UNIVERSIDADE FEDERAL DE SERGIPE
CENTRO DE CIÊNCIAS EXATAS E TECNOLOGIA
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO

# Quality Concerns in Public Procurement Data

Dissertation Work

Fernando Henrique Moura de Oliveira

São Cristóvão – Sergipe

2024

UNIVERSIDADE FEDERAL DE SERGIPE
CENTRO DE CIÊNCIAS EXATAS E TECNOLOGIA
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO

Fernando Henrique Moura de Oliveira

# Quality Concerns in Public Procurement Data

Master's Dissertation submitted to the Graduate Program in Computer Science of the Federal University of Sergipe as a partial requirement for obtaining the Master's degree in Ciência da Computação.

Orientador: Dr. Glauco de Figueiredo Carneiro

São Cristóvão – Sergipe
2024

**FICHA CATALOGRÁFICA ELABORADA PELA BIBLIOTECA CENTRAL**
**UNIVERSIDADE FEDERAL DE SERGIPE**

O48q

Oliveira, Fernando Henrique Moura de
Quality concerns in public procurement data / Fernando Henrique Moura de Oliveira ; orientador Glauco de Figueiredo Carneiro . - São Cristóvão, 2024.
98 f.; il.

Dissertação (mestrado em Ciência da Computação) – Universidade Federal de Sergipe, 2024.

1. Compras (Serviço público). 2. Informações eletrônicas governamentais. 3. Inteligência arficial. 4. ChatGPT. I. Carneiro, Glauco de Figueiredo orient. II. Título.

CDU 004.8

**Ata da Sessão Solene de Defesa da Dissertação do Curso de Mestrado em Ciência da Computação-UFS. Candidato:Fernando Henrique Moura de Oliveira**

Em 09 dias do mês de dezembro do ano de dois mil e vinte quatro, com início às 14hs, realizou-se na Sala de Seminários do PROCC da Universidade Federal de Sergipe, na Cidade Universitária Prof. José Aloísio de Campos, a Sessão Pública de Defesa de Dissertação de Mestrado do candidato **Fernando Henrique Moura de Oliveira**, que desenvolveu o trabalho intitulado: **" Quality Concerns in Public Procurement Data",** sob a orientação do Prof. Dr. **Glauco de Figueiredo Carneiro.** A Sessão foi presidida pelo Prof. Dr. **Glauco de Figueiredo Carneiro** (PROCC/UFS), que após a apresentação da dissertação passou a palavra aos outros membros da Banca Examinadora, Prof. Dr. **Flávio Dusse** (UCSAL) e, em seguida, o Prof. Dr. **Paulo Salgado Gomes de Mattos Neto** (Procc/UFS) e o Prof. Dr. **Edward David Moreno Ordonez** (Procc/UFS). Após as discussões, a Banca Examinadora reuniu-se e considerou o mestrando **APROVADO**. Atendidas as exigências da Instrução Normativa 05/2019/PROCC, do Regimento Interno do PROCC (Resolução 67/2014/CONEPE), e da Resolução nº 04/2021/CONEPE que regulamentam a Apresentação e Defesa de Dissertação, e nada mais havendo a tratar, a Banca Examinadora elaborou esta Ata que será assinada pelos seus membros e pelo mestrando.

Cidade Universitária "*Prof. José Aloísio de Campos*", 09 de dezembro de 2024.

gov.br Documento assinado digitalmente
GLAUCO DE FIGUEIREDO CARNEIRO
Data: 11/12/2024 11:42:33-0300
Verifique em https://validar.iti.gov.br

**Prof. Dr. Glauco de Figueiredo Carneiro**
**(PROCC/UFS)**
**Presidente**

gov.br Documento assinado digitalmente
FLAVIO DUSSE
Data: 09/12/2024 17:16:36-0300
Verifique em https://validar.iti.gov.br

**Prof. Dr Flávio Dusse**
**(UCSAL)**
**Examinador Externo**

gov.br Documento assinado digitalmente
PAULO SALGADO GOMES DE MATTOS NETO
Data: 11/12/2024 16:14:55-0300
Verifique em https://validar.iti.gov.br

