your whole life, with salary. In UK Civil Service there are always vacancies advertised across the Government. Internal vacancy, only with ONS employees, or when Government employee can apply for the vacancy, or for anyone from the UK. It is internal, across Government and external recruitment.” [05_ONS]

“And that's because we have a lot of methodological experience with survey data, data that comes from samples, that has an assumption of randomness. It's hard. (...)” [02_INE]

Influenced by the Environment dimension, there are some constraints related to human resources factors. One of those is hiring data science professionals to work for NSOs. The main constraint is competition with the private sector. Hiring and keeping a professional can be a steep challenge.

However, when the NSO is considered an attractive place to work in the public sector of a specific country, it can attract professionals from other public bodies.

“So, we are in a competition for those Big Data skills. So, that is a constraint. And lots of about the people are attracted to do data infrastructure. So, that is one constrain. Second constrain, in terms of workforce we have... I think we have a good balance of technical skills and the main skills. A good example, demography and health, so on.”

[03_ONS]

This constraint on hiring top-level professionals is a threat for the start of the ABD process. Knowledge about Big Data that comes into the NSOs is based on the skills of the employees. The time and money required to hire a Big Data expert may be less than the period required to prepare an internal organisation professional. By the way, those skills are necessary to prepare adequate procurement orders in the acquisition process of Big Data technologies. If an NSO does not have skilled professionals in Big Data, the adoption of those technologies and methodologies can fail in the early stages because the procurement and the project can be designed wrongly.

“My background: I have spent time in Industry as a start-up chief executive, and for companies and those are in academia using a background in physics (undergraduate). But then, I used what we would now describe as deep learning (PhD in artificial intelligence and neuroscience) to develop control brains for robots, and there is a link with Big Data. (...) So, I joined ONS at the same time as joining the public sector civil service in the UK government, when I applied for the role of Campus Director for the Data Science Campus. (...) So, my role in ONS is being fairly stable and I have spent six months in the Department of Health setting up another group to work on data analysis and data science in COVID.” [03_ONS]

Human resources and skills can be summarised as:

1. Why is the human resources factor necessary for ABD in NSOs? Because there is a competition for skills in the labour market and NSOs are affected by the lack of skilled professionals in Big Data.

[ONS; INE; IBGE]

2. How does the human resources factor support ABD in NSOs? It supports ABD by hiring skilled professionals in the government or the labour market and training staff in Big Data skills. [ONS;

INE; IBGE]

3. Requirements for a supportive human resources factor for ABD:

- a. Development of new professional profiles adequate to job descriptions related to Big Data. [ONS]
- b. Constant skills development and training as part of the ABD's maintenance process. [ONS; INE]

107. Improved design and execution efficiencies

Improved design and execution efficiencies can be defined as the efficiencies gains related to the improvements in the NSOs after ABD (Ram et al., 2019). Considering this definition, the factor is part of ABD because the perception of staff about the improvements from Big Data can support the adoption expansion into the NSO. So, *improved design and execution efficiencies* act on ABD through the discoveries generated by this adoption and the gains from it in the NSOs routines.

NSOs embrace innovation and interpret ABD as an opportunity for improvement. In this context, the motivation for innovation bypasses resistance.

“Now, for us, it's a constant discovery, because in this case of administrative records, when we're receiving them, we're handling them for the first time. We're actually chewing them up and figuring them out.” [03_INE]

As a result, the *improved design and execution efficiency* factor can be summarised as follows:

1. Why is improved design and execution efficiency part of ABD in NSOs? Because the perception of staff about the improvements from Big Data can support the adoption expansion into the NSO. [INE]
2. How does improved design and execution efficiency affect ABD in NSOs? The improved design and execution efficiencies act on ABD through the discoveries generated by this adoption and the gains from it in the NSOs routines. [INE]
3. Requirement for the improved design and execution efficiency factor in ABD - dealing with innovation as a series of challenges and discoveries, rather than as a problem or risk. [INE]

108. Information security culture

Salleh et al. (2015) understand that *information security culture* denotes the totality of patterns of behaviour in an organisation that contributes to the protection of information of all kinds. So,

information security culture is a relevant factor for ABD because there is a constant concern about the risk of data leaks or cyber-attacks on NSOs operations.

This factor supports ABD through the development of a security culture for innovation projects, such as ABD. In fact, security and data privacy are strong values for the NSOs staffs. In this culture, security and privacy come ahead of any adoption project, such as Big Data.

“Our Statistical confidentiality is something that is deeply important. The protection of individual data is really fundamental. It's a concern for us.” [03_INE]

“(…) we have here, almost 2 rhythms or 2 thresholds. The first one related to the RGPD that we are complying with. It is more specific one of security standards, of access to information, recording who ascends to the information.” [02_INE]

“The change that I perceive most eminent is the issue of information security.”
[04_IBG]

Regarding the relevance of this factor for NSOs, *information security culture* relies on:

1. Why is information security culture a relevant factor for ABD? Because there is a constant concern about the risk of data leaks or cyber-attacks in NSOs. [ONS; INE; IBGE]
2. How does an information security culture support ABD? ? It prevents data leaks and cyber-attacks, generating a safety context for the ABD. [ONS; INE; IBGE]
3. Requirements of information security culture for ABD? The adoption of the security and data privacy as dogmas by NSOs' staff. [ONS; INE; IBGE]

109. Information System strategy orientation

“The firm's information system strategy prioritises Big Data usage (e.g., information strategy, information governance policy)” (Sun et al., 2018, p.198). According to this definition, the *information system strategy* factor is important for ABD in NSOs because it demands integration of the technological areas, and isolated adoption initiatives tends to fail. In this context, the *information system strategy* factor works for ABD through the overcoming of silo effects in the NSOs departments.

Indeed, *information system strategy orientation* for NSOs embraces organisational integration while contending with the silo effect in technological areas. In the same way, training, like organisational

integration, is an important component of this strategy. So, the training effort proposes updating skills, making the workforce able to deploy the *IS strategy orientation* for ABD.

"By the way, we are having training in technologies for Big Data. It's been a constant. Every year courses come up where we...the idea is that technicians have, are familiar with those kinds of concepts, of the technologies themselves, Hadoop, and things like that." [02_INE]

The interviewees reinforced that the initial stage of an IS strategy is decentralised and moves to a centralised mode when it scales, as part of the successful ABD process. There is an internal pillar to manage the adoption process regarding two axes: security and priority.

Additionally, the interviewees signalled that there is a combination of internal, external, and hybrid solutions. The IT strategy emerges as a relevant factor in organising this adoption process, indicating what needs to be internal and what can be external or hybrid.

"It depends. We have an in-house team and a Data Science Campus. (...) But we are currently developing an IS strategy which help us to discern where and when we have to pull different levels, using different mechanisms. I don't think we have any problem with hybrid solutions. I think it is a question of pace, really. Our customers want us moving at a very fast pace and that inevitably means we have to be quite radical in the way we're thinking about this." [04_ONS]

On the contrary, a poor and outdated *IS strategy orientation* inhibits ABD. One example is the strategic focus on an outdated view of the data process centre perspective, which ignores the data-driven orientation. This poor strategy also impacts the procurement of Big Data solutions.

(...) I keep seeing this difficulty in IBGE. It does not have a policy regarding cloud usage, and I have a concern with our IT area because it is taking responsibility for taking care of this. I think it is dangerous. (...) But we inherited a mentality from something called DPC, the Data Processing Centre." [08_IBG_DGC]

"For example, buying software like Stata, we had a process that's been going on for almost 10 years! We were able to buy it just recently." [03_IBG_DPE]

"So, it must be said that, in the medium term, not being in the strategic planning can have an impact, because for the (Big Data) Hub to grow, more than it has already

grown, it will need structure. If the HUB is not included in the strategic planning, it will be more complicated for it to obtain resources to expand." [09_IBG_HUB]

The main topics about the *information system strategy* for ABD in NSOs are listed below.

1. Why is information system strategy important for ABD in NSOs? Because ABD demands integration of the technological areas, and isolated adoption initiatives tend to fail. [ONS; INE]
2. How does information system strategy work for ABD? It works through overcoming the silo effects in the NSOs departments. [ONS; INE]
3. Requirements for the information system strategy to support ABD:
 - a. Focus on system integration. [ONS]
 - b. Continuous training for updating skills. [ONS]
4. Mandatory avoidances for the information system strategy:
 - a. Focus on data process centre perspective (outdated). [IBGE]
 - b. Bureaucratic procurement policy — 10 years for a software license acquisition. [IBGE]

110. Organisation Performance

Based on Yadegaridehkordi et al. (2018b) and Picoto et al. (2021), *organisation performance* refers to the organisations that transform their business processes and systems to adopt Big Data applications, or, in the following stage, the positive results derived from this adoption.

Considering the definition above, *organisation performance* is necessary for ABD because the related projects demand measures and evaluations to verify if the adoption brings positive impact for the NSO deliveries. So, this factor works for ABD through measuring the performance of the adopted Big Data technologies and methodologies.

In this way, *organisational performance* is associated with data quality and the skills of employees in the context of NSOs. More precise measurements deployed in real-time can represent an improved quality of NSO operations. So, the delivery of economic research can follow this trend and be available to society faster than it is nowadays. A more precise and faster data collection means an improvement in the service level of NSOs. This enhancement has a positive impact on organisational performance that supports ABD expansion into others NSO departments.

"I think that they are starting to use alternative data sources to see whether they can measure the economic activity more precisely, more in a real-time sense." [01_ONS]

"The ability to use those data sources and the skills in-house, to delivering and developing Official Statistics ..." [03_ONS]

The *organisation performance* can be evaluated in terms of its balanced capacity to shift from formal to informal adoption processes, regarding the quality of the procedures established by the organisation.

"In general, I think that it's more about centralized activities. It seems that they have established kind of organizational procedures to have this alternative resources." [01_ONS]

In summary, adoption can start in an informal way. Although, it will just advance if it is fitted in a formal, defined, and centralised framework. This is not a blind rule, but a necessity to preserve the quality of data.

Additionally, *organisation performance* is verified and measured in the set of requirements to evaluate the Official Statistics outcomes. In this context, NSO members deal naturally with those requirements, probably because there is a path of past achievements used as motivational references.

"We ranked in the top 3 states in the European Union with the highest percentage of internet response. Ten years ago, it was also another very important innovation. (...) My perception is that the broad majority of technicians are receptive to these changes." [04_INE]

"Regional Service of the Azores or the Regional Directorate of Madeira have their organizational structures and have to meet a set of requirements to be part of, because they are these delegated bodies." [03_INE]

"At the level of INE, we have factors that may be more internal, which is that we have people with enough know-how, people in sufficient numbers, (...) It is really a question of responding to... the demanding calendars to which we are subject and making information more pertinent. " [03_INE]

Organisation performance for ABD in NSOs can be summarised as follows:

1. Why is organisation performance necessary for ABD? Because ABD projects demand measures and evaluations to verify if the adoption brings positive impact for NSO deliveries. [ONS; INE]
2. How does organisation performance work on ABD? It works through the measurements of ABD performance. [ONS; INE]
3. Requirements for organisation performance to support ABD:
 - a. Intense use of frameworks to start adoption processes. [ONS]
 - b. Regular evaluations of the ABD effort. [ONS; INE]
 - c. Association with data quality and the human resources and skills strategies. [ONS; INE]

111. Organisational structure

According to Sun et al., (2018, p. 198) *organisational structure* implies that *"a well-organised structure is well-suited to the adoption of Big Data (e.g., cross-organisational collaboration structure, IT departments, and staff configurations)."* In the case of NSOs, *organisational structure* is important because ABD demands the rebalancing of power in the structure, breaking down the silo effect dominance. So, this factor can allow ABD through the promotion of flexible structures, task teams, project teams, and achieving agility in the deliveries.

The *organisational structure* for ABD in NSOs includes a dedicated area, department, business unit, or sector to run and spread the adoption. Because of the use of Big Data, the organisational structure can be flexible in order to maintain the performance and deliverables.

In the Portuguese case, there is an example of flexible structure for ABD. INE adopts the matrix management model preserving the possibility of task forces and departments working with members in different Portuguese regions. Related to the ABD process, a specific core area was created in the INE structure to support the process of internalising new types of data, such as administrative records.

"The creation of a unit within the department that also changed its name. It was a department I called - which is the largest department - 'information collection' and it was renamed 'Data management and collection department'." [01_INE]

"INE has a delegation in each administrative region: the North, the Centre, the Alentejo, the Algarve and the Lisbon area. (...) obviously, INE had an evolution in the way it organized its resources. This innovation refers to the matrix management of

the resources they had. Independently of their geographical location, the resources are shares with other regional units.” [03_INE]

In the same way, there is a registered evolution in ONS regarding the organisational structure, with the creation of the Data Science Campus. For five years, ONS boosted its internal teams’ efforts to develop its own Big Data systems and methodologies. This procedure established a data environment and an appropriate *organisational structure*, overcoming the constrain of silos effect. Thus, this empowerment of the team’s work affected the learning culture, promoting collaboration rather than competition. In the end, ONS shows an organisational readiness for innovation and technological adoption in a redefined *organisational structure*.

“Over the last five years in ONS, we have really moved from separate teams, working on their own systems, using data that they have acquired themselves and using tools platforms that they have either built themselves or commissions, to build a mixture of commercial supplies and internal digital teams.” [03_ONS]

Another important fact is that external agents, such as regulatory agencies, have influenced organisational structure. External supervision can be a promoter, not an inhibitor, of the adoption. In this context, the forces to innovate are stronger than internal scepticism.

“I think we are very open. Many of our systems have reached an age we know we can’t keep them going for much longer. So, we’re having to be quite radical. So, our Finance Ministries are pushing to minimize risk. Again, they want to get continuity of service. They don’t want breaks. They don’t want mis deliveries of statistics. They want to know if we are addressing these issues.” [04_ONS]

As cited before, ONS has the Data Science Campus. It has a directory status. In this case, the five years since the creation of this unit represent a period of maturity for data science in ONS. As an initiative to embrace Big Data and an organisational redesign, the Data Science Campus is a milestone for ONS’ organisational structure. Different from other NSOs, ONS has a dedicated directory to promote ABD, regarding the respective technologies and methodologies.

In fact, the accumulated experience with Big Data usage in ONS can be verified with (a) the data science campus initiative, (b) the set of institutional cooperation with other research centres, (c)

the use of Economic Statistics Centre of Excellence² (ESCoE) as an administrative innovation tool, and (d) internal processes, like collecting, storing, and processing data. In this case, it is feasible to determine how ONS accumulates experience with Big Data usage in *organisational structure* redesigning. So, the organisational structure is not an inhibitor to innovation at NSO.

In addition to the dedicated structure for ABD, there is the hybrid nature of this unit, which works to promote the inbound and outbound adoption in the organisation. This is the Brazilian example, where the Big Data hub promotes ABD to internal and external partners. However, on the inbound scene, this hub demands more support to achieve its mission.

"And as for the hub issue, I think it's fundamental. What I think relevant is not to claim to be leadership. Firstly, let's learn. The word "hub" means arrivals and departures. So, for me, it would be arrogant trying to be the provider..." [04_IBG]

"The old plan was to create discussion hubs and we tried to force that which didn't work. Again, because of the silos." [02_IBG]

Without an adequate *organisational structure*, the ABD tends to fail. Regarding this relevance, the *organisation structure* implies:

1. Why is organisational structure important for ABD? Because ABD demands the rebalancing of power in the structure, breaking down the silo effect dominance (a constraint). [ONS; INE; IBGE]
2. How does organisational structure allow ABD? Promoting flexible structures, task teams, project teams, and achieving agility in the deliveries. [ONS; INE]
3. Requirements for the organisational structure to support ABD:
 - a. Organisational chart review. [ONS; INE]
 - b. Requirement for matrix or flexible structures. [ONS; INE]
4. Results from organisational structure to ABD includes the creation of dedicated structures, such as Data Science campus (ONS), Department of data management and collection (INE), and Big Data Hub (IBGE).

² Economic Statistics Centre of Excellence, or ESCoE, is an initiative involving ONS and National Institute of Economic and Social Research (NIESR). The proposal is to develop a centre of expertise for the analysis of emerging and future issues in measuring the modern economy, including the ABD. In <https://www.ons.gov.uk/aboutus/whatwedo/programmesandprojects/economicstatisticstransformation/economicstatisticscentreofexcellenceescoe>. Accessed in 01/01/2023.

112. Organisational Data Environment

Organisational data environment “refers to the extent to which data resources are managed in an organisation” (Verma and Bhattacharyya, 2017). In the context of NSOs, *organisation data environment* refers to the organisation’s adaptation to deal with new data sources or new data management processes.

This factor is mandatory for ABD in NSOs because data resources are at the core of ABD. In this context the organisational environment, where the data will be used, can define the adoption success, or failure. So, the *organisational data environment* enables the adoption of Big Data through NSO staff, making the employees capable to use any type of data under the criteria of control, security, and quality.

“Once again, the example of the Tax Authority: it extracts information and sends this information to us, many times without meta-information. There is no metadata; there is nothing. We have to create that. We have to clean that up, put those bases in a way that is workable. (...) For example, after going through the information systems, the data could go directly to the so-called Subject Matter Department, which deals, for example, with Social Statistics.” [01_INE]

“We have to estimate when the data comes from surveys, find ways to generalize the data to the whole population...because the data is done based on sample surveys. Using geographic coordinates, for example, to portray stretches and firms, it is something we've started to do.”[02_INE]

The *organisational data environment* is also based on the postal address database. It is a massive dataset with address records for housing, including Brazilian, Portuguese, and British citizens. This is the base for strategic and operational planning to adopt Big Data.

“The address database is one of adoption. But it is much more used in the technological part of IBGE, the part of the databases, of the data collected to support Census operations.” [06_IBG]

At the same time, the dedicated *organisation data environment* for Big Data is the drive of data structure modernization. ONS exemplifies this fact through the Data Science Campus, who acts as the maintainer of the ONS *organisation data environment*. This is a change in the organisational structure, developing an area with directory status. Not only in this context but also in the

organisation learning culture, the Data Science Campus created a specific training environment and culture for ONS members. There are two outcomes from this initiative: (i) a reinforcement of technological competence through training, and (ii) the development of a renewed organisational data environment, including unreleased data types and formats.

"And I think one defining moment was probably years ago. They decided to establish a so-called Data Science Campus. So, that not only their employees could be trained to understand the Big Data concepts, basic techniques. But also, they want to attract some kind of external collaboration within other institutions. And that's one encouraging thing that I have been quite happy to find in ONS. "[01_ONS]

Organisation data environment allows the NSO to move from a structured data driven organisation to a non-structured data driven one. In fact, Big Data sources are composed, also, by non-structured data. In the same way, NSOs staffs must be able to address the databases and data-lakes as the operational core of their activities. The ability to use all types of data represents the effective ABD by staff members of an NSO.

"The ability to use any type of data, whether it is survey data or large volume, administrative records, or unstructured text data, such as documents, or unstructured data, such as images and video." [03_ONS]

Then, the factor named *organisational data environment* presents those elements in the context of ABD in NSOs:

1. Why is organisational data environment mandatory for ABD in NSOs? Because data resources are at the core of ABD. The organisational environment is where the data will be used, defining the adoption success, or failure. [ONS; INE; IBGE]
2. How does organisational data environment enable an NSO to adopt Big Data? This factor enables the NSO staff to use any type of data under the criteria of control, security, and quality. [ONS; INE]
3. Requirements of organisational data environment to enable ABD in NSOs:
 - a. Move from a structured data driven NSO to a non-structured data driven one. [ONS; INE]
 - b. Address the databases/data-lakes as the operational core of the NSO. [ONS; INE]
4. Benefits of the organisational data environment for ABD: capability to adapt to new data sources, types, and formats. [ONS; INE]

113. Organisational Learning Culture

An *organisational learning culture* means “the learning characteristics and orientation of an organization, especially in the process of complex technology adoption” (Salleh et al., 2015). In fact, learning and training are necessary to prepare NSOs staff for the ABD process. For this reason, *organisation learning culture* is critical because this factor drives the training and learning policies. Those policies internalize Big Data skills for NSO staff.

ABD advances in an NSO through the *organisation learning culture* when it prioritizes quality rather than quantity of training, with incentives for long-term programs and academic backgrounds acquisitions (e.g.: master’s in data science).

NSO members remarked on the constant investment in training, especially in topics related to Big Data skills. In this way, an NSO's commitment to learning is also reflected in the acquisition of academic backgrounds by the employees. An example is the effort lead by ONS to support their staff achieving post-graduate degrees in Big Data technologies and methodologies.

“The training management that happened at ONS was quite strong. ONS launched and funded the master’s in data analytics of government. I was the manager of that program. There was also a fully funded trainee program. That was a brilliant initiative promoted by ONS.”[05_ONS]

In the same way, INE promotes continuous training programs to internalise Big Data skills in the Portuguese staff. In both case, British and Portuguese, the *organisation learning culture* works as a motor of ABD. Without Big Data skills, there is no staff with knowledge to run the adoption.

“They are training (the staff). For example, now, INE has a team in training. It is precisely in a course that is Data Science and Artificial Intelligence. (...) We have a lot of internal training (...)”[01_INE]

“In fact, we are having training in technologies for Big Data. It has been a constant. Every year courses come up... the idea is to make those technicians familiar with those kinds of concepts, of the technologies, such as Hadoop, and things like that.”
[02_INE]

Along with the training effort, there is the development of dedicated structures to promote Big Data learning skills in the workforce. As the example of ONS shows, the tipping point was the Data

Science Campus' creation. This is one structural achievement that boosted the organisational learning culture for Big Data in the British case.

Furthermore, ONS's effort to train staff in data science is another initiative that has yielded benefits in two dimensions: technological and organisational. Technology improved in terms of accuracy with the fast achievement of an outcome: GDP estimation.

“Yes. So, it started, along with training. And it wasn't something restricted. It was something continuous, including change the mindset of the workforce.” [05_ONS]

On the organisational dimension, the staff's mindset shifted during ABD in ONS. At the same time, the data science specialists took part in the ONS operations. This action noted that it was possible to manage an organisational reality where traditional statisticians and data scientists could work together.

One example of those changes is the migration from MS-Excel to STATA and from Stata to “R” and Python. This is an example of how the learning culture for Big Data influences the daily routines. So, the capacity to embrace changes and innovation faster is a benefit obtained from the *organisational learning culture* for Big Data.

“So, productivity! We took the productivity production system out of Excel, moved it into Stata, because many of our economist use Stata through university... this happened only four or five years ago. But today, most economists come out of university with “R” or Python. They're not using Stata. That seems a very sensible transition to make four or five years ago. Now, I'm having to think about transitioning and moving that system into “R”, or Python. Similarly, a lot more use of cloud” [04_ONS]

Despite the *organisation learning culture* being part of the NSOs' routines, if this factor is not well-set for ABD, the results are frustrating. Long-term learning projects have to substitute, or at minimum, work together with short-term training programs because the immersion provokes engagement beyond the basics. In IBGE, there is a focus on quantity rather than quality, with the adoption of short-term programmes with reduced outcomes.

“I think IBGE quickly needs a training program for Data Science” [08_IBG_DGC]

“Big Data has to get away from the idea that it is something in the future. No! We have to bring it to the present of IBGE.” [03_IBG_DPE]

"Because when we talk about learning in Data Science, it's long-term learning. And it requires something that, in the culture of IBGE, is a problem, the interpersonal relationships. It is the issue of communication, of exchange." [10_IBG_TRN]

The *organisation learning culture* is therefore another critical factor for ABD in NSOs. It can be summarized by the listed elements below.

1. Why is an organisation learning culture critical for ABD? Because this factor drives the training and learning policies. Those policies internalize Big Data skills for the NSO staff. [ONS; INE]
2. How does organisation learning culture supports ABD in NSOs? It supports ABD through prioritization of quality rather than quantity of training, with incentives for long-term programs and the acquisition of academic backgrounds (e.g.: master's in data science). [ONS; INE]
3. Benefit for an NSO when it has an organisational learning culture focused on Big Data: the capacity to embrace changes and innovation faster. [ONS]
4. Wrong focus of organisational learning culture for Big Data: quantity rather than quality. Only short-term trainings do not promote staff engagement for ABD. [IBGE]

114. Organisational readiness

Ghaleb et al. (2021) define *organisational readiness* as *"the extent to which financial, technological, and skilled human resources are accessible to an organisation wishing to embrace a technology."* *Organisational readiness* can be understood as the capacity of an organisation to use resources or look for them to adopt Big Data.

Organisational readiness is relevant because it allows the NSO staff to seek resources for a successful ABD. So, this factor works for ABD through process improvements and training policies.

It includes the redefinition of professional profiles, for example. In fact, *organisational readiness* presents singularities for each NSO because IBGE, INE, and ONS work in different environments.

However, there are common elements for this organisational factor including resource-seeking capacity, process improvements, testing and experimentalism, and organisational restructuring.

Resource-seeking includes human, financial, and technological assets to support ABD. This seeking is a collaborative effort by various NSO units with the common goal of innovating.

"(...) we have other people from various areas, such as Geoinformation, from methodology, from computer development. A lot of people involved in collecting information, both from businesses and from households." [02_INE]

Equally important is the integration of the initiatives, process improvements, and focused training. Those actions are key to improving the *organisational readiness* for ABD. As example, ONS intends to create its own adoption process for Big Data. This process runs on experimental statistics, which means adopting external technologies and methodologies, adapting them to the internal process, and testing the viability of the outcomes. In this description, the creation of experimental statistics to adopt Big Data techniques is an example of *organisational readiness*.

"For example, for the first group, in terms of involving NLP, they want to see whether the dynamics of construction markets can be measured. So, they can measure the construction output statistics better. For that specific series, it was not qualified as National Statistics status. But with the improvement, and we are happy to see that, they increase the quality. So, the previous downgraded series, upgraded again to be a National Statistics series again. So that's one immediate impact of the Big Data."

[01_ONS]

Organisation readiness is also presented in the capacity of an organisation to arrange resources, as partnerships, to evaluate and innovate. ONS ran this path with the development of organisational structures to deploy innovations about Big Data. One example of the dedicated structure is the ESCoE (Economic Statistics Centre of Excellence) to exchange knowledge about ABD for Official Statistics.

"I also set up along the way the Economic Statistics Centre of Excellence (ESCoE). So, my professional journey isn't so much the production of economic statistics. It is more the research end, although I do some production around capital data and productivity data." [04_ONS]

The lack of integrated ABD initiatives, associated with failed collective effort to access funding results in failures and delays of the adoption in the respective NSO. This is a damage not only to an organisation, but also to a government and the citizens.

"I see a lot of initiatives that are solitary. Sometimes you have great initiatives from people alone within a state unit." [10_IBG]

"In my opinion, we've made a lot of purchases that we didn't need, for the promise of technology to solve all problems. What solves problems is process improvement and training people. There is no point in putting a supercomputer on the table without knowing how to turn it on." [07_IBG]

In summary, *organisational readiness* includes those elements:

1. Why is organisational readiness relevant for ABD? Because it allows the NSO staff to seek resources for a successful adoption. [ONS; INE; IBGE]
2. How does organisational readiness work for ABD? It promotes process improvements and training policies. [ONS]
3. Requirements for organisational readiness relevant for ABD:
 - a. Collective effort to arrange resources for ABD. [ONS; INE]
 - b. Demanded integrated ABD initiatives. [ONS; INE]
 - c. Testing and experimentation of technologies and methodologies. [ONS; INE]
4. Benefits of organisational readiness for ABD: organisational restructure focused on ABD, allowing the NSO to offer faster deliveries. [ONS; INE]

115. Project Management readiness

Based on Bakici et al. (2022), *project management readiness* means the organisational expertise for working with project management, a portfolio of projects, project frameworks, and different sizes of projects in a project-oriented organisational structure. Considering this definition, *Project management readiness* is a central factor because it refers to a mandatory set of skills to adopt Big Data, the project management techniques.

