

Rim, Maria J.; Kwon, Youngsun

Conference Paper

Collecting, generating and analyzing national statistics with AI: what benefits and costs?

32nd European Conference of the International Telecommunications Society (ITS):
"Realising the digital decade in the European Union – Easier said than done?", Madrid,
Spain, 19th - 20th June 2023

Provided in Cooperation with:

International Telecommunications Society (ITS)

Suggested Citation: Rim, Maria J.; Kwon, Youngsun (2023) : Collecting, generating and analyzing national statistics with AI: what benefits and costs?, 32nd European Conference of the International Telecommunications Society (ITS): "Realising the digital decade in the European Union – Easier said than done?", Madrid, Spain, 19th - 20th June 2023, International Telecommunications Society (ITS), Calgary

This Version is available at:

<https://hdl.handle.net/10419/278015>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

Collecting, generating and analyzing national statistics with AI: what benefits and costs?

Maria J. Rim*

Korea Advanced Institute of Science and Technology, majorim@kaist.ac.kr

Youngsun Kwon

Korea Advanced Institute of Science and Technology, yokwon@kaist.ac.kr

ABSTRACT

The paper addresses the increasing adoption of digital transformation in public sector organizations, mainly focusing on its impact on national statistical offices. The emergence of data-driven strategies powered by artificial intelligence (AI) disrupts the conventional labour-intensive approaches of NSOs. This necessitates a delicate balance between real-time information and statistical accuracy, leading to exploring AI applications such as machine learning in data processing. Despite its potential benefits, the cooperation between AI and human resources requires in-depth examination to leverage their combined strengths effectively. The paper proposes an integrative review and multi-case study approach to comprehensively contribute to a deeper understanding of the benefits and costs of AI adoption in national statistical processes, facilitate the acceleration of digital transformation, and provide valuable insights for policymakers and practitioners in optimizing the use of AI in collecting, generating and analyzing national statistics.

Keywords: Digital transformation, national statistics, artificial intelligence, human resources, data-driven strategy

1 INTRODUCTION

Processing raw data into information is at the core of digital transformation strategies (Klimczak & Fryczak, 2022). The amount of data generated by the public sector, businesses and individuals is growing at a rate that could not be processed without technological advances (Yung, Wesley et al., 2018). Moreover, the necessity of data-driving strategies at the organizations' core has imposed the adoption of machine-intensive and automation strategies powered by artificial intelligence (AI) technologies (Chu & Poirier, 2015; UNECE, 2021; Yung, Wesley et al., 2018). The advances in AI tools and the more affordable cost of technologies enabling AI were expected to have more impact on the widely used AI in organizations. Prior research shows that organizations' efforts are falling short of being engaged in core practices supporting widespread AI adoption (Fountaine et al., 2019).

Historically, processing data into information as a core practice in an organization with national and international official significance has been the job of statistical institutes in each country (UN, 2022). These institutions oversee capturing, analyzing, and generating official national statistics. Unlike other government agencies, National Statistics

Offices (NSOs) are governed by the Fundamental Principles of Official Statistics (UN Statistical Commission), making them internationally comparable. At the same time, each organization's technological development is governed by each country's context. Understanding the differences and similarities between NSOs worldwide generates a unique perspective for data-driving strategies in digital transformation in the public sector.

On the other hand, data-driving strategies for adopting machine-intensive and automation strategies are particularly relevant for NSOs as their traditional, labour-intensive statistics-generating methods may face disruption from AI technologies necessary for exponential data growth. As NSOs explore the application of machine learning in tasks such as editing, imputation, categorization, and coding to support their conventional processes, they must carefully navigate the trade-off between real-time information and the accuracy and reliability of their existing statistical outputs (Julien, Claude et al., 2020).

Furthermore, contrary to the expectation of reduced labour, more qualified personnel were required, from data preparation to algorithm testing (Chu & Poirier, 2015; Julien, Claude et al., 2020; UNECE, 2021). Therefore, a collaboration between AI and human resources has the potential benefit to improve efficiency and reduce errors in national statistics tasks, but the cost related to training people and data (Chu & Poirier, 2015; Julien, Claude et al., 2020; UNECE, 2021). To address these needs, we explore the theory of digital transformation by AI, the skills required for AI data production strategies, and the effects of AI adoption through a contrast of existing knowledge and an exploratory case study of the adoption of AI in national statistics offices.

We propose using an integrative review method combined with a multi-case study to comprehend the context in which AI has been utilized for statistics offices. The integrative review allows for the analysis of the progression of digital transformation and its immediate implications for practice and policy. This comprehensive approach provides a framework for the optimal transition path from ongoing digitization to a sustainable digital transformation of NSOs. Additionally, the multi-case study allows us to establish a foundation for developing an early-stage theory and constructing measures related to the alternative uses of AI in data-driven strategies, as well as the skills required for human resources to collaborate with AI in a national statistic production. We focused on the machine learning project by the UNECE High-Level Group for the Modernisation of Official Statistics, which aims to gain insights and address the research questions related to the benefits and costs of using AI to collect, generate, and analyze national statistics.

In the following, section 2) provides the theoretical background of digital transformation, AI and data-driven strategies in NSOs by conducting an integrative review; section 3) describes fifteen machine learning cases in NSOs using the theoretical background found in the previous section. Finally, section 4) concludes by synthesizing the findings and propositions that can be tested in future research.

2 DIGITAL TRANSFORMATION AND ARTIFICIAL INTELLIGENCE LITERATURE REVIEW

Considering the increasing indispensability of digital technology in the public sector, this section aims to synthesize factors affecting the benefits and costs of using AI in an organization, and factors that are more likely to transform the business model when adopting AI as a production strategy. To do this, we undertake an extensive literature analysis and carefully synthesize and structure prior research exploring digital transformation and artificial intelligence (AI) as main data-driven strategies.

Given the exponential growth of literature in the field, employing an integrative review approach is considered an effective strategy (Hanna, 2018; Verhoef et al., 2021; Vial, 2019). An essential aspect of the integrative literature review is its ability to comprehend scientific advancements and theoretical developments and provide space for implementing practical and policy-oriented outcomes (Whittemore & Knafl, 2005). Moreover, this sub-category of systematic review

combines systematic techniques with the flexibility of including different types of research, experimental or no experimental, qualitative or quantitative (Torraco, 2005, 2016). Furthermore, it provides a holistic point of view for the research concepts (Torraco, 2005, 2016).

