BEPP-DS: Building Evidence-Based Public Policies with Data Science

José Antonio F. Macedo[a], *Rossana Maria de Castro Andrade*[b], *Regis P. Magalhães*[a], *Lívia A. Cruz*[a], *José Florêncio Q. Neto*[a], *Samir B. Chaves*[a], *Mauricio Feijo B. M. Filho*[c], *Joaquim José Escola*[d], *Amanda Sousa*[b], *Pedro Almir M. Oliveira*[b,e], *Davyson S. Ribeiro*[b], *Paulo V. A. Fabrício*[b]

[a] Insight Lab, Federal University of Ceará, Brazil
jose.macedo@insightlab.ufc.br, ORCID: 0000-0002-0661-2978.

[b] Group of Computer Networks, Software Engineering and Systems (GREat), Federal University of Ceará, Brazil,
rossana@ufc.br, ORCID: 0000-0002-0186-2994.

[c] Federal University of Ceará, Brazil,
mauricio.benevides@ufc.br, ORCID: 0000-0002-8652-4328

[d] Trás-os-Montes and Alto Douro University, Portugal,
jescola@utad.pt ORCID: 0000-0002-6676-6928.

[e] Laboratory of Innovation and Scientific Computing (LICC), Federal Institute of Maranhão, Pedreiras, Brazil,
pedro.oliveira@ifma.edu.br, ORCID: 0000-0002-3067-3076.

Submitted: 31 January 2025, Revised: 26 March 2025, Accepted: 21 April 2025, Published: 26 May 2025

Abstract. The exponential growth of data and the advancement of computational tools have made Data Science (DS) an essential discipline for addressing complex societal challenges. In the public sector, Evidence-Based Public Policies (EBPP) leverage data-driven insights to enhance governance transparency, efficiency, and effectiveness. However, the integration of Data Science into policymaking presents challenges, including data quality, interdisciplinary collaboration, and institutional resistance. This paper introduces BEPP-DS, a structured methodology for developing EBPP using DS principles, emphasizing transparency, reproducibility, and scalability. The methodology is informed by real-world applications such as Big Data Social and Big Data Fortaleza, which illustrate how data-driven strategies improve policy design, implementation, and monitoring. BEPP-DS defines a structured framework, from problem identification to policy evaluation, ensuring data-driven decision-making in governance. The methodology provides a replicable model for governments seeking to harness Data Science in policy formulation. Future work includes expanding AI-driven analytics and strengthening citizen engagement in data governance.

Keywords. Data Science, Evidence-Based Public Policies, E-Government, Public Administration, Big Data, Policy Analytics, Digital Governance, Decision Support Systems
Practical report, DOI: https://doi.org/10.59490/dgo.2025.1045

1. Introduction

The exponential growth of data and developments in computational tools made Data Science (DS) a key field to tackle multidimensional societal challenges (Silva et al., 2022). Helping to derive actionable insights from data, DS has established itself as a vital force in many fields, such as medicine, education, and urban administration (Santos et al., 2023). Lately, applications of DS in governance have been exemplary, especially in the framing and implementation of Evidence-Based Public Policies (EBPP) (Anderson et al., 2005).

There is an increasing need for governments worldwide to make policies effective, equitable, and more transparent, with full involvement of relevant stakeholders (Saltelli et al., 2020). Evidence-based approaches allow

the systematic use of data within policy-making to ensure decisions are evidence-based, as opposed to intuition or tradition. However, for developing DS into public policy-making, there are related challenges –issues of data availability and quality, an interdisciplinary way of collaboration, and ethical, systematic methodologies guiding how knowledge is extracted within the government institutions' peculiarities (MacArthur et al., 2022).

Such is the demand for data-driven solutions in Brazil, whose socio-economic setting has also entrenched regional disparities. Despite all challenges, there is an emerging consensus on the need to structure and formalize DS use in government to work toward more effective public policies ultimately. This work describes a methodology for constructing EBPP using DS based on principles of transparency and non-discrimination.

Given this context, the question that guides this research is: how to structure a replicable methodology for building evidence-based public policies, using Data Science principles, that is applicable in different governmental contexts and capable of promoting more informed and effective decisions?

This paper describes the methodology that profited from the lessons learned while implementing large-scale data platforms such as Big Data Social (Sucupira Furtado et al., 2023a) and Big Data Fortaleza (Batista et al., 2024). Both examples give evidence of the potential of Data Science for developing governance that is efficient in processes, reproducible in decisions, and scalable in policies.