In this context, *Project management readiness* influences ABD through the extensive experience in long-term project, allowing the NSO staff to execute ABD effectively. This effective adoption represents a journey from strategic definition until the operational adoption.

As a planned effort, the adoption involves planning, budgeting, procurement, and deployment associated with training. Those elements are part of project management effort for the ABD.

An example of the *Project management readiness* is the ONS case. It has a project-oriented portfolio. So, this orientation signals the capacity to work under a project management methodology, improving the success of the ABD process in NSOs.

"(...) If we put in the time in terms of the project research, basically, my connection is purely project-based. I'm a kind of advisor. But whenever they have some project, I will work full time for a specific period." [01_ONS_PRO]

ABD demands a mix of *human resources and skills* associated with *technology competencies*. Both factors working under the project management orientation. This blending occurs at various stages of the ABD project. So, *project management readiness* allows the NSOs to be able to deal with the *complexity* of the adoption process.

"To underpin that, we need the technology as in computing power, the software, the methods that enable the ratio of the data. But also, the analysis of the data has been integrated. Those adoptions are advanced by projects." [02_ONS]

As a critical factor, *project management readiness* is also present in the development of new Official Statistics with the use of Big Data. For instance, the new methodologies and techniques have to follow project stages: (1) understanding the kind of research, (2) testing experimental statistics, (3) ending up being used, (4) deployment, and (5) regular production.

Those stages demand a set of actions to run the ABD project, including: (a) to define how to visualize the research question; (b) to identify the method to best support the answer searching; (c) to define the methodologies and technologies to be adopted; (d) to observe the results.

Regarding the duration, *project management readiness* demands the ability to work with short- and long-term projects. As an example, INE has an agenda dedicated to long-term projects with the European Union, including one related to ABD. Phases, checkpoints, short-term deliveries, and deadlines are daily-basis elements for the NSO staff during ABD.

"We have the big project to move the Census to an administrative basis." [01_INE]

" We have there, more in terms of experimenting. We are in a European project called precisely Big Data, which is a project that is in several lines and one of them has to do with Smart farming" [02_INE]

Even though an NSO plans to adopt Big Data, the lack of *project management readiness* is an obstacle to the success of this journey. If ABD is not the central objective of the corporate projects, this undervaluation leads to no engagement and isolated initiatives. Additionally, managers without skills in project management techniques worsen the delays.

“The Big Data adoption is slow, also because the project is not institutional...(…) It's a fictional project...like 'let's use it!' and it dies in the sequence. It's a concentrated effort in a coordination where you have people interested.” [02_IBG_PRO]

Project management readiness for ABD in NSOs includes:

1. Why is Project management readiness a central factor for ABD in NSOs? This factor refers to a mandatory set of skills to adopt Big Data, relating to project management techniques. [ONS; INE]
2. How is Project management readiness mandatory for ABD in NSOs? The extensive experience in long-term projects allows the NSO staff to execute ABD effectively. [ONS; INE]
3. Requirements for Project management readiness to support ABD in NSOs:
 - a. Adoption of institutional/strategic projects for the adoption, sponsored by top management. [ONS; INE]
 - b. The ability to simultaneously work with short- and long-term projects. [ONS; INE]
4. Avoidance of Big Data using Project management readiness: investment in not integrated Big Data projects. [IBGE]

116. Technological capabilities

As defined by Park and Kim, (2021, p. 287) *technological capabilities* refers to "*the degree to which firms have the ability to operate Big Data systems in terms of technological readiness.*"

Regarding this definition, the *technological capability* is important for the ABD in NSOs because it allows the NSO staff to handle fast changes, keeping the production of Official Statistics. So, the know-how to deal with changes supports ABD.

In fact, *technological capability* supports ABD through the development of the ability to work within standardized frameworks by the NSO staff.

Considering that Big Data is a mix of technologies and methodologies, the adoption of standardized frameworks to develop Big Data solutions represents a stage for the adoption by NSOs. One of those frameworks is the General Statistical Business Process Model (GSBPM), orientated to define processes and responsibilities in the production of Official Statistics.

“We use it in statistics which is GSBPM, General Statistical Business Process Model. It was developed in conjunction with the UNECE. So, we can see how important those international entities are.” [02_INE]

Meanwhile, the learning curve for technological capabilities in NSOs includes international cooperation with the United Nations' global platform for statistics and data. Additionally, cooperation with other NSOs improves the accumulated knowledge about Big Data technologies and, consequently, *technological capability*.

“So that's got a first point. And an example of us would be using the shipping GPS data which is not as the ONS identification system information. We use this to produce rapid estimates that can feed to an import/exports analysis. (...) And it is made available for the UN Global platform for statistics and data. Which we are working on in collaboration with other parts, with other ONSs across the world.” [03_ONS]

Again, the relevance of Big Data methodologies emerged during the interviews. So, the critical initial point for ABD is the capacity to ask the right question about a survey. Indeed, the accumulated knowledge about Big Data improves the capacity of NSO members to define targets for the improvement of the Official Statistics through the construction of experimental statistics. Therefore, *technological capabilities* and competences are reinforced through international partnerships and learning to follow best practices.

“The second one is the old problem: what information do I have that shines lights on the questions that I'm trying to answer? So, this is always the same problem of all the sources.” [03_ONS]

Another example of the technological capability refers to the ability to handle drastic changes using technology. A case in point, the staffs from the three NSOs (IBGE, INE, and ONS) changed quickly to the home office work regime during the COVID-19 pandemic. As an asset, this capacity is a differentiator to run ABD projects.

“I think that a remarkable thing for me in this Pandemic impact. It was the way that the IBGE reinvented itself. (...) Of course, technology helped. But I will be honest, we were not prepared for this, and we had to prepare ourselves. I think that this resilience is fundamental. Sampling techniques, statistical methods, forms, are not enough if you don't have people in the house with this capacity. So, I think that was also a very important moment.” [03_IBG]

Additionally, *technological capability* brings a benefit for the NSO who adopts Big Data: the capacity to use different Big Data sources and Big Data structures. This innovative use allows the NSO to move from the experimental to the Official Statistics outcomes.

Technological capability factor includes features such as:

1. Why is technological capability important for ABD in NSOs? Because it allows the NSO staff to handle fast changes keeping the production of Official Statistics. The know-how to deal with changes supports ABD. [ONS; INE; IBGE]
2. How does technological capability support ABD in NSOs? It supports ABD through the development of the ability to work within standardized frameworks by the NSO staff. [ONS; INE]
3. Requirements for technological capability supporting ABD:
 - a. Best practices - the adoption of frameworks, such as the General Statistical Business Process Model (GSBPM). [INE]
 - b. Reinforcement of international partnerships and exchange of best practices with others NSOs. [ONS; INE; IBGE]
4. Benefit of the technological capability for ABD: the use of different Big Data sources and Big Data structures to delivery experimental and Official Statistics. [ONS]

117. Top management support

Top management support is a determinant factor for ABD. Based on Park & Kim (2019), *top management support* refers to the degree to which high-level management recognises the importance of Big Data, sponsoring and supporting the adoption process. Considering this definition, *top management support* is important because top managers are the main sponsors of ABD in NSOs.

The three cases (ONS, INE, IBGE) showed that *top management support* determines the adoption or avoidance of Big Data in NSOs through top-down decisions. They can propose policies and programs for the adoption or impose barriers to avoid Big Data in NSOs.

This factor allows the staff to innovate in the proposal to achieve a successful ABD. In the context of NSOs, task groups worked as the initial spark for the adoption. Each task group had a specific theme, and all of them interacted with each other. If it were necessary, a task group might access the support of *top management* to achieve its objectives. In fact, the initial adoption process was decentralised, as represented by the task groups.

“(...) it was kind of separated task groups that they set up their own theme. So normally, this kind of group is within specific stream of statistics. (...). And for some sector analysis teams, and they also want to have their own priority. So, they may

want to contact different groups and they may put their base in their own way with some kind of support from the top management team.” [01_ONS]

(...) The transformation launched in 2016, running in a Top-down mode.” [05_ONS]

Additionally, *top management support* can take place in a collective way with the association of managers in different NSO's departments supporting ABD, as occurred in INE. This collective approach shows the communal feature of top management's support action.

“More from the managerial level. I think there is already more care in terms of the management, even for the implementation of future projects, to take this issue of Big Data into account.” [05_INE]

However, the absence of *top management support* can block the ABD process. Indeed, this lack of an effective *top-management support* can allow the emergence of a limited adoption process with a reduced impact.

“I think that the managerial development. I see that it is an urgent need for IBGE to develop this culture, this digital mentality, this mentality of social response as well.” [10_IBG]

“But I think what's missing is the lack of a top-down decision. When the top management decided to implement an area, or create an area, implement a study on the use of Big Data, the resource will show up. Because I don't consider that the IBGE is completely understaffed. And even less, it doesn't have expertise. I think if the staff does, expertise does. But this has to be a top-down project to gain resources and support.” [09_IBG]

In summary, *top management support* for ABD in NSOs includes:

1. Why is top management support determinant for ABD in NSOs? Because top managers are the main sponsors of ABD in NSOs. [ONS; INE; IBGE]
2. How does top management support determine the adoption or avoidance of Big Data in NSOs: Top managers can promote the adoption or the avoidance of Big Data through top-down decision. They can propose policies and programs for the adoption or impose barriers to avoid Big Data in NSOs. [ONS; INE; IBGE]
3. Requirements for top management to engage in ABD:
 - a. Development of task-groups to spread ABD among NSO staff. [ONS; INE]

- b. Centralized internal environment control during the adoption process. [ONS; INE]
 - c. Freedom environment (almost anarchic) during the early stages of adoption process. [ONS; INE]
 - d. Incentive to the development of managerial skills in NSO staff. [IBGE]
4. Benefit of top management support for ABD: emergence of a collective way to adopt and spread Big Data solutions into NSO departments. [INE]
5. ABD avoidance in NSOs: lack of top management support to run adoption projects. [IBGE]

1X1. Future Vision

This is the first emergent organisational factor identified during this research. *Future vision* refers to the intention of the members about the desired and proposed future scenarios for an NSO regarding ABD. Considering this definition, the *future vision* is important because it defines the selected vision of the NSO to redesigned internal and external organisational structures for ABD.

In this context, *future vision* supports ABD in NSOs through the definition of strategic horizons. This definition formats the strategic plans for ABD.

Under this factor's scope, there are three clear visions or targets: (i) the association between administrative records and traditional surveys; (ii) the evolution of Census operations; and (iii) internal and external cooperation.

(i) The association between administrative records and traditional surveys

This was a consensual point in all three NSOs about the future vision.

" But this is the future line of combining administrative records with IBGE surveys."

[01_IBG]

"Basically, the future is to integrate data assets, rather than one type of data. The integration of different types of data." [02_ONS]

"What is the big bet of the last few years? To use this very varied panoply of administrative records to assist the production of Official Statistics." [03_INE]

(ii) The evolution of Census operation

By shifting the Census format or spreading the use of data sources, NSOs are designing their future. Indeed, the evolution facilitated by ABD is an ongoing project with an expected outcome in the coming years.

"And our goal is no more CENSUS in a conventional way. Probably, this (Census 2021) will be the last conventional one."[01_INE_PRE]

"I can give you an example in relation to the next Census. So, in ten years - in 2031 - we are already in a phase of data collection via administrative sources, achieving uniformity and avoiding to do the in-person Census."[05_INE]

(iii) Internal and external cooperation

Future vision calls for internal cooperation to eliminate the silo effect within IBGE as well as the creation of a regulatory body to oversee Official data.

"The creation of a Statistical authority to supervise the NSO. This Statistical authority will deploy a standardization and normalization for the Official Statistics." [02_IBG]

For NSOs, the *future vision* about ABD refers to:

1. Why is future vision important for ABD in NSOs? Because it defines the selected vision of the NSO to redesigned internal and external organizational structures for ABD. [ONS; INE]
2. How does future vision support ABD in NSOs? It defines strategic horizons and this definition formats the strategic plans for ABD. As an example, INE has a future vision for 2031. [INE]
3. Requirements for a supportive future vision for ABD:
 - a) Evolution of the data collection: surveys + Big Data sources. [ONS; INE; IBGE]
 - b) Creation of a regulatory agency for Official Statistics. [IBGE]
4. Benefits for the NSO of developing a future vision for ABD:
 - a) The capacity to design and control the future. [ONS; INE; IBGE]
 - b) Achievement of a designed future, with censuses using mostly administrative records. [ONS; INE]

1X4. Potential Pervasiveness

This is another emergent factor identified in this research. *Potential pervasiveness* means the possibility, potentiality, or demand from other areas to adopt Big Data at a later stage. It is a relevant factor for ABD in NSOs because it shows the potential advance of Big Data technologies and methodologies in the NOS's internal processes.

Potential pervasiveness influences positively the adoption showing to top management the potential processes improvements in distinct NSO's departments through ABD.

Regarding the context of NSOs, there is a large potential to spread Big Data technologies and methodologies into departments and areas. Aside from the possibility of failure in some specific sectors, the adoption effort must be carried out across the entire organisation.

“I think they probably start to think about adopting Big Data in nearly all the departments. So, all the statistics sections. So, all the Official Statistics areas. So, they may not adopt it in the end, but they are testing on all fronts.” [01_ONS]

There is a clear potential to spread Big Data in all Official Statistics areas. The incentive for adoption is dispersed on "all fronts," as an interviewee remarked. So, the pervasiveness has been built with a mindset change and runs on an organisational culture able to learn new paradigms and adopt innovations. The mindset change is a key to opening the organisation for ABD and, consequently, pervasiveness.

In the British case, the pervasiveness of Big Data is also organised into ONS by a specific structure, the ESCoE. Universities, NIESR, and ONS are all part of this scientific consortium. It brings together researchers from those organisations to propose solutions to ONS-posed problems. As an open innovation system, the partners develop the solutions and propose them to ONS.

“The Economic Statistics Centre of Excellence (ESCoE) is a consortium of leading institutions led by the National Institute of Economic and Social Research (NIESR) with King’s college London, Innovation Foundation Nesta, the University of Cambridge, Warwick business School (University of Warwick), and Strathclyde Business School.” [01_ONS]

Guided by topics that demand innovation, the ONS staff attempt to develop solutions. If this alternative is not available, the partners are called to provide support. This is an example where external solutions are adopted and spread into the ONS business cycle. This is referred to as a pervasiveness out-in, as it stems from a collaboration with the external environment and spreads into ONS.

ESCoE, as a pervasiveness driver, has specific key roles such as to (i) deliver research and analysis that enables ONS to meet its vision of delivering world-class economic statistics, (ii) deliver an international annual conference to facilitate the sharing of best practice and world-class challenge to the work of ESCoE, (iii) work in collaboration with the ONS Data Campus to develop new techniques."

"I think with the so-called ESCOE wants to define the research themes that may be interested in their teams while in advanced before they do the research. So, they first get the topic, in internal discussions. They find the topic that they want to work with."

[01_ONS]

Summarising the *potential pervasiveness*, this factor includes:

1. Why is potential pervasiveness an important factor for ABD in NSOs? It is important because it shows the potential advance of Big Data technologies and methodologies in the NOS's internal processes. [ONS; INE]

2. How does potential pervasiveness influence ABD in NSOs? It influences adoption positively showing top management the potential process improvements in distinct NSO's departments through ABD. [ONS; INE]

3. Requirements for potential pervasiveness to become actual pervasiveness:

a. Mindset change, enabling the NSO staff to learn new paradigms and adopt innovations. [ONS]

b. Definition of a pervasiveness driver, like ESCoE in ONS. [ONS]

c. Pervasiveness out-in, where a collaboration with the external environment spreads into NSO's departments. [ONS]

4. Potential benefit of potential pervasiveness: the adoption of Big Data technologies and methodologies for "all the sections," in "on all fronts." [ONS]

1X6. Cultural Blockers (CB)

This is the last emergent organisational factor found in this research. *Cultural blockers* foster avoidance. It refers to the groups and practises related to the avoidance of Big Data or any other type of innovation in NSOs.

Considering this definition and nature, this factor is a constrain for ABD because it promotes scepticism combined with an attempt to keep the status quo, avoiding Big Data in NSOs. In this way, resistance is verified as a visible effort to preserve the status quo against Big Data opportunities, including internal conflicts for power and influence.

In the same way, cultural *blockers* promote the avoidance of Big Data in soft or hard levels, from isolated initiatives to hidden groups acting to keep the status quo. The soft level is related to scepticism and low resistance in more isolated occurrences. The hard level, on the other hand, refers to the strong opposition to ABD, which includes negligence, internal anti-adoption lobbying, and influence network building for avoidance.

So, in the context of NSOs, staff resistance to change, in the form of scepticism, can be overcome gradually, considering the examples from other departments. If the adoption of ABD does not evolve, this scepticism will remain at a high level.

“Initially there were some puzzlingly form. I shouldn't say the generational gap, but some senior guys that were a little bit sceptical about the usefulness. They put probably more trust in their own experiences in dealing with these kinds of data.”

[01_ONS]

Indeed, there is opposition for ABD in NSOs staff. As an example, in the initial stage of ABD inside ONS, or, in the interviewee's words, in the "*baby stage*" of adoption, there were "*probably still some top guys that are not fully convinced about the Big Data thing.*" If top management's support is not strong enough to change this cultural resistance, the adoption effort can fail. Fortunately, the interviews showed this is a past scenario for ONS.

ABD is a process that involves cultural changes. Despite the efforts of NSOs to overcome resistance, it still exists in some sectors. Interviewees indicate the resistance is more located in groups of employees from the "generational gap" than in a specific sector. Senior workers are conservative and against the change from traditional Official Statistics methods to Big Data methods and technologies.

“They don't want to see that Big Data is the only way, but there should be the traditional way for the traditional statistician to work, and I think that kind of working stream was still ongoing.” [01_ONS]

“I'm not saying that there can't be some resistance here and there. But it is natural, even because they refer to, sometimes, technical resistances, (...)” [02_INE]

One solution to overcome *cultural blockers* is an intense and widespread training effort.

“So, I think this kind of traditional data is still with the advantage. But people, specially trained for some kind of qualitative, with quality to the background, they

may still have difficulty understanding the Big Data, and there might be some future communication tasks to be done. “[01_ONS]

In this context, how do we protect NSOs and other public agencies from *cultural blockers*? ONS showed the solution: job position rotation. In the UK, a civil servant is motivated to change the job position every five years. Human resources and skills are shared with other governmental bodies. *Organisational readiness* improves with the ability to deal with changes. At the same time, *technological capabilities* are embraced because the employee needs to learn new skills. Finally, the sludge of resistance is reduced.

"I stayed about eight months in the position when they opened an internal competition vacancy at ONS - this is very common here and so I will explain to you that people don't stay long in one department, it is very common here to have loans, the loans, between Governments. It is a possibility to develop in your career and to get promotions, it is not like in Brazil where you stay in a position all your life and with that salary the most you will get is a position of trust for a short period of time. Here there are always vacancies advertised across the whole Government in the public service and the vacancy can be internal, only with ONS employees, or any Government employee can apply for the vacancy, or for anyone from the UK. It's internal, across Government and external recruitment. So, it has a lot of movement and it is very well regarded, there is a lot of chance to be promoted and if you are looking for a vacancy in a higher position, you wait for a vacancy at that level to be advertised and to apply, you have the interview. Every time you want to change, you go through this selection. And then you have a well-defined rule of how the process is, the interview, the presentation... it is all used across all departments. (...)

A combination of skills. Using what you already know from the statistician's experience and bringing in the innovation of machine learning and data visualization, which you don't learn how to deal with by being a statistician. And it is not about being a data scientist or being a statistician, it is about bringing together the skills that are essential to modernize the ONS and modernize the Government. How do we do this? By bringing in statisticians, social researchers, data scientist, mathematicians, computational scientist, IT. It is multidisciplinary. (...) The first thing we need are the skills, without the basics there is no progress, so we needed the analyst - data scientist

or whatever you want to call it - and we defined who was this professional, what were the skills, what was the behaviour, it took years to establish this consensus in the Government."[05_ONS_TRN]

In the case of Brazil, IBGE provides some examples of *cultural blockers* on different levels. There is a conservative and blocking organisational culture established in that NSO. Part of this culture emerged from a group of senior professionals who resist innovating and ABD. This group is composed by senior statisticians who are sceptical about the potentialities of Big Data. The solution is a drastic innovation effort, starting with the hiring of new professionals from the labour market.

"I realized that people here in the IBGE kind of got restricted to the question of usability of a system, like, 'I just want to push a button'. (...) So, I think that the culture of the IBGE... I call it as an outdated culture, from the 'Anthropocene' age..., a "Fossil" culture. It is an organizational culture from 1980s years (...)"[10_IBG]

"I think that in fact there still exists within IBGE, as there is anywhere in the world: people who end up being owners of a work developed within the institution. (...) the culture, more hinders than favours it."[08_IBG_DGC]

"Because we know that the IBGE's members are conservative. You define the law, say what you can and what you can't. (...) So this has always worried me, I said, "what is the point of me theorizing and doing nothing?" But this is the IBGE culture."

[01_IBG_PRO]

As an organisational factor for the avoidance of Big Data in NSOs, the *cultural blockers* factor refers to:

1. Why do cultural blockers cause avoidance of ABD in NSOs? Because this factor promotes scepticism combined with an attempt to keep the status quo, avoiding ABD in NSOs. [ONS; INE; IBGE]
2. How do cultural blockers avoid Big Data in NSOs? They promote avoidance in soft or hard levels, from isolated initiatives to hidden groups. [IBGE]
3. Cultural blockers' actions require internal struggle for power and influence, including conservative groups of the NSO staff. [IBGE]
4. Critical damage: process obsolescence and outdated Official Statistics. [IBGE]
5. Solution to block cultural blockers: workforce renewal and engagement promotion. [IBGE]

6.1.3. Environment factors

The literature review presented twelve TOE factors in the Environment dimension related to the adoption of Big Data (ABD). As presented in Chapter 4, this set of factors was used as a code book to analyse the content of the interviews (Saldaña, 2013).

During the interview analysis, those twelve factors were validated by the three NSO (ONS, INE, IBGE) members. Just one original factor was discarded because it did not receive any mention. "IS Fashion" did not appear in the environmental context of ONS, INE, and IBGE. Maybe "IS fashion" is typical for companies working in a private environment with a reduced level of cooperation. Conversely, NSOs are in another reality: public-governmental and embedded in a cooperative environment.

Additionally, one original factor emerged from the content analysis: "international-institutional context". This factor means "the relevance of the NSO in the GSS network, which is understood as the interactions, cooperation, and presence in international organisms or in partnerships with other NSOs.

Table 23 summarises the findings for the Environment dimension, following the same structure as the previous tables. Afterwards, the detailed analysis of each Environment factor is presented.

201. Augmented availability of Big Data related technology (AABDRT)

Based on Ram et al. (2019), the *augmented availability of Big-Data-related technology* (AABDRT) is one of the driving factors of ABD. This factor can be defined as the availability of off-the-shelf or open-source data management applications, and how this availability is important to improve the use and analyses of Big Data.

This is a factor with influence on ABD because NSOs suffer with cuts in their budgets, and the alternative to keep running technological projects includes the availability of open-source technologies, including Big Data solutions.

So, AABDRT influences ABD through the adoption of open-source technologies in the NSOs internal processes. In the case of INE, those technologies are identified as programming languages, "R" and "Python".

"But some people started with "R" and are better suited to "R". "R", today, is a sustainable project. Ten years ago, I wouldn't have advised it because it was still at a previous stage..."[01_INE]

In the case of ONS, the change to the use of open-source technologies demanded a mind-set shift. Currently, the British staff keeps using those types of technologies. The usage always follows the security and privacy protocols for the Official Statistics.

"But today, most economists come out of university with "R" or Python. They're not using Stata. And so, whilst that seems a very sensible transition to make four or five years ago. Now, I'm having to think about transitioning and moving that system into R or Python. Similarly, a lot more use of cloud!" [04_ONS]

AABDRT is considered an environmental factor because the open-source solutions are available and accessible in zones and countries where this type of technology is allowed. If a government forces the exclusive use of licensed software in its organisations, AABDRT tends to be blocked by the high costs of licences. This constraint can impact Big Data project in the organisations with budget restrictions, as NSOs.

For ABD in NSOs, *augmented availability of Big Data related technology* (AABDRT) is summarised as:

1. Why is AABDRT an influencing factor for ABD in NSOs? Because NSOs suffer with budget cuts, and the alternative to keep running technological projects includes the availability of open-source technologies, including Big Data solutions. [ONS; INE; IBGE]

2. How does AABDRT influence ABD in NSOs? It refers to the adoption of open-source technologies in the NSOs' internal processes, such as programming languages: "R" and Python. [ONS; INE; IBGE]

3. Requirements for the AABDRT in NSOs during the ABD process:

a. Security and stability of open-source applications. [ONS]

b. Current use by the staff, substituting licensed software. [ONS; IBGE]

Table 23 – Environment factors: Summarised findings

Environment Factors	Consolidated Findings	Findings per case		
		ONS-UK	INE-PT	IBGE-BR
201. Augmented availability of Big Data related technology (AABDRT)	1. Why is AABDRT an influencing factor for ABD in NSOs? Because NSOs suffer with budget cuts, and the alternative to keep running technological projects includes the availability of open-source technologies, including Big Data solutions. [ONS; INE; IBGE] 2. How does AABDRT influence ABD in NSOs? It refers to the adoption of open-source technologies in the NSOs' internal processes, such as programming languages: "R" and Python. [ONS; INE; IBGE] 3. Requirements for the AABDRT in NSOs during the ABD process: a. Security and stability of open-source applications. [ONS] b. Current use by the staff, substituting licensed software. [ONS; IBGE]	1; 2; 3a; 3b.	1; 2.	1; 2; 3b.
202. Competitive Pressure	1. Why is competitive pressure a sensitive factor for ABD in NSOs? Because NSOs are competing for the citizens' attention to the Official Statistics (OS). If the attention for OS is reduced, NSOs lose relevance. [ONS; INE; IBGE] 2. How does competitive pressure affect ABD in NSOs? ABD becomes mandatory to offer faster deliveries for citizens and keep their attention in Official Statistics. [ONS; INE; IBGE] 3. Major top competitors of NSOs for the citizens' attention: fake news, misinformation, and reputational attacks against the NSOs. [ONS; INE; IBGE] 4. Requirements for NSOs to overcome competitive pressure: a. Adoption of Big Data technologies and methodologies. [ONS; INE; IBGE] b. Faster delivery of the outcomes. [ONS; INE; IBGE]	1; 2; 3; 4a; 4b.	1; 2; 3; 4a; 4b.	1; 2; 3; 4a; 4b.
203. Government support/laws and policy	1. Why are government support, laws, and policies relevant for ABD in NSOs? Because laws can promote supporting rules for ABD in Official Statistics rather than constrain this adoption. [ONS; INE; IBGE] 2. How do government support, laws, and policies influence ABD in NSOs? Governmental action, through laws, policies, and regulations, interfere with the Official Statistics deliveries and data access by NSOs. [ONS; INE; IBGE] 3. Requirements for a positive influence of government support, laws, and policies in ABD: a. Creation or maintenance of a Statistical Regulatory Authority as the key player in the National Statistical System. [IBGE] b. Demand for complementary laws for Official Statistics and administrative records access. [ONS; INE; IBGE]	1; 2; 3b.	1; 2; 3b; 3c.	1; 2; 3a; 3b; 3c.