Drawing inspiration from Torraco's (2005, 2016) and Whitemore and Knaf's (2005) approach, this paper will employ the following phases of the integrative review: (1) define the concept of digital transformation and artificial intelligence to provide the scope and the boundaries of the review, (2) search the literature, including the techniques for determining relevant primary sources, (3) select and classify the relevant literature and display the results of this review.

2.1 Background and concepts

Although AI-related tools such as machine learning (ML) are similar for any industry, some factors must be interpreted in context. For example, while accuracy for new technology is vital to an internal user of any AI system in any industry, an external user of national statistics will prefer the ability to explain and time consistency as a parameter for evaluating the performance of a new tool (Julien, Claude et al., 2020). Therefore, the meaning and importance of AI-related concepts and their impact on the business model (digital transformation) must be defined for this paper in the context of National Statistic Offices (NSOs).

2.1.1 National statistics "business", data-driven strategies, and artificial intelligence.

National official statistics are those disseminated by the ensemble of statistical organizations and units within a country (OECD, 2002). Among the organizations involved in producing national statistics, the National Statistic Offices (NSOs) provide the government, citizens, corporations, and organizations (national and international) the relevant data about the economy, demography and society within a country and its international relationships (OECD, 2002). NSO oversees a country's statistical production by managing its information, application, and technologies (UN, 2022). In this regard, the Statistical business process model covers defining needs, processing data, disseminating results, archiving information, and evaluating the process (Karlberg, 2017). Although technological advances affect all these activities, this paper focuses on data-driven strategies related to the NSO's core production phase, the processing data step, which includes collecting, generating and analyzing data into statistical outputs.

While conventional analogue data processing relies on statistical methods such as sampling before data collection, digital data, on the other hand, typically emerges as a noisy by-product of non-statistical processes (Yung, Wesley et al., 2018). However, within this digital data lies human activity signals that hold the potential for extracting official statistics (Yung, Wesley et al., 2018). The advancements in AI-driven by ML techniques, the presence of extensive quantities of digital data and powerful IT infrastructures can potentially transform official statistics (Chu & Poirier, 2015; Yung, Wesley et al., 2018).

Publications related to the adoption of AI in the public sector in general (Desouza et al., 2020; Duan et al., 2019; Janssen et al., 2020; Wirtz et al., 2019) and in NSOs in particular (Chu & Poirier, 2015; Yung, Wesley et al., 2018) could either describe the specific algorithm and its implications or the context of the decision to adopt the new technology and its repercussions. The following sections address the second group of publications and exclude papers that only talk about AI methods or algorithms.

2.1.2 Digital transformation in National Statistics Offices.

According to Kavadias et al. (2016), introducing a new technology is not enough to revolutionize an industry. The real drivers of transformation lie in the intersection between adopting new technology and emerging market demands. In the

case of NSOs, this means identifying the key factors that link AI trends to the needs of stakeholders. While AI has thus far been used in NSOs to support data-driven strategies (i.e. digitalization), true digital transformation requires discovering new roles and creating new value (PARIS21, 2022).

Digitalization generally denotes integrating digital technology as a tool in the operational process. Therefore, AI tools are related to the increased digitalization of the organization and are less dependent on the industry. For example, an image recognition AI tool can diagnose the health industry (Wirtz et al., 2021) and estimate crop yields in statistics-related industries (Chu & Poirier, 2015). Conversely, digital transformation signifies a shift in business operations where the core of the business model revolves around a continuous culture of digital innovation (Saldanha, 2019). Consequently, digital transformation is a more industry-dependent concept than digitalization because business model varies by industry, so its definition is preferable to take from articles related to statistical production:

Digital transformation is to produce action-oriented knowledge for sectoral experts and decision-makers through improved collection, processing, aggregation and contextualization of raw data into statistical information (PARIS21, 2022, p. 29)

Digital transformation aims to...looking holistically at the processes at stake within the institution. A digital transformation is a fundamental change in the way statistics are produced (Paris21, 2022, p. 27).

From this definition, changes in the technology available to produce action-oriented knowledge affect the expectations of statistical users from the public, private, and civil society (UN, 2022). In this regard, Mergel et al. (2019) point out that increasingly more users expect high-quality statistics, more detail and more immediate. The traditional production of statistics contemplates improving computer systems to evolve according to users' needs (UN, 2003); however, because AI competes at the core of the purpose of statistical offices, this new technology is changing the entire landscape of NSOs (MacFeely, 2016). Leveraging the potential of information technologies can drive innovation of new products and processes (UN, 2022), but it must be aligned with the business model of NSOs for its sustainable digital transformation.

2.2 Literature search, screening and skimming stage.

Having defined the concepts of digital transformation and AI, which will be considered in the remainder of this paper, a review of the ex-ante literature allows us to understand how different authors perceive the changes AI produces in today's organizations' life. Therefore, a query was conducted with the concept of 'Digital Transformation' AND 'Artificial Intelligence' (or 'AI' or 'Machine Learning') in the Web of Science, resulting in 1,566 articles. For this review, articles from 2018 to 2023 were chosen to ensure that the concept has fully evolved as a transformative force in the business model.

Initially, to refine the search and focus on articles specifically addressing digital transformation in the public sector and a field closely related to NSOs, a topic analysis was performed using the BERTopic tool with the KMean model, resulting in six clusters (out of the 1,485 papers with abstracts). Clusters associated with industries unrelated to NSOs, such as health, education, and finance, were excluded. Consequently, only two topics remained: 'Digital technology using AI' and 'Digital transformation and business model' (n=580).

Papers included in these two topics were classified by title and abstract. Eligible criteria include references to uses of AI, skills need, and AI impact on human resources in the public sector, national statistics business model, or data-driven strategies (n=166). These papers were carefully read to determine if their contribution applies to a country, organizational, and individual level (n=15).

The clusters and topics identified, and the distribution of the papers are shown in Table 1:

Table 1: AI+ Digital transformation: Topics and selected papers.