The rest of the paper is organized as follows: Section 2 gives an overview of the theoretical underpinning of Evidence-Based Public Policies and Data Science, showing their meeting point in governance. Section 3 reviews the related methodologies and pinpoints the literature gaps. Section 4 describes the proposed methodology, core components, and activities in depth. Section 5 demonstrates how the methodology works in real applications and discusses its impact. Finally, Sections 6, 7 and 8 discuss lessons learned and future opportunities, especially integrating Artificial Intelligence into data-driven policy-making.

2. Background

This section provides the basics necessary to understand the conjunction of Evidence-based Public Policies and Data Science, considering e-government. Understanding these domains allows us to comprehend better the opportunities and challenges of leveraging data science into governance practices, particularly in a context such as Brazil.

2.1. Evidence-based Public Policies

Evidence-based Public Policies are strategies and interventions designed and implemented based on empirical data and scientific research (Anderson et al., 2005). This approach ensures that rigorous evidence, rather than intuition or tradition, is the primary determinant of whether a policy is effective, efficient, and equitable. It is based on the central premise that structured methodologies that critically appraise and synthesize sound data substantially aid public decision-making and thus enable governments to handle societal challenges systematically and transparently.

EBPP adoption is rapidly expanding as the data and analytical tools base improve. Globally, the call for governments to account for their actions and deliver increased public service has grown. However, embedding evidence in policy faces several challenges: problems related to the quality of the data, resistance at the institutional level, and requirements for collaboration across disciplines. These hurdles notwithstanding, with increased calls for transparency and accountability, EBPP has become an indispensable framework for modern governance.

2.2. Data Science For Governments

Data Science (DS) is an interdisciplinary field that deals with methods and processes for extracting meaningful insights from data, combining statistical analysis, computational techniques, and domain expertise (Van Der Aalst and van der Aalst, 2016). It is pivotal in addressing complex societal challenges; it offers tools and methodologies to help uncover patterns, predict outcomes, and support decision-making processes.

The data science life cycle consists of five phases: data collection, cleaning, analysis, interpretation, and dissemination (Rahul and Banyal, 2020). Data collection involves aggregating information from diverse sources, while preprocessing ensures the data is clean and structured. In the analysis phase, analytical techniques like machine learning and statistical modeling are applied to derive actionable insights. These insights are then interpreted in the context of the problem domain and disseminated to stakeholders to inform decisions. This iterative cycle allows continuous improvement and adaptation, particularly suited for dynamic environments like governance.

2.3. e-Government in Brazil

e-Government is the use of digital technologies to improve the delivery of government services, enhance citizen engagement, and streamline administrative processes (Silcock, 2001). In Brazil, e-Government initiatives represent a unique opportunity to reduce persistent regional disparities and enhance public sector efficiency (Musafir, 2018). With more than 200 million people and a growing internet penetration rate, Brazil has the potential to use digital platforms to expand access to public services and enhance transparency.

However, adopting e-Government in Brazil is still dogged by various factors: deficiencies in rural digital infrastructure, socioeconomic disparities, and bureaucratic resistance. Nevertheless, projects like "Big Data Fortaleza" (Batista et al., 2024) and "Big Data Social"(Sucupira Furtado et al., 2023a) represent the transformational use of data science integrated into governance. These projects show how data-driven strategies could improve decision-making, resource allocation, and citizen outcomes.

3. Related Works

This section presents the related works studied to understand models and their respective results presented by Evidence-Based Public Policies and Data Science. Understanding these works allows us to better understand the challenges of leveraging data science in governance practices, especially in social contexts present in Brazil.

(Sucupira Furtado et al., 2023b) demonstrated how Ceará's digital transformation integrated technological innovation into public policy cycles to address social challenges and promote sustainable development. Tools like big data platforms and mobile apps enabled data-driven policymaking, improving efficiency, transparency, and responsiveness. Collaboration among public, academic, and private sectors (e.g., the IRIS Lab) enhanced institutional capacity and policy legitimacy. Focus on vulnerable populations aligned with SDGs, reducing inequalities through targeted interventions.

(Pereira et al., 2021) highlighted the role of standardized information structures in creating efficient and accessible policies. Interoperability between government systems improved evidence-based decision-making and digital inclusion by simplifying access to services. Standardized taxonomies enabled better policy monitoring and evaluation, ensuring adaptive and inclusive governance.