Environment Factors	Consolidated Findings	Findings per case		
		ONS-UK	INE-PT	IBGE-BR
204. Industry features	1. Why is the industry features factor critical for ABD in NSOs? Because the Official Statistics industry works based on specific features, as international cooperation rather than competition. [ONS; INE; IBGE] 2. How do these industry features support ABD in NSOs? They support the adoption through the partnerships where NSOs acquire Big Data expertise with their fellows. [ONS; INE; IBGE] 3. Requirements for the industry features factor for ABD in NSOs: a. Keeping NSO autonomy relative to the Central government. [ONS; INE; IBGE] b. Participation in international and national forums for acquiring Big Data expertise. [ONS; IBGE] 4. Benefits of the industry features factor for ABD in NSOs: a. GSS defines general rules of the industry of Official Statistics, supporting NSOs. [ONS; INE; IBGE] b. Findings and innovations in ABD are shared among the NSOs. [ONS; INE; IBGE]	1; 2; 3a; 3b; 4a; 4b.	1; 2; 3a; 4a; 4b.	1; 2; 3a; 3b; 4a; 4b.
205. Institutional-based trust	1. Why is institutional-based trust fundamental for ABD in NSOs? Because only in a trusting relationship will citizens allow the NSO to collect their data for using it in Official Statistics outcomes. Without data, there is no ABD. [ONS; INE; IBGE] 2. How does institutional-based trust create conditions for ABD in NSOs? Through institutional-base trust, NSOs keep the links with the society and citizens, preserving the social relevance of Official Statistics, and allowing ABD. [ONS; INE; IBGE] 3. Requirement for institutional-based trust supporting ABD in NSOs: a. Autonomy and independence from the Central Government. [ONS] b. Privacy and security culture and policies to manage citizens' data. [ONS; INE; IBGE] c. Risk management for the data privacy and security. [ONS; INE; IBGE]	1; 2; 3a; 3b; 3c.	1; 2; 3b; 3c.	1; 2; 3b; 3c.
206. Information System fashion	No relevance in this research			
207. Market turbulence	1. Why is market turbulence a risk for ABD in NSO? Because it provokes disruptions and plan changes, including in ABD projects. [INE; IBGE] 2. How does market turbulence impact ABD in NSOs? This factor influences ABD through the cuts and blocks in the current operation mode of NSOs. [INE; IBGE] 3. Requirement for market turbulence prevention by NSOs: a. Use of innovative data collection methods in the Official Statistics operations. [INE] b. ABD in the Official Statistics business architecture to insure resilience during incidents. [INE]		1; 2; 3a; 3b.	1; 2.
208. National Institutional Context	1. Why is the national institutional context critical for ABD in NSOs? Because a country with a variety of research centres that follow a scientific-driven trajectory, as ABD, can provide a hub of interactions, collaboration, and exchanges for supporting NSOs. [ONS] 2. How does the national institutional context shape ABD in NSOs? It shapes ABD through the national political-administrative structure and the demographic distribution on the national territory, influencing the NSO structure. [INE] 3. Requirements for the national institutional context to promote ABD in NSOs: a. Development of a governmental networks for collaborating on ABD with the national public administration. [ONS] b. Knowledge-transfer environment between research centres and the NSO. [ONS]	1.	2.	3a; 3b.

Environment Factors	Consolidated Findings	Findings per case		
		ONS-UK	INE-PT	IBGE-BR
209. Partner Pressures	1. Why is the partner pressures factor relevant to ABD in NSOs? Because this pressure nudges NSOs towards ABD, creating a constraint for the non-adopters. [ONS; INE; IBGE] 2. How do partner pressures nudge NSOs to adopt Big Data? It nudges in passive and active modes, making the NSO offer its Big Data expertise and forcing it to follow earlier adopters. [INE] 3. Requirements for partner pressures to nudges ABD: a. Active partnership with others in NSOs (technical knowledge exchange). [ONS; INE; IBGE] b. Passive partner under the control of international organisations (funding and resources). [ONS; INE; IBGE] c. Technological and methodological benchmarking with other NSOs. [ONS; INE; IBGE]	1; 3a; 3b; 3c.	1; 2; 3a; 3b; 3c.	3a; 3b; 3c.
210. Risk of Outsourcing	1. Why is the risk of outsourcing complementary for ABD in NSOs? Because outsourcing works as alternative, not a central solution, to run Big Data projects when NSOs do not have the full expertise. [ONS; INE; IBGE] 2. How does the risk of outsourcing affect ABD in NSOs? It affects ABD under a controlled supervision by NSOs' staffs, with those employees managing the outsourcing's legacy and knowledge. [INE; IBGE] 3. Requirements to avoid the risk of outsourcing and use outsourcing as a supplement for ABD: a. Controlled outsourcing: short-term projects. [INE; IBGE] b. Considering outsourcing as an alternative, not a full solution. [INE; IBGE] c. Legacy and expertise are internalised, not kidnapped. [INE; IBGE]	1	1; 2; 3a; 3b; 3c.	1; 2; 3a; 3b; 3c.
211. Security and privacy regulations	1. Why are security and privacy regulations normative to ABD in NSOs? Those organizations base their adoption on security and privacy rules and laws to keep the citizens' data protected from hackers. [ONS; INE; IBGE] 2. How do security and privacy regulations influence ABD in NSOs? They affect ABD through the limitations and definitions of personal data collection and use. [ONS; INE; IBGE] 3. Requirement for security and privacy regulations to support ABD in NSOs: the creation of a specific regulations for the Big Data collection and use by NSOs. [INE; IBGE] 4. Consequences of security and privacy regulation for NSOs: a. Limitation in the regulation as an inhibitor for ABD. [INE] b. Regulation is necessary; overregulation is damaging and a barrier for ABD in NSOs. [ONS; INE; IBGE]	1; 2; 4b.	1; 2; 3; 4a; 4b.	1; 2; 3; 4b.
212. Uncertainty and risk concern	1. Why are the uncertainty and risk concern an issue for ABD in NSOs? Because in the face of uncertainty, NSOs can delay or avoid ABD, maintaining the status quo. [INE; IBGE] 2. How does the uncertainty and risk concern affect ABD in NSOs? This factor affects NSOs through the avoidance of Big Data in function of the risky scenarios or impacts in the internal operational environment. [IBGE] 3. Examples of the uncertainty and risk concern for the ABD in NSOs: a. COVID-19 pandemic. [IBGE] b. Damage to public image. [IBGE] c. Loss of capillarity. [IBGE] d. Loss of workforce related to retirements. [INE; IBGE] e. Ageing effect. [INE]		1; 3d; 3e.	1; 2; 3a; 3b; 3c; 3d.

Environment Factors	Consolidated Findings	Findings per case		
		ONS-UK	INE-PT	IBGE-BR
2X5. International – Institutional Context	1. Why and how does the international-institutional context factor affect ABD in NSOs? This factor is critical because it allows the understanding about the external context and how it influences and has been influenced by the respective NSO, especially in the ABD journey. [ONS; INE; IBGE] 2. Requirement for the international-institutional context for ABD in NSOs: to work as the marketplace for technological and methodological Big Data solutions. [ONS; INE; IBGE] 3. Benefits from the international-institutional context for NSOs: to be in a relevant position in the GSS, developing ties with others NSOs. [ONS; INE; IBGE]	1; 2; 3.	1; 2; 3.	1; 2; 3.

202. Competitive Pressure

Based on El-Haddadeh et al. (2021, p. 404), it is possible to define *competitive pressure* as *"the actions and movements from the competitors that influence the organisation's adoption of Big Data."* Even though NSOs are not private companies, *competitive pressure* is a sensitive factor for ABD in NSOs because NSOs are competing for citizens' attention to the Official Statistics (OS). If the attention for OS is reduced, NSOs lose relevance.

In this competitive environment, the delivering-time for Official Statistics is critical. Citizens tends to pay less attention to outdated statistics and indices. So, ABD becomes mandatory to offer faster deliveries for citizens and keep their attention on Official Statistics. If an NSO is not adopting Big Data, it is entering a risky zone.

Since the last decade, NSOs are dealing with fake news and misinformation. Simultaneously, there is competition for trust in Official Statistics. In this case, organized groups contest the reputation of NSOs as an honest data supplier and a privacy manager, as remarked by one of the interviewees. So, competition is for citizens' attention and trust.

"The INE's social image is positive, faithful, and reliable. So, the information that reaches our citizens is not fake news information. I think we are very well at this image level." [02_INE_DPE]

"There are lawyers because the creation of the LGPD in Brazil. As a result, this lawyers movement scares me a little bit by putting the credibility of IBGE in question. Because IBGE has a credibility of more than 85 years, respecting the statistical confidentiality. IBGE has never failed. I think that we would have to be more proactive to inhibit these lawyers from going ahead with this satanization of the IBGE. I see a risk in this.'." [01_IBG_PRO]

For ABD in NSOs, *competitive pressure* factor is summarised as:

1. Why is competitive pressure a sensitive factor for ABD in NSOs? Because NSOs are competing for the citizens' attention to the Official Statistics (OS). If the attention for OS is reduced, NSOs lose relevance. [ONS; INE; IBGE]
2. How does competitive pressure affect ABD in NSOs? ABD becomes mandatory to offer faster deliveries for citizens and keep their attention in Official Statistics. [ONS; INE; IBGE]

3. Major top competitors of NSOs for the citizens' attention: fake news, misinformation, and reputational attacks against the NSOs. [ONS; INE; IBGE]

4. Requirements for NSOs to overcome competitive pressure:

a. Adoption of Big Data technologies and methodologies. [ONS; INE; IBGE]

b. Faster delivery of the outcomes. [ONS; INE; IBGE]

203. Government support/laws and policy

According to Ghaleb et al. (2021), governmental support/laws and policy can be defined as institutional support or restriction for organizations adopting Big Data. As significant drivers, laws, legislations, and funding affect the adoption of novel technologies, such as Big Data, by all companies and enterprises in a country.

In fact, the *government support, laws, and policies* factor is relevant because the laws can promote supporting rules for ABD in Official Statistics, rather than constrain this adoption. Regarding this relevance, this factor influences ABD in NSOs through the laws, policies, and regulations, interfering with Official Statistics deliveries and data access by NSOs.

The adoption of Big Data technologies and methodologies cannot be addressed by NSOs independently. In the British case, when ONS proposes that an experimental statistic is elevated to an official one, it requires the approval of the UK Statistical Authority. This regulation is an example of government support for ONS adopting Big Data. On the other hand, this governmental arm works as an administrative system to protect the quality of Official Statistics.

“But you can’t get to a process where you are producing Official Statistics without having gone through their technology adoption and their approval of the methods because that needs to be communicated and it needs to be very transparent (for the UK Statistical Authority).” [02_ONS]

ONS members do not understand the set of laws as a restrictive factor but more as a stabilising factor to support ABD. Laws and acts can work as magnets to attract researchers, funds, and institutional support for ABD.

“I think there are probably not many restrictions or limitations from understanding. I think, in general, from the UK Government perspective, they set up a so-called AI plan,

and Big Data related activities within the government bodies are kind of encouraged. I hope this kind of encouragement could be properly supported by real money after that.” [01_ONS]

However, laws can inhibit ABD in others national contexts, sometimes because of restrictions or a lack of legal updates, as reported by INE and IBGE.

“(…) we need to have that legal support, in the sense that the legislation provides that INE can use all that information. We have that in the National Statistical System law (…) But still, it is INE's intention to propose a specific law for administrative data use.” [04_INE]

“I think the legislation is very old. It facilitates, but it is very old. The statistics legislation, I think, is from the 1960s. The LGPD, I think, it got in the way. I think it has a comprehension problem. I mean, we would have to do a more active action. There is a confusion between privacy, which is the LGPD side, and statistical confidentiality, which is very close to privacy.” [01_IBG]

For ABD in NSOs, the factor *government support, laws, and policies* are summarised as follows:

1. Why are government support, laws, and policies relevant for ABD in NSOs? Because laws can promote supporting rules for ABD in Official Statistics rather than constrain this adoption. [ONS; INE; IBGE]
2. How do government support, laws, and policies influence ABD in NSOs? Governmental action, through laws, policies, and regulations, interfere with the Official Statistics deliveries and data access by NSOs. [ONS; INE; IBGE]
3. Requirements for a positive influence of government support, laws, and policies in ABD:
 - a. Creation or maintenance of a Statistical Regulatory Authority as the key player in the National Statistical System. [IBGE]
 - b. Demand for complementary laws for Official Statistics and administrative records access. [ONS; INE; IBGE]

204. Industry features

According to Verma and Bhattacharyya (2017, p. 361), *industry features* or *industry types* refer to "the sector to which the business belongs." In this research there is an industry with specific

features, the Official Statistics industry. So, this factor is critical for ABD in NSOs because the Official Statistics industry works based on specific features, as international cooperation rather than competition.

In fact, the *industry features* factor supports ABD in NSOs through the partnerships where NSOs acquire Big Data expertise with their fellows. NSOs adopt several partnerships in an international context, showing this is a common feature. International practitioners include the UN Committee of Experts on Big Data, the World Bank, the IMF, UNECE, CBS-ND, Uganda Stats, and Senegal Stats. Those organisations receive and offer knowledge about ABD in NSOs. This set of partnerships is a feature of the Global Statistics System (GSS), considering the Official Statistics as a type of industry.

“We, as INE, are the representatives of Portugal in those decision-making bodies (GSS, UN, ..).”[01_INE]

A specific example of international partnership is the agreement where ONS supported IBGE to run the Brazilian Census Operation. Under the lens of both staffs, this partnership was rich and relevant. As remarked during the interviews, when an NSO is supporting a partner to solve a problem in Official Statistics, it is helping itself and the whole GSS. This conclusion comes from the understanding that Official Statistics issues are similar for all members of the GSS, the NSOs.

“So, both. We are very involved in the at the UN Committee of experts for Big Data and Data Science and have been since that group started. We've been very involved in the UN Global platform for Big Data, supporting four international programs in other countries working on Big Data and Data Science approaches. So, we have very strong partnerships with the UN teams and others NSOs, as IBGE or UNECE, for example. (...) In a regional level, we have strong relationships with IMF, World Bank, all of those groups. So, the first thing is that we have a two-way relationship. We are often looking at examples of what other countries are doing and learning from them. We hope that exchanging some of our examples, we can help others NSOs. So, that's the first thing. The second one is we often working in partnership with other countries such as example with CBS Netherlands, with Stats Canada, on using different approaches for mobile data and satellite imagery. And similar, we do a lot of partnership programs with pieces of work. Because, in heart, we, as NSOs, try wanna solve the same problems and some of the Big Data sources are global. So, solving in

one area involves solving in everywhere. So, we have some very nice examples of this where work to use satellite imagery for agricultural crops statistics production for the Big Data program on the UN Global platform. We now have tools that are being used by Senegal, by Uganda and many other countries to produce estimations that ONS consider outputs.” [03_ONS]

NSOs are parts of the respective National Governmental cores. To avoid losing autonomy, this proximity must be balanced. In this environment, the autonomy and relevance of the Official Statistics are maintained. ONS will continue to maintain this dual perception.

“But I will reflect to observe the role of a NSO is to provides robust, impartial, and independent Official Statistics. It is absolutely at the hearts of governments and public sector decision-making. So, that means we need to understand the decisions that are being made in order that we can produce analysis that helps people make those decisions. So, in a way, we have to be close but not too close. I think that balance would always shift, and it will be different for different things.” [03_ONS]

Another feature is the partnership with national governmental agencies, as remarked in the Brazilian case. Governmental bodies are seen as the innovation generators and can share their contributions with the NSO, at the national level.

“So today there is the Secretary of Innovation and Digital Government, and it pushes this innovation from the Federal Government. It is an instance that acts from top to bottom. So, if they come and say that they want all public service working with 100% of the virtual machines, we will have to adhere to this.” [08_IBG]

For ABD in NSOs, the *industry feature* factor is summarised as:

1. Why is the industry features factor critical for ABD in NSOs? Because the Official Statistics industry works based on specific features, as international cooperation rather than competition. [ONS; INE; IBGE]

2. How do these industry features support ABD in NSOs? They support the adoption through the partnerships where NSOs acquire Big Data expertise with their fellows. [ONS; INE; IBGE]

3. Requirements for the industry features factor for ABD in NSOs:

a. Keeping NSO autonomy relative to the Central government. [ONS; INE; IBGE]

b. Participation in international and national forums for acquiring Big Data expertise. [ONS; IBGE]

4. Benefits of the industry features factor for ABD in NSOs:

a. GSS defines general rules of the industry of Official Statistics, supporting NSOs. [ONS; INE; IBGE]

b. Findings and innovations in ABD are shared among the NSOs. [ONS; INE; IBGE]

205. Institutional-based trust

Based on Sun et al. (2018) and Yadegaridehkordi et al. (2018), this factor refers to the trust that an organization has in the face of the society, considering the innovation and adoption process it uses to deliver outcomes. The trust is a mandatory input for an NSO to work with citizens in the respective country. The reason for this relevance is because only in a trusting relationship will citizens allow the NSO to collect their data for using it in the Official Statistics outcomes.

Through the institutional-base trust, NSOs keep the links with society and citizens, preserving the social relevance of the Official Statistics. The loss of the institutional-base trust can drastically affect the NSO operating in a country.

As a result of the institutional-based trust, NSOs receive autonomy from society to deal with data management. In fact, NSOs are working in a risky manner regarding the privacy and security of citizen data. It is possible to understand this process as a two-way road, where in one direction goes the responsibility to keep sensitive data safe and, on the other, the autonomy delegated by national authorities to NSOs for dealing with citizens' data.

In the case of ONS, there is an expressive autonomy, independent of the political party which is the incumbent. This autonomy improves *institutional-based trust*.

“Our autonomy is a real benefit to the political system.” [04_ONS]

“We cannot say that it's completely autonomy to any organization. Even the Central Banks cannot say that they are wholly independent of politics. But to my observation, I think it (ONS) is still quite independent. (...) ONS still publish statistics according to their own schedule, even the status may influence the collective discussions at Westminster parliament.” [01_ONS]

Another relevant point is that the risk of trust loss in front of society is a concern for NSOs. It is a risk related to the *institutional-based trust*. Surrounding the NSOs, including the ONS, the care with privacy and security of citizens' data represents the greatest concern related to *institutional-based trust*, or in the interviewee's words, "*the existential threat*".

"If you lose the public, if you lose the support, then you got a real problem. I think we recognize that this is probably the existential threat for the NSOs. This is the big risk: if we get this wrong, we lose our legitimacy to do this type of work with Official Statistics. Security has to come first." [04_ONS]

"ONS is not subordinated to any ministry." This sentence summarises the ONS's features as a governmental agency with autonomy and managerial freedom. This autonomy reinforces social trust in ONS outcomes. In this context, citizens understand that an autonomous NSO is a trustworthy organisation, because the influence of the incumbent government is reduced.

"It (ONS) continues with autonomy, with independence. We are not subordinated to any Ministry" [05_ONS]

In the other cases, INE and IBGE work to preserve their image as reliable institutions for the citizens. This preservation involves data protection and ensures confidentiality in a continuous process of performance improvement.

" This is also due a lot to this trust relationship. That is, people know that when they are giving data to INE, it is only for that purpose. It has nothing to do with other purposes." [03_INE]

"Now, the extent to which society recognizes the LGPD and believes in the law, this issue of trust... in this sense, the IBGE, as a public institution, is also impacted by this law. The question is about the trust that society has in the government. So, both the law and the IBGE are together. I think the law even helps because we can cite the law as a reference." [10_IBG]

For ABD in NSOs, the *institutional-based trust* factor is summarised as:

1. Why is institutional-based trust fundamental for ABD in NSOs? Because only in a trusting relationship will citizens allow the NSO to collect their data for using it in Official Statistics outcomes. Without data, there is no ABD. [ONS; INE; IBGE]

2. How does institutional-based trust create conditions for ABD in NSOs? Through institutional-base trust, NSOs keep the links with the society and citizens, preserving the social relevance of Official Statistics, and allowing ABD. [ONS; INE; IBGE]

3. Requirement for institutional-based trust supporting ABD in NSOs:

- a. Autonomy and independence from the Central Government. [ONS]
- b. Privacy and security culture and policies to manage citizens' data. [ONS; INE; IBGE]
- c. Risk management for the data privacy and security. [ONS; INE; IBGE]

206. Information System fashion

Sun et al. (2018, p.198) define this factor as *"information is obtained through external communication channels by focusing on an organisation's peers and perceived experts, such as vendors and customers (e.g., managerial interactions, social channels, and trends)"*. However, in this research, IS fashion was not present in the interviews and documental research. For this reason, it has been discarded.

207. Market turbulence

Based on Sun et al. (2018), Baig et al., (2019), and Yadegaridehkordi et al. (2018), market turbulence refers to changes in citizens' demands, the operational environment, and government actions with impacts on the NSO's business architecture. In this research, the factor name might be revised to "environmental turbulence", but we opted to maintain the term used in the current literature.

In this context, the market turbulence factor is a risk for ABD in NSOs because it provokes disruptions and plan changes, including in ABD projects. So, this factor influences ABD through the cuts and blocks in the current operation mode of the NSOs.

Although NSOs are not commercial organizations, market volatility also affects them. An example was the COVID-19 pandemic and its impact on the Census data collection in Brazil and Portugal. Fortunately, internet data collection diminished the impact of the lockdowns in the Portuguese case.

" With the Covid-19 pandemic and the limitations related to the confinement, the data collection by internet actually also worked in that sense of the people confined to respond to the Census. It is better, because, as we know, in the population Census, one of the big problems is the mobility of people." [03_INE_DPO]

"Let's say that the pandemic brought up the issue of Big Data in practice. It is not solved, but the pandemic showed the necessity of adoption. So, the most emblematic case - the case of cellular telephony for telephone interviews – took place during the pandemic." [01_IBG_PRO]

Thus, considering market turbulences are unpredictable, the adoption of Big Data arises as a preventive insurance for NSOs to deal with those occurrences and incidents. More than a matter of option, ABD is a matter of survival for NSOs in actual turbulent times.

For ABD in NSOs, *market turbulence* is summarised as:

1. Why is market turbulence a risk for ABD in NSO? Because it provokes disruptions and plan changes, including in ABD projects. [INE; IBGE]
2. How does market turbulence impact ABD in NSOs? This factor influences ABD through the cuts and blocks in the current operation mode of NSOs. [INE; IBGE]
3. Requirement for market turbulence prevention by NSOs:
 - a. Use of innovative data collection methods in the Official Statistics operations. [INE]
 - b. ABD in the Official Statistics business architecture to insure resilience during incidents. [INE]

208. National Institutional Context

Initially proposed by Bakici et al. (2022), this factor can be understood as the relevance of the interactions, cooperation, and institutional ties in the domestic environment that are promoting or inhibiting the adoption of Big Data by an organisation. In this sense, the *national institutional context* is a determinant environmental factor because a country with a variety of research centres that follow a scientific-driven trajectory, as ABD, can provide a hub of interactions, collaboration, and exchanges for supporting NSOs.

This factor influences NSOs in the ABD journey through the national political-administrative structure and the demographic distribution on the national territory. So, NSOs adjust their firm size

and organisational structure in function of the national institutional context, making this a shaper factor.

In this way, the availability of a highly capacitated workforce is part of the national-institutional context. A context where there are top professionals available to engage in new projects is positive for ABD in NSOs.

As a technological and organisational innovation, ABD demands high-qualified professionals from the labour market. Additionally, top professionals insert updated knowledge into NSOs.

“On the research part, related to the methodology, yes, there were. They started talking with Research Centres, with PhDs, like the Alan Turing Institute, which was the first partner, to collaborate with scientists on Data Science Campus who were not from the public service. They were from the private sector to teach us this way of working.” [05_ONS]

This knowledge transfer occurs in an environment based on cooperation and exchange with other national research institutes. An example of the national context for ABD, is the ONS case. That NSO works with its partners to enrich the staff with new Big Data skills. Initiatives such as the NIESR research centre and the ESCoE, which bring researchers to ONS, show that context is relevant for supporting ABD.

“And another is the National Institute of Economic and Social Research, the NIESR, to support the establishment a joint research centre. We called it ESCoE. So, basically, ESCoE want to bring the researchers to the wider domain, to work together with them.” [01_ONS]

Another impact of the *national institutional context*, the political-administrative structure, is considered in the Portuguese case. Despite the reduced dimensions, Portugal presents some specificities related to the mainland's occupation and administrative structure. With the autonomous regional administration, this country has regional organisations working with the central government. The DREM (Regional Statistics Department of Madeira), in the Madeira archipelago, works as a delegation of INE. An additional specificity of Portugal is the demographic distribution, which is highly concentrated in the coastal zone. Beyond the singularities, INE maintains autonomy and independence from the central administration to run Official Statistics and adopt new technologies, including Big Data.

"INE has in its statute technical independence and autonomy, which in my view has been guaranteed over the years. "[02_INE]

"We (DREM) work here as a delegation of INE. So, the information from Madeira that goes to Eurostat is all transmitted by INE." [05_INE]

For ABD in NSOs, the *national institutional context* is summarised as:

1. Why is the national institutional context critical for ABD in NSOs? Because a country with a variety of research centres that follow a scientific-driven trajectory, as ABD, can provide a hub of interactions, collaboration, and exchanges for supporting NSOs. [ONS]
2. How does the national institutional context shape ABD in NSOs? It shapes ABD through the national political-administrative structure and the demographic distribution on the national territory, influencing the NSO structure. [INE]
3. Requirements for the national institutional context to promote ABD in NSOs:
 - a. Development of a governmental networks for collaborating on ABD with the national public administration. [ONS]
 - b. Knowledge-transfer environment between research centres and the NSO. [ONS]

209. Partner Pressures

Based on Sun et al. (2018, p.198), ABD can be advanced in order to follow partners and maintain the firm's internal balance with them (e.g., the readiness of suppliers in external collaborations). This factor is relevant because this pressure nudges NSOs for ABD, creating a constraint for the non-adopters. It works in passive and active modes to pressure NSOs.

Firstly, those organisations suffer pressure from other research institutes to advance and propose new technologies and methods for ABD in Official Statistics. Secondly, NSOs, as members of GSS, look to exert dominance over new methods and technologies for ABD. In this environment, each NSO looks for a top position among its peers.

"(...) nobody yet, including Stats Nederland, has got to a position where the perfect replacement between Big Data methods and the survey methods. (...) We (ONS) work closely with the National Health System. I wouldn't say that their innovations disrupt our processes at all, on the contrary ... we probably innovate in a similar way and that is why we can easily work with them." [02_ONS]

Partner pressure also works as a two-way street. In the case of INE, the NSO interacts with its partners in active and passive modes. Regarding knowledge and processes, INE works as a pressure agent to improve quality in other NSOs, like the Brazilian IBGE.