**Prof. Dr. Paulo Salgado Gomes de Mattos Neto**
**(PROCC/UFS)**
**Examinador Interno**

gov.br Documento assinado digitalmente
EDWARD DAVID MORENO ORDONEZ
Data: 10/12/2024 17:38:50-0300
Verifique em https://validar.iti.gov.br

**Prof. Dr. Edward David Moreno Ordonez**
**(PROCC/UFS)**
**Examinador Interno**

gov.br Documento assinado digitalmente
FERNANDO HENRIQUE MOURA DE OLIVEIRA
Data: 10/12/2024 18:20:26-0300
Verifique em https://validar.iti.gov.br

**Fernando Henrique Moura de Oliveira**
**Candidato**

# Acknowledgements

This dissertation is dedicated to my parents, Armando and Expedita, whose education has been the cornerstone of my life. I am deeply grateful to my beloved wife, Livia, for her unwavering support and companionship through all times, and to my dear daughter, for her loving affection. I thank God, who has made the realization of this dream possible. I acknowledge Professor Glauco for his guidance and all the other professors from PROCC whom I had the honor of being a student; here, my gratitude is particularly extended to Professor Leonardo.

# Abstract

Dados de compras governamentais, apesar de seu potencial para melhorar a transparência e a tomada de decisão, enfrentam desafios significativos relacionados à qualidade, incluindo descoberta, acessibilidade e usabilidade. Essas questões impedem uma análise efetiva e a adoção de estratégias orientadas por dados nos processos de contratação. Este estudo busca preencher essas lacunas ao caracterizar os principais desafios na qualidade dos dados de compras governamentais e explorar o papel dos Grandes Modelos de Linguagem ou *Large Language Models* (LLMs), como o ChatGPT, em superá-los.

A pesquisa utiliza um protocolo duplo: um estudo de mapeamento sistemático para identificar e classificar desafios e soluções existentes, e um protocolo experimental exploratório que aproveita os LLMs para avaliar sua capacidade de melhorar a descoberta, acessibilidade e usabilidade dos dados. Por meio de prompts personalizados e modelagem de interação, o estudo demonstra como os LLMs podem auxiliar usuários, como fornecedores governamentais, a navegar em conjuntos de dados de contratação complexos, traduzir jargão técnico e melhorar a descoberta de dados.

Descobertas importantes incluem a identificação das principais lacunas de qualidade nos dados de compras governamentais e a validação dos LLMs como ferramentas eficazes para enfrentar esses desafios. As contribuições desta pesquisa são o desenvolvimento de uma taxonomia para a qualidade dos dados de compras governamentais, nas aplicações de LLMs neste contexto e nas recomendações para integrar ferramentas avançadas de IA nos fluxos de trabalho do setor público. Essas percepções abrem caminho para futuros estudos que visem otimizar ainda mais os processos de compras governamentais por meio de soluções impulsionadas por IA.

**Palavras-chave:** Dados abertos, Dados Abertos Governamentais, Compras governamentais, Inteligência Artificial, Grandes Modelos de Linguagem, ChatGPT.

# Abstract

Public procurement data, despite its potential to enhance transparency and decision-making, faces significant challenges related to quality, including discoverability, accessibility, and usability. These issues hinder effective analysis and the adoption of data-driven strategies in procurement processes. This study seeks to address these gaps by characterizing the primary challenges in public procurement data quality and exploring the role of Large Language Models (LLMs), such as ChatGPT, in overcoming them.

The research employs a twofold protocol: a systematic mapping study to identify and classify existing challenges and solutions, and an exploratory experimental protocol leveraging LLMs to assess their capacity to improve data usability and accessibility. Through tailored prompts and interaction modeling, the study demonstrates how LLMs can assist users, such as government suppliers, in navigating complex procurement datasets, translating technical jargon, and enhancing data discoverability.

Key findings include the identification of major quality gaps in procurement data and the validation of LLMs as effective tools for addressing these challenges. The contributions of this research lie in the development of a taxonomy for public procurement data quality, the demonstration of LLM applications in this context, and recommendations for integrating advanced AI tools into public sector workflows. These insights pave the way for future studies to further optimize procurement processes through AI-driven solutions.