(Furtado et al., 2023) presented the importance of integrating data and advanced analytics to support the creation of more effective and targeted public policies. By consolidating information on social vulnerabilities, such as income, housing, and nutrition, the system made it possible to identify populations at greatest risk and prioritize government actions in a well-founded manner. This approach demonstrates how digital tools can strengthen the diagnosis of social problems, offering an empirical basis for the formulation of policies that are more aligned with the real needs of the population. A central feature of this work is its ability to personalize public policy recommendations based on detailed analyses of specific conditions. Through interactive dashboards, public managers can access solutions adapted to meet particular demands, such as the provision of financial aid, housing policies, or access to government programs to support society. This personalization reflects a more human and strategic approach to policymaking, by focusing on practical results and immediate impacts for the most vulnerable populations, maximizing the efficiency and social impact of government interventions.

(Schröer et al., 2021) presents CRISP-DM (Cross-Industry Standard Process for Data Mining) highlighting its relevance as a methodological framework capable of guiding the creation of evidence-based public policies. Divided into six phases – business understanding, data understanding, data preparation, modeling, evaluation,

and implementation –, this model offers a systematic approach to analyze large volumes of data, transform information into useful knowledge, and translate analytical insights into concrete actions. This methodology has a strong potential for application in the public sector, where data analysis can improve the formulation of policies aimed at complex social problems. The initial phase of business understanding is essential to align the data analysis objectives with public policy priorities. It allows managers to identify critical issues, such as social inequality or low coverage of health services, and define clear goals for interventions. By establishing these connections between data and objectives, the model enables the creation of more strategic policies aligned with the real needs of the population. This approach highlights how structuring data in analytical projects can improve the diagnosis of problems and direct resources to areas of greatest impact. The data preparation and modeling phases are equally important to the policymaking process, as they ensure the quality and usability of the analyzed data. By consolidating information from different sources, such as government databases or sectoral statistics, CRISP-DM helps to create a comprehensive overview of the challenges faced by society. In addition, the analytical models developed at this stage allow for scenario simulations and predictions of the effects of policies before their implementation, reducing uncertainty and optimizing expected results. This is especially relevant in contexts such as health, education and public safety, where well-planned interventions can generate significant transformations.

The analysis of the presented studies reveals a convergence around the strategic use of digital technologies and structured methodologies to strengthen the process of creating and implementing public policies. However, they also highlight significant gaps that need to be addressed to ensure a comprehensive and sustainable implementation. These studies underscore both the central role of technology in modernizing public governance and the challenges associated with inter-institutional integration, data privacy, continuous monitoring, and social participation.

The use of structured frameworks demonstrates the importance of well-defined analytical processes to guide data-driven decision-making. This approach allows for the identification of critical problems, the definition of clear goals, and the evaluation of the potential impact of policies before their implementation. However, inter-institutional integration, necessary to consolidate data and promote a holistic view of social demands, still faces challenges such as standardizing information and defining responsibilities among government agencies. Overcoming these challenges is crucial to maximize the effectiveness of public interventions.

Furthermore, the application of digital transformation in specific contexts reinforces the need to align technology and governance with strategic planning that prioritizes inclusion, sustainability, and data privacy protection. However, the studies leave open questions about how to ensure the financial and operational sustainability of these solutions and how to balance the use of data with guarantees of security and ethical treatment. They also highlight gaps in post-implementation monitoring. Time is needed to assess the long-term impacts of these strategies derived from public policies. These studies emphasize that achieving modern and sustainable public governance requires the integration of technology, ethics, inter-institutional collaboration, and citizen participation. This integration is essential to ensure efficiency, equity, and significant social impact.

4. BEEP-DS

Given our experience in government Big Data projects, it was possible to observe and extract good practices to compose a methodology for Building Evidence-Based Public Policies from Data Science (BEEP-DS). The construction of the BEPP-DS methodology followed a set of guiding requirements, derived from the analysis of real cases and the limitations observed in previous projects. The main requirements include: (i) clarity and traceability of decisions; (ii) minimization of rework and data reprocessing; (iii) integration of different professional profiles with defined roles; (iv) feasibility of replication in different institutional contexts; and (v) support for impact assessment. Based on these requirements, it was possible to design the workflow represented in Figure 1.