"The Censuses, of course, do not have the numbers of Brazil - and, by the way, INE's President and I participated in the meeting of the IBGE's Census team. It was very grateful to share these experiences with Mr. President of IBGE and the fellow Census professionals."[04_INE]

On the other hand, INE acts as a passive agent, receiving protocols and norms from international partners such as Eurostat. Furthermore, the external partner pressure comes from the funding support sent by the European Union.

"Eurostat works as a kind of a disseminating vehicle for good practices. In that sense, the adoption of innovation is done through dissemination. (...) There's no point invented in INE and not shared and compared with all countries in this European Union. So, this question of the Eurostat coverage is always a fundamental symbol."

[02_INE]

For ABD in NSOs, *partner pressures* are summarised as:

1. Why is the partner pressures factor relevant to ABD in NSOs? Because this pressure nudges NSOs towards ABD, creating a constraint for the non-adopters. [ONS; INE; IBGE]
2. How do partner pressures nudge NSOs to adopt Big Data? It nudges in passive and active modes, making the NSO offer its Big Data expertise and forcing it to follow earlier adopters. [INE]
3. Requirements for partner pressures to nudges ABD:
 - a. Active partnership with others in NSOs (technical knowledge exchange). [ONS; INE; IBGE]
 - b. Passive partner under the control of international organisations (funding and resources). [ONS; INE; IBGE]
 - c. Technological and methodological benchmarking with other NSOs. [ONS; INE; IBGE]

210. Risk of Outsourcing

Baig et al. (2019, p.10) consider the *risk of outsourcing* as "hiring from outside firms to implement Big Data adoption-related services. Privacy, security, reputation, transaction, compliance, and contractual risks are associated with outsourcing."

This factor is important and complementary for NSOs because outsourcing works as alternative, not a central solution, to run Big Data projects when NSOs do not have the full expertise. For this reason, the risk of outsourcing is considered controlled by NSOs' staffs, mainly when these employees are supervising the outsourcing of Big Data projects, its legacy and knowledge.

"And outsourcing - this is where the hybrid model comes in, where all this makes sense. However, outsourcing is not a full solution." [01_IBG]

"It makes no sense to hire a technical expert in pairing or in Web Scraping, when he/she doesn't leave a legacy for the IBGE. This expert has to train the staff. This is fundamental to do." [03_IBG]

In this environment, outsourcing is not realized as a risk by the NSOs, but a necessity for them in ABD projects. According to the British interviewees, ONS addresses outsourcing in some stages of the ABD projects in order to reduce costs.

"If the ONS thinks that they cannot do their own, it starts the outsourcing project with a list of potential tenders, summarized in a document, and put it to ESCOE. In sequence, ONS lets ESCOE to kind of advertise/call in a way that external researchers can bid for the funding." [01_ONS]

"It ends up being the universities always the easiest partner to build those bridges (outsourcing)." [03_INE]

The outsourcing type varies in each case. In Brazil, IBGE hired consultants to execute specific operations in the demographic Census and in web scraping projects, without a long-term contract. Similarly, INE adopts outsourcing for programming hours and specific Big Data solutions. In a wider perspective, outsourcing in part of the ABD journey in ONS. For the British staff, outsourcing can present, not a risk, but gain in speed in the advancement of ABD projects.

For ABD in NSOs, the *risk of outsourcing* is summarised as:

1. Why is the risk of outsourcing complementary for ABD in NSOs? Because outsourcing works as alternative, not a central solution, to run Big Data projects when NSOs do not have the full expertise.
[ONS; INE; IBGE]

2. How does the risk of outsourcing affect ABD in NSOs? It affects ABD under a controlled supervision by NSOs' staffs, with those employees managing the outsourcing's legacy and knowledge. [INE; IBGE]

3. Requirements to avoid the risk of outsourcing and use outsourcing as a supplement for ABD:

a. Controlled outsourcing: short-term projects. [INE; IBGE]

b. Considering outsourcing as an alternative, not a full solution. [INE; IBGE]

c. Legacy and expertise are internalised, not kidnapped. [INE; IBGE]

211. Security and privacy regulations

Security and privacy regulations, as an environmental factor, are defined by Ahmad et al. (2015) as *"organisational concerns in ensuring compliance to security and data privacy regulations."* This is a normative factor for ABD in NSOs because those organizations based their adoption on security and privacy rules and laws to keep the citizens' data from hackers or anyone (even governments) with evil intentions. So, the security and privacy regulations factor affect ABD through the limitations and definitions of personal data collection and use.

Furthermore, while the TOE framework divides the adoption factors in a strictly linear manner, some of them work in a collaborative mode. In the context of NSOs, "security and privacy regulations" and "uncertainty and risk concern" appeared associated with each other in the interviews.

The NSO's business architecture relies on a concern for the risk to citizens' privacy and whether access to individual data is regulated. Communication with citizens is the adopted strategy to overcome the risk of a lack of confidence, sensitising them to the importance of OS.

"So, a great example of something which, virtually, everyone has it is a mobile phone. Not quite, but some levels: the data, or the movements is held through the telecom companies. We can often analyse that in partnership with them in a secure, ethical, and responsible way. But it gives a grey extra power on understanding, or movement of population. We need to be talking into the public about what we do on that dataset."

[03_ONS]

The Data Protection Act and the Digital Economy Act, both British laws, restrict ONS's action in Big Data collection from citizens. This law establishes guidelines for ONS's access to citizens' data and

their privacy. Additionally, this access and the procedures related to citizens' data has an extra-layer of regulation defined by the UK Statistical Authority. So, it is a controlled environment to deploy innovations, mainly about sensitive data, and Big Data. Security and privacy are the core themes of the concerns for ONS and others NSOs.

"I think we have the Digital Economy Act which complements the Data Protection Act and creates the pathways to allow data sharing. I think there are some challenges we need to overcome. So, one of the areas our data scientists are working on is the creation of synthetic data, (...) The Data Protection Act sets a high bar and we're use Data Science to work at how we get over that bar and meets those requirements"

[ONS_DPE]

In fact, *security and privacy regulations*, associated with *security and privacy* (technological factor) and *information security culture* (organisational factor), compose a central value for NSOs in the ABD processes. It is mandatory to preserve citizens' confidence during data collection operations for Official Statistics, both in small and also in Big Data. However, the current coverage of regulations can constrain ABD. This reality gives rise to a demand: a specific regulation for ABD in NSOs, including administrative records and personal data, where privacy is maintained, and adoption can advance.

"Looking from the perspective of IBGE, I perceive a more vehement need for attention to the issue of data confidentiality - sensitive data, as we call it. We, really, have to be preserved the confidentiality, and this is a very central issue for IBGE, which does not leak data. NSOs have requirements frameworks and all of them follow international good practices." [04_IBG]

"But there's a legal level - the question is 'how do I protect the confidentiality and comply with the RGPD?' So, there are other big questions: 'Where are the data storage servers?' 'Are they in Portugal?' 'Are they not?' In fact, there is the question of data sovereignties." [01_INE]

For ABD in NSOs, the *security and privacy regulation* factor is summarised as:

1. Why are security and privacy regulations normative to ABD in NSOs? Those organizations base their adoption on security and privacy rules and laws to keep the citizens' data protected from hackers. [ONS; INE; IBGE]

2. How do security and privacy regulations influence ABD in NSOs? They affect ABD through the limitations and definitions of personal data collection and use. [ONS; INE; IBGE]
3. Requirement for security and privacy regulations to support ABD in NSOs: the creation of a specific regulations for the Big Data collection and use by NSOs. [INE; IBGE]
4. Consequences of security and privacy regulation for NSOs:
 - a. Limitation in the regulation as an inhibitor for ABD. [INE]
 - b. Regulation is necessary; overregulation is damaging and a barrier for ABD in NSOs. [ONS; INE; IBGE]

212. Uncertainty and risk concern

Uncertainty and risk concern, according to Sun et al. (2018), refers to unexpected environmental events and unpredictable occurrences that limit the adoption of Big Data by organisations, such as NSOs. This factor impacts ABD because in the face of uncertainty, NSOs can delay or avoid ABD, maintaining the status quo. So, the *uncertainty and risk concern* factor affects NSOs through the avoidance of Big Data in function of the risky scenarios or impacts in the internal operational environment.

The following examples refer to the uncertainty and risks related to NSO operations: unexpected events, such as the COVID-19 pandemic; public image damage; loss of operational capacity; workforce loss due to retirements; and workforce aging. Some of those risks are unpredictable and others predictable, but hard to be managed by NSOs.

"The pandemic brought the issue of Big Data in practice, which is what we didn't have." [01_IBG]

"I think that if IBGE loses the collection network, agencies, in my opinion, IBGE ends. Then, other data generators can delivery Official Statistics in its place. What differentiates IBGE is this collection network, the 570 agencies." [01_IBG]

"In the European countries, in general, there is the ageing of the population as a fact. There's no denying it. So, the ageing of the population is in a mirror directly, and it has a direct mirror on Public Administrations." [02_INE]

For ABD in NSOs, the *uncertainty and risk concern* factor is summarised as:

1. Why are the uncertainty and risk concern an issue for ABD in NSOs? Because in the face of uncertainty, NSOs can delay or avoid ABD, maintaining the status quo. [INE; IBGE]
2. How does the uncertainty and risk concern affect ABD in NSOs? This factor affects NSOs through the avoidance of Big Data in function of the risky scenarios or impacts in the internal operational environment. [IBGE]
3. Examples of the uncertainty and risk concern for the ABD in NSOs:
 - a. COVID-19 pandemic. [IBGE]
 - b. Damage to public image. [IBGE]
 - c. Loss of capillarity. [IBGE]
 - d. Loss of workforce related to retirements. [INE; IBGE]
 - e. Ageing effect. [INE]

2X5. International–Institutional Context

This is the emergent environmental factor identified by this research. *International-institutional context* means the relevance of the NSO in the Global Statistical System network. Relevance is defined as the interactions, collaboration, and presence in international organisms or in partnerships with other NSOs. This factor is critical for ABD in NSOs because it allows the understanding about the external context and how it influences and has been influenced by the NSO, especially in the ABD journey.

Innovation in NSOs includes international cooperation with GSS. Additionally, cooperation with other NSOs improves the accumulated knowledge about Big Data technologies and methodologies.

“And it is made available for the United Nation Global Platform for Statistics and Data. Which we are working on in collaboration with other parts, with other NSOs across the world.” [03_ONS]

Regarding the position in GSS, NSO members realise the relevance of positions in this international network. The interviewees understand GSS is a relevant node in the global network of Official Statistics, including to support other NSOs.

“I think, to a large degree, we’re a leader rather than a follower in GSS.”
[04_ONS_DEX]

“So, we have a good network with all of those, including with IBGE, (...) despite the Brexit. We use that effort as good opportunities to learn. These nets allow ONS tries to adopt or share with others NSOs what we are doing. (...).” [02_ONS_DPE]

The multiple partnerships that ONS has maintained around the world are an example of how valuable the international context is to this NSO. International projects include the United Nations Committee of Big Data Experts, World Bank, the IMF, UNECE, CBS-ND, Uganda statistics, Senegal statistics, CBS statistics, Canada statistics, and IBGE.

Those organisations receive and offer knowledge, data, and new scenarios for ABD in Official Statistics. This set of partnerships indicates that ONS is not on the periphery of the GSS network. In the same way, IBGE works to influence GSS members with its best practices and results. Additionally, INE attempts to influence members of other international networks, such as OCDE and UNECE.

“Of course, IBGE participates in congress, events, groups, international discussions, of the UN and others, such as the World Bank, IDB. IBGE, participates in many decisions.” [06_IBG_COC]

“In which areas do INE and UNECE work? We use it in statistics which is GSBPM, General Statistical Business Process Model. It was developed in conjunction with the UNECE. So, we can see how important those international entities are. And there are others. For example, the OECD. The OECD, in fact, is a very important partner. Now, EUROSTAT is with whom we have more links.” [02_INE_DPE]

For ABD in NSOs, the *international-institutional context* factor is summarised as:

1. Why and how does the international-institutional context factor affect ABD in NSOs? This factor is critical because it allows the understanding about the external context and how it influences and has been influenced by the respective NSO, especially in the ABD journey. [ONS; INE; IBGE]
2. Requirement for the international-institutional context for ABD in NSOs: to work as the marketplace for technological and methodological Big Data solutions. [ONS; INE; IBGE]
3. Benefits from the international-institutional context for NSOs: to be in a relevant position in the GSS, developing ties with others NSOs. [ONS; INE; IBGE].

6.2. Summarizing Factor analysis

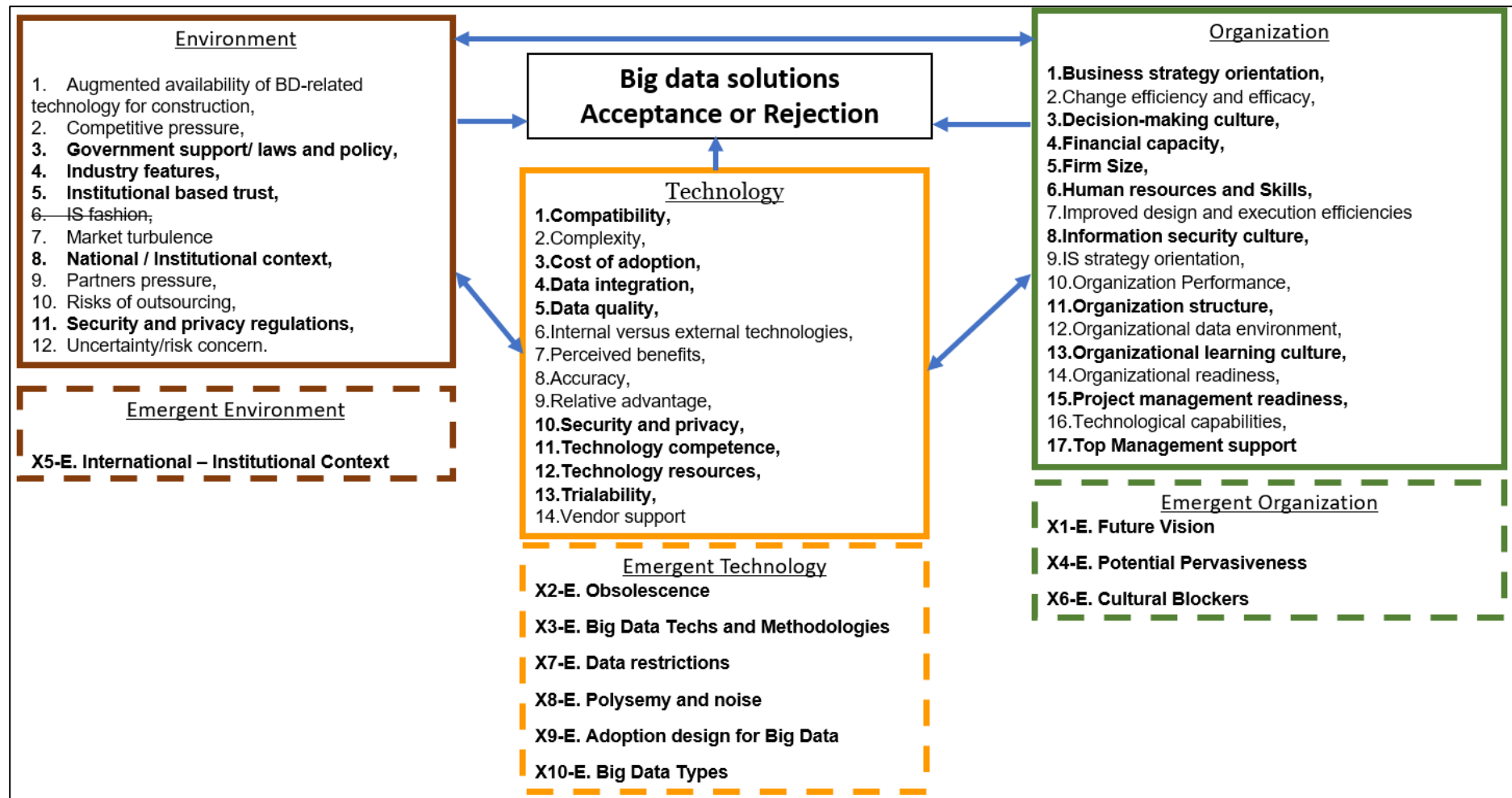
The interviews revealed a rich and complex landscape for ABD in NSOs. Indeed, the answers reinforced the idea that using Big Data for Official Statistics is not exclusively a technological issue. There are methodological, environmental, and, mainly, organisational factors that influence an adopter organisation.

One relevant finding in this case was the set of processes, features, requirements, and stages described by NSO members about how to adopt and deploy Big Data technologies and methodologies. Despite it being a non-conclusive journey, it keeps a clear focus on the organisational objectives, based on robust strategic planning, including IS strategy, business strategy, and human resources strategy.

Additionally, the environment really matters for NSOs and their efforts to adopt Big Data. Complex national and international contexts influence and are influenced by the NSO's actions. Regarding NSO's work on and for GSS, they use and receive benefits from this worldwide network.

Related to the new factors for the TOE framework, this is a contribution from the NSOs' perspective to this research. It is possible to assert that ABD in NSOs has specific features and details. The market and competition lens can create an analytical bias to understand how an NSO adopts Big Data. This is the reason why a specific study like this needs to take place. Figure 23 illustrates the advance of the TOE framework from the original one to the NSOs cases. The dashed boxes represent the emergent factors or codes. In the same perspective, the scratched words are the discard factors with no representation in the interviews.

In summary, TOE's three dimensions presented some factors as more relevant than others, considering the testimonials of the NSO members. Relevance includes the highlights attributed by NSO members during the interviews. Intentionally, the following Figure 23 shows the more relevant factors in bold type. The fact that thirty-three factors were highlighted indicates that the three dimensions present critical points about ABD in NSOs. Through this set of relevant factors, it is possible to develop the discussions, as explained in Chapter 7. Some are more critical than others, and those critical factors are identified graphically, in the TOE model, in Figure 23.



Legend: Continuous rectangles depict TOE factors found in the Literature review. Dashed rectangles depict emergent TOE factors resulting from this research. The colours represent the three TOE dimensions: Environmental factors are brown; Organisational factors are green; Technological factors are orange. Arrows represent influences from one factor to each other. Bold format refers to factors with stronger influence on the adoption or avoidance process. Not bold format refers to factors with weaker influence on the adoption or avoidance process.

Figure 23 – The adoption of Big Data in NSOs under the TOE Framework perspective

Chapter 7 - Findings and discussions

The interview analysis brought relevant findings for the theory and practices related to the adoption of Big Data (ABD) in National Statistics Offices (NSOs). In fact, those findings can fill up some theoretical gaps identified by previous researchers. One example is the proposal by Baker (2012) for research adopting the TOE framework in studies involving organisational networks. In this doctoral research, TOE is used to understand the distinct mode of adoption of Big Data in NSOs with different positions in the Global Statistical System network.

The three NSOs studied revealed that the Official Statistics business sector has idiosyncrasies regarding ABD relative to other business sectors. Different from previous studies, like McMahon *et al.* (2020) or Shahbaz *et al.* (2019), where ABD was scrutinised in aggregate national or regional contexts, the current study examined three individual NSOs. Following this comparative perspective, the types of data and technologies adopted by NSOs were also considered. More than creating a checklist of similarities and differences, this comparative analysis allowed for the understanding that ABD includes some specifications for each organisation, especially in the public sector. Related to those specificities, the business sector really matters regarding general factors, such as security levels of the core operations where ABD runs.

Another critical finding that contests the current knowledge about ABD refers to the impact on organisational processes. Sometimes the idea of ABD is presented with the adjective *disruptive*, referring to the idea of a turbulent and rowdy process. However, the analysis showed that the Official Statistics sector, as a whole, assimilates Big Data innovations not in a substitutive but complementary way to advance the delivery capacity of NSOs. Contrary to the idea of discarding the status quo in an NSO, ABD proposes to improve what is currently being done, not substitute it totally.

Still, two other findings about the TOE framework necessitate further discussion. Firstly, the relevance of some factors over others. In the quantitative studies examined in the literature review (Ghaleb et al., 2021; Gangwar, 2018; Lai et al., 2017), some factors presented more relevance after statistical analysis. In the same way, in this qualitative research, from the original group of TOE factors, associated with the emergent factors, a set appeared more influential and determinant in promoting ABD in NSOs. As previously stated, one purpose of this study is to fill some gaps in

academic research. Then, the correspondence between qualitative and quantitative studies can advance by verifying the relevance of TOE factors.

Secondly, the association of factors from different dimensions is another finding from this doctoral research. On the path to understanding how TOE translates the ABD process, it was possible to verify the association of Technology, Organization, and Environment factors in a scheme. This fabric of factors helps explain why ABD succeeds or fails in the Official Statistics context.

Another critical finding refers to how the organisational design of each NSO constrains or promotes ABD. In this aspect, the organisational singularity is heavily determinant.

In this chapter, those findings are discussed in seven subsections: (7.1) Dimensional precedence; (7.2) The most important factors for the adoption of Big Data in National Statistics Offices; (7.3) The idiosyncrasies of adoption of Big Data processes in NSOs; (7.4) Interdimensional association of factors for the adoption of Big Data; (7.5) Adoption or avoidance of Big Data? A cumulative process; (7.6) Big Data for Official Statistics: Beyond the Disruption.

7.1. Dimensional precedence

During the literature review stage, the bibliographic sources presented the three dimensions of the TOE framework in an associated and complementary mode. Each dimension has specific features and factors. From the original study by DiPietro *et al.* (1990), to the most recent one by Bakici *et al.* (2021), the TOE dimensions are shown as a static structure regarding the factors and their influence on the adoption process.

In the context of ABD by NSOs, it was possible to figure out an additional view of those dimensions. Following what the framework name (T-O-E) implies, the Technology dimension comes first because of its relevance. So, it is necessary to highlight that the Technology dimension comes first when relevance is the primary lens.

However, in this research, a temporal perspective arose from the data analysis. From this temporal perspective, the Environment dimension drastically influences the Organisation dimension. This finding is strongly relevant considering the previous studies (Salleh and Janczewski, 2016; Baig *et al.*, 2019) did not emphasized this aspect in TOE.

So, following Figure 23, in a clockwise direction, the analysis can start in the Environment dimension, move on to the Organisation dimension, and end in the Technology one.

According to this perspective, the Environment creates or develops conditions to influence the adopting organisation's decision to avoid or adopt Big Data solutions. These influences can coexist with adoption, but in this study, they represent an initial set of environmental factors necessary for the adoption to be prompted.

In cases where integration, exchanges, and associations had been promoted and stimulated by the environment, ABD advanced in an established pattern within a positive context. On the other side, an aggressive and disconnected environment blocks and disturbs the adoption process within a negative context. In relation to those two different contexts, positive and negative, one social actor, the government, appears as a determinant agent to interfere with the conditions for adoption processes. In this research, the factor of *government support/ laws and policy* is prevalent because the NSOs work in the Public Administration of each country. So, the government is the primary stakeholder when control is considered.

This control from the government is part of the factor *Industry features* and is associated with the *Government support/laws and policy* factor. Under the State's control, an NSO is influenced by the government's public policies and laws. In this way, ONS, IBGE, and INE suffer substantial environmental influence by the respective national government. Influence can be felt by the operational freedom to advance with administrative data uses by a NSO, or, in the opposite, with drastic budget cuts promoted by these governments.

Depending on the *national context* — another environmental factor — NSOs can work in a positive or negative environment. During the interviews, national subjects arose, such as "Brexit" in the case of ONS, the "National Budget Fail" (Chumbo do Orçamento) in INE's reality, or even the "political instability" that affected IBGE. Those national episodes are fingerprints proving how the *national context* influences the *government support/laws and policies* factor, consequently influencing the NSOs regarding the *Industry features*.

Then, a favourable context can promote integrations between an NSO and partner organisations, as in the cases of ONS and INE. Both work closely with universities, research centres, and think tanks. These collaborations benefit the NSOs in some organisational factors, such as *human resources and skills* with the development of new skills by the staff or the *organisation learning culture* with the learning of new technologies and methodologies related to Big Data.

Additionally, a positive Environmental context influences other Organisational Factors. To begin, an NSO with an integrated *Business strategy orientation* is more open to innovation and accepting of

changes or challenges than that of an isolated NSO. In the middle of this strategy, the *Decision-making culture* is more democratic and decentralised, maintaining internal harmony. With an open strategy and a flexible culture, *Financial capacity* can be improved with good practices learned from the organisational partners. In the same way, an *Information security culture* and the ability to work with project frameworks, as well as the *Project management readiness* are improved with knowledge exchanges between NSOs and their organisational partners. One positive example is the case of INE, where the integration with European Organisms (UNECE, EUROSTAT, European Commission) brought expertise, references, and technologies to advance Official Statistics using Big Data, while improving their technology capability.

In summary, taking a temporal analysis, it makes sense to interpret ABD in NSOs in a clockwise orientation regarding the TOE dimensions. Additionally, understanding that this temporal perspective is a feature of the Official Statistics business sector can allow its application in other studies about NSOs.

7.2. Most relevant factors for the adoption of Big Data in National Statistics Offices

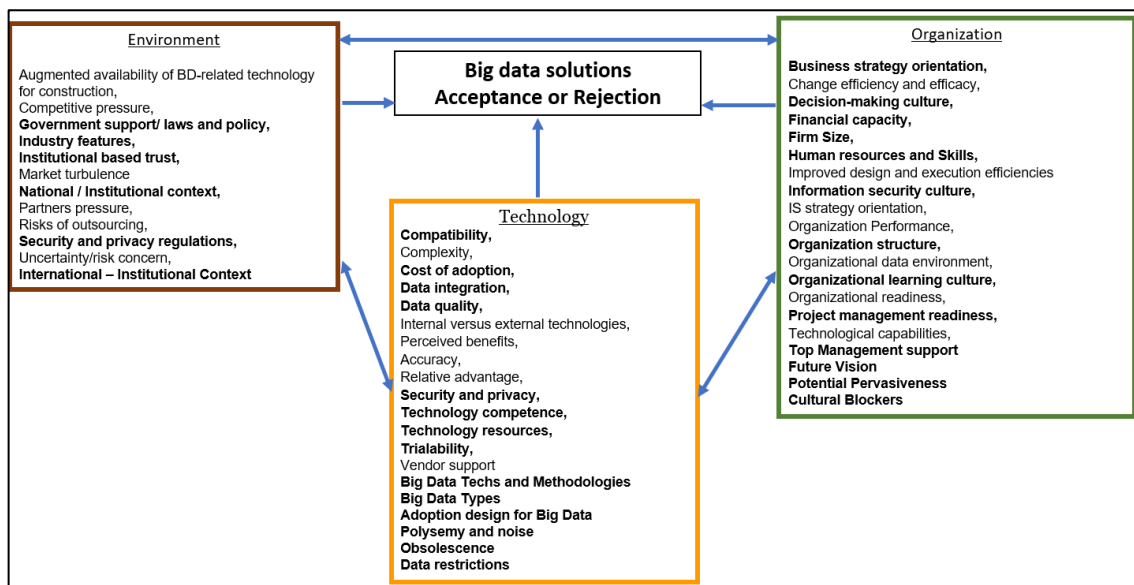
In the early stage of this research, during the Literature review, the references brought a set of forty-three factors into the TOE framework, with fourteen in the Technology, seventeen in the Organisation, and twelve in the Environment dimension, as presented in Figure 10 (page 92) in chapter 3. These forty-three factors represent the current stage of the theory related to ABD through the lens of the TOE framework.