TOPIC LABELS	NO SELECTED	SELECTED	TOTAL
Digital technology using AI	341	6	349
Digital transformation technology business	244	9	253
Digitalized processes in different industries	273		273
Digital transformation financial organizations	305		305
Health Digital Solutions	206		206
Education and Learning in the Digital Age	99		99
Total	1470	15	1485

Lastly, extensive research was conducted on the latest publications from statistical offices and international organizations to address the topic of AI and Digital Transformation comprehensively applied to NSOs. This step was prompted primarily by the impact of COVID-19, which has led many NSOs to initiate their digitization and digitalization strategies. A few NSOs from developed countries have also ventured into digital transformation strategies. As a result, this integrative review includes an additional four articles from organizations closely associated with NSOs Digital Transformation of National Statistical Offices (PARIS21, 2022), Handbook on Management and Organization of National Statistical Systems (UN, 2022), Handbook on Methodology of Modern Business Statistics (Karlberg, 2017), Machine Learning for Official Statistics (UNECE, 2021).

2.3 Selected articles

Data-driven strategies, such as AI, involve exploring novel roles and value creation (PARIS21, 2022). However, at an early stage, the organization goes through an initial awareness phase in which adoption and use do not show significant changes in the business model but allow the researcher to perceive the alternative uses of the technology, the skills needed and the possible impacts on production.

This literature review summarizes AI tools' uses, human resources skills, and effects in NSOs. The selected papers will be condensed into two steps of digital transformation. The first cluster of selected papers, titled 'Digital technology using AI,' demonstrates how AI can generate value in the initial production of new statistics during the 'initial awareness' phase. The second cluster, 'Digital transformation technology,' discusses the sustainable implementation of this change, aiming for full integration into the business model.

2.3.1. Adoption and use of AI in a Digital Transformation

Selected papers in the cluster "Digital technology using AI" encompass authors that perceive AI as a novel tool capable of impacting an organization's existing conditions and context. The context of analysis or unit of investigation in the selected papers varies between the broader country level, the organizational level and the individual level. When examining an NSO, it is crucial to consider the potential impact across these levels, which is why in this section, articles are categorized based on the level they focus on (country, organization, or individual).

At a country's level, digital transformation may positively affect the country's economic development, labour productivity and aggregated employment (Dąbrowska et al., 2022; Elia & Margherita, 2022). In such cases, the uses of AI are related to cognitive enterprises that create new value by leveraging data, workflow, and people using AI, big data

analytics, and advanced analytics (Dąbrowska et al., 2022). Even if it is not clear if the development of the country and the positive impact on the labour market is the cause or effect of a successful digital transformation, the initial awareness of the use of AI in NSO must include a careful analysis of the country context in terms of data, workflow, people and their respective laws.

At the organizational level, AI impacts differently at different stages of adoption. Generally, an organization can be classified based on its maturity in three stages: sensing-perceiving, adopting a reacting, and envisioning (Elia & Margherita, 2022). In addition, it is crucial to identify key dimensions of the organization related to cognition (decisions at the management level), routine (the place of employees) or organizational forms (culture) that will affect the way the organization goes through the digital transformation by evolutive or transformative, responding or shaping the ecosystem (Volberda et al., 2021). Unlike other technologies, AI has particular requirements that add the need to define roles and responsibilities between humans and artificial intelligence (Sandkuhl & Rittelmeyer, 2022). Unfortunately, for the sake of generalization, authors referring to this level of analysis have opted for a case study. Still, as we demonstrate in the following section, the stages of maturation can also be used to analyze AI adoption in NSO.

At the individual level, various factors associated with digital transformation encompass perceptions and attitudes towards new technology, skills and training, workplace dynamics, and adaptability. The Handbook of Management of NSO (UN, 2022) emphasizes the importance of understanding the necessary skills for present and future needs and critical principles like data access, exchange capabilities, innovative production methods, and continuity. Hence, IT management should foster agile, flexible, and collaborative workplaces (UN, 2022). Analyzing factors at the individual level, such as leadership models, human resources management, and integration with working groups, becomes crucial (Trenerry et al., 2021). Managers need to differentiate between simple decision-making, which is more likely to be replaced by AI, and skills related to complex decision-making, knowledge, and handling pressure, which will be more likely to be augmented by AI (Giraud et al., 2022). Lastly, top management support and organizational readiness are as critical as establishing trust between employees and AI (Giraud et al., 2022).

Proposition 1: Given each country's context in terms of economic development (data-driven technologies), workflow (labour productivity), and people (aggregated employment), the benefits and costs of AI adoption are conditional on the presence of the following factors at each stage of adoption maturity: Skills and training, Workplace dynamic, Adaptability, Collaboration, Complexity of decision solved by AI.

2.3.2. Digital transformation technology and business model

In section 2.3.1, the authors discuss the use of AI mentioning keywords such as digital, technology, AI usage, transformation, and development. However, for AI adoption to undergo sustainable implementation and become fully integrated into operations, it needs to evolve into an influential tool capable of driving changes in the business model. Authors on the topic 'Digital transformation and business model' use keywords related to business, novelty, digitalization, impact, process, and change. Therefore, the second group of selected papers are more related to business strategies at each level of analysis country, organization, and individuals.

At the country's level, authors prefer ICT-related variables as a proxy of digital transformation (Eom et al., 2022; Małkowska et al., 2021). For external organizational factors, government R&D policies (R&D subsidies, research) and internal factors, the appropriability strategy (patents, copyrights), the information source strategy (consultants, universities), and the R&D cooperation strategy (Eom et al., 2022; Małkowska et al., 2021). The measures used in these papers help to understand the context of each country where an NSO will produce their digital transformation. There is no

clear evidence of what factor can accelerate the process, but the readiness of a country's infrastructure is a good starting point for encouraging all industries (including those related to data processing) to move towards the pace of digital transformation.

At the organizational level, two examples were found related to the impact of AI in the public sector. The case study in Norwegian municipalities shows that employees' good reception tools were related to citizens' consultation and system/data quality integration (Mikalef et al., 2019). These two are among the most crucial factors to advance in AI uses in the public sector (Mikalef et al., 2019). In the case of expert consultation in organizations from Spain's public administration, Sobrino-Garcia (2021) evaluated the insufficient legislation to make AI safer to avoid negative societal impacts in opacity, legal uncertainty, biases, or personal data protection.