Based on the experiences of the Federal University of Ceará and the resources provided by the Fundo Nacional de Desenvolvimento da Educação (FNDE), the data product creation process is structured into six main roles: the requester, the data scientist, the business analyst, the data engineer, the developer, and the data manager.

The requester is the FNDE representative responsible for submitting a data product request. They present a strategic question already approved by FNDE, meaning that the business need has been previously analyzed

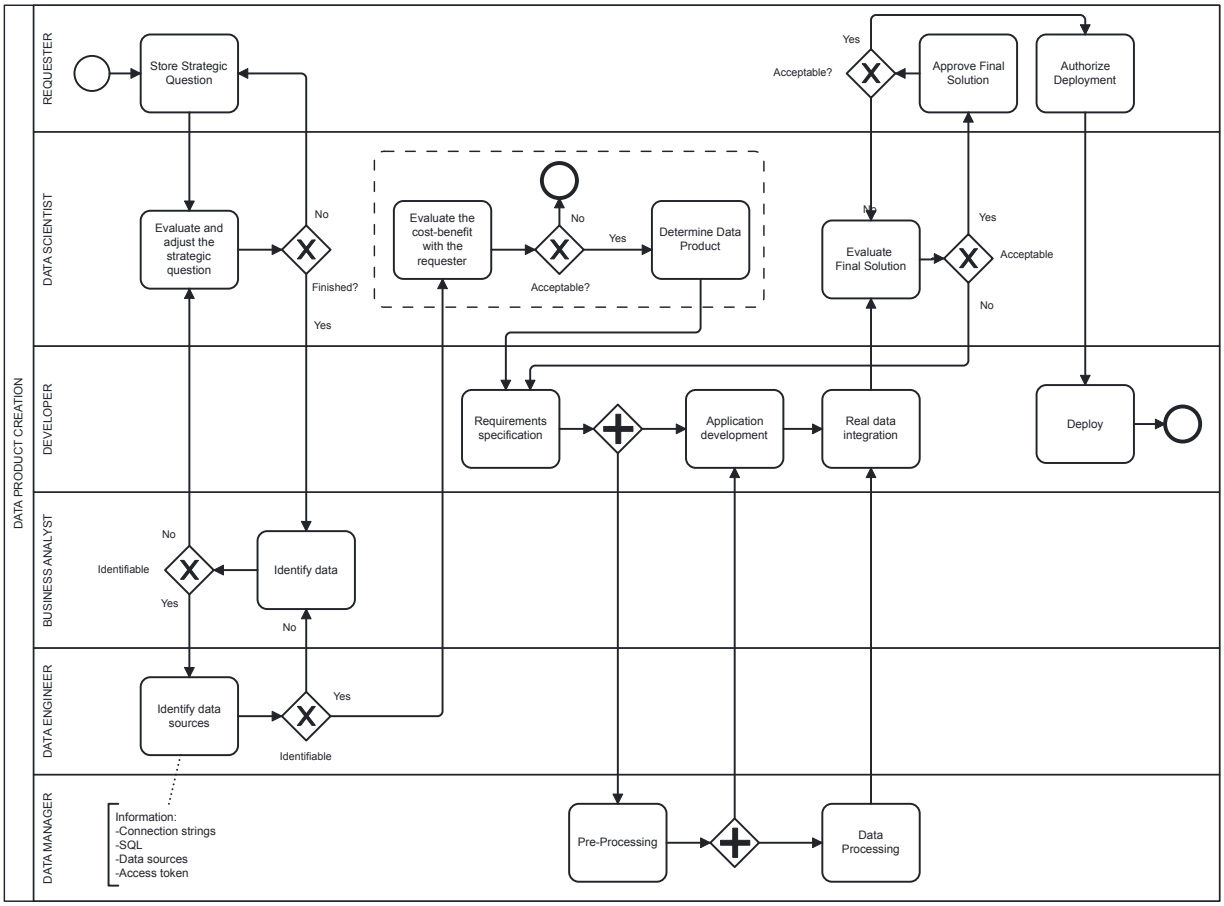

Fig. 1 – BPM Diagram for data product creation.

before the process has been started. Consequently, the submitted question would be structured and ready to be addressed.