**Keywords:** Open data, Open Government Data, Public procurement, Artificial Intelligence, Large Language Models, ChatGPT.

# List of Figures

# List of Tables

# List of abbreviations and acronyms

| | |
|---|---|
| AI | Artificial Intelligence |
| ACM | Association for Computing Machinery |
| API | Application Programming Interface |
| BIM | Building Information Modeling |
| BERT | Bidirectional Encoder Representations from Transformers |
| CKAN | Comprehensive Knowledge Archive Network |
| CoT | Chain of Thought |
| CMS | Content Management System |
| CSV | Comma-Separated Values |
| DAOs | Decentralized Autonomous Organizations |
| DMS | Document Management System |
| DKAN | Drupal-based Open Data Platform |
| EU | European Union |
| EN | European Norm |
| FAIR | Findable, Accessible, Interoperable, Reusable |
| FAPITEC | Fundação de Amparo à Pesquisa e Inovação Tecnológica (Foundation for Research and Technological Innovation) |
| GDP | Gross Domestic Product |
| GQM | Goal Question Metric |
| GPT | Generative Pre-trained Transformer |
| IEEE | Institute of Electrical and Electronics Engineers |
| ICT | Information and Communication Technologies |
| ISO | International Organization for Standardization |
| IT | Information Technology |
| LLM | Large Language Model |
| MEAT | Most Economically Advantageous Tender |
| NASA | National Aeronautics and Space Administration |
| NGO | Non-Governmental Organization |
| NRW | North Rhine-Westphalia |

| | |
|---|---|
| OGD | Open Government Data |
| OKF | Open Knowledge Foundation |
| OECD | Organisation for Economic Co-operation and Development |
| OPEN | Open, Public, Electronic and Necessary |
| PICO | Population, Intervention, Comparison, Outcome |
| PeP | Electronic Public Procurement Platform |
| PRISMA | Preferred Reporting Items for Systematic Reviews and Meta-Analyses |
| RBAC | Role-Based Access Control |
| REBUS-PLS | A software tool used for structural equation modeling |
| RQ | Research Question |
| SIASG | Sistema Integrado de Administração de Serviços Gerais (Integrated System for General Services Administration) |
| SME | Small and Medium-sized Enterprises |
| SRQ | Specific Research Question |
| TCU | Tribunal de Contas da União (Federal Court of Accounts) |
| TED | Tenders Electronic Daily |
| UFS | Universidade Federal de Sergipe (Federal University of Sergipe) |
| URL | Uniform Resource Locator |

# Contents

# 1
# Introduction

This chapter presents an overview of the key concepts and motivations behind a dissertation focused on public procurement data. The contextualization offers a concise history of open data and its global adoption, with particular attention to its implementation in the public sector. The research objectives are then outlined, defining the study's scope and contributions. The chapter concludes with a preview of the dissertation's structure, providing an overview of the content and focus of the subsequent chapters.

## 1.1 Contextualization

According to OpenDataSoft (OpenDataSoft, 2024), the concept of open data originated in 1995, initially focusing on geophysical and environmental data sharing. It was officially established in 2007 by internet activists in Sebastopol, California, who, inspired by the open source movement, gathered with the objective of advocating for U.S. federal legislation to foster open data initiatives[1]. This goal was achieved with the introduction of the Open Government Directive[2] in 2009, which required federal agencies to share their data openly through the DATA.GOV[3] portal. The initiative was further bolstered by the Open, Public, Electronic and Necessary (OPEN) Government Data Act[4], which was passed in 2019.

As detailed by (ALI; ALEXOPOULOS; CHARALABIDIS, 2022), the open data initiative consists of three main stages: preparation, launch, and extension/sustainment. The preparation phase concludes with the establishment of a data catalog or platform, overseen by data curators, which shares open data with users and the community. Various open data platforms are

[1] https://www.opendatasoft.com/en/blog/open-data-anniversary-ten-years-after-the-sebastopol-meeting/
[2] https://obamawhitehouse.archives.gov/open/documents/open-government-directive
[3] https://data.gov/
[4] https://www.govinfo.gov/content/pkg/PLAW-115publ435/html/PLAW-115publ435.htm

available to enhance the accessibility of open data catalogs and initiatives tailored for a diverse range of stakeholders. Governments, businesses, and NGOs utilize these platforms to create open data portals, which facilitate the sharing of public data.