Other authors see AI adoption that may add value to organizational activities, boundaries and goals (Holmström, 2022), either in the present or future. Business model organizations can be affected quickly by changes in the process or artefacts coming from external pressure (Mergel et al., 2019). Conversely, internal pressure has more impact in the long term if it involves changes in the organizational culture, competencies, and mindset (Mergel et al., 2019).

At the individual level, Ahn and Chen (2022) conducted a study examining the understanding and perceptions of government employees regarding various AI technologies, including their long-term societal impact, relationship with human workers, and appropriate functions within the government. Adding to this analysis, Fedorets et al. (2022) explored the direct and indirect effects of technological exposure on job quality, considering changes in the workplace, working hours, autonomy, and stress levels. Furthermore, Kolade and Owoseni (2022) discovered that the percentage of German employees at risk of automation decreased from 47% to 15% when analyzing tasks rather than occupations. Lastly, in addition to employee engagement with AI, one notable difference between AI-related technologies and other ICT tools is the increasing involvement of stakeholders in the adoption process, enhancing the value of final services and products. The significance of networks becomes crucial due to internal skill and budget constraints (van Noordt & Misuraca, 2022).

Proposition 2: Given each country's context in terms of ICT measurements, Government R&D, Appropriability strategy, Information source strategy, and R&D cooperation, the benefits and costs of AI adoption are more likely to transform the business model if the following factors are considered: Citizen consultation, System/Data quality integration, Internal legislation by inducing a business model project including Job Quality, Automation, and Stakeholders involvement analysis.

The following table summarizes the key factors mentioned in the selected papers.

Table 2: Key factors from the selected papers.

	DIGITAL TECHNOLOGY USING AI	DIGITAL TRANSFORMATION TECHNOLOGY BUSINESS
COUNTRY	-Country context related to data, workflow and people. (Dąbrowska et al., 2022; Elia & Margherita, 2022)	-ICT measurements -Government R&D -Appropriability strategy -Information source strategy -R&D cooperation (Eom et al., 2022; Małkowska et al., 2021)

Table 2: Key factors from the selected papers (continued).

	DIGITAL TECHNOLOGY USING AI	DIGITAL TRANSFORMATION TECHNOLOGY BUSINESS
ORGANIZATION	<ul style="list-style-type: none"> -Stages of digital maturation -Roles and responsibilities <p>(Elia & Margherita, 2022; Sandkuhl & Rittelmeyer, 2022; Volberda et al., 2021)</p>	<ul style="list-style-type: none"> -Citizen consultation -System/Data quality integration -Internal legislation <p>(Mergel et al., 2019; Mikalef et al., 2019; Sobrino-García, 2021)</p>
INDIVIDUAL	<ul style="list-style-type: none"> -Skills and training -Workplace dynamic -Adaptability -Collaboration -Complexity of decision solved by AI <p>(Giraud et al., 2022; Trenerry et al., 2021; UN, 2022)</p>	<ul style="list-style-type: none"> -Job quality -Automation -Stakeholders involvement <p>(Ahn & Chen, 2022; Fedorets et al., 2022; Kolade & Owoseni, 2022; van Noordt & Misuraca, 2022)</p>

3 MACHINE LEARNING IN NATIONAL STATISTICS OFFICES CASE STUDY

We propose a multiple case study to explore, with concrete examples, how the benefits, costs and subsequent success of adopting artificial intelligence (AI) to collect, generate or analyze data in National Statistical Offices (NSOs) were affected by the factors and relationships obtained in the above literature review. Using case study methodology, we focus on understanding the dynamics within a single setting by combining different sources of information such as archives and observations (Eisenhardt, 1989).

We follow two simplified steps based on Eisenhardt (1989): 1) Initial definition of research settings, 2) Analyzing the cases by developing an early-stage theory and constructing measures related to the alternative uses of AI in data-driven strategies, as well as the skills required for human resources to collaborate with AI in a national statistic production.

3.1 Research settings and data collection

Public and Private organizations producing statistics are challenging the role of NSOs, prompting NSOs to explore alternative data sources to supplement or replace traditional data collection or to generate new analysis tools (Yung, Wesley et al., 2018).

In 2015, a workshop on the Modernization of Statistical Production presented the Machine Learning Documentation Initiative (Chu & Poirier, 2015) as the beginning of the compilation of international community efforts to promote the integration of machine learning (ML) practices within statistical offices. The UNECE High-Level Group for the Modernisation of Official Statistics (HLG-MOS) continued this work and launched 2019 the Machine Learning Project. At first, 23 participants from 13 organizations participated by developing projects related to ML applications in their country's NSO. The last reports had grown to more than 120 members from different countries, and national and international organizations (Julien, Claude et al., 2020).

From the list of ML projects on the official page of HLG-MOS, fifteen cases were selected among those who presented the report between 2020 and 2021. Information referencing uses, skills and effects on workers, accuracy, reliability and timely were extracted. Projects that do not contain complete information regarding the goal, cost, benefits, and current and future steps were excluded.

The 15 cases were organized by AI tool for classification and coding (cases 1-6, 11 and 13), editing and imputation (cases 7-9, and 15), and imagery (cases 10, 12, and 14); and by three stages of development: exploratory (cases 1-4, 7 and 14), proof of concept (cases 5, 6, 8, 9, 12, 13, and 15), and in production (cases 10 and 11). Only projects "in production" were considered successfully developed. Finally, cases aiming to complement or improve the performance of existing procedures (cases 1-11) were more frequent than those introducing new services (cases 12-15).

3.2 Data analysis

Many national statistical offices (NSOs) are investigating how ML can increase the relevance and quality of official statistics. The demand for trusted information is growing rapidly, considering the increasingly accessible technologies and numerous competitors.

Initially, the HLG-MOS project saw the potential for the production of statistics to be more efficient by automating specific processes or assisting humans in carrying out the process. Cases with these goals that were successful in making it part of the regular production mentioned key factors such as collaboration, simplicity and a solid business project (cases 10, 11). In other words, projects that only improve existing cost, time or any other measurable goal but do not consider the whole ecosystem remains in the exploratory or proof of concept stage. For a project to significantly impact creating new value for the NSO, a solid business project must be presented to the organization.