The data scientist, in charge of the next step, is an information technology professional experienced in data analysis. Their role is to assess whether the strategic question meets essential criteria such as objectivity, well-defined scope, and feasibility within a reasonable timeframe. If necessary, the data scientist collaborates with the requester to refine the question. The data scientist is also supposed to accompany the whole data product creation process, trying to keep the efforts aligned with the goals of the requester. Once the question is sufficiently clear and viable, it is passed on to the business analyst. At this point, the process aligns with Business Understanding in CRISP-DM (Figure 2), where the objectives are identified, and the success criteria are established. Although the analysis of predictions is carried out by the data scientist, it will be clear that decisions about the development of data products for public policies require a broader vision. The technique envisaged, although essential, may not fully capture the social or strategic value of an initiative. Therefore, it is proposed that this stage evolve into a collegiate assessment, also involving experts from the business area and representatives of public management, allowing for the consideration of not only computational or technical effort aspects, but also the potential impact of the policy to be strengthened — for example, in terms of urgency, population coverage or mitigation of social risks.

The business analyst identifies data sources needed to address the strategic question, leveraging institutional knowledge. If data is unavailable or costly, they return the request with feedback; otherwise, they forward it (with documentation) to the data engineer. This aligns with CRISP-DM's Data Understanding phase. The data engineer handles technical infrastructure (connection strings, queries, access tokens). If obstacles arise (e.g., external data ownership), they escalate to the analyst and data scientist. Successfully sourced data is documented, mirroring CRISP-DM's Data Preparation.

The data scientist then evaluates feasibility (strategic fit, technical viability, cost-benefit) with the requester. If

Fig. 2 – Phases of the CRISP-DM Process Model for Data Mining

approved, they jointly define the product (dashboard, chart, etc.), initiating development (Modeling in CRISP-DM). The developer builds the product using a sample dataset (provided by the data manager), while the manager prepares the final dataset. The data scientist validates the product with real data, requesting adjustments if needed (Evaluation phase).

Finally, the requester reviews the product. If approved, the team deploys it, and the data manager catalogs it for use (Deployment in CRISP-DM). Rejection triggers reevaluation by the data scientist.

5. Usage Scenarios

This section describes two projects in which we obtained the necessary experiences to propose the method discussed in this paper. Big Data Social and Big Data Fortaleza are platforms built to help public managers make decisions based on analytical data. Although created in different scenarios, both faced challenges, such as the lack of a data culture in government sectors and the absence of processes to ensure the continuity of the platforms based on the management of the construction of new analytics.

5.1. Big Data Social

Big Data Social platform is a centralized hub for consolidating, managing, and analyzing diverse government databases related to low-income families enrolled in the “Mais Infância” welfare program. By integrating multiple sources, the system improves governance by generating key indicators and highlighting the social policies that directly benefit these families. These households are a priority for state agencies, as they often face significant socioeconomic hardships, predominantly living in rural areas and including one child under six.

To build the database, a state-led survey was conducted, gathering data from more than 150,000 families on topics such as food security, health, child welfare, housing conditions, income, and education. The Big Data Social platform not only stores this vast dataset but also provides advanced tools for data analysis, including analytics, data mining, and interactive dashboards. The platform maps the geographic locations of these families, providing decision-makers with valuable insights into how current policies impact families’ quality of life.

The execution of this project gave us experience with the challenges inherent in constructing evidence-based public policies. Although several data science processes can be applied to address such challenges, the lessons learned in this type of project are valuable for avoiding future mistakes and optimizing the results achieved.

5.2. Big Data Fortaleza

The Big Data Fortaleza platform was designed to create a robust data ecosystem that supports evidence-based public policy formulation, implementation, and monitoring. By centralizing and processing large volumes of

data from diverse governmental and institutional sources, the platform enables in-depth analysis and pattern recognition, optimizing public management. Its methodology is grounded in Data Science principles tailored for the public sector, with a strong emphasis on the BEPP-DS framework.

Aligned with e-Government initiatives and international data-driven policy standards, the platform adopts a modular approach encompassing data collection, preprocessing, exploratory analysis, predictive modeling, and interactive visualization. Machine learning and artificial intelligence enhance its capabilities, allowing for scenario forecasting and strategic recommendations that improve resource allocation and impact assessment. Built with interoperability and data governance best practices, the platform facilitates seamless integration across agencies, preventing data silos and promoting collaboration.

Validation across key public administration areas—such as urban mobility, health, education, and security—demonstrated the platform's effectiveness in integrating and cross-referencing previously fragmented databases, offering a comprehensive view of urban and social challenges. However, implementation faced hurdles, including institutional resistance to data technologies, the need for technical upskilling among public managers, and constraints in data quality and availability. To overcome these challenges, a participatory approach was adopted, engaging stakeholders and fostering a data-driven governance culture. The application of BEPP-DS proved instrumental in standardizing processes for extracting actionable insights from heterogeneous data sources, ensuring reliable and replicable outcomes.