Then, this doctoral research verified if those factors were also applicable in NSO contexts. Many of them adhered to ABD for Official Statistics. In this way, it was possible to verify those factors in the interviewees' answers, mainly in relevant statements about some factors, such as *Cost of Adoption*, *Complexity*, *Accuracy*, *Security and Privacy*, *Government support/ laws and policy*, *Institutional based trust*, and *Top Management Support*.

Indeed, the convergence of the current set of factors extracted from the literature and the cases results revealed two preliminary critical findings. Firstly, the TOE framework is an adaptative and resilient framework with over three decades of existence. This theoretical construction retains the ability to explain adoption of a specific technology or methodology. Secondly, the collective construction of the studies using TOE offers the possibility to test the set of factors in a specific context, such as Official Statistics. This testing capacity offers the possibility to compare

organisations regarding the same factors. In this way, just one environmental factor, *Information System fashion*, was discarded from the original set because it is more relevant in studies where the organisations operate in the private sector.

Further, a third finding, results from the emergence of new factors in this research. In total, ten new factors arose in the interviews. In the context of ABD in NSOs, six technological, three organizational, and one environmental new factors were identified. Summarizing, fifty-one factors, in the three dimensions are therefore considered, as presented in Figure 24.



Legend: Rectangles depict TOE dimensions and their respective factors. The colours represent the three TOE dimensions: Environmental factors are brown; Organisational factors are green; Technological factors are orange. Arrows represent influences from one factor to each other. Bold format refers to factors with stronger influence on the adoption or avoidance process. Normal type format refers to factors with weaker influence on the adoption or avoidance process

Figure 24 – TOE framework for the adoption of Big Data in National Statistics Offices: most relevant factors.

The emergence of six technological factors signalled ABD, which includes specific subjects represented by *Adoption design for Big Data*, *Big Data Types*, and *Big Data Techs and Methodologies*. Other new technological factors, such as *Polysemy and Noise*, and *Obsolescence*, provide information about the organisation's technological readiness to adopt or avoid Big Data solutions. If a technology team does not agree about what Big Data and its solutions mean, this polysemic noise will constrain the adoption process. Additionally, if the processes, equipment, and installations are not updated, the adoption will probably fail or be postponed. Finally, the last factor refers to data availability from a critical perspective. How would it be possible to adopt and deploy Big Data solutions without data? As a result, the *Data restrictions* factor imposes the constraints,

prohibitions, and bureaucracies faced by NSOs when trying to gain access to the datasets required to advance with the adoption of Big Data.

Despite the fact that Big Data is predominantly considered a technological theme, this research verified its intrinsic association with the organisational reality. So, these ten organisational factors played a critical role in the content analysis: *Business strategy orientation*, *Decision-making culture*, *Financial capacity*, *Firm size*, *Human resources and skills*, *Information security culture*, *Organisational structure*, *Organisational learning culture*, *Project management readiness*, and *Top management support*. In this dimension, those factors explained the relevant stages of adopting Big Data technologies and methodologies.

However, this adoption process is extensive and complex. So, three additional factors were identified and added to the original Organisational set. *Future vision*, *Potential pervasiveness*, and *Cultural blockers* added more exploratory capacity to the TOE framework. In this way, *Future vision* is associated with the *Business strategy orientation* and represents the desired scenario for NSOs after the ABD. In sequence, *potential pervasiveness* considers the capacity to spread ABD in other departments beyond those that adopted it initially. This new factor is associated with the *Firm size*, *Organisational structure*, and *Organisational learning culture*.

Finally, the *Cultural Blockers* factor introduces a unique aspect of NSO culture. Some strongholds of resistance related to ABD exist in those organisations. This resistance is promoted by professionals afraid of losing relevance in their current organisational roles and represents a divided context where part of the staff wants and demands Big Data. At the same time, another group tries to avoid and even block this adoption.

Last but not least, the Environment Dimension discarded one factor, *IS Fashion*, and added a new one, *International-Institutional Context*, regarding the influence of the Global Statistics System (GSS) Network on the NSOs. Those organisations attempt to work in harmony with the guidelines of the GSS. In fact, this new factor can contribute to future studies of adoption in transnational enterprises or global trade ecosystems.

In summary, some factors showed more relevance than others to the ABD in NSOs. Those factors can be considered more critical to understanding how NSOs adopt ABD. So, the last version of the TOE framework for this study is presented with some factors marked in bold type to show the strong relevance relative to the other factors, as presented in Figure 24.

7.3. The idiosyncrasies of adoption of Big Data processes in NSOs

Previous studies (Yadegaridehkordi et al., 2018; Ram et al., 2019) confirmed that companies adopt Big Data in particular ways, concluding that a factor can be relevant for a type of industry and irrelevant for others in different activities. We uncovered a number of aspects that are idiosyncratic to NSO's.

Under the temporal perspective of the TOE dimensions, there is a second important finding about ABD in NSOs. As Public Administration bodies, those organisations work under a differentiated procurement regime. While a private company can hire a Big Data specialist and, in sequence, buy related technologies, NSOs run the acquisition process in a bureaucratic way.

In the three countries where the studied NSOs are located, extensive regulations define how a public organisation can acquire products and services. Aside from the necessity of maintaining transparency for those processes, the rigour and number of stages have a negative impact on the ABD effort due to the nature of the public procurement process.

A well-specified description of the purchased item must be provided regarding the procurement process. If this stage is not well designed, a typical process can take a decade to be concluded, as in the case of IBGE in the purchase of one software license. So, the organisational factor of *Human resources and Skills* is determinant for running the ABD process in this research context. If an NSO has a staff with experts in Big Data methodologies and technologies, including Data Scientists, Engineering and Analysts, this procurement process can advance safely. The reason for this perspective is simple: top-skilled staff are more able to prepare a well-detailed procurement process for acquiring Big Data solutions. So, this scenario shows the link between two factors: *Human resources and skills* and *Technological competence*.

In terms of *Technology competence*, which necessitates *Human resource and skills*, the adoption or avoidance of Big Data is more a result of the Organisational dimension than a single technological process. Under this cause-and-effect relationship, ABD in NSOs differs from ABD in private companies because a second cause is the hiring process in the Public Sector.

Following public hiring processes, NSOs suffer, as reported in the interviews, from the bureaucracy of hiring top-skilled professionals. In this way, the British case presented a more balanced situation on this issue. When ONS demands a Data Scientist, for example, it is possible to attract this

professional from other public organisations, overcoming the slow process of hiring an employee from the private sector. Similarly, INE attempts to do the same.

In the context of NSOs, ABD is strongly influenced by *financial capacity* in addition to *Human resources and skills*. Those organisations work under budgetary limits as public bodies. If an NSO does not press the Central Government to acquire a substantial budget, its *Financial Capacity* is restricted, and the hiring effort fails.

Therefore, two organisational processes are vital to a successful ABD in NSOs: procurement and hiring. If an NSO has highly skilled staff, this group will be able to deploy a productive process for acquiring ABD solutions. Furthermore, there is a link between the hiring capacity and *financial capacity* since the ability to hire requires the availability of funds. This connection therefore represents a series of crucial elements that form the core of the ABD process.

7.4. Interdimensional association of factors for the adoption of Big Data

Maroufkhani et al. (2022) followed a quantitative method, using structural equations, to understand the interactions among TOE factors from different dimensions. Bakici et al. (2022), in turn, studied ABD following a qualitative method, but did not consider the interdimensional association of factors. This research attempts to fill the gap in the literature regarding the explanation of the interactions among TOE factors using a qualitative method.

Considering the qualitative approach, it was possible to verify that the interviewees understood the factors' interactions as a group of chains, similar to a cause-effect evolution. So, this qualitative analysis indicated the existence of layers among factors. Those interactions can be identified as a set of processes that build up to a central process of ABD.

The temporal perspective allowed by this qualitative approach to highlight a sequential relationship among the different factors along the ABD process. Some factors, in one dimension, link with other factors from another dimension. This verification is not found in quantitative studies on ABD, such as those analysed in the Literature Review (e.g., Baig et al., 2021; Park and Kim, 2021; Maroufkhani et al., 2020; Ghaleb et al., 2021).

As an integrative/adjudicative research (Cronin and George, 2020), this study remarks that factors work in an associative way to achieve ABD. Following this integrative perspective, the adoption process in NSOs is defined as an association of factor chains. Until the conclusion of the analysis, it was possible to verify the relevance of four chains of factors that have a key influence in the

adoption of Big Data technologies and methodologies. Three of those chains represent sequences of factors that, together, compose a further critical sequence, working with *Technology competence*, *Technology resources*, and *Trialability*.

Similar to branches in a tree, the three factors chains act like nodes in a main chain where the adoption process is finally concluded. In the NSOs context, the assimilation of *Big Data types* and the deployment of *Adoption design for Big Data* factor result in the use of *Big Data Technologies* and *Methodologies* during the Official Statistics daily processes. So, the factor chains labelled *Technology competence*, *Technology resource*, *Trialability* – that ultimately lead to a further ABD chain, are explained in Figures 25, 26, and 27. In those graphic representations, yellow balloons represent Technology factors. Green is the colour of Organisation factors, and brown refers to the Environment.

It should be noted that when a balloon shows the factor name and the "+" sign, it is an indication that this factor has a preceding chain/branch. Just to simplify visualization, the preceding chain is not expressed in all graphical structures.

The first chain of factors identified was the *Technology competence* branch, represented in Figure 25. This critical technological factor is a result of a sequence started with the *Top management support* for the ABD. If there is no real commitment from the *Top management support*, ABD in an NSO can be kept at the speech and narrative, not will translate into the practices. In the cases of INE and the ONS, the *top management support* plays this organisational role and promotes the adoption, including ensuring another critical factor, *Financial capacity*. Committed managers fight against external stakeholders to obtain budgets and funding for ABD. Those budget-seeking and other management efforts are allocated to investments in training and hiring. In the case of training for Big Data solutions, successful long-term training programs were launched in the cases of the ONS and INE. This training policy is part of the *Organizational learning culture*, a culture orientated towards learning the innovation related to *Big Data technologies and methodologies*.

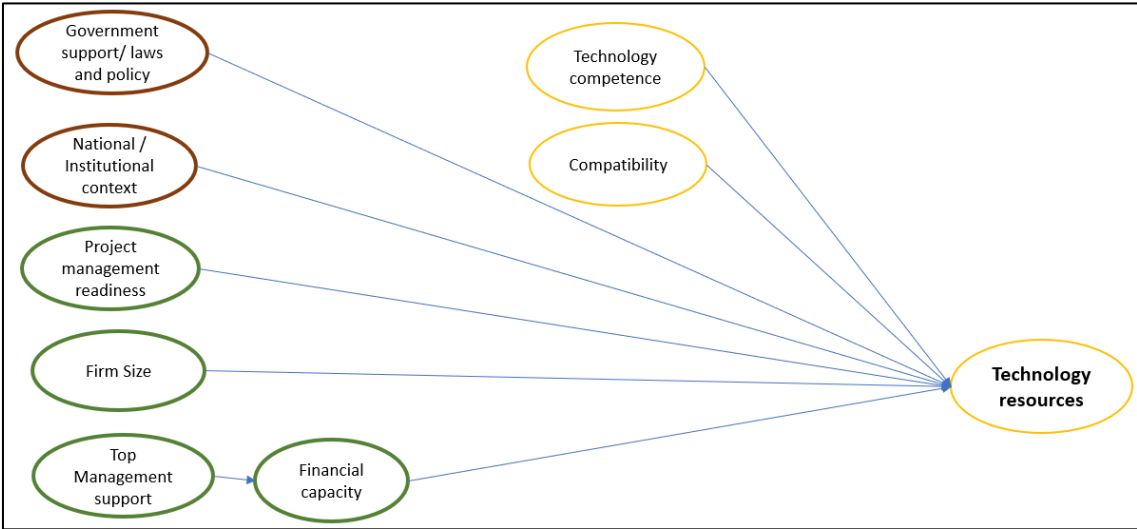


Legend: Circles depict factors influencing or being influenced by previous factors. Arrows represent influences from one factor to the next. The colours represent the two TOE dimensions: Organisational factors are green; Technological factors are orange.

Figure 25 – The factor chain for the Technology Competence

In fact, the knowledge obtained within this culture boosts staff competence. A staff prepared to adopt Big Data present skills in Data Science, Data Engineering, Data Analysis, and Business Analytics. All of them have to be associated with the traditional Statistics background. This association between staff and skills is related to the *Human resources and skills* factor. Sometimes, staff can take a long time to assimilate the skills. So, in this scenario, hiring professionals is advanced, despite the bureaucracy of the Public Service to recruit new employees. Finally, *Technology competence* can be achieved with employees able to deploy core actions. Those critical actions involve proposing, adopting, planning, deploying, and establishing *Big Data technologies and methodologies* to core activities at the initial stage. This factor, *Technology competence*, is a condition for the adoption processes. The lack of a clear follow-up of this chain of factors damages the ABD, generating a procrastination cycle, as observed in the IBGE.

Another branch identified leads up to *Technology resources*. This is influenced, among other factors, by *Technology competence*. Only with a well-established *Technology competence* is it possible to develop a procurement process for acquiring *Technology resources* for the ABD. In the same chain, the *Compatibility* between the traditional and new data sources is also critical for acquiring new technological infrastructure. So, new data sources adjustments are necessary to the current Official Statistics' dissemination schedule. Those factors links in this chain are expressed in Figure 26.

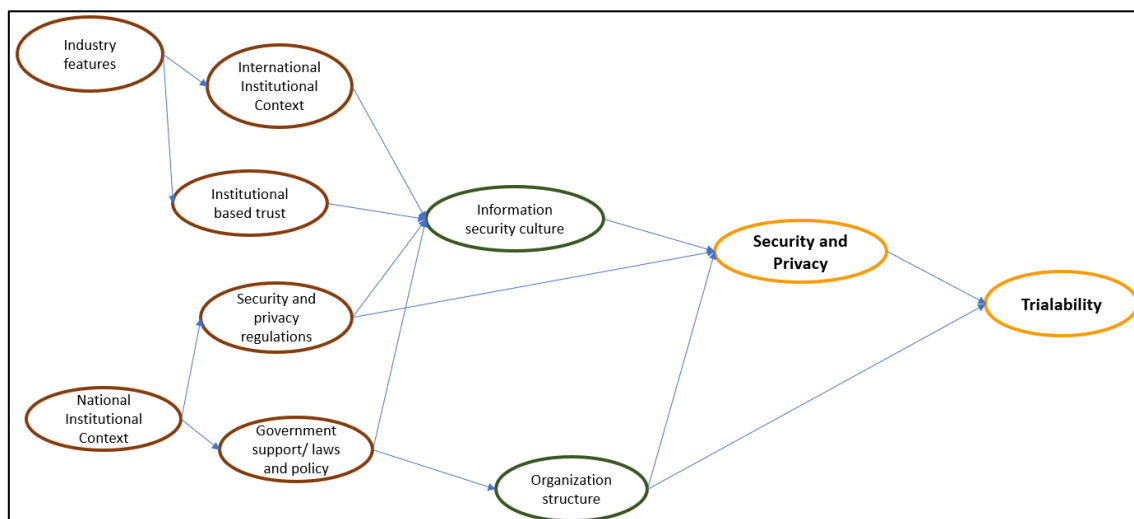


Legend: Circles depict factors influencing or being influenced by previous factors. Arrows represent influences from one factor to the next. The colours represent the three TOE dimensions: Environmental factors are brown; Organisational factors are green; Technological factors are orange.

Figure 26 - The factor chain for the Technology Resources

Additionally, two environmental factors influence the *Technology resources*. *Government support/laws and policies* can be a barrier for ABD. This link results from the NSO demands for complementary laws for Official Statistics and administrative records accesses. Still about legal issues, an NSO can require an update in the Statistical Law to achieve greater convergence with the ABD. If the Government denies this legal support, ABD will be marginal and peripheral. As a result, the NSO will be faced with a lack of access to administrative records, one of the data sources for the ABD. In advance, *National/institutional context* represents a critical factor in the acquisition process. A country suffering from domestic instability or under a poor Governmental network of collaboration can constrain the NSOs from advancing in the adoption process.

Regarding *Technology resources* – consisting of power computing, storage and software licenses – are affected by *Project management readiness*, *Firm size*, *Top management support* and *Financial capacity*. As evidenced in the interviews, NSOs require long-term projects to learn how to acquire adequate *Technology resources*, indicating *project Management readiness* as very relevant to advance this factor. Supposing an NSO has extensive territorial and national-scale coverage, as in the Brazilian case, the *Firm size* interferes because several departments will struggle for resources, and ABD can be considered less relevant. Additionally, the *Top management support* is significant because of the influence on the *Financial capacity*. As an interviewee stated, “*Big Data is very expensive*”. If an NSO does not plan to update its *Technology resources*, ABD will be part of the collective imaginary in that organisation but go no further. In this way, *Technology Resources* links all three dimensions in TOE and represent a necessary condition for the ABD in NSOs.



Legend: Circles depict factors influencing or being influenced by previous factors. Arrows represent influences from one factor to the next. The colours represent the three TOE dimensions: Environmental factors are brown; Organisational factors are green; Technological factors are orange.

Figure 27 - The factors chain for the Trialability

Trialability is another critical technological factor for ABD in NSOs, shown in Figure 27. NSOs members understood that *Trialability* has as its inner elements (a) the existence of a previous trailable mentality before trialability, (b) the use of a trial projects pipeline, (c) orderable, peer-reviewable, and reusable methods and procedures, (d) adopting new methodological approaches. Those four elements result from the influence of nine factors in the three dimensions of the TOE framework: six Environmental factors, two Organizational and one Technological.

Factors related to *Security and Privacy* are very important for *Trialability* because if experiments with ABD are seen to pose risks to data security and privacy, ABD will be avoided. *Security and Privacy* factors are present in the three dimensions. Those factors impact *Trialability* because either in the Environment, with regulations about *Security and Privacy*, or in the organisational policy for *Security and Privacy*, or in Technological processes to keep *Security and Privacy*, those two values are strongly critical for NSOs. As data generators, those public organisations are guardians of statistical secrets related to the citizens' privacy.

In the Organisation dimension, parallel with *Security and Privacy*, *Organisational structure* influences *Trialability*. If an NSO has a dedicated structure for controlling and auditing *security and privacy* processes, this supports the trial effort to adopt new Big Data solutions.

About the Environment dimension, *Industry features* and *National institutional context* bring complementary influences for the *Trialability*. National Statistics Offices base their business on *Institutional based trust*, which is an asset to be preserved by the *Information security policy*. Additionally, the *International institutional context* acts as a mindset maker on the *Information security* policy in the NSOs.

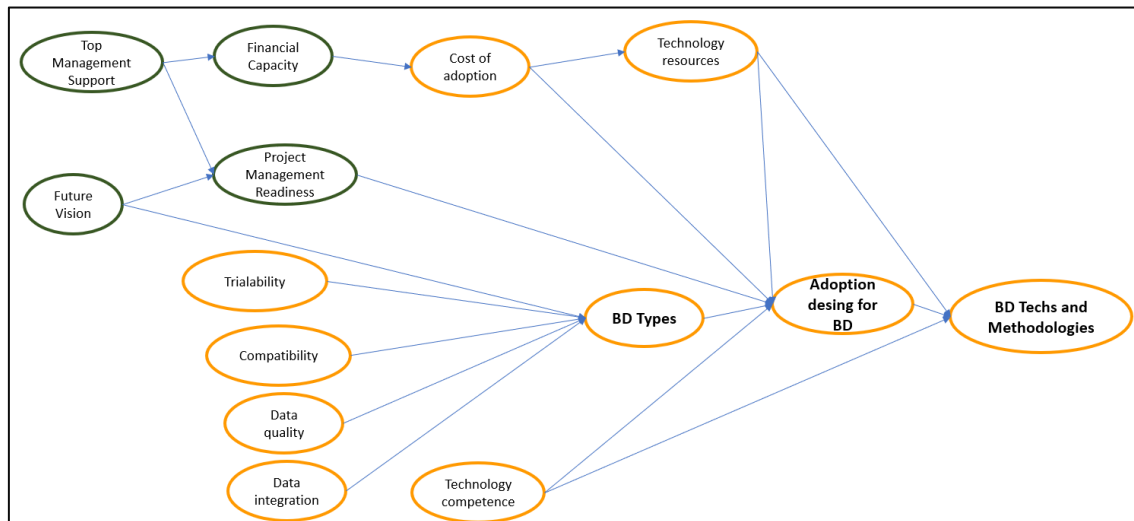
In the domestic scene, the *national/institutional context* brings the *security and privacy* regulations and the public policies or governmental actions. In the end, the regulations and laws affect the internal culture and the structure of NSOs. Suppose one of those organisations is firmly committed to updated security procedures. In that case, its staff will be confident to test, trial, propose and demand Big Data technologies and methodologies. So, those innovations will not be considered a risk for the statistical secrets or cybersecurity.

As a result of the financial, managerial and security concerns, ABD includes *Big Data types* to be addressed, the *Big Data technologies and methodologies* the NSO will run in the current use of Big Data. In this core factors chain (Figure 28), *Top management support* is critical, not only because

of *Financial capacity*, as explained before, but also related to *Project management readiness*. As interviewees highlighted, the *Top management support* promotes *Project management readiness*.

This Organisational factor, *Top management support*, advances with a top-down decision to adopt Big Data. In summary, the strategic decision from the top managers is translated into deploying task groups of adoption, stimulating collective ways for engaging the staff, and controlling the initial effort related to adoption actions.

Still in the Organisation dimension, the *Future vision* guides and sometimes imposes courses of action, including the types of Big Data an NSO will add to its core operational processes. Additionally, *Future vision* interferes with the *Project management readiness* when the staff takes part in extensive collective experience from long-term projects about the ABD.



Legend: Circles depict factors influencing or being influenced by previous factors. Arrows represent influences from one factor to the next. The colours represent two TOE dimensions: Organisational factors are green; Technological factors are orange.

Figure 28 - The factors chain for the adoption of Big Data: types, design, technologies, and methodologies

The Technology dimension presents specific links of influence in the adoption process. Administrative records and their harmonisation to be used as official data is one demand from the *Data quality*. In parallel, *Data integration* demands the ability to use (i) all types of data, including non-structured data, and (ii) the adoption of data measurements more precisely in real-time.

Regarding the NSOs having an extensive portfolio of Official Statistics, the Big Data has to be compatible with the current surveys. So, *Compatibility* is critical because the first milestone for an adopting NSO is to find how compatible traditional and new data sources are. In sequence, it is necessary to verify if the new technological and methodological procedures can deploy new

outcomes for Official Statistics or not. Advancing from those mentioned challenges, the adopting NSO still has to fit the new outcomes into the current Official Statistics' dissemination schedule. It is far from an easy journey.

In the end, a robust *Financial capacity* enables the NSO to manage the *Cost of adoption* and acquire necessary *Technology resources*. Associated with those three factors, the *Technology competencies* factor allows the NSO to design a real and possible adoption process. Additionally, it makes deploying *Big Data technologies and methodologies* into the NSO's routines feasible.

According to the interviews and the literature review, adopting Big Data is not a state or a checkbox question to mark as "adopted" or "avoided." The adoption takes time and advances gradually. So, when an NSO like ONS adopts methodologies such as Natural Language Process or includes administrative records in its portfolio of Official Statistics, it is possible to admit this organisation has adopted Big Data in its current stage.

In the Portuguese case, when an NSO uses Artificial Intelligence to support the demographic Census data collection or develops a national-wide administrative database with administrative records, it can be considered an adopter. In this scenario, it becomes clear that the adoption process of Big Data is advancing from an initial to an intermediate stage. In fact, the understanding and verification of an NSO as a Big Data adopter is present in its daily processes and projects, not in a signed document or manifesto generated by a board of directors.

In fact, ONS and INE are Big Data adopters, but at different stages. The case of IBGE is different because it is not in the initial stage of ABD, but in an avoidance position. This reality is an unintentional result of the lack of *Top management support* initiatives. Furthermore, individual projects without sponsorship are associated with sloppy *Financial capacity* to run long-term projects. A limited *Future vision* drives all those failures.

In summary, this research analysed three different NSOs at three distinct stages of ABD, including non-adoption, which we are here calling avoidance of Big Data. The TOE factors explain why ONS and INE run in one direction and IBGE is stuck in a different one, as described in the next section.

7.5. Big Data Adoption or Big Data Avoidance? A cumulative process

The decision to become a Big Data adopter is not taken by a single top manager in a single day. Previous studies have explored the critical importance of the adoption of Big Data for firms (e.g.,

Bakici et al., 2022; Park & Kim, 2019). However, specifically how this adoption advances into the organisations has not yet been studied.

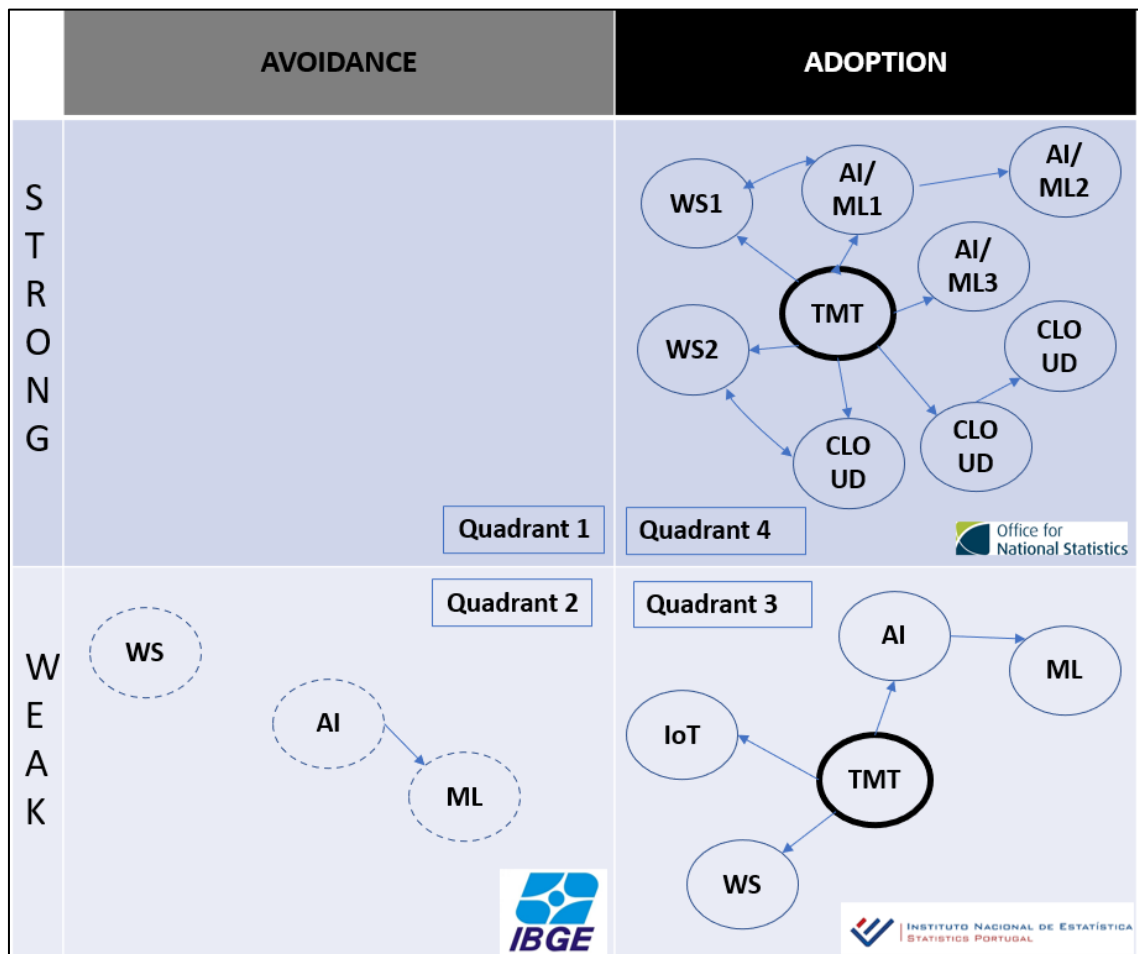
During this research it was possible to understand that ABD is a long and complex process that run along gradual stages. The adoption or avoidance of Big Data in NSOs can be positioned in a matrix representing those stages. The matrix results from the intersection of two dimensions: the attitude towards Big Data adoption, that can be “Avoidance” or “Adoption”; and the respective intensity, that can be “strong” or “weak”. The first quadrant refers to a strong avoidance attitude, and the second indicates cases where avoidance is weakened. In the third quadrant, there is adoption but with weak intensity. Finally, the fourth quadrant indicates a high adoption level.