The majority of cases are in exploratory or proof of concept stage. The key factors mentioned are related to the tool's performance, such as cost, time and manual work reduction and increased accuracy (cases 1-9). Because the next steps involve testing with other tools (cases 2, 3, 5 - 7, and 9), evaluation (cases 4 and 8) and studying how to incorporate them into the workflow process (case 1), these projects highlight the importance of developing employees' skills and training, consultation with internal users, and a system/data quality validation.

Therefore, cases 1 to 11 show evidence related to *Proposition 1*. The benefits generated by the adoption of AI in the stage of collection (imagery), generation (coding) and analysis (editing and imputing) in the initial stage of the project (exploration or proof of concept) were given by the improvement in accuracy in some cases from 76% to 90% for tasks that were previously performed manually (case 1). Moreover, improved work dynamics were generated by the reduced manual work for tasks easily automated by the ML allocating the human resource to complete the system only for the cases in which the AI did not perform as expected (case 2). In addition, internal and external collaboration was perceived as a benefit when skills were available to the organization (cases 3 and 5) and a future cost when they were not (cases 6 and 7). The most advanced projects in the production phase show benefits in terms of increased precision and time reduction thanks to the organizational capacity to generate strong teams capable of evaluating the role of the project as a solid business case developing a quick validation of the quality of the system to respond to specific needs (cases 10-11). In terms of cost, each organization evaluates the necessity to complete the IT infrastructure (such as cloud storage or virtual machine), IT maintenance, staff training, and quality assurance requirements.

Lastly, if successfully incorporated into production, new products or services add value to the organization. Among the selected cases, the case 14 explores new services, and cases 12, 13 and 15 are in the proof of concept stages of new

¹ See the list of selected case studies Annex 2

services. The new services examples are related to predicting during non-census years (case 14) or exploitation on a new source of data such as register-based (case 15) or web-scaping (case 13), or imagery (case 12). The key factors mentioned in these projects are related to external consultation and storage infrastructure.

Cases from 12 to 15 show evidence for *Preposition 2*. The benefits related to transforming the business model were the exploration of new sources of raw data to generate new values that complement the existing statistics during non-survey periods. The consultation with stakeholders ensures that the added value is evaluated and considered by specific users who refine the business model to incorporate the new system into production, satisfying specific needs. However, since the projects are not yet in production, a complete evaluation of the impact on the business model is impossible.

Table 3 summarizes the benefits and costs mentioned in the 15 cases. Overall successful cases reported significant benefits for using ML in NSO, even when it comes to improvements to existing production.

Table 3: Summary of cases analyzed by purpose, goal, and project stage.

Cases	Purpose of the BM			Goal		Project stage			Benefits (B)							Cost (C)		
	Collect	Generate	Analyze	Upgrade existing process	New product	Exploratory	Proof of Concept	In production	Accuracy, Precision	New value-added	Collaboration	Time reduction	Manual work reduction	Stakeholder involvement	Simplicity	Solid Business Model Project	Skills and training	Storage infrastructure
1		X		X		X			B								C	C
2		X		X		X						B	B					
3		X		X		X			B		B						B	
4		X		X		X			B									
5		X		X			X		B		B						B	
6		X		X			X		B								C	
7		X		X		X			B								C	
8		X		X			X		B		B							
9		X		X			X		B									
10	X			X				X	B		B	B			B	B	B	C
11		X		X				X	B		B	B		B		B		
12	X				X		X			B								C
13		X		X			X			B								
14	X			X		X				B								C
15		X		X			X			B								

4 CONCLUSION

We proposed in this paper to combine an integrative review of digital transformation and artificial intelligence (AI) with the study of machine learning cases applied in National Statistics Offices (NSO) in different countries. In prior works, authors perceive AI as a novel tool capable of impacting an organization's existing conditions and context, mentioning the importance of identifying the new skills and training demand, the impact in the workplace, and collaboration. However, in this study, we found that in most cases of machine learning uses in AI, these factors affect the awareness stage, and the AI projects remain in exploratory or proof of concept stages for at least a year. Moreover, for a project to successfully undergo production, it is necessary to address a solid business model project, involve stakeholders, and integrate the system/data quality team to validate the AI tool.

On the other hand, although our study attempted to include as many cases as possible, our research has the limitation of considering cases in countries whose development is higher than average so that developing countries could include as main factors issues related to infrastructure, which in our analysis did not play a significant role. Nevertheless, we mentioned factors at the country and organizational level that could be used to conduct a broader study that includes a more significant number of countries.

Finally, we provided a new point of view on the place of AI in NSO that are traditionally intensive in human resources. In addition, we contribute to a deeper understanding of the benefits and costs of AI adoption in national statistical processes by introducing factors that provide valuable insights for policymakers and practitioners in optimizing the use of AI and human resources in the public sector.

Annex 1: Phases of the business process model, digitalization and digital transformation