The integration of open data with the principles of Open Government[5] led to the establishment of Open Government Data (OGD) initiatives(TANG; JIANG, 2020). According to (ABDELRAHMAN, 2021), OGD refers to information that government entities make freely accessible to the public. This data encompasses a wide array of datasets, including but not limited to financial records, demographic statistics, census reports, legislative proceedings, and environmental data collected by public organizations or agencies.

Open Government Data initiatives face significant barriers that hinder their advancement. As highlighted by (ABDELRAHMAN, 2021), a main challenge is the persistent culture of secrecy within governments and the absence of comprehensive open data policies, which makes data sharing difficult. Legal issues, such as privacy laws and conflicting data access regulations, further complicate efforts. Poor data quality, characterized by outdated or inaccurate information and insufficient metadata or search functionality, also impedes the effective use of data. Furthermore, there are accessibility issues, such as limited APIs and non-machine-readable formats in government repositories (CARDOSO; CARNEIRO; MENESES, 2021), and interoperability problems, including data granularity issues (CARNEIRO et al., 2021). Data reuse is constrained by partial restrictions, and disorganized data impedes efficient utilization (NUNES; MORENO; CARNEIRO, 2023).

Among the various types of Open Government Data available, public procurement data stand out. According to (RIBEIRO, 2018), public procurement can be defined as a formal process in which government organizations acquire goods and services. This process involves the specification of requirements, the selection of suppliers, the analysis of proposals, the drafting and awarding of contracts, the resolution of conflicts and complaints, and encompasses all phases of contractual management. Public procurement is relevant due to its role in fostering competitive bidding and promoting public accountability.

## 1.2 Motivation

According to the Organisation for Economic Co-operation and Development (OECD) report[6], public procurement plays a crucial role in influencing the quality of life of the population, a fact that was particularly evident during the COVID-19 crisis. The report indicates that public spending on procurement as a percentage of Gross Domestic Product (GDP) rose from 11.8% to 12.9% in 2021 among OECD member countries. Meanwhile, in Brazil, government investments in public procurement in that same year amounted to approximately 15% of GDP. In 2023,

[5] https://www.opengovpartnership.org/stories/how-about-defining-open-government-principles/
[6] https://doi.org/10.1787/ce2208f6-en

Brazil's public procurement reached around R$236 billion[7], representing a significant portion of the federal, state, and municipal government budgets.

The motivation for investigating public procurement data quality stems from its significant economic impact on society and the growing reliance on public data portals by governments to promote transparency, participatory governance, and accountability. Stakeholders such as citizens, journalists, and auditors demand accessible and reliable data to scrutinize government operations and improve public oversight.

## 1.3 Problem Statement

Public procurement encompasses legal frameworks, open data repositories, and critical information necessary for achieving fairness and efficiency in procurement processes. The integrity of the data involved is crucial to maintain transparency, accountability, and to prevent fraudulent activities, significantly impacting governments, suppliers, and the public. However, the poor quality of procurement data, particularly regarding discoverability, accessibility, and usability, impedes comprehensive analysis and informed decision-making, undermining the efficiency of procurement procedures (SOYLU et al., 2022a). This lack of quality is not just a matter of incomplete or inaccurate records; it extends to the very systems that are meant to provide access to this information, creating significant barriers for all stakeholders. Despite its critical importance, stakeholders face substantial challenges in maintaining data quality and performing effective analysis (RODRÍGUEZ et al., 2019). These challenges are multifaceted, including technological issues with data integration, inadequate documentation practices, complex regulatory environments, and data quality issues such as inconsistencies and inaccuracies.

Furthermore, the reliability of public procurement data is essential for civil society to monitor government expenditures. Nonetheless, this reliability is often compromised by incomplete datasets and a lack of standardization (MENDES; VOIGT, 2022a). These issues not only impede the efficiency of procurement procedures but also limit the capacity of stakeholders—including citizens, journalists, and suppliers—to perform their roles in scrutinizing government operations. The lack of data quality hinders the ability to hold governments accountable and ensure transparency in public spending.