The flat ellipses inside the quadrants refers to the Big Data technologies adopted by the NSOs. Dashed ellipses represent the use of those technologies in an avoidance context, and regular ellipses are located in the context of adoption. Inside the ellipses there are the acronyms for each Big Data technologies (AI/ML, Cloud, WS, IoT). If a NSO adopts the respective technology in more than one area or process, the acronym is presented with a number to show the adoption is advancing through the departments. Each NSO presented different set of those technologies. The bold ellipses refer to the top management team (TMT), who advances the top management support factor for the ABD. If a NSO has this bold ellipse, its strategic planning supports the ABD. If there is not the bold ellipse, the NSO do not support the ABD, and allows just isolated initiatives. This matrix is a finding from this research, resulting from the observation of the three cases studied, and it is labelled “Avoidance vs. Adoption of Big Data Matrix in National Statistics Offices”, exhibited in Figure 29.

In the cases of INE and ONS, both adoption processes started and still run continuously under a number of interactions with national and international partners, including the deployment of long-term projects. Indeed, the analysis of the interviews revealed the factors that influenced the adoption of Big Data in INE and ONS. In a different context, IBGE demonstrated reasons to avoid, in a weak way, the use of Big Data. Regarding the different stages of those three NSOs, it is possible to consider that each is on a specific position in the Figure 29. In the United Kingdom, ONS presented a strong adoption level of Big Data, while INE is at the weak one. In the Brazilian case, IBGE was positioned in the weak avoidance level of Big Data.

The classification of the three NSOs looked for understanding if the avoidance vs. adoption in Official Statistics was a continuum or a matrix. Figure 29 is a matrix because the avoidance and

adoption presented four degrees/stages. An NSO without any initiative for ABD is in an intensive avoidance stage (quadrant 1). In this research, none of the NSOs presented this extreme feature. In sequence, the Brazilian case is different (quadrant 2), considering IBGE has isolated initiatives for ABD. It suffers with the lack of strategic sponsoring by top management, represented in Figure 29 by a solid black circle with the acronym “TMS” (*Top management support*). In this way, IBGE is not struggling to avoid ABD, but attempting to advance despite the internal resistance (*Cultural blockers*).



Legend: AI - Artificial Intelligence; ML - Machine Learning; IoT - Internet of Things; WS - Web Scraping; CLOUD - Cloud Computing. The numbers (ML1, ML2, ML3...) depict the order of adoption by departments. Dashed circles: isolated Big Data adoption initiatives, without Top Management support. Continuous circles: adoption of Big Data technologies by departments. Bold circles: Top Management support for the adoption of Big Data. One-way arrows: support for Big Data adoption. Two-way arrows: knowledge exchange between department.

Figure 29 - Avoidance vs. Adoption Matrix for Big Data in National Statistics Offices

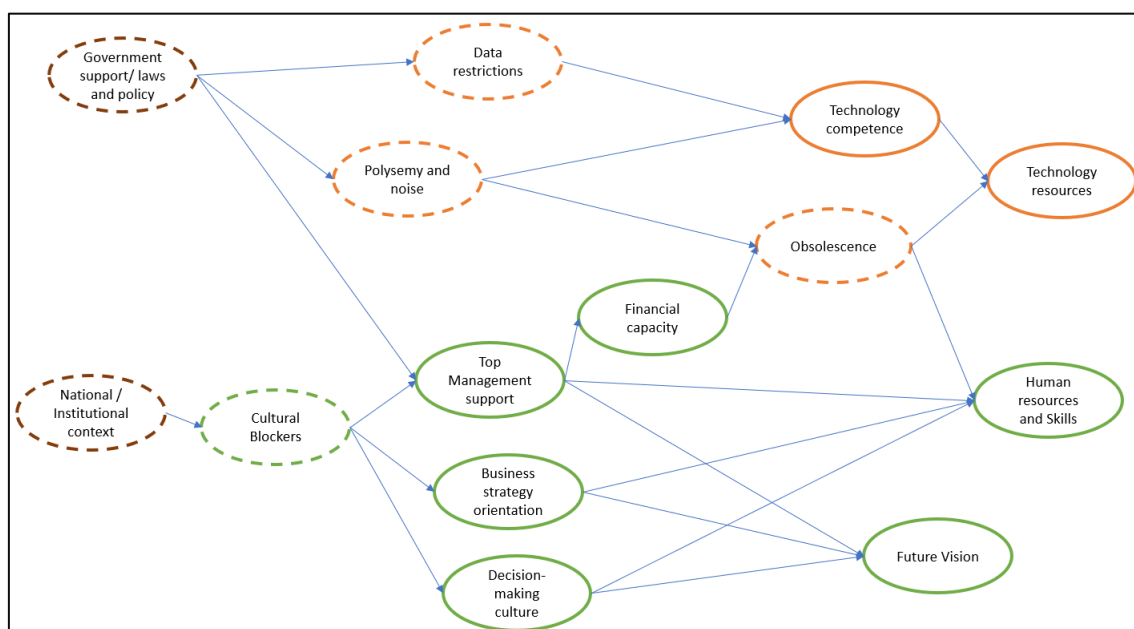
In the right side of the matrix, INE is in the bottom (quadrant 3) because its initiatives for ABD are strategically supported and are spilling over to different departments. However, bureaucratic obstacles constrain the advance of ABD, like some Portuguese laws (*Government support/laws and policy*).

In the case of ONS (quadrant 4), there is the representation of the most advanced stage of ABD. This NSO uses ABD in the Official Statistics to improve the processes, and at the same time, invests in experimental statistics to design new outcomes and deliveries.

Considering the results from the three cases, NSOs avoid or adopt Big Data in a flow. The orientation to understand the matrix must follow this sequence quadrant 1 (strong avoidance) » quadrant 2 (weak avoidance) » quadrant 3 (weak adoption) » quadrant 4 (strong adoption). Future studies can verify if a NSO is able to jump from quadrant 1 to quadrant 3 strictly. In this research, the finding showed this is a step-by-step movement.

The four different stages are listed as (i) strong avoidance, (ii) weak avoidance, (iii) weak adoption, and (iv) strong adoption, are explained by factors in the three dimensions of the TOE. They not only explain how NSOs progress along the path of ABD, but also how that progresses can be halted.

From the analysis of the cases, six factors emerged as exerting a specific negative effect on ABD. They are therefore considered constraining factors, as shown in Figure 30. In this illustration, those constraining factors are depicted in dashed circles. The interactions of this sextet work as a hidden engine that seizes the initiative skills of the NSOs members, blocking their efforts to adopt Big Data techniques and methodologies



Legend: Dashed circles depict avoidance of the specified factors. Continuous circles depict factors being influenced by previous factors. Arrows represent influences from one factor to the next. The colours represent the three TOE dimensions: Environmental factors are brown; Organisational factors are green; Technological factors are orange.

Figure 30 - How factors can block the ABD in NSOs

Firstly, in the Environment dimension, *Government support/laws and policy* can impose restrictions on access to administrative datasets, consequently inhibiting attempts by the NSO to adopt Big Data. In the same perspective, an outdated legal set for Official Statistics, as in the Brazilian case, contributes to the prohibition idea that advancing on ABD journey is forbidden. The reason for this restriction is that if it is not allowed to use administrative datasets, there is nothing to do beyond that. When there is obsolete legislation, it provokes polysemy about what Big Data is, and using administrative records is considered a legal taboo. As a result of strict data restriction and a polysemy effect resulting from outdated laws, the *Government support/laws and policy* factor also harms *Top management support*. In this situation, the selected board or the top management team are under a dome of poor legal support to advance and attempt any ABD.

Still, regarding environmental factors, the *National institutional context*, when unstable, can damage the *Top management support*. An unstable *National institutional context* means a scenario where the government and society do not attribute relevance to the Official Statistics. In this reality, employment in the NSO is not recognised as relevant. In addition, constant instabilities impose restrictions on managerial actions. This type of context promotes the generation of the most damaging of the factors — *Cultural blockers*.

Secondly, organisational factors might play critical roles in avoiding Big Data in NSOs. Mainly, the *Cultural blockers* represent the domain of an avoidance culture that is extremely retrograde and where any attempt to promote change, including ABD, is frustrated. In fact, *Cultural blockers* can be so caustic that sometimes employees associate informally to preserve the outdated status quo in the respective NSO. Much more than an isolated fact, this creation of an "avoidance team" in the NSO is an aggregation process generated from an unstable *National institutional context* associated with neglected *Governmental support*. An unstable *National institutional context* means a scenario where the government and society do not attribute relevance to the Official Statistics. However, the damages can be controlled in other cases when *Cultural blockers* are restricted and do not hold *Top management support*. As a result, their effects are marginal, with a reduced impact on ABD.

As a result of the power of *Cultural blockers'* power, *Top management support* can be influenced to avoid ABD. Consequently, this influence causes draining of the *financial capacity* for innovation, thus promoting an adverse selection in the human resources hiring processes. As a result of this constraint, an obscured *future vision* is established, imposing a lack of an innovation perspective.

In the same way, the *decision-making culture* is shifted to a passive one, where decentralised initiatives and proposals are discouraged. In this scenario, skilled employees attempt to abandon their job position at the NSO, aggravating the *Human resources and skills* situation.

Thirdly, the Technology dimension can constrain ABD with *Polysemy and noise*, *Data restriction*, and *Obsolescence*. In an avoidance orientation, those three factors collapse the *technology competence* formation process and disable the development of long-term projects or necessary procurement processes. In this scenario, the *technology resources* receive just maintenance, not updates, and ABD is present only in the top management speeches, not in practice.

Then, the avoidance reality (strong or weak) is a cumulative process developed during years of poor *government support*, lack of an updated legal support, a national context where Official Statistics are not considered relevant, and mainly the active domain of *cultural blockers*. In this way, the constraining scenario represents a sum of factors illustrated in the Figure 30.

According to Orlikowski (2000, p.422), hierarchical, individualistic, and competitive institutional conditions make an organisation adopt a limited use of technology. As a consequence, the status quo is preserved. In the same article, the author discusses other adopter organisations. Two types are represented in this research: "*collective problem solving*" and "*improvisation*." As evidenced by *collective-problem-solving* organisations, communal and cooperative actions improve a status quo oriented towards innovation and adoption. More advanced, the organisational type called "*improvisation*" works with cooperative projects deployed by team tasks under a learning orientation, with the proposal to transform and modernise the status quo.

Those three different institutional conditions, proposed by Orlikowski (2000), can be adapted to those research findings. In the stage of soft avoidance, IBGE presents hierarchical, individualistic, and competitive institutional conditions that materialised in the cultural blockers. This reality avoids ABD, allowing only limited and individual uses of some Big Data technologies by persistent members without corporative impacts. In the soft adoption stage, INE works in a communal orientation, using the matrix structure, achieving results with collaboration from teams in different cities and technical areas. In this case, the preserved status quo is the incremental-innovative one, positioning the Portuguese NSO in the middle level of adoption maturity for ABD. Finally, in the intensive adoption stage, the ONS developed its outcomes based on cooperative task teams with a target: to transform the status quo and make this NSO a leader in Big Data in Official Statistics.

To summarise, adoption or avoidance decisions are the result of interactions among factors over time. In this way, the definition of rejecting or accepting Big Data is the results of a cumulative

process influenced by the Technology, Organisation, and Environment dimensions. As a result, the ABD fails or advances not because the NSO decided this at one time. In fact, the internal and external institutional conditions allow or inhibit the NSO from adopting and advancing in the ABD journey. Without those conditions, the decision for ABD can be taken, but the deployment will remain an organisational illusion.

7.6. Big Data for Official Statistics: beyond disruption

According to Ghaleb et al. (2021), Big Data technologies, like the Cloud or IoT, can be considered disruptive technologies when adopted by a novice organisation. In the same way, Gangwar (2018) considered Big Data technologies disruptive in the manufacturing and service sectors. However, this doctoral thesis, unlike the current literature, finds that ABD in the NSO context is not disruptive. After the data analysis, this doctoral research proposes a different interpretation of ABD for National Statistics Offices. In the three case studies, IoT, Cloud, Web Scraping, and AI/Machine Learning, as Big Data technologies, were considered powerful and useful complements for Official Statistics production.

An ONS interviewee mentioned this transition from the disruptive to the complementary interpretation, explaining that, initially, Big Data was considered a new beginning for Official Statistics. After a while, ONS members realised that Big Data was a useful supplement to, not a replacement for, current Official Statistics.

So, the main change that the adoption of Big Data brings is the improvement in the production of the Official Statistics outcomes. This perspective can change with the progress of the adoption. Under the current view of the NSOs, Big Data does not apparently mean changes in the types of Official Statistics outcomes for the short term. Even with the use of Big Data, the price index, unemployment rate, industrial production rate, and demographic variation are still relevant Official Statistics.

According to this contemporary understanding, ABD brings changes to "how" (processes) Official Statistics are produced, not "what" (types) Official Statistics are. In fact, the three cases have confirmed this understanding. From the use of AI algorithms to support data collection in the demographic Portuguese Census to the expanded use of cloud computing in ONS and the limited application of web scrapping in IBGE, Big Data acts as a booster to support NSO members in their day-to-day processes.

In summary, ABD is welcomed by Official Statistics operators as a valuable asset to improve their delivery to society with more accuracy in real-time and with less demand on material and human resources. Indeed, Big Data is understood as an incremental innovation, not a breakthrough one. This positive perception encourages the adoption of Big Data as long as it does not jeopardise *institutional based trust*. On the contrary, Big Data can reinforce it.

Chapter 8 - Contributions, Limitations and Propositions

In this conclusion stage, it is necessary to remember how this research advanced and the limitations it suffered. Not only those aspects but also the contributions are summarised in this last chapter. Finally, the possibilities for future studies are proposed under the perspective that this exploratory result is just a starting point.

Adopting or avoiding Big Data technologies and methodologies represents a critical step for National Statistics Offices (NSOs). The seven chapters presented the relevance of Big Data and how it can contribute to Official Statistics. In Chapter 4 this question was proposed: "How and why do National Statistical Offices adopt or avoid Big Data technologies and methodologies?"

The answer to the research question was obtained following a specific methodological design. Based on the adjudicative perspective proposed by Cronin and George (2020), the methodology advanced a systematic literature review (LR). This LR was developed using a knowledge-synthesis orientation, integrating studies about Big Data adoption, including several business sectors, organisational types, and methodologies. As an outcome of the LR, the TOE factors were reviewed and, in some cases, redefined to add juxtapositions. This application of the knowledge-synthesis orientation contributes to reinforcing the approach developed Cronin and George (2020).

A multiple case-study approach was followed to pursue the empirical study. The case selection for this research followed three different criteria: governmental nature, geographic location, and position in the Global Statistical System (GSS). As a result, ONS (United Kingdom), INE (Portugal), and IBGE (Brazil) were selected. In fact, the differences between the cases offered a rich opportunity for comparisons and a broad understanding about ABD.

Using the study about TOE, a semi-structured interview guide was developed. It was applied to twenty representants from the three NSOs. In sequence, the interviews and the collected documents were analysed using codification as proposed by Saldaña (1993). This methodological strategy brought relevant findings, such as the specific features of ABD in the NSO context, and the stage each NSO is at in the ABD process. Conclusions emerged from those findings and are detailed in the next pages.

The overall answers found are, firstly, that NSOs adopt Big Data to avoid obsolescence and keep the relevance of Official Statistics for citizens and societies. Related to the mode of adoption, NSOs

adopt Big Data in incremental stages, initializing with selected deliveries, and in sequence, spreading this adoption for the other surveys and Census operations.

Related to avoidance, NSOs' resistance to ABD involves the attempt to keep the status quo, focusing on the preservation of internal outdated power structures. A combination of environmental and organisational factors affects the technological context, blocking the development and acquisition of Big Data technologies and methodologies.

In the end, this research brings a clear message for the guardians of Official Statistics: the adoption of Big Data is not a choice but a matter of survival for those organisations and their stakeholders.

8.1. Summarizing the journey

The initial idea for this research occurred seven years ago, at the beginning of 2016. Since that period, the demand for Big Data in the NSOs has been a common thought among those organisations. Over the years, this demand has remained high, and the aspiring future for Official Statistics using Big Data is a moving target.

Regarding this constant search for Big Data adoption, NSOs realised that, as a technological target, Big Data is not a push-button solution. Beyond this, Big Data is a rich asset for Official Statistics in the 21st century, and as an innovation, it demands adaptation and flexibility. Unlike previous studies, this research revealed an incremental orientation to the adoption process, not a disruptive one.

After forty-eight months, three international technical visits, three international conferences, twenty interviews, three NSOs accessed, and more than three hundred papers analysed, the result was satisfactory. Those numbers and efforts supported the documentary and content analysis.

During this trajectory, it was hard to identify how to compare an NSO with continental coverage with another located in an archipelago and another in a small European country. Even so, it was possible to advance with the comparability using the TOE framework.

Additionally, the participation of the PhD student in several seminars and workshops allowed a broad comprehension of how Big Data is powerful and valuable for Official Statistics. This understanding signalled to the researcher that ABD must overcome obsolescence risks. So, it is mandatory that every NSO adopt Big Data instead of avoiding it.

8.2. Contributions

This study incorporates both theoretical and practical contributions. On the first branch, the results propose an advance in the current theory. In the context of practice, the findings are helpful for NSOs to develop or update their strategies and processes.

Firstly, at the theoretical level, this research verified a novel feature about TOE framework, the dimensional precedence. In the context of NSOs, the Environmental dimension is the starting point influencing the adoption of Big Data. Key-factors like *government support/ laws and policy*, *National context*, and *Industry features* trigger the respective NSO in the direction of Big Data adoption or avoidance. In fact, this research showed that the decision to include IoT, Cloud, Web Scraping, and AI in the Official Statistics activities is a sequential and cumulative process. This process is drastically affected by the external environment, the Environmental dimension, and, in sequence, by the Organisational dimension. In the final process stage, the Technological dimension demonstrates a conditioned influence on the adoption of Big Data. So, the clockwise direction is necessary to explain the ABD in the NSOs reality.

Also at the theoretical level, this study contributes by asserting the relevance of some factors over others. In the three dimensions – Environmental, Organisational, and Technological –, there are subsets of factors that influence ABD more than others. *Institutional based trust* and *International - Institutional Context* are examples of the critical Environmental factors for ABD in NSOs. *Top Management Support*, *Cultural Blockers*, and *Potential Pervasiveness* are part of the determining Organisational factors subset. Last, but not least, *Polysemy and noise*, *Adoption design for Big Data*, and *Obsolescence* refers to the group of Technological factors with relevant influence on ABD in NSOs. As a qualitative research, this doctoral thesis complements the research about TOE framework for ABD, demonstrating that there are two levels of factor influence: some more and some less relevant.

Considering this relevance of factors from the three dimensions, one more finding emerged during the research: the interdimensional association. ABD in NSOs showed that factors interact with each other in chain of ties, inducing inputs to adopt or avoid Big Data technologies. This finding shows the factors work together, and not isolated in the respective dimension.

Regarding practical contributions, three specific findings were obtained: (i) idiosyncrasies of Big Data adoption processes in NSOs; (ii) adoption as a cumulative process over time, not a single decision at a certain point; and (iii) ABD as a complementary/incremental process. Beyond the

academic results, those three findings show solid orientations for NSO managers and staff to advance in their ABD journey.

Idiosyncrasies of adoption of Big Data processes in NSOs include the critical role of the procurement departments, as part of the business architecture. This finding is embedded in the Organisational factor *Human resources and Skills* because a highly skilled procurement staff is able to drive the procurement process to acquire Big Data technologies. In the same path, the hiring process is also another particularity of NSOs as public organisations. Submitted to the Central Government control, those organisations are restricted in hiring employees or consultants, different from private companies. Those specifications strengthen the ABD specific landscape in NSOs.

In this particular reality, the adoption of Big Data, or its avoidance, results from a cumulative process of interactions between the Top Management Team and the staff, influenced by the TOE factors remarked in Chapter Seven. This aggregation of interactions and decisions showed that ABD presents distinct intensities of avoidance or adoption. NSOs can adopt Big Data in a strong or weak mode. Strong adoption means spreading Big Data technologies in several departments and processes. A weak adoption signifies the use of Big Data is restricted to a few core processes, or in some experimental efforts. In the opposite direction, strong avoidance means the adoption of no Big Data technology whatsoever, whereas weak avoidance one refers to isolated adoptions without the support from the Top Management team.

The last finding refers to the adoption as an incremental/complementary process, not a disruptive one. This is the way the three NSOs understand and deploy ABD in their business architecture and processes. Progressive efforts to adopt the Big Data, technology by technology, reinforce this incremental orientation. When an NSO adopts Big Data, it adopts one technology, such as Web Scraping, or at most, two technologies in specific processes under an experimental mode. In this effort, the technological factor *Trialability* is critical. If an NSO supports and reinforces *Trialability*, it increased the possibility to advance ABD successfully. So, firstly, the NSO adopts Big Data experimentally. Based on the results, if satisfactory, the new technology is added to the Official Statistics processes. So, there is no evidence in this research that ABD disrupts the Business Architecture or dispels the original pipeline of each NSO in a dramatically mode. A managed transition appears as the rule of ABD in NSOs.

Beyond those six main findings, this doctoral research brought additional contributions. Firstly, the resilience of TOE was proven in this research, guiding the analysis with its logic of three dimensions and allowing findings about the specific context of Official Statistics. It is necessary to highlight that this context of Official Statistics had not been previously studied using the TOE framework.

Second, the TOE framework demonstrated flexibility to be adopted in this research as it was adopted in previous articles. One example is the proposition offered by Baker (2012) about applying TOE in adoption studies involving specific business sectors. In this doctoral research, TOE was feasible and intensively valuable for understanding how NSOs have adopted Big Data. TOE is thus an aggregative framework that can be applied to research agendas over time.

Regarding originality, the Literature Review (LR) followed a critical perspective, allowing the construction of an integrative compilation of the TOE factors. This compilation advanced to a review of the TOE framework through the consolidation of terms and concept for each factor. Indeed, the consolidation was necessary in the face of the verified polysemy related to the factors. In summary, a set of reviewed factors was identified and defined, and this set was adopted to understand how and why NSOs adopt or avoid Big Data. The concept of Big Data also presented polysemy and was therefore revised, in general, and specifically for the context of NSOs.

Based on the empirical study, this doctoral research advanced ten new TOE factors never mentioned before in the literature. This group of factors – including six new technological, three organisational, and one environmental – brought new possibilities to the evolution of the studies about ABD under the lens of the TOE framework. Likewise, the contribution of new factors showed the explanatory capacity of TOE in a complex qualitative study beyond the traditional quantitative papers where it is used to support structural equations.

As a result of this four-and-a-half-year research, the contributions of this research advanced on specific topics related to the adopters, the NSOs. One finding concerns the vital and urgent demand for adopting Big Data in the Official Statistics, considering this as a business sector. It is a matter of survival, not a choice. Avoidance is harmful for the NSOs persisting on it.

Related to additional findings, there is the concerns about ABD stages. Each NSO advances in the adoption effort and related organisational changes along specific stages. This process takes time – in some cases, over five years – as demonstrated in the Portuguese and British cases. As identified by this research, there are four phases for the ABD in NSOs: (i) strong avoidance, (ii) weak avoidance, (iii) weak adoption, and (iv) strong adoption. Each phase of adoption seems to be

related to a specific maturity level. Some NSOs are at a high maturity level, others in the middle, and some at a low one. However, there are NSOs that are at a rejection level, which is avoidance.

Finally, the last finding confirms that the adoption of Big Data need not have a disruptive impact on NSOs. When organizations adopt Big Data, the pace is incremental, gradually aggregating the new technologies and methodologies into the existing portfolio of outcomes. In some previous studies, Big Data and disruption were considered synonyms.

8.3. Limitations

This doctoral research was subjected to structural and situational limitations. In the first category, structural limitations, the methodological restrictions are explained in the following paragraphs. In addition, adverse events impacted the planned evolution of the study.

Regarding the methodological limitations, the multiple-case study focused on three specific realities: British, Portuguese, and Brazilian. As defined in the methodological chapter, the selection of the cases attempted to offer generalisations for the findings. However, this generalisation must be tested in a future qualitative study with other NSOs. So, the impossibility of checking the generalisation of those conclusions is a structural limitation of this research.

Although the use of the TOE framework to address the adoption of Big Data proved highly valuable in this research, other frameworks may be used and tested to understand how ABD advances in the context of Official Statistics and NSOs. Given the temporal dimension identified in this study, a Process Perspective (Langley et al., 2013) and Maturity models (Chen and Nath, 2018) are examples of analytical tools to be used in future studies.

As for practical constraints, the research plan was initially based on a schedule of forty-eight months, with the conclusion estimated for the 2022 second semester. However, the COVID-19 pandemic had a significant impact on this planning. The lockdowns in 2020 and 2021; the flight reductions between Brazil, Portugal, and the United Kingdom in 2021; and the sanitary barriers-imposed limitations on the research. Almost all interviews advanced remotely, contrasting with the original intention to conduct them in person. Because of the restrictions, ONS, INE, and IBGE had to postpone this researcher's access because those organisations were dealing with their reviews during the three Census calendars. Actually, it is possible to measure that the pandemic caused a delay of one semester in this research.

8.4. Future studies

An exploratory and qualitative study can be a starting point for future investigations like this doctoral research. A research agenda is proposed in this last section; understanding ABD in National Statistics Offices is a relevant theme for those organisations, governments, citizens, and also for the academy.

This research agenda can be advanced with a quantitative study applied to other NSOs, analysing ABD in their processes and outcomes. In this scenario, the universe of 198 NSOs should be considered. A clear objective can be proposed for this future research: identify if there are specific patterns for ABD in NSOs.

In another study, ABD can be verified in the complete portfolio of surveys or only in Census operations. However, if the second proposal is considered, only NSOs where the Census took place in the last three years should be included, considering that Big Data became part of the Census recently. All surveys or just the Census operation can allow the verification of best practices for ABD, generating rules to advance this adoption in the current decade for all GSS members.