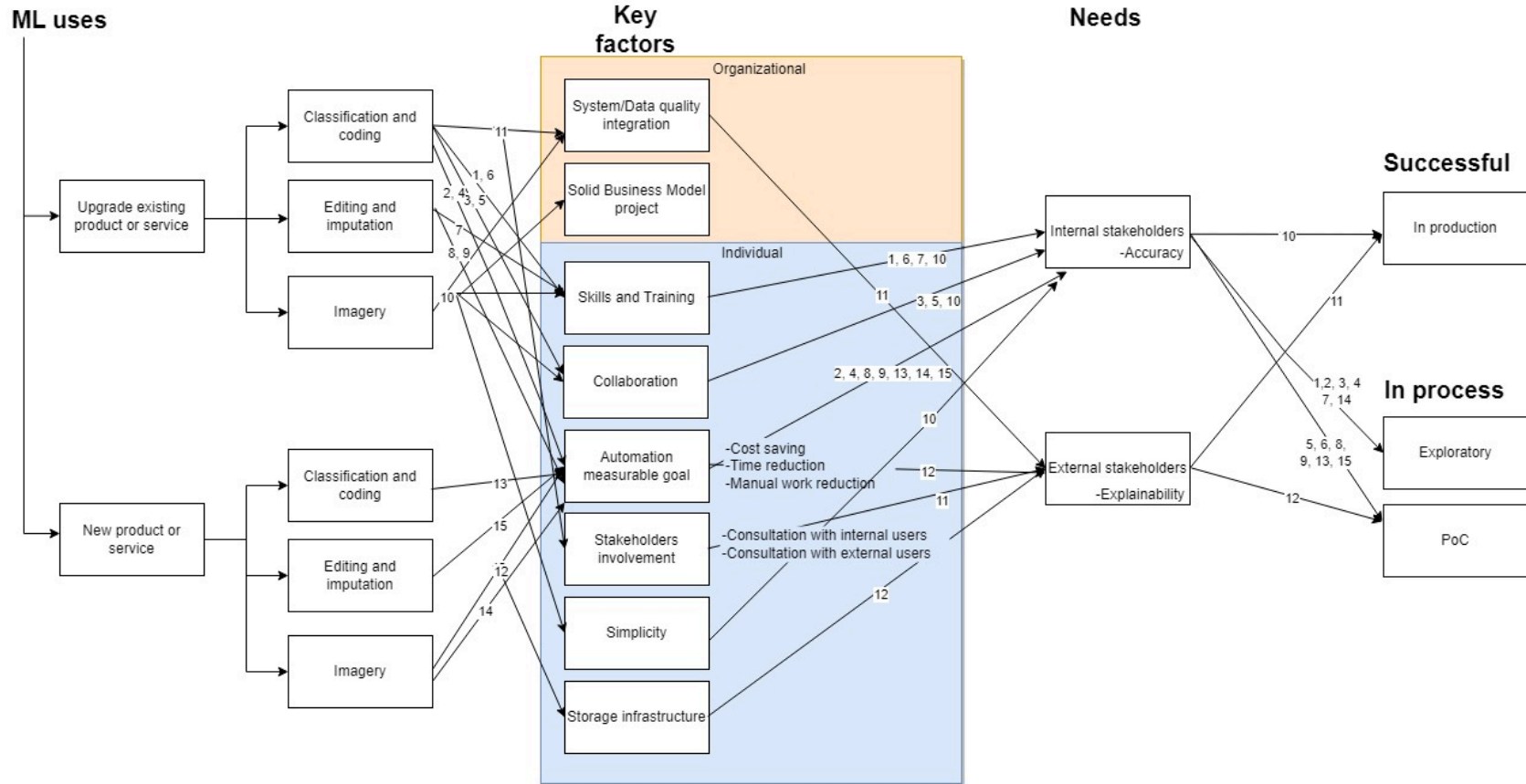
STATISTICAL BUSINESS PROCESS MODEL			DIGITALIZATION	DIGITAL TRANSFORMATION
PHASE	SUB-PHASE	TASK	New AI tools	Emerging demands
Preparatory	Specify needs			
	Design			
	Build			
Production	Collect	Create frames & select sample, Set up collection	SVM, Logistic regression, Random Forest, Neural Networks, XGBoost, K-NN, Naïve Bayes, Decision trees, Fasttext	high-quality statistics, more detail and more immediate
	Process	Integrate data, Classify & code, Edit & impute, Derive new variables and units	Naive Bayes, MissForest, Weighted k-nearest-neighbours, Extremely Randomized Trees, LeNet	
	Analyze	Prepare & validate output, Interpret & explain output	Convolutional Neural Network	
	Disseminate			
Saving for the future	Achieve			
	Summarize and formulate an action plan	Evaluate		

Annex 2: Detail of the cases analyzed in the program UNECE High-Level Group for the Modernization of Official Statistics (HLG-MOS)

CASE	COUNTRY, YEAR	TITLE	AUTHOR	GOAL	AI TYPE	STATE	FUTURE
1	Mexico (2020)	Occupation and Economic activity coding using natural language processing	José Alejandro Ruiz Sánchez, Jael Pérez Sánchez, Adrián Pastor	(Upgrade) Complementing current process	Classification and coding	Exploratory	Expand the dataset, incorporate into the workflow process
2	Iceland (2020)	Automatic coding of occupation and industry in social statistical surveys	Anton Örn Karlsson	(Upgrade) Complementing current process	Classification and coding	Exploratory	Test other ML tools
3	Serbia (2020)	ML pilot study – coding textually described data on economic activity collected from Labour Force Survey	Nevena Pavlovic, Sinisa Cimbajevic, Branko Josipovic, Dusica Zecevic	(Upgrade) Complementing current process	Classification and coding	Exploratory	Test other ML tools
4	United Kingdom (2021)	Automatic coding of standard industrial and occupational classifications	Thanasis Anthopoulos, Data Science for public goods, Data science campus	(Upgrade) Complementing current process	Classification and coding	Exploratory	Evaluation
5	Norway (2020)	Standard Industrial Code Classification by Using Machine Learning	Thivyesh Ahilathan, Tatsiana Pekarskaya	(Upgrade) Complementing current process	Classification and coding	PoC	Test, new tools
6	Poland (2020)	ECOICOP clasiffication	Marta Kruczek-Szepel, Krystyna Piątkowska	(Upgrade) Complementing current process	Classification and coding	PoC	Test, new tools
7	Poland (2021)	Multiple imputation through machine learning in a survey of sport clubs	Sebastian Wójcik, Agnieszka Gienza	(Upgrade) Complementing current process	Edit and imputation	Exploratory	Test, new tools
8	Germany (2020)	Machine learning for imputation	Florian Dumpert	(Upgrade) Complementing current process	Edit and imputation	PoC	Evaluation
9	United Kingdom (2020)	Editing of LCF (Living Cost and Food) Survey Income data with Machine Learning	Claus Sthamer	(Upgrade) Complementing current process	Edit and imputation	PoC	Exploratory

CASE	COUNTRY, YEAR	TITLE	AUTHOR	GOAL	AI TYPE	STATE	FUTURE
10	Australia (2020)	An ML application to automate an existing manual process through the use of aerial imagery. Numerous areas throughout the ABS will benefit from the development of this ML application.	Daniel Merkas and Debbie Goodwin	(Upgrade) Complementing current process	Imagery	Production	Improve quality
11	Canada (2020)	Industry and occupation coding	Isaac Ross and Justin J. Evans	(Upgrade) Complementing current process, fast	Classification and coding	Production	Integrate QC solution into our systems, try other ML tools
12	Switzerland (2020)	Arealstatistik Deep Learning (ADELE)	Claudio Facchinetti	(New) methods and/or new data sources	Imagery	PoC	Validation improve
13	Brazil (2020)	Apply ML techniques to classification and aggregation of web scraped price data	Vladimir G. Miranda, Lincoln T. da Silva	(New) New data source web scraping	Classification and coding	PoC	Exploratory
14	Mexico (2020)	ML application to use satellite data in combination with census data to produce new information	Abel Coronado; Jimena Juárez; Ricardo Bucio	(New) Prediction during non-census years	Imagery	Exploratory	Evaluation
15	Italy (2020)	Imputation of the variable "Attained Level of Education" in Base Register of Individuals	Fabrizio De Fausti, Marco Di Zio, Romina Filippini, Simona Toti, Diego Zardetto	(New) Register based	Edit and imputation	PoC	Test other ML tools