The platform's deployment significantly enhanced municipal analytical capabilities, leading to optimized resource allocation, reduced administrative redundancies, and stronger policy monitoring and evaluation mechanisms. Interactive dashboards and automated reports improved data accessibility for both public managers and civil society, while increased transparency allowed citizens to track government actions more closely, promoting accountability. These advancements reinforced the shift toward responsive, data-informed governance.

The insights derived from Big Data Fortaleza contributed to refining the BEPP-DS methodology, establishing best practices for data science in public policy. The project highlighted the importance of robust data governance, including standardized processes, clear institutional responsibilities, and continuous data validation. As the methodology evolved, the platform incorporated advanced predictive models and expanded interoperability with national and international databases. Equally critical was the emphasis on continuous training for public officials, ensuring effective tool utilization to drive tangible societal benefits.

Looking ahead, Big Data Fortaleza aims to broaden its impact by leveraging open data and artificial intelligence to refine public policy formulation on a national scale, further cementing its role as a catalyst for data-driven governance.

6. Discussion

The BEPP-DS methodology was designed to address gaps observed in applying traditional data science frameworks to public policy contexts. While CRISP-DM is widely recognized for its six-phase structure—from business understanding to deployment—its direct application in government settings can be challenging due to insufficient emphasis on public-sector specifics, such as institutional complexity and the need for transparency and accountability. In comparison, BEPP-DS adapts and expands CRISP-DM principles by incorporating stages tailored to public-sector realities. For instance, while CRISP-DM focuses on business understanding, BEPP-DS emphasizes strategic questions aligned with political and social priorities. Additionally, BEPP-DS introduces a data governance framework addressing privacy, ethics, and responsible data use—aspects often overlooked in more generalized frameworks.

Although it has been applied in projects like Big Data Social and Big Data Fortaleza, a systematic evaluation of its effectiveness and adaptability across different government contexts is still needed. This includes further case studies, comparative analyses with other methodologies, and performance metrics to demonstrate its benefits and limitations. Moreover, implementing BEPP-DS requires significant institutional maturity and technical capacity. Public organizations with less developed structures may struggle with full adoption. Therefore, adaptations and ongoing support are crucial to ensure long-term effectiveness and sustainability.

BEPP-DS represents progress in integrating data science into public policy, offering a more contextualized and

sensitive approach to public-sector needs. However, its consolidation as a standard methodology will depend on further validation, adaptations to diverse institutional realities, and strengthening the technical capabilities of involved organizations.

7. Lessons Learned

Based on the BEPP-DS and the usage scenarios described in section 5, it was possible to systematize a set of lessons learned. This section discusses these lessons using the following structure: Each lesson has a title, a description of the context, the challenge faced, and the actions taken to overcome it.

Identifying the Real Problems of Each Institution: Understanding the unique challenges institutions face is crucial to implementing data-driven solutions effectively. In many cases, institutions can not clearly understand their operational hurdles, leading to poorly defined goals and misallocated resources. To address this, it is recommended that in-depth diagnostic assessments be conducted before initiating any project. These assessments should focus on understanding institutional workflows, key performance indicators, and stakeholder expectations.

Collaboratively Constructing Strategic Questions to Be Addressed: One of the most significant challenges in evidence-based policy-making is defining actionable and strategic questions. These questions must be relevant, measurable, and aligned with the institution's goals. A collaborative approach involving all stakeholders, including policymakers, data scientists, and end-users, ensures that the questions address real needs. Facilitation workshops and iterative refinement processes are recommended to align these diverse perspectives.

Establishing a Pilot to Validate the Methodology Quickly: Large-scale projects often face risks associated with untested methodologies. Pilots serve as controlled environments to test assumptions, refine processes, and gather feedback. By implementing a pilot phase, organizations can identify potential bottlenecks or risks. It is advised to select a small, representative sample for the pilot to maximize its relevance and scalability.

Systematizing the Recording of New Strategic Questions: In dynamic environments, new challenges and opportunities continuously emerge. Without a systematic approach, these new strategic needs might go unnoticed or be poorly documented. Implementing a structured repository or a formal process for capturing and reviewing new strategic questions ensures that the organization can adapt its policies and strategies effectively over time.