Another potential research topic is examining the role of networks on ABD in NSOs where Big Data Hubs were deployed. In this line, NSOs from Brazil, China, Rwanda, and the United Arab Emirates could be studied qualitatively, following a similar orientation to this doctoral research. As a central objective, the verification of a spill-over effect to other regional partners can advance.

The role of networks could also be analysed by studying the impact of the Global Statistical System (GSS) on the ABD process. In this future study, it would be possible to investigate how and why some NSOs are more influenced than others by this global network. In this proposal, quantitative research can be adequate to discover if GSS influence works as a supportive environmental factor for the ABD.

8.5. Final Consideration: NSOs, an organisational species at risk of extinction?

Besides this doctoral research focus on the adoption of Big Data, it is necessary to understand the emergent findings about the adopters, the National Statistics Offices (NSOs). This group of organisations is dealing with threats, as Cardoso (2022) remarked. One is obsolescence, and the other is loss of social relevance, as highlighted by Allin (2021).

In this multiple-case study, it became clear that NSOs can face those threats when ABD is at a middle or high maturity level. So, those adopters' control and administer the changes and demands presented by governments and citizens. A demonstration of this capability is the perception of INE and ONS that Official Statistics production has to be faster and cheaper than it is nowadays. With reduced expenses to deliver an outcome, those NSOs plan to keep their relevance for the population, in spite of the escalation of fake news.

On the other side, IBGE is still operating from an outdated perspective, concluding the Demographic Census with three years of delay and at a cost of around half a billion euros³. Contrary to this, the IBGE members realise the Brazilian Official Statistics have to advance and be modernised, but this perception is not reflected in corporate initiatives. In cases similar to the Brazilian, NSOs are at considerable risk of losing social relevance. If this scenario is fulfilled, NSOs can be substituted by private data generation, compromising not only Official Statistics in some countries but also the credibility of governments and nations.

Another risk refers to the Global Statistical System. In a scenario where some members of this network will go extinct or are unable to keep the exchange flow of knowledge going, the GSS will suffer a disruption, and this system can start to collapse.

The GSS members must struggle to avoid the extinction of any NSO. This proposition can come by influencing traditional organisations such as the OECD and UN to have their members' input on the mandatory preservation of the National Statistics Offices in each country. Concerning preservation, this can be examined by peers. In this context, inspectors from the GSS would verify whether an NSO was achieving the Global patterns and principles of the Official Statistics during an in-field inspection. A similar approach has been taken by the International Atomic Energy Agency (IAEA)⁴. Those inspections would attribute a quality certification to an NSO, provided the organisation runs the Official Statistics regularly, just like the IAEA does with atomic energy uses in many countries.

³ Reference value of the exchange rate between Brazilian Real (BRL) and Euro (EUR) in the launching day of the demographic Census in Brazil, August 1st 2022. Retrieved from <https://www.bcb.gov.br/conversao>, in January 10th 2023

⁴ Retrieved from <https://www.iaea.org/topics/verification-and-other-safeguards-activities>, in January 10th 2023.

9. References

- Aguinis, H., Ramani, R. S., & Alabduljader, N. (2020). Best-Practice Recommendations for Producers, Evaluators, and Users of Methodological Literature Reviews. *Organizational Research Methods*, 26(1), 1–31. <https://doi.org/10.1177/1094428120943281>
- Ahuja, M. K., & Thatcher, J. B. (2005). Moving beyond Intentions and toward the Theory of Trying: Effects of Work Environment and Gender on Post-Adoption Information Technology Use. *MIS Quarterly*, 29(3), 427–459. <https://www.jstor.org/stable/10.2307/25148691>.
- Ajzen, I. (1991). The theory of planned behavior. *Organizational Behavior and Human Decision Processes*, 50(2), 179–211. [https://doi.org/10.1016/0749-5978\(91\)90020-T](https://doi.org/10.1016/0749-5978(91)90020-T)
- Al-Rahmi, W. M., Yahaya, N., Aldraiweesh, A. A., Alturki, U., Alamri, M., Bin Saud, M. S., Kamin, Y. Bin, Aljeraiwi, A. A., & Alhamed, O. A. (2019). Big Data Adoption and Knowledge Management Sharing: An Empirical Investigation on Their Adoption and Sustainability as a Purpose of Education. *IEEE Access*, 7, 47245–47258. <https://doi.org/10.1109/ACCESS.2019.2906668>
- Al Hadwer, A., Tavana, M., Gillis, D., & Rezanian, D. (2021). A Systematic Review of Organizational Factors Impacting Cloud-based Technology Adoption Using Technology-Organization-Environment Framework. *Internet of Things*, 15, 100407, 1–10. <https://doi.org/10.1016/j.iot.2021.100407>
- 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>
- Almeida, P. (2022, November 1st). Com três meses de coleta , Censo 2022 tem 66 % da população contabilizada. *CNN Brasil*.
- Alves-mazzotti, A. J., & Gewandsznajder, F. (2002). O Método nas Ciências Naturais e Sociais: Pesquisa Quantitativa e Qualitativa. In *O Método nas Ciências Naturais e Sociais: Pesquisa Quantitativa e Qualitativa* (2nd ed.). Pioneira.
- Amer, K., & Noujaim, J. (2019). *The Great Hack*. NETFLIX.
- Angst, C., & Agarwal, R. (2016). Adoption of Electronic Health Records in the Presence of Privacy Concerns: The Elaboration Likelihood Model and Individual Persuasion. *MIS Quarterly*, 33(2), 339–370.
- Angwar, H. (2018). Understanding the determinants of big data adoption in India: An analysis of the manufacturing and services sectors. *Information Resources Management Journal*, 31(4),

- 1–22. <https://doi.org/10.4018/IRMJ.2018100101>
- Baig, M. I., Shuib, L., & Yadegaridehkordi, E. (2019). Big data adoption: State of the art and research challenges. *Information Processing & Management*, 56(6), 1–18. <https://doi.org/10.1016/j.ipm.2019.102095>
- Baig, M. I., Shuib, L., & Yadegaridehkordi, E. (2021a). A Model for Decision-Makers' Adoption of Big Data in the Education Sector. *Sustainability*, 13(24), 13995. <https://doi.org/10.3390/su132413995>
- Baig, M. I., Shuib, L., & Yadegaridehkordi, E. (2021b). A Model for Decision-Makers' Adoption of Big Data in the Education Sector. *Sustainability*, 13(24), 1–29. <https://doi.org/10.3390/su132413995>
- Baker, J. (2012). The Technology–Organization–Environment Framework. In *Information systems theory*, 231–245. Springer. https://doi.org/10.1007/978-1-4419-6108-2_12
- Bakici, T., Nemeh, A., & Hazir, O. (2022). Big Data Adoption in Project Management: Insights From French Organizations. *IEEE Transactions on Engineering Management*, 1–15. <https://doi.org/10.1109/TEM.2021.3091661>
- Bandura, A. (1985). Model of Causality in Social Learning Theory. In *Cognition and Psychotherapy* (pp. 81–99). Springer. https://doi.org/10.1007/978-1-4684-7562-3_3
- Bardin, L. (2016). *Análise de Conteúdo* (1st). Almedina Brasil.
- Bayerl, P., Lauche, K., & Axtell, C. (2016). Revisiting group-based technology adoption as a dynamic process: the role of changing attitude-rationale configurations. *MIS Quarterly*, 40(3), 775–784.
- 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>
- Bollineni-Balabay, O., van den Brakel, J., & Palm, F. (2015). Multivariate State Space Approach to Variance Reduction in Series with Level and Variance Breaks Due to Survey Redesigns. *Journal of the Royal Statistical Society Series A: Statistics in Society*, 179(2), 1–26. <https://doi.org/10.1111/rssa.12117>
- Bonner, W. B. (2013). History and IS – Broadening Our View and Understanding: Actor–Network

- Theory as a Methodology. *Journal of Information Technology*, 28(2), 111–123.
<https://doi.org/10.1057/jit.2013.6>
- Bose, I., & Leung, A. C. M. (2019). Adoption of identity theft countermeasures and its short- And long-term impact on firm value. *MIS Quarterly: Management Information Systems*, 43(1), 313–327. <https://doi.org/10.25300/MISQ/2019/14192>
- Bowen, G. A. (2009). Document Analysis as a Qualitative Research Method. *Qualitative Research Journal*, 9(2), 27–40. <https://doi.org/10.3316/QRJ0902027>
- Bremser, C. (2018). Starting points for big data adoption. *26th European Conference on Information Systems: Beyond Digitization - Facets of Socio-Technical Change, 2018*, 1–14.
https://aisel.aisnet.org/ecis2018_rp/32
- Bremser, C., Piller, G., & Rothlauf, F. (2017). Strategies and Influencing Factors for Big Data Exploration. Full Paper. *Twenty-Third Americas Conference on Information Systems*, 1–10.
<https://aisel.aisnet.org/amcis2017/AdoptionIT/Presentations/4/>
- Briner, R. B., Denyer, D., & Rousseau, D. M. (2009). Evidence-Based Management: Concept Cleanup Time? *Academy of Management Perspectives*, 23(4), 19–32.
<https://doi.org/10.5465/AMP.2009.45590138>
- Brock, V., & Khan, H. U. (2017). Big data analytics: does organizational factor matters impact technology acceptance? *Journal of Big Data*, 4(21), 1-28. <https://doi.org/10.1186/s40537-017-0081-8>
- Brown, S., & Venkatesh, V. (2005). Model of Adoption of Technology in Households: A Baseline Model Test and Extension Incorporating Household Life Cycle. *MIS Quarterly*, 29(3), 399–426.
- Burrell, G., & Morgan, G. (1979). Assumptions about the Nature of Social Science. In *Sociological Paradigms and Organisational Analysis: Elements of the Sociology of Corporate Life* (1st). Ashgate Publishing Limited.
- Cabrera-Sánchez, J.-P., & Villarejo-Ramos, Á. (2019). Fatores que afetam a adoção de análises de Big Data em empresas. *RAE - Revista de Administração de Empresas*, 59(6), 415–429.
- CABRERA-SÁNCHEZ, J.-P., & VILLAREJO-RAMOS, Á. F. (2019). Factors affecting the adoption of big data analytics in companies. *Revista de Administração de Empresas*, 59(6), 415–429.
<https://doi.org/10.1590/s0034-759020190607>
- Cabrera-Sánchez, J. P., & Villarejo-Ramos, Á. F. (2020). Acceptance and use of big data techniques in services companies. *Journal of Retailing and Consumer Services*, 52(101888), 1-8.

- <https://doi.org/10.1016/j.jretconser.2019.101888>
- Cardoso, F. dos S., Varajão, J., & Carvalho, A. (2021). The adoption of Big Data technologies - A challenge for National Statistics Offices. *American Conference of Information Systems 2021*, 1-5.
https://aisel.aisnet.org/amcis2021/data_science_decision_support/data_science_decision_support/13
- Chau, P. Y. K., & Tam, K. Y. (1997). Factors affecting the adoption of open systems: An exploratory study. *MIS Quarterly: Management Information Systems*, 21(1), 1–20.
<https://doi.org/10.2307/249740>
- Chen, H. M., Kazman, R., Schütz, R., & Matthes, F. (2017). How Lufthansa capitalized on big data for business model renovation. *MIS Quarterly Executive*, 16(1), 19–34.
- Chen, L., & Nath, R. (2018). Business analytics maturity of firms: an examination of the relationships between managerial perception of IT, business analytics maturity and success. *Information Systems Management*, 35(1), 62–77.
<https://doi.org/10.1080/10580530.2017.1416948>
- Chua, Lim, Soh, & Sia. (2012). Enacting Clan Control in Complex IT Projects: A Social Capital Perspective. *MIS Quarterly*, 36(2), 577. <https://doi.org/10.2307/41703468>
- Côrte-Real, N., Oliveira, T., & Ruivo, P. (2017). Assessing business value of Big Data Analytics in European firms. *Journal of Business Research*, 70(February), 379–390.
<https://doi.org/10.1016/j.jbusres.2016.08.011>
- Cresswell, J., & Cresswell, J. D. (2018). *Research Design - Qualitative, Quantitative, and Mixed Methods Approaches* (5th). SAGE Publications, Inc.
- Cronin, M. A., & George, E. (2023). The Why and How of the Integrative Review. *Organizational Research Methods*, 26(1), 1–25. <https://doi.org/10.1177/1094428120935507>
- Daradkeh, M. K. (2019). Determinants of visual analytics adoption in organizations: Knowledge discovery through content analysis of online evaluation reviews. *Information Technology and People*, 32(3), 668–695. <https://doi.org/10.1108/ITP-10-2017-0359>
- Data Science Campus. (2023). *About us*. Data Science for the Public Good.
<https://datasciencecampus.ons.gov.uk/about-us/>
- Davis, F. D., Bagozzi, R. P., & Warshaw, P. R. (1989). User Acceptance of Computer Technology: A Comparison of Two Theoretical Models. *Management Science*, 35(8), 982–1003.
[https://doi.org/0025-1909/89/3508/0982\\$01.25](https://doi.org/0025-1909/89/3508/0982$01.25)

- Davis, F. D. J. (1986). *A technology acceptance model for empirically testing new end-user information systems: Theory and results*. Massachusetts Institute of Technology. <http://dspace.mit.edu>
- DePietro, R., Wiarda, E., & Fleischer, M. (1990). The context for change: Organization, technology and environment. In L. G. Tornatzky & M. Fleischer (Eds.), *The processes of technological innovation* (pp. 151–175). Lexington Books.
- DeLone, W. H., & McLean, E. R. (2003). The DeLone and McLean Model of Information Systems Success: A Ten-Year Update. *Journal of Management Information Systems*, 19(4), 9–30. <https://doi.org/10.1080/07421222.2003.11045748>
- DePietro, R., Wiarda, E., & Fletscher, M. (1990). The Context for Change: Organization, Tehcnology, and Environment. In L. Tornatzky & M. Fleischer (Eds.), *The Processes of Technological Innovation* (pp. 151–175). Lexington Books. [papers2://publication/uuid/88F69569-33FC-441F-9A3E-EAD0AED4AD11](https://publication/uuid/88F69569-33FC-441F-9A3E-EAD0AED4AD11)
- Diretoria Regional de Estatística da Madeira. (2023). *Mission, Vision and Values*. Mission, Vision and Values. <https://estatistica.madeira.gov.pt/en/drem/instituicao/missao-gb.html>
- Doolin, B., & Lowe, A. (2002). To Reveal is to Critique: Actor–Network Theory and Critical Information Systems Research. *Journal of Information Technology*, 17(2), 69–78. <https://doi.org/10.1080/02683960210145986>
- Duan, W., Gu, B., & Whisnton, A. (2017). Informational Cascades and Software Adoptio on the Internet: An Empirical Inves. *MIS Quarterly*, 110(9), 1689–1699. <https://www.jstor.org/stable/20650277>.
- Eisenhardt, K. M. (1989). Building Theories from Case Study Research. *The Academy of Management Review*, 14(4), 532–550. <https://www.jstor.org/stable/258557>
- El-Haddadeh, R., Osmani, M., Hindi, N., & Fadlalla, A. (2021). Value creation for realising the sustainable development goals: Fostering organisational adoption of big data analytics. *Journal of Business Research*, 131, 402–410. <https://doi.org/10.1016/j.jbusres.2020.10.066>
- European Commision. (2022). *Data Protection in the EU*. Data Protection in the EU. https://commision.europa.eu/law/law-topic/data-protection/data-protection-eu_en.
- Eurostat. (2023). *Who we are*. About Us. <https://ec.europa.eu/eurostat/web/main/about-us/who-we-are>
- Félix, B. M., Tavares, E., & Cavalcante, N. W. F. (2018). Fatores críticos de sucesso para adoção

- de Big Data no varejo virtual: Estudo de caso do Magazine Luiza. *Revista Brasileira de Gestao de Negocios*, 20(1), 112–126. <https://doi.org/10.7819/rbgn.v20i1.3627>
- Ferfolja, T., & Burnett, L. (2002). Getting Started on Your Literature Review. *The Learning Centre, UNSW*, 1–4. <http://www.lc.unsw.edu.au/onlib/pdf/Litrev.pdf>
- Fichman, R. G., & Kemerer, C. (1999). The Illusory Diffusion of Innovation: An Examination of Assimilation Gaps. *Information Systems Research*, 10(3), 255–275. <https://doi.org/10.1047/7047/99/1003/0255>\$05.00
- Fishbein, M., & Ajzen, I. (1977). *Belief, attitude, intention, and behavior: An introduction to theory and research*. Philosophy and Rhetoric. <http://philarchive.org>
- Fisher, C., Buglear, J., Lowry, D., Mutch, A., & Tansley, C. (2007). Researching and Writing a Dissertation: A Guidebook for Business Students. In *Pearson Education Limited* (2nd ed.). Pearson Education Limited.
- Francisco, R., & Meirelles, D. (2016). Technologies for Distance Education and the Development of Technological Competence: a Case Study in an Institution of Higher Education. *Revista de Estudios Para El Desarrollo Social de La Comunicación*, 12, 150–169. <https://doi.org/10.15213/redes.n12.p150>
- Gallivan, M. J. (2001). Organizational Adoption and Assimilation of Complex Technological Innovations: Development and Application of a New Framework. *Data Base for Advances in Information Systems*, 32(3), 51–85. <https://doi.org/10.1145/506724.506729>
- Gandra, A. (2022, July 31). IBGE lança Censo 2022 nesta segunda. *Geral*. <https://agenciabrasil.ebc.com.br/geral/noticia/2022-07/ibge-lanca-censo-2022-nesta-segunda>
- Gangwar, H. (2018). Understanding the determinants of big data adoption in India: An analysis of the manufacturing and services sectors. *Information Resources Management Journal*, 31(4), 1–22. <https://doi.org/10.4018/IRMJ.2018100101>
- Gangwar, H. (2020). Big Data Analytics Usage and Business Performance: Integrating the Technology Acceptance Model (TAM) and Task Technology Fit (TTF) Model. *Electronic Journal of Information Systems Evaluation*, 23(1), 45–64. <https://doi.org/10.34190/ejise.20.23.1.004>
- Ghaleb, E. A. A., Dominic, P. D. D., Fati, S. M., Muneer, A., & Ali, R. F. (2021). The Assessment of Big Data Adoption Readiness with a Technology–Organization–Environment Framework: A Perspective towards Healthcare Employees. *Sustainability*, 13(15), 1–33.

- <https://doi.org/10.3390/su13158379>
- Gilbert, M., & Cordey-hayes, M. (1996). Understanding the process of knowledge transfer to achieve successful technologicla innovation. *Technovation*, 16(6), 301–312.
- Goodhue, D. L., & Thompson, R. L. (1995). Task-Technology Fit and Individual Performance. *MIS Quarterly*, 19(2), 213-236. <https://doi.org/10.2307/249689>
- Günther, W. A., Rezazade Mehrizi, M. H., Huysman, M., & Feldberg, F. (2017). Debating big data: A literature review on realizing value from big data. *The Journal of Strategic Information Systems*, 26(3), 191–209. <https://doi.org/10.1016/j.jsis.2017.07.003>
- Harari, Y. (2017). *Homo Deus - A Brief History of Tomorrow*. Vintage Publishing. https://www.bookdepository.com/Homo-Deus-Yuval-Noah-Harari/9781784703936?ref=pd_detail_1_sims_b_p2p_1
- Harari, Y. (2019). *21 Lessons for the 21st Century*. Vintage Publishing. <https://www.bookdepository.com/21-Lessons-for-the-21st-Century/9781784708283>
- Harzing, A.-W., & Alakangas, S. (2016). Google Scholar, Scopus and the Web of Science: a longitudinal and cross-disciplinary comparison. *Scientometrics*, 106(2), 787–804. <https://doi.org/10.1007/s11192-015-1798-9>
- Hatch, M. J., & Cunliffe, A. L. (2013). *Organization Theory - Modern, Symbolic and Postmodern Perspectives*. Oxford University Press.
- Hiebl, M. R. W. (2021). Sample Selection in Systematic Literature Reviews of Management Research. *Organizational Research Methods*, 109442812098685. <https://doi.org/10.1177/1094428120986851>
- Ho, C. K. Y., Ke, W., Liu, H., & Chau, P. Y. K. (2020). Separate versus joint evaluation: The roles of evaluation mode and construal level in technology adoption. *MIS Quarterly: Management Information Systems*, 44(2), 725–746. <https://doi.org/10.25300/MISQ/2020/14246>
- Iacovou, C. L., Benbasat, I., & Dexter, A. S. (1995). Electronic data interchange and small organizations: Adoption and impact of technology. *MIS Quarterly: Management Information Systems*, 19(4), 465–485. <https://doi.org/10.2307/249629>
- Instituto Brasileiro de Geografia e Estatísticas. (2020a). *Censo Agropecuário 2017*. Sobre. <https://censoagro2017.ibge.gov.br/sobre-censo-agro-2017.html>
- Instituto Brasileiro de Geografia e Estatísticas. (2020b). *Postponing the Brazilian Census form 2020 to 2021*. Highlights. <https://www.ibge.ge.gov.br/en/highlights/27164-censo-2020-adiado-para-2022.html>

Instituto Brasileiro de Geografia e Estatísticas. (2021). *Coletânea Legislação*. Memória. https://memoria.ibge.gov.br/images/pdf/memoria/coletanea_legislacao_ibge.pdf

Instituto Brasileiro de Geografia e Estatísticas. (2022a). *Brazilian Census postponement from 2021 to 2022*. Highlights. <https://www.ibge.gov.br/en/highlights/30572-postponement-of-the-population-census.html>

Instituto Brasileiro de Geografia e Estatísticas. (2022b). *Recenseamento do Brasil em 1872* (G. Leuzinge (ed.)). https://biblioteca.ibge.gov.br/visualizacao/livros/liv25477_v1_br.pdf

Instituto Brasileiro de Geografia e Estatísticas. (2023a). *Linha do tempo*. Memória IBGE. <https://memoria.ibge.gov.br/linha-do-tempo.html>

Instituto Brasileiro de Geografia e Estatísticas. (2023b). *O que é*. População. <https://www.ibge.gov.br/estatisticas/sociais/populacao/31876-dimensionamento-emergencial-de-populacao-residente-em-areas-indigenas-e-quilombolas-para-acoes-de-enfrentamento-a-pandemia-provocada-pelo-coronavirus.html?edicao=31877&t=o-que-e>

Instituto Brasileiro de Geografia e Estatísticas. (2023c). *Recenseamentos Gerais e Estatísticas Populacionais no Brasil*. Memória IBGE. <https://memoria.ibge.gov.br/historia-do-ibge/historico-dos-censos/censos-demograficos.html>

Instituto Brasileiro de Geografia e Estatísticas. (2023d). *Regional Hub of Big Data in Brazil*. UN Big Data Regional Hub in Brazil. <https://hub.ibge.gov.br/>

Instituto Nacional de Estatística (2011). *E-Censos*. Censos 2011. https://censos.ine.pt/xportal/xmain?xpid=CENSOS&xpgid=censos2011_e-censos

Instituto Nacional de Estatística (2022). *Mensagem do Presidente do INE*. Censos 2021. https://censos.ine.pt/xportal/xmain?xpgid=censos21_menpresidente&xpid=CENSOS21&xlang=pt

Instituto Nacional de Estatística (2023). *Presentation*. Statistic Council. https://www.ine.pt/xportal/xmain?xpid=CSE&xpgid=cse_main&cont_cse=51354&cse_smenu.boui=3109257&xlang=en

INE. (2021a). *Edital para Recrutamento Recenseadores*. Instituto Nacional de Estatísticas. https://www.ine.pt/xportal/xmain?xpid=INE&xpgid=ine_rh_reccensos&contexto=rc&selTab=tab4&xlang=pt

Instituto Nacional de Estatísticas (2021b). *História dos Censos em Portugal*. Censos Em Portugal. https://censos.ine.pt/xportal/xmain?xpid=INE&xpgid=censos_historia_portugal

Instituto Nacional de Estatísticas (2021c). *Recenseamento Agrícola - Análise dos principais*

- resultados - 2019* (Instituto Nacional de Estatísticas (ed.)). Instituto Nacional de Estatísticas.
<https://censoagro2017.ibge.gov.br/sobre-censo-agro-2017.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
- Jackson, K., & Bazeley, P. (2019). *Qualitative Data Analysis with NVIVO* (First). SAGE Publications Ltd.
- Jung, J. H., Bapna, R., Ramaprasad, J., & Umyarov, A. (2019). Love unshackled: Identifying the effect of mobile app adoption in online dating. *MIS Quarterly: Management Information Systems*, 43(1), 47–72. <https://doi.org/10.25300/MISQ/2019/14289>
- Kaplan, A. (1964). *The conducto fo inquiry: Methodology for behavioral science*. Chandler.
- Karahanna, E., Straub, D. W., & Chervany, N. L. (1999). Information technology adoption across time: A cross-sectional comparison of pre-adoption and post-adoption beliefs. *MIS Quarterly: Management Information Systems*, 23(2), 183–213. <https://doi.org/10.2307/249751>
- Kim, S. S., & Son, J.-Y. (2017). Out of Dedication or Constraint? A Dual Model of Post-Adoption Phenomena and its Empirical Test in the Context of Online Services. *MIS Quarterly*, 110(9), 1689–1699.
- Köhler, T. (2016). From the Editors: On Writing Up Qualitative Research in Management Learning and Education. *Academy of Management Learning & Education*, 15(3), 400–418. <https://doi.org/10.5465/amle.2016.0275>
- Kohli, R., & Tan, S. S. L. (2016). Electronic health records: How can is researchers contribute to transforming healthcare? *MIS Quarterly: Management Information Systems*, 40(3), 553–573. <https://doi.org/10.25300/MISQ/2016/40.3.02>
- Komiak, S. Y. X., & Benbasat, I. (2006). The effects of personalization and familiarity on trust and adoption of recommendation agents. *MIS Quarterly: Management Information Systems*, 30(4), 941–960. <https://doi.org/10.2307/25148760>
- Kothari, C. R. (2004). *Research Methodology: Methods and Techniques* (Second). New Age International Publishers.
https://www.researchgate.net/publication/269107473_What_is_governance/link/548173090cf22525dcb61443/download%0Ahttp://www.econ.upf.edu/~reynal/Civilwars_12December2010.pdf%0Ahttps://think-asia.org/handle/11540/8282%0Ahttps://www.jstor.org/stable/41857625
- 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.
- Kunisch, S., Menz, M., Bartunek, J. M., Cardinal, L. B., & Denyer, D. (2018). Feature topic at organizational research methods: How to conduct rigorous and impactful literature reviews? *Organizational Research Methods*, 21(3), 519–523. <https://doi.org/10.1177/1094428118770750>
- Kvale, S. (1996). *InterViews - An Introduction to Qualitative Research Interviewing*. SAGE Publications, Inc.
- 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*, 34(3), 387–394. <https://doi.org/10.1016/j.ijinfomgt.2014.02.002>
- Lai, Y., Sun, H., & Ren, J. (2018a). Understanding the determinants of big data analytics (BDA) adoption in logistics and supply chain management: An empirical investigation. *International Journal of Logistics Management*, 29(2), 676–703. <https://doi.org/10.1108/IJLM-06-2017-0153>
- Lai, Y., Sun, H., & Ren, J. (2018b). Understanding the determinants of big data analytics (BDA) adoption in logistics and supply chain management. *The International Journal of Logistics Management*, 29(2), 676–703. <https://doi.org/10.1108/IJLM-06-2017-0153>
- Langley, A., Smallman, C., Tsoukas, H., & Van de Ven, A. H. (2013). Process Studies of Change in Organization and Management: Unveiling Temporality, Activity, and Flow. *Academy of Management Journal*, 56(1), 1–13. <https://doi.org/10.5465/amj.2013.4001>
- Lundblad, J. (2003). A Review and Critique of Rogers' Diffusion of Innovation Theory as It Applies to Organizations. *Organization Development Journal*, 21(4), 50–64.
- Maroufkhani, P., Iranmanesh, M., & Ghobakhloo, M. (2022). Determinants of big data analytics adoption in small and medium-sized enterprises (SMEs). *Industrial Management & Data Systems*. 123(1). 278-301. <https://doi.org/10.1108/IMDS-11-2021-0695>
- Maroufkhani, P., Wan Ismail, W. K., & Ghobakhloo, M. (2020). Big data analytics adoption model for small and medium enterprises. *Journal of Science and Technology Policy Management*, 11(4), 483–513. <https://doi.org/10.1108/JSTPM-02-2020-0018>
- Martinez-Mosquera, D., & Luján-Mora, S. (2019). Framework for big data integration in e-government. *DYNA*, 86(209), 215–224. <https://doi.org/10.15446/dyna.v86n209.77902>
- McGuire, A. (2016). Researching Big Data Skills Using a Mixed-Methods Approach. *SAGE Research Methods Cases*, 1-16. <https://doi.org/10.4135/9781526419620>