Diagram 1: AI uses in NSOs and key factors for adopting AI tools in statistical production



Own elaboration based on reports from HLG-MOS projects: 1: Ruiz Sánchez, Pérez Sánchez, and Pastor (Mexico, 2020), 2: Karlsson (Iceland, 2020), 3: Pavlovic, Cimbajjevic, Josipovic, Zecevic (Serbia, 2020), 4: Anthopoulos (UK, 2021), 5: Ahilathasan, Pekarskaya (Norway, 2020), 6: Kruczek-Szepel, Piątkowska (Poland, 2020), 7: Wójcik, Gienza (Poland, 2021), 8: Dumpert (Germany, 2020), 9: Sthamer (UK, 2020), 10: Merkas and Goodwin (Australia, 2020), 11: Ross and Evans (Canada, 2020), 12: Facchinetti (Switzerland, 2020), 13: Miranda, da Silva (Brazil, 2020), 14: Coronado, Juárez, Bucio (Mexico, 2020), 15: De Fausti, Di Zio, Filippini, Toti, Zardetto (Italy, 2020)

REFERENCES

- Ahn, M. J., & Chen, Y.-C. (2022). Digital transformation toward AI-augmented public administration: The perception of government employees and the willingness to use AI in government. *Government Information Quarterly*, 39(2), 101664. <https://doi.org/10.1016/j.giq.2021.101664>
- Chu, K., & Poirier, C. (2015). *Machine Learning Documentation Initiative. Topic (ii): Enterprise Architecture and its role in the Modernisation of Statistical Production*.
- Dąbrowska, J., Almpantopoulou, A., Brem, A., Chesbrough, H., Cucino, V., Di Minin, A., Giones, F., Hakala, H., Marullo, C., Mention, A.-L., Mortara, L., Nørskov, S., Nylund, P. A., Oddo, C. M., Radziwon, A., & Ritala, P. (2022). Digital transformation, for better or worse: A critical multi-level research agenda. *R&D Management*, 52(5), 930–954. <https://doi.org/10.1111/radm.12531>
- Desouza, K. C., Dawson, G. S., & Chenok, D. (2020). Designing, developing, and deploying artificial intelligence systems: Lessons from and for the public sector. *Business Horizons*, 63(2), 205–213. <https://doi.org/10.1016/j.bushor.2019.11.004>
- Duan, Y., Edwards, J. S., & Dwivedi, Y. K. (2019). Artificial intelligence for decision making in the era of Big Data – evolution, challenges and research agenda. *International Journal of Information Management*, 48, 63–71. <https://doi.org/10.1016/j.ijinfomgt.2019.01.021>
- Eisenhardt, K. M. (1989). Building Theories from Case Study Research. *The Academy of Management Review*, 14(4), 532–550. <https://doi.org/10.2307/258557>
- Elia, G., & Margherita, A. (2022). A conceptual framework for the cognitive enterprise: Pillars, maturity, value drivers. *Technology Analysis & Strategic Management*, 34(4), 377–389. <https://doi.org/10.1080/09537325.2021.1901874>
- Eom, T., Woo, C., & Chun, D. (2022). Predicting an ICT business process innovation as a digital transformation with machine learning techniques. *Technology Analysis & Strategic Management*, 0(0), 1–13. <https://doi.org/10.1080/09537325.2022.2132927>
- Fedorets, A., Kirchner, S., Adriaans, J., & Giering, O. (2022). Data on Digital Transformation in the German Socio-Economic Panel. *Jahrbücher Für Nationalökonomie Und Statistik*, 242(5–6), 691–705. <https://doi.org/10.1515/jbnst-2021-0056>
- Fountaine, T., McCarthy, B., & Saleh, T. (2019, July 1). Building the AI-Powered Organization. *Harvard Business Review*. <https://hbr.org/2019/07/building-the-ai-powered-organization>
- Giraud, L., Zaher, A., Hernandez, S., & Akram, A. A. (2022). The impacts of artificial intelligence on managerial skills. *Journal of Decision Systems*, 0(0), 1–34. <https://doi.org/10.1080/12460125.2022.2069537>
- Hanna, N. (2018). A role for the state in the digital age. *Journal of Innovation and Entrepreneurship*, 7(1), 5. <https://doi.org/10.1186/s13731-018-0086-3>
- Holmström, J. (2022). From AI to digital transformation: The AI readiness framework. *Business Horizons*, 65(3), 329–339. <https://doi.org/10.1016/j.bushor.2021.03.006>
- Janssen, M., Brous, P., Estevez, E., Barbosa, L. S., & Janowski, T. (2020). Data governance: Organizing data for trustworthy Artificial Intelligence. *Government Information Quarterly*, 37(3), 101493. <https://doi.org/10.1016/j.giq.2020.101493>