Creating Automated Flows to Keep Data Products Updated: Data products often lose relevance if not updated regularly. Many institutions rely on manual processes, which are prone to delays and errors. Automating data Extraction, Transformation, and Loading (ETL) processes ensures that data products remain current and accurate. Institutions should invest in robust automation tools and ensure adequate training for their technical teams.

Promoting a Data Culture Within Institutions: Adopting data-driven methodologies requires more than technical tools; it demands a cultural shift. Many public institutions encounter resistance to change due to their lack of familiarity with data practices. Building a data culture involves continuous education, transparent communication, and promoting data-driven decision-making at all organizational levels.

Establishing Standards for System Integration: Fragmented systems often lead to inefficiencies and missed opportunities for holistic insights. Integration challenges are particularly acute in multi-agency environments where data interoperability is essential. Developing and enforcing standards for system integration can address these challenges. Such standards should include data formats, APIs, and security protocols to ensure seamless collaboration.

Defining Data Architectures Independent of Business Domains: Rigid architectures tied to specific business domains often limit scalability and adaptability. Designing modular and domain-agnostic data architectures enables institutions to reuse data and infrastructure for various applications. This approach facilitates integration with external systems and future-proofing against evolving organizational needs.

8. Final Remarks

The BEPP-DS methodology introduced in this paper integrates Data Science into creating Evidence-Based Public Policies, focusing on transparency, scalability, and efficiency. Drawing from real-world initiatives like Big Data Social and Big Data Fortaleza, the methodology tackles the distinct challenges of implementing data-driven solutions within the public sector. It provides a structured framework that guides every stage of developing data products—from identifying key strategic questions to delivering actionable insights—helping improve decision-making in governance.

The study's findings underscore the importance of collaborative problem definition, systematic testing through pilot projects, and embedding data-driven practices within institutions. Applying the BEPP-DS methodology demonstrated its effectiveness in overcoming common challenges public organizations face, such as fragmented data, limited data culture, and a lack of automated processes. It also enabled policymakers to prioritize and implement targeted interventions, promoting evidence-based solutions to address critical societal issues.

Future work should aim to enhance the BEPP-DS methodology by incorporating Artificial Intelligence and Machine Learning for predictive analytics and real-time decision-making. Another valuable direction is developing approaches to increase citizen participation, enabling stakeholders to contribute to data collection, validation, and policy evaluation. Additionally, efforts must address ethical and privacy concerns to ensure that data-driven governance adheres to fairness, accountability, and inclusive principles.

Acknowledgement

- **Funding or Grant**: This work was supported by the National Fund for Educational Development (FNDE) in partnership with the Federal University of Ceará (UFC).
- **Data/Software Access Statement**: The datasets used in this study were derived from government sources and are not publicly available due to privacy and confidentiality concerns. However, processed results and methodological insights can be requested from the corresponding author.
- **Contributor Statement**: Conceptualization and Methodology: José Antônio F. Macedo, Rossana Maria de Castro Andrade; Data Curation and Analysis: Regis P. Magalhães, Lívia A. Cruz, José Florêncio Q. Neto, Samir B. Chaves; Software Development: Davyson S. Ribeiro, Paulo V. A. Fabrício; Writing - Original Draft: Pedro Almir M. Oliveira, José Florêncio Q. Neto, Samir B. Chaves, Davyson S. Ribeiro; Writing - Review and Editing: Mauricio Feijo B. M. Filho, Joaquim José Escola, Amanda Sousa, Paulo V. A. Fabricio.
- **Use of AI**: During the preparation of this work, the authors used ChatGPT and Copilot to assist in drafting and refining some textual sections. After using these tools, the authors reviewed, edited, and validated the content to ensure accuracy and originality, taking full responsibility for the final publication.
- **Conflict Of Interest (COI)**: There is no conflict of interest to declare.

References

Anderson, L. M., Brownson, R. C., Fullilove, M. T., Teutsch, S. M., Novick, L. F., Fielding, J., & Land, G. H. (2005). Evidence-based public health policy and practice: Promises and limits. *American journal of preventive medicine*, *28*(5), 226–230.

Batista, É., Andrade, R. M., Santos, I. S., Nogueira, T. P., Oliveira, P. A., Lelli, V., & Oliveira, V. T. (2024). Fortaleza city hall strategic planning based on data analysis and forecasting. *Congresso Ibero-Americano em Engenharia de Software (CIbSE)*, 433–436.