- McMahon, P., Zhang, T., & Dwight, R. (2020). Requirements for Big Data Adoption for Railway Asset Management. *IEEE Access*, 8, 15543–15564. <https://doi.org/10.1109/ACCESS.2020.2967436>
- Memon, S., Changfeng, W., Rasheed, S., Pathan, Z. H., Saddozai, S. K., Yixin, Q., & Yanping, L. (2016). Adoption of Big Data Technologies for Communication Management in Large Projects. *International Journal of Future Generation Communication and Networking*, 9(10), 73–82. <https://doi.org/10.14257/ijfgcn.2016.9.10.07>
- Ministério do Planejamento. (2023). *Unidades do Ministério do Planejamento*. Institutional. <https://www.gov.br/economia/pt-br/aceso-a-informacao/institucional/planejamento>
- Miranda, M. Q., Farias, J. S., Schwartz, C. D. A., Pascualote, J., & Almeida, L. De. (2016). Technology adoption in diffusion of innovations perspective: introduction of an ERP system in a non-profit organization. *RAI Revista de Administração e Inovação*, 13(1), 48–57. <https://doi.org/10.1016/j.rai.2016.02.002>
- Mital, M., Chang, V., Choudhary, P., Papa, A., & Pani, A. K. (2018). Adoption of Internet of Things in India: A test of competing models using a structured equation modeling approach. *Technological Forecasting and Social Change*, 136, 339–346. <https://doi.org/10.1016/j.techfore.2017.03.001>
- Morakanyane, R., Grace, A., & O'Reilly, P. (2017). Conceptualizing Digital Transformation in Business Organizations: A Systematic Review of Literature. *Bled Proceedings*, 427–443. <https://doi.org/10.18690/978-961-286-043-1.30>
- Moss, L. (2023). Census data reveals LGBT+ populations for first time. *BBC News*. <https://www.bbc.com/news/uk-64184736>
- Moura, I., Dominguez, C., & Varajão, J. (2021). Information systems project team members: factors for high performance. *The TQM Journal*, 33(6), 1426–1446. <https://doi.org/10.1108/TQM-07-2020-0170>
- Müller, O., Fay, M., & vom Brocke, J. (2018). The Effect of Big Data and Analytics on Firm Performance: An Econometric Analysis Considering Industry Characteristics. *Journal of Management Information Systems*, 35(2), 488–509. <https://doi.org/10.1080/07421222.2018.1451955>
- Noor, K. B. M. (2008). Case Study: A Strategic Research Methodology. *American Journal of Applied Sciences*, 5(11), 1602–1604.
- O'Neil, C. (2016). *Weapons of Math Destruction*. Penguin Books.

- Office for National Statistics. (2022a). *Data and analysis from Census 2021*. Legislation and Policy.
- Office for National Statistics. (2022b). *Story of Census*. Office for National Statistics.
<https://www.ons.gov.uk/visualisation/storyofthecensus>
- Office for National Statistics. (2023a). *About Us*. What We Do. <https://www.ons.gov.uk/aboutus>
- Office for National Statistics. (2023b). *Home*. Organisations.
<https://www.gov.uk/government/organisations/office-for-national-statistics>
- Office for National Statistics. (2023c). *Release Plans*. Release Plans.
https://www.who.int/emergencies/diseases/novel-coronavirus-2019/media-resources/science-in-5/episode-63—omicron-variant?gclid=Cj0KCQiAgaGgBhC8ARIsAAAYLfHI9hHol1v4EQWHLKaKB8d1p36GoxgZGbeHwzS4IH_2Bozu5KualtYaAlBjEALw_wcB
- Olszak, C. M., & Zurada, J. (2020). Big Data in Capturing Business Value. *Information Systems Management*, 37(3), 240–254. <https://doi.org/10.1080/10580530.2020.1696551>
- Orlikowski, W. J. (2000). Using Technology and Constituting Structures: A Practice Lens for Studying Technology in Organizations. *Organization Science*, 11(4), 404–428.
<https://doi.org/10.1287/orsc.11.4.404.14600>
- Orwell, G. (2013). *1984*. Penguin Books. https://www.bookdepository.com/Nineteen-Eighty-Four-George-Orwell/9780141393049?redirected=true&utm_medium=Google&utm_campaign=Base1&utm_source=PT&utm_content=Nineteen-Eighty-Four&selectCurrency=EUR&w=AF7DAU99ZT93JOA8YC13
- Paradza, D., & Daramola, O. (2021). Business intelligence and business value in organisations: A systematic literature review. *Sustainability*, 13(20), 3–27.
<https://doi.org/10.3390/su132011382>
- Park, J. H., & Kim, Y. B. (2021). Factors Activating Big Data Adoption by Korean Firms. *Journal of Computer Information Systems*, 61(3), 285–293.
<https://doi.org/10.1080/08874417.2019.1631133>
- Paul, J., Modi, A., & Patel, J. (2016). Predicting green product consumption using theory of planned behavior and reasoned action. *Journal of Retailing and Consumer Services*, 29, 123–134.
<https://doi.org/10.1016/j.jretconser.2015.11.006>
- Pavlou, P., & Fygenson, M. (2006). Understanding and Predicting Electronic Commerce Adoption:

- An Extension of the Theory of Planned Behavior. *MIS Quarterly*, 30(1), 115–143.
- Penna, G. O., Silva, J. A. A. da, Neto, J. C., Temporão, J. G., & Pinto, L. F. (2020). PNAD COVID-19: um novo e poderoso instrumento para Vigilância em Saúde no Brasil. *Ciência & Saúde Coletiva*, 25(9), 3567–3571. <https://doi.org/10.1590/1413-81232020259.24002020>
- Petter, S., Delone, W., & McLean, E. R. (2013). Information systems success: The quest for the independent variables. *Journal of Management Information Systems*, 29(4), 7–62. <https://doi.org/10.2753/MIS0742-1222290401>
- Phillips-Wren, G., Iyer, L. S., Kulkarni, U., & Ariyachandra, T. (2015). Business Analytics in the Context of Big Data: A Roadmap for Research. *Communications of the Association for Information Systems*, 37, 448–472. <https://doi.org/10.17705/1CAIS.03723>
- Picoto, W. N., Crespo, N. F., & Carvalho, F. K. (2021). The influence of the technology-organization-environment framework and strategic orientation on cloud computing use, enterprise mobility, and performance. *Review of Business Management*, 23(2), 278–300. <https://doi.org/10.7819/rbgn.v23i2.4105>
- Pillai, R., & Sivathanu, B. (2020). Adoption of internet of things (IoT) in the agriculture industry deploying the BRT framework. *Benchmarking*, 27(4), 1341–1368. <https://doi.org/10.1108/BIJ-08-2019-0361>
- Podsakoff, P. M., MacKenzie, S. B., & Podsakoff, N. P. (2016). Recommendations for Creating Better Concept Definitions in the Organizational, Behavioral, and Social Sciences. *Organizational Research Methods*, 19(2), 159–203. <https://doi.org/10.1177/1094428115624965>
- Pramana, S., Mariyah, S., & Takdir. (2021). Big data implementation for price statistics in Indonesia: Past, current, and future developments. *Statistical Journal of the IAOS*, 37(1), 415–427. <https://doi.org/10.3233/SJI-200740>
- Pullinger, J. (1997). The Creation of the Office for National Statistics. *International Statistical Review*, 65(3), 291–308. <https://doi.org/10.1111/j.1751-5823.1997.tb00310.x>
- Queiroz, M. M., & Farias Pereira, S. C. (2019). Intention to adopt big data in supply chain management: A Brazilian perspective. *Revista de Administração de Empresas*, 59(6), 389–401. <https://doi.org/10.1590/S0034-759020190605>
- Quetelet, A. (2013). *A Treatise on Man and the development of this faculties*. 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>
- Ram, J., Afridi, N. K., & Khan, K. A. (2019). Adoption of Big Data analytics in construction: development of a conceptual model. *Built Environment Project and Asset Management*, 9(4), 564–579. <https://doi.org/10.1108/BEPAM-05-2018-0077>
- Ramadoss, R., & Elango, N. M. (2015). Proactive exploratory testing methodology during enterprise application modernization. *International Journal of Engineering and Technology*, 7(2), 673–681.
- Ravin, Y., & Leacock, C. (2000). Polysemy: an overview. In Y. Ravin & C. Leacock (Eds.), *Polysemy: Theoretical and computational approaches* (pp. 1-29.). Oxford Press Inc.
- República Federativa do Brasil. (2018). *Lei Geral de Proteção de Dados Pessoais (LGPD)*. http://www.planalto.gov.br/ccivil_03/_Ato2015-2018/2018/Lei/L13709.htm
- Retana, G. F., Forman, C., Narasimhan, S., Niculescu, M. F., & Wu, D. J. (2018). Technology support and post-adoption IT service use: Evidence from the cloud. *MIS Quarterly: Management Information Systems*, 42(3), 961–978. <https://doi.org/10.25300/MISQ/2018/13064>
- Rogers, E. M. (1962). *Diffusion of innovations* (1st ed.). Free Press of Glencoe.
- Rogers, E. M. (1983). *Diffusion of innovations* (3rd ed.). Free Press of Glencoe.
- Rogers, E. M. (1995). *Diffusion of innovations* (4th ed.). Free Press.
- Rogers, E. M. (2004). A prospective and retrospective look at the diffusion model. *Journal of Health Communication*, 9, 13–19. <https://doi.org/10.1080/10810730490271449>
- Royal Statistical Society. (2014). *The Data Manifesto*. <https://rss.org.uk>.
- Royal Statistical Society. (2022). *History*. About. <https://rss.org.uk/about/history/>
- Saheb, T. (2020). An empirical investigation of the adoption of mobile health applications: integrating big data and social media services. *Health and Technology*, 10(5), 1063–1077. <https://doi.org/10.1007/s12553-020-00422-9>
- Saldaña, J. (2013). The Coding Manual for Qualitative Research. In *The Coding Manual For Qualitative Researchers* (Sepond). SAGE Publications Ltd.
- Salleh, K. A., & Janczewski, L. (2016). Technological, Organizational and Environmental Security and Privacy Issues of Big Data: A Literature Review. *Procedia Computer Science*, 100, 19–28. <https://doi.org/10.1016/j.procs.2016.09.119>

- Salleh, K. A., Lech, J., & Beltran, F. (2015). SEC-TOE Framework : Exploring Security Determinants in Big Data Solutions Adoption. *Pacific Asia Conference on Information Systems*, 1-11. <https://aisel.aisnet.org/pacis2015/203/>
- Santos, C., Santos, V., Tavares, A., & Varajão, J. (2020). Project Management in Public Health : A Systematic Literature Review on Success Criteria and Factors. *Portugues Journal of Public Health*. <https://doi.org/10.1159/000509531>
- Sarker, S., Sarker, S., & Sidorova, A. (2006). Understanding Business Process Change Failure: An Actor-Network Perspective. *Journal of Management Information Systems*, 23(1), 51–86. <https://doi.org/10.2753/MIS0742-1222230102>
- Sarker, S., & Valacich, J. (2010). An Alternative to Methodological Individualism: A Non-Reductionist Approach to Studying Technology Adoption by Groups. *MIS Quarterly*, 34(4), 779–808.
- Saunders, M., Lewis, P., & Thornhill, A. (2009). *Reserach Method for Business Students* (5th).
- Schiavoni, C., Palm, F., Smeekes, S., & van den Brakel, J. (2021). A dynamic factor model approach to incorporate Big Data in state space models for official statistics. *Journal of the Royal Statistical Society. Series A: Statistics in Society*, 184(1), 324–353. <https://doi.org/10.1111/rssa.12626>
- Schwab, K. (2016). *The Fourth Industrial Revolution* (1st ed.). Crown Business.
- Scott, S. V., & Wagner, E. L. (2003). Networks, negotiations, and new times: the implementation of enterprise resource planning into an academic administration. *Information and Organization*, 13(4), 285–313. [https://doi.org/10.1016/S1471-7727\(03\)00012-5](https://doi.org/10.1016/S1471-7727(03)00012-5)
- Shahbaz, M., Gao, C., Zhai, L., Shahzad, F., & Hu, Y. (2019). Investigating the adoption of big data analytics in healthcare: the moderating role of resistance to change. *Journal of Big Data*, 6(1), 6. <https://doi.org/10.1186/s40537-019-0170-y>
- Shahbaz, M., Gao, C., Zhai, L., Shahzad, F., & Arshad, M. R. (2020). Moderating Effects of Gender and Resistance to Change on the Adoption of Big Data Analytics in Healthcare. *Complexity*, V. 2020, 1–13. <https://doi.org/10.1155/2020/2173765>
- Sheng, J., Amankwah-Amoah, J., & Wang, X. (2017). A multidisciplinary perspective of big data in management research. *International Journal of Production Economics*, 191(November 2016), 97–112. <https://doi.org/10.1016/j.ijpe.2017.06.006>

- Silva, J., Hernández-Fernández, L., Torres Cuadrado, E., Mercado-Caruso, N., Rengifo Espinosa, C., Acosta Ortega, F., Hernández P, H., & Jiménez Delgado, G. (2019). Factors affecting the big data adoption as a marketing tool in SMEs. *Communications in Computer and Information Science*, 1071, 34–43. https://doi.org/10.1007/978-981-32-9563-6_4
- Silveira, D. (2023, January 25). IBGE prevê para abril divulgação dos resultados definitivos do Censo 2022. *Portal G1, Economia*. <https://g1.globo.com/economia/noticia/2023/01/25/ibge-preve-para-abril-divulgacao-dos-resultados-definitivos-do-censo-2022.ghtml>
- Sinha, A., Kumar, P., Rana, N. P., Islam, R., & Dwivedi, Y. K. (2019). Impact of internet of things (IoT) in disaster management: a task-technology fit perspective. *Annals of Operations Research*, 283(1–2), 759–794. <https://doi.org/10.1007/s10479-017-2658-1>
- Soto-Acosta, P. (2020). COVID-19 Pandemic: Shifting Digital Transformation to a High-Speed Gear. *Information Systems Management*, 37(4), 260–266. <https://doi.org/10.1080/10580530.2020.1814461>
- Serviço Regional de Estatística dos Açores. (2023). *O que somos O que fazemos Como fazemos*. About Us. https://srea.azores.gov.pt/Conteudos/Relatorios/lista_relatorios.aspx?idc=307&idsc=6209&lang_id=1
- Stake, R. (1995). *The Art of Case Study Research*. Sage Publications, Inc.
- Statistics Canada. (2020). *The Integration of Web-Scraped Data into the Clothing and Footwear Component of the Consumer Price Index* (Issue 62).
- 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. (2018a). 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>
- Sun, S., Hall, D. J., & Cegielski, C. G. (2020). Organizational intention to adopt big data in the B2B context: An integrated view. *Industrial Marketing Management*, 86(September 2019), 109–121. <https://doi.org/10.1016/j.indmarman.2019.09.003>
- Sun, Z., & Huo, Y. (2021). The Spectrum of Big Data Analytics. *Journal of Computer Information Systems*, 61(2), 154–162. <https://doi.org/10.1080/08874417.2019.1571456>
- Taylor, S., & Todd, P. A. (1995). Understanding information technology usage: A Test of Competing Models. In *Information Systems Research*, 6 (2), 144–176.

- <https://www.jstor.org/stable/23011007>
- 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*. E/CN.3/2020/24. New York.
- Thompson, R. L., Higgins, C. A., & Howell, J. M. (1991). Personal Computing: Toward a Conceptual Model of Utilization. *MIS Quarterly*, 15(1), 125. <https://doi.org/10.2307/249443>
- Trigo, A., Varajão, J., Barroso, J., Soto-Acosta, P., Molina-Castillo, F. J., & Gonzalez-Gallego, N. (2013). Enterprise information systems adoption in Iberian large companies: Motivations and trends. In *Enterprise Resource Planning: Concepts, Methodologies, Tools, and Applications*. 2 (3), 1035–1052. <https://doi.org/10.4018/978-1-4666-4153-2.ch060>
- Trigo, A., Varajão, J., & Sousa, L. (2022). DevOps adoption: Insights from a large European Telco. *Cogent Engineering*, 9(1), 1–31. <https://doi.org/10.1080/23311916.2022.2083474>
- Tushman, M. L., & Anderson, P. (1986). Technological Discontinuities and Organizational Environments. *Administrative Science Quarterly*, 31(3), 439–465. <https://doi.org/10.2307/2392832>
- Tushman, M., & Nadler, D. (1986). Organizing for Innovation. *California Management Review*, 28(3), 74–92. <https://doi.org/10.2307/41165203>
- 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/>
- UK Statistics Authority. (2023). *About the Authority*. About the Authority. <https://uksa.statisticsauthority.gov.uk/about-the-authority/>
- UNESCO. (2023). *What is UNESCO?* Brief. <https://www.unesco.org/en/brief>
- United Nations. (2014). Fundamental Principles of Official Statistics. In *Sixty-eighth session of General Assembly*, 9 (January).
- United Nations. (2021). *UN Data*. <http://data.un.org/Explorer.aspx>
- United Nations. (2022). *Home*. SDG Indicators. <https://unstats.un.org/sdgs/>
- United Nations. (2023). *Statistics Division*. Department Of Economic and Social Affairs. <https://unstats.un.org/UNSDWebsite/>
- United Nations Statistics Division. (2019). *About us*. Home. <https://unstats.un.org/UNSDWebsite/about/>
- United Nations Statistics Division. (2023a). *Catalog*. Big Data Training Catalog.

- <https://unstats.un.org/bigdata/task-teams/training/catalog/Catalog>
- United Nations Statistics Division. (2023b). *Census Documents*. World Population and Housing Census Programme. <https://unstats.un.org/unsd/demographic-social/census/document-resources/>
- United Nations Statistics Division. (2023c). *Impact of COVID-19*. World Population and Housing Census Programme. <https://unstats.un.org/unsd/demographic-social/census/COVID-19/>
- United Nations Statistics Division. (2023d). *Millenium Development Goals Indicators*. Millenium Development Goals. <https://unstats.un.org/unsd/mdg/Default.aspx>
- United Nations Statistics Division. (2023e). *UN Regional Hubs for Big Data*. UN Big Data. <https://unstats.un.org/bigdata/regional-hubs.cshtml>
- Urbinati, A., Bogers, M., Chiesa, V., & Frattini, F. (2019). Creating and capturing value from Big Data: A multiple-case study analysis of provider companies. *Technovation*, 84–85(July 2018), 21–36. <https://doi.org/10.1016/j.technovation.2018.07.004>
- Valério, N., & Bastien, C. (2010). *O INE: Desafios do Passado, Desafios do Futuro*. ISBN 978-989-25-0076-8. INE, I.P., Lisboa.
- Van Campenhoudt, L., Marquet, J., & Quivy, R. (2017). *Manuel de recherche en sciences sociales* (5ed ed.). Dunod.
- Van den Brakel, J., & Kreig, S. (2016). Small area estimation with state space common factor models for rotating panels. *Journal of the Royal Statistical Society Series A: Statistics in Society*, 179(3), 763–791.
- Van Den Brakel, J., Söhler, E., Daas, P., & Buelens, B. (2017). Social media as a data source for official statistics; the dutch consumer confidence index. *Survey Methodology*, 43(2), 183–210.
- Varajão, J., Trigo, A., & Silva, T. (2021). A Web Tool for Information Systems Researchers. *ISRI - Information Systems Research Constructs and Indicators*, 9(1), 54–67. <https://doi.org/https://doi.org/10.1633/JISTaP.2021.9.1.4>
- Venkatesh, V., Morris, M. G., Davis, G. B., & Davis, F. D. (2003). *Quarterly*. 27(3), 425–478.
- Verma, S., & Bhattacharyya, S. S. (2017). Perceived strategic value-based adoption of Big Data Analytics in emerging economy: A qualitative approach for Indian firms. *Journal of Enterprise Information Management*, 30(3), 354–382. <https://doi.org/10.1108/JEIM-10-2015-0099>
- Vial, G. (2019). Understanding digital transformation: A review and a research agenda. *The Journal of Strategic Information Systems*, 28(2), 118–144.

- <https://doi.org/10.1016/j.jsis.2019.01.003>
- Wang, Y. S., & Liao, Y. W. (2008). Assessing eGovernment systems success: A validation of the DeLone and McLean model of information systems success. *Government Information Quarterly*, 25(4), 717–733. <https://doi.org/10.1016/j.giq.2007.06.002>
- Webster, J., & Watson, R. T. (2002). Analyzing the past to prepare for the future : Writing a literature review. *MIS Quarterly*, 26(2), xiii–xxiii. <https://doi.org/10.2307/4132319>
- Welch, C., Rumyantseva, M., & Hewerdine, L. J. (2016). Using Case Research to Reconstruct Concepts. *Organizational Research Methods*, 19(1), 111–130. <https://doi.org/10.1177/1094428115596435>
- Wong, L. W., Tan, G. W. H., Lee, V. H., Ooi, K. B., & Sohal, A. (2020). Unearthing the determinants of Blockchain adoption in supply chain management. *International Journal of Production Research*, 58(7), 2100–2123. <https://doi.org/10.1080/00207543.2020.1730463>
- World Health Organization. (2022). *Episode #63 - Omicron varian*. Home. https://www.who.int/emergencies/diseases/novel-coronavirus-2019/media-resources/science-in-5/episode-63—omicron-variant?gclid=Cj0KCQiAgaGgBhC8ARIsAAAYLfHI9hHol1v4EQWHLKaKB8d1p36GoxgZGbeHwzS4IH_2Bozu5KualtYaAlBjEALw_wcB
- Wunderlich, P., & Veit, D. J. (2019). Adoption of Sustainable Technologies: A Mixed-Methods Study of German Households. *MIS Quarterly*, 43(2), 1–20.
- Yadegaridehkordi, E., Hourmand, M., Nilashi, M., Shuib, L., Ahani, A., & Ibrahim, O. (2018). Influence of big data adoption on manufacturing companies' performance: An integrated DEMATEL-ANFIS approach. *Technological Forecasting and Social Change*, 137(July), 199–210. <https://doi.org/10.1016/j.techfore.2018.07.043>
- Yin, R. K. (2014). *Case Study Research - Design and Methods*. 5th. Sage Publications.
- Ylijoki, O., & Porras, J. (2016). Conceptualizing Big Data: Analysis of Case Studies. *Intelligent Systems in Accounting, Finance and Management*, 23(4), 295–310. <https://doi.org/10.1002/isaf.1393>
- Ylijoki, O., & Porras, J. (2018). What managers think about big data. *International Journal of Business Information Systems*, 29(4), 485–501. <https://doi.org/10.1504/IJBIS.2018.096034>
- Zhang, L.-J. (2012). Editorial: Big Services Era: Global Trends of Cloud Computing and Big Data. *IEEE Transactions on Services Computing*, 5(4), 467–468.

<https://doi.org/10.1109/TSC.2012.36>

Appendix 1 - Interview script to NSO members

Introduction

Firstly, thank you for your time. I inform you this interview is part of my PhD research process. It is an investigation anchored at the University of Minho in Portugal, and Presbyterian University Mackenzie, in Brazil.

Secondly, the National Statistics Offices' adoption of Big Data technologies is the focus, and your perspective concerned about Office for National Statistics is fundamental to this research.

Additionally, I highlight that interview privacy is absolutely ensured. In this context, the content's confidentiality are guided by those three legal registers: the British Data Protection Act 2018, the European Data Protection Law, and the Brazilian Data Protection Law.

Thus, I ask your permission to record our interview. The registers will be transcribed in an anonymized format, and the recorders will be destroyed after the PhD research conclusion.

"May I record your interview?"

"Thank you! I am starting the interview record."

<Start recording>

Block 1 – Profile Data

What is your name?–

Briefly, tell me about your background?

How many years do you working time at [NSO NAME]?

What is your current job position and how many you work in this position?

Block 2 – Professional's journey

1. Please, tell me briefly about your professional journey in [NSO_NAME].
2. Which were the main technological changes you witnessed or participated in this professional journey, including in the last five years?
3. What were the generated benefits to [NSO_NAME] after those changes?

Block 3 – Big Data technologies

4. What is the understanding of the Big Data concept to you and your colleagues at [NSO_NAME]?
5. Have you been working with Big Data tech/applications? If so, which?
6. Which of those techs/applications do you consider more relevant to the official statistics?
7. How have those applications been adopted, deployed and used in [NSO_NAME] processes and products?
8. Strictly in your area/department, how is Big Data Technology adopted and used?
9. What are other areas/departments potentially adopting Big Data technologies? Why them?
10. Are these Big Data technologies developed in-house by the [NSO_NAME] staff or bought from IT suppliers, such as Big Techs? Or there are hybrid solutions?
11. In your opinion, which are other technological factors to support or constrain Big Data technologies in the [NSO_NAME]?

Block 4 – Organisational factors

12. Is there an established and formalised process to spread technology adoptions into the [NSO_NAME] departments? Or is the current procedure an informal one?
13. Is it a centralised or a decentralised decision-making process that defines the technology Adoption in [NSO_NAME]?
14. How would you classify the [NSO_NAME] organisational culture relative to technological adoption in a range from “tightly closed” to “highly opened”? Are there silos effects?
15. Can financial costs be considered central factors in determining technology adoption, or the budget is not an onerous restriction?
16. Are there other organisational factors to support or constrain Big Data technologies in the [NSO_NAME]?

Block 5 – Environmental factors

17. How do the Global Statistical System, and other international organizations influence the Big Data adoption into [NSO_NAME]? Why/how?
18. Considering external influences, is the Data Protection [LOCAL LAW NAME] a limitation to the spread of Big Data use in Official Statistics?

19. [OPENED QUATION 1]

20. [OPENED QUATION 2]

21. Are there other external factors to support or constrain Big Data technologies in the [NSO_NAME] ?

[Closing sentence: "I will stop the record now. Thank you so much for your time and answers."]