- Julien, Claude, Choi, InKyung, Deeben, Eric, Yung, Wesley, & , Eric Deeben, Wesley Yung and Alex. (2020). *HLG-MOS Machine Learning Project—Machine Learning for Official Statistics—UNECE Statswiki*. <https://statswiki.unece.org/display/ML/HLG-MOS+Machine+Learning+Project>
- Karlberg, M. (2017, August 2). *Handbook on Methodology of Modern Business Statistics* [Text]. CROS - European Commission. https://ec.europa.eu/eurostat/cros/content/handbook-methodology-modern-business-statistics_en
- Klimczak, K. M., & Fryczak, J. M. (2022). Datafication – economization and monetization of data. In *Digital Finance and the Future of the Global Financial System*. Routledge.
- Kolade, O., & Owoseni, A. (2022). Employment 5.0: The work of the future and the future of work. *Technology in Society*, 71, 102086. <https://doi.org/10.1016/j.techsoc.2022.102086>
- MacFeely, S. (2016). The Continuing Evolution of Official Statistics: Some Challenges and Opportunities. *Journal of Official Statistics*, 32. <https://doi.org/10.1515/jos-2016-0041>
- Małkowska, A., Urbaniec, M., & Kosała, M. (2021). The impact of digital transformation on European countries: Insights from a comparative analysis. *Equilibrium*, 16, 325–355. <https://doi.org/10.24136/eq.2021.012>
- Mergel, I., Edelman, N., & Haug, N. (2019). Defining digital transformation: Results from expert interviews. *Government Information Quarterly*, 36(4), 101385. <https://doi.org/10.1016/j.giq.2019.06.002>
- Mikalef, P., Fjørtoft, S. O., & Torvatn, H. Y. (2019). Artificial Intelligence in the Public Sector: A Study of Challenges and Opportunities for Norwegian Municipalities. In I. O. Pappas, P. Mikalef, Y. K. Dwivedi, L. Jaccheri, J. Krogstie, & M. Mäntymäki (Eds.), *Digital Transformation for a Sustainable Society in the 21st Century* (pp. 267–277). Springer International Publishing. https://doi.org/10.1007/978-3-030-29374-1_22
- OECD. (2002). *Measuring the Non-Observed Economy—A Handbook* -. <https://www.oecd.org/sdd/na/measuringthenon-observeconomy-ahandbook.htm>
- PARIS21. (2022). *Digital Transformation of National Statistical Offices*. OECD. <https://doi.org/10.1787/ee4b1b85-en>
- Saldanha, T. (2019). *Why Digital Transformations Fail: The Surprising Disciplines of How to Take off and Stay Ahead*. Berrett-Koehler Publishers, Incorporated. <http://ebookcentral.proquest.com/lib/kaist/detail.action?docID=5785676>
- Sandkuhl, K., & Rittelmeyer, J. D. (2022). Use of EA Models in Organizational AI Solution Development. In D. Aveiro, H. A. Proper, S. Guerreiro, & M. de Vries (Eds.), *Advances in Enterprise Engineering XV* (pp. 149–166). Springer International Publishing. https://doi.org/10.1007/978-3-031-11520-2_10
- Sobrino-García, I. (2021). Artificial Intelligence Risks and Challenges in the Spanish Public Administration: An Exploratory Analysis through Expert Judgements. *Administrative Sciences*, 11(3), Article 3. <https://doi.org/10.3390/admsci11030102>
- Torraco, R. J. (2005). Writing Integrative Literature Reviews: Guidelines and Examples. *Human Resource Development Review*, 4(3), 356–367. <https://doi.org/10.1177/1534484305278283>
- Torraco, R. J. (2016). Writing Integrative Reviews of the Literature: Methods and Purposes. *International Journal of Adult Vocational Education and Technology (IJAVET)*, 7(3), 62–70. <https://doi.org/10.4018/IJAVET.2016070106>
- Trenerry, B., Chng, S., Wang, Y., Suhaila, Z. S., Lim, S. S., Lu, H. Y., & Oh, P. H. (2021). Preparing Workplaces for Digital Transformation: An Integrative Review and Framework of Multi-Level Factors. *Frontiers in Psychology*, 12. <https://www.frontiersin.org/articles/10.3389/fpsyg.2021.620766>
- UN. (2022). *Handbook on Management and Organization of National Statistical Systems*. Capacity Development // UNSD. <https://unstats.un.org/capacity-development/handbook/index.cshtml>

- UNECE. (2021). *Machine Learning for Official Statistics*. <https://unece.org/sites/default/files/2022-09/ECECESSTAT20216.pdf>
- van Noordt, C., & Misuraca, G. (2022). Exploratory Insights on Artificial Intelligence for Government in Europe. *Social Science Computer Review*, 40(2), 426–444. <https://doi.org/10.1177/0894439320980449>
- Verhoef, P. C., Broekhuizen, T., Bart, Y., Bhattacharya, A., Qi Dong, J., Fabian, N., & Haenlein, M. (2021). Digital transformation: A multidisciplinary reflection and research agenda. *Journal of Business Research*, 122, 889–901. <https://doi.org/10.1016/j.jbusres.2019.09.022>
- 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>
- Volberda, H. W., Khanagha, S., Baden-Fuller, C., Mihalache, O. R., & Birkinshaw, J. (2021). Strategizing in a digital world: Overcoming cognitive barriers, reconfiguring routines and introducing new organizational forms. *Long Range Planning*, 54(5), 102110. <https://doi.org/10.1016/j.lrp.2021.102110>
- Whittemore, R., & Knaf, K. (2005). The integrative review: Updated methodology. *Journal of Advanced Nursing*, 52(5), 546–553. <https://doi.org/10.1111/j.1365-2648.2005.03621.x>
- Wirtz, B. W., Langer, P. F., & Fenner, C. (2021). Artificial Intelligence in the Public Sector—A Research Agenda. *International Journal of Public Administration*, 44(13), 1103–1128. <https://doi.org/10.1080/01900692.2021.1947319>
- Wirtz, B. W., Weyerer, J. C., & Geyer, C. (2019). Artificial Intelligence and the Public Sector—Applications and Challenges. *International Journal of Public Administration*, 42(7), 596–615. <https://doi.org/10.1080/01900692.2018.1498103>
- Yung, Wesley, Karkimaa, Jukka, Scannapieco, Monica, Barcarolli, Giulio, Zardetto, Diego, Ruiz Sanchez, José Alejandro, Braaksma, Barteld, Buelens, Bart, & Burger, Joep. (2018). *The use of machine learning in official statistics*. UNECE Machine Learning Team. https://www.google.com/search?q=The+use+of+machine+learning+in+official+statistics&rlz=1C1CHBD_enAR963AR963&oq=The+use+of+machine+learning+in+official+statistics&aqs=chrome..69i57j33i160j33i22i29i30.1351j0j7&sourceid=chrome&ie=UTF-8
- PARIS21 (2022), Digital Transformation of National Statistical Offices, OECD Publishing, Paris, <https://doi.org/10.1787/ee4b1b85-en>.
- United Nations (2003), Handbook of Statistical Organization: The Operation and Organization of a Statistical Agency, Third., vol. Serie F, No 88. New York: UN.