Furtado, L. S., Moura, G., Vasconcelos, D. J., Fernandes, G. S., Cruz, L. A., Magalhães, R. P., & Coelho da Silva, T. L. (2023). An analytical citizen relation management system (czrm) for social vulnerability mapping and policy recommendation in brazil. *Decision Support Systems*, *172*, 113995. DOI: https://doi.org/https://doi.org/10.1016/j.dss.2023.113995.

MacArthur, B. D., Dorobantu, C. L., & Margetts, H. Z. (2022). Resilient government requires data science reform. *Nature Human Behaviour*, *6*(8), 1035–1037.

Musafir, V. E. N. (2018). Brazilian e-government policy and implementation. *International E-Government Development: Policy, Implementation and Best Practice*, 155–186.

Pereira, G., Monteiro, I., Vasconcelos, D., Braz, L., & Silva, C. (2021). Classificação taxonômica de categorias de serviços públicos para aplicações digitais. *Anais do IX Workshop de Computação Aplicada em Governo Eletrônico*, 119–130. DOI: https://doi.org/10.5753/wcge.2021.15982.

Rahul, K., & Banyal, R. K. (2020). Data life cycle management in big data analytics. *Procedia Computer Science*, *173*, 364–371.

Saltelli, A., Bammer, G., Bruno, I., Charters, E., Di Fiore, M., Didier, E., Nelson Espeland, W., Kay, J., Lo Piano, S., Mayo, D., Pielke Jr, R., Portaluri, T., Porter, T. M., Puy, A., Rafols, I., Ravetz, J. R., Reinert, E., Sarewitz, D., Stark, P. B., … Vineis, P. (2020). Five ways to ensure that models serve society: A manifesto. *Nature*, *582*(7813), 482–484. DOI: https://doi.org/10.1038/d41586-020-01812-9.

Santos, I. S., Oliveira, P. A. M., Oliveira, V. T., Nogueira, T. P., Dantas, A. B. O., Menescal, L. M., Batista, É., & Andrade, R. M. C. (2023). Big Data Fortaleza: Plataforma Inteligente para Políticas Públicas Baseadas em Evidências. *Workshop de Computação Aplicada em Governo Eletrônico (WCGE)*, 200–211. DOI: https://doi.org/10.5753/wcge.2023.230796.

Schröer, C., Kruse, F., & Gómez, J. M. (2021). A systematic literature review on applying crisp-dm process model [CENTERIS 2020 - International Conference on ENTERprise Information Systems / ProjMAN 2020 - International Conference on Project MANagement / HCist 2020 - International Conference on Health and Social Care Information Systems and Technologies 2020, CENTERIS/ProjMAN/HCist 2020]. *Procedia Computer Science*, *181*, 526–534. DOI: https://doi.org/https://doi.org/10.1016/j.procs.2021.01.199.

Silcock, R. (2001). What is e-government. *Parliamentary affairs*, *54*(1), 88–101.

Silva, W. C. P., Macedo, J. A. F. D., & De Queiroz Neto, J. F. (2022). Usando um modelo de classificação para a adequada implantação do patrulhamento policial para o enfrentamento à assaltos a bancos no nordeste do Brasil. *Revista Brasileira de Ciências Policiais*, *13*(9), 185–205. DOI: https://doi.org/10.31412/rbcp.v13i9.845.

Sucupira Furtado, L., Da Silva, T. L. C., Ferreira, M. G. F., De Macedo, J. A. F., & De Melo Lima Cavalcanti Moreira, J. K. (2023a). A framework for Digital Transformation towards Smart Governance: Using big data tools to target SDGs in Ceará, Brazil. *Journal of Urban Management*, *12*(1), 74–87. DOI: https://doi.org/10.1016/j.jum.2023.01.003.

Sucupira Furtado, L., da Silva, T. L. C., Ferreira, M. G. F., de Macedo, J. A. F., & de Melo Lima Cavalcanti Moreira, J. K. (2023b). A framework for digital transformation towards smart governance: Using big data tools to target sdgs in ceará, brazil [Digital Technologies in Urban Planning and Urban Management]. *Journal of Urban Management*, *12*(1), 74–87. DOI: https://doi.org/https://doi.org/10.1016/j.jum.2023.01.003.

Van Der Aalst, W., & van der Aalst, W. (2016). *Data science in action*. Springer.