## 1.4 Objective

The primary objective of this study is to systematically characterize the quality concerns associated with public procurement data, focusing on the dimensions of discoverability, accessibility, and usability. This characterization aims to provide a comprehensive understanding of the current state of public procurement data quality, which is essential for enhancing transparency,

[7] http://paineldecompras.economia.gov.br/processos-compra

accountability, and efficiency in government procurement processes. To achieve this objective, the study will address the following specific aims:

1. Assess public procurement data quality: Investigate the current methodologies utilized to evaluate data quality.

2. Identification of issues: The study will investigate the reported issues concerning discoverability, accessibility, and usability in the context of public procurement data. By examining practical challenges faced by stakeholders, the research aims to provide insights into specific areas that impact on transparency and efficiency.

3. Propose best practices for data improvement: Present practical recommendations and strategies to improve the discoverability, accessibility, and usability of public procurement data, thereby contributing to enhanced management practices and stakeholder engagement.

4. Contribute to improving data quality: The insights generated will be instrumental in informing policymakers, researchers, and practitioners, thereby promoting a more transparent, competitive, and accountable procurement environment.

The methodology outlined in Chapter 2 provides the structured framework for this investigation, combining theoretical and empirical approaches to address the objectives.

## 1.5 Topics Excluded from Analysis

This research on characterizing the quality concerns of public procurement data does not encompass the specific areas or subjects that are intentionally excluded from the research focus:

1. Detailed analysis of public procurement processes and policies beyond their impact on data quality. The focus of this study is on evaluating the quality of public procurement data, rather than conducting a comprehensive examination of the procurement processes themselves.

2. In-depth examination of procurement practices and data quality in specific sectors or regions not covered by the selected studies. The study provides a broad overview of the landscape, but does not explore the nuances of data quality concerns in every possible sector or geographic region.

3. Detailed technical implementation details of the recommended best practices for improving public procurement data quality. The study outlines high-level strategies and recommendations, but does not provide step-by-step guidance for their technical implementation.

4. Longitudinal studies tracking the evolution of public procurement data quality concerns over time. The current study is focused on the current state of research, without a specific emphasis on the temporal dynamics of data quality issues.

By clearly delineating these aspects as out of scope, the study maintains a focused approach on characterizing the quality concerns of public procurement data, without attempting to address every possible dimension related to this topic.

## 1.6 Contributions

This dissertation makes significant contributions by systematically characterizing the quality concerns associated with public procurement data, focusing on discoverability, accessibility, and usability. It evaluates various methodologies for assessing data quality, identifying gaps in current research and best practices through a systematic mapping. The study highlights the potential of Large Language Models (LLMs), such as ChatGPT, to enhance data clarity and relevance, thereby supporting suppliers in navigating complex procurement processes. Additionally, it presents a user-centered interaction model to guide stakeholders in effectively engaging with procurement data.

## 1.7 Funding Acknowledgment

This research was supported by a grant from the Fundação de Apoio à Pesquisa e à Inovação Tecnológica do Estado de Sergipe (FAPITEC), under the reference number 019203.05537/2023-3.

## 1.8 Dissertation Structure

This dissertation is organized into six chapters. The introductory chapter (Chapter 1) outlines the motivations driving this research and offers a concise summary of the study's objectives. Chapter 2 outlines the methodology, which combines a systematic mapping of the literature and an exploratory study leveraging Large Language Models (LLMs) to address research questions related to public procurement data quality. Chapter 3 offers an overview of the foundational areas and key concepts that are pivotal to this dissertation, detailing the fields of Open Government Data (OGD), public procurement data and Large Language Models (LLMs). Chapter 4 provides an in-depth exploration of the quality issues pertaining to public procurement data, with a particular emphasis on key dimensions including discoverability, accessibility, and usability. Building upon the insights from Chapter 4, Chapter 5 presents an exploratory study that investigates the utilization of ChatGPT to enhance the quality of public procurement data. The concluding remarks are presented in Chapter 6, where the main contributions of this dissertation are discussed, and directions for potential future research are outlined. Figure 1 illustrates the structure of our dissertation, designed to assist readers in easily navigating through its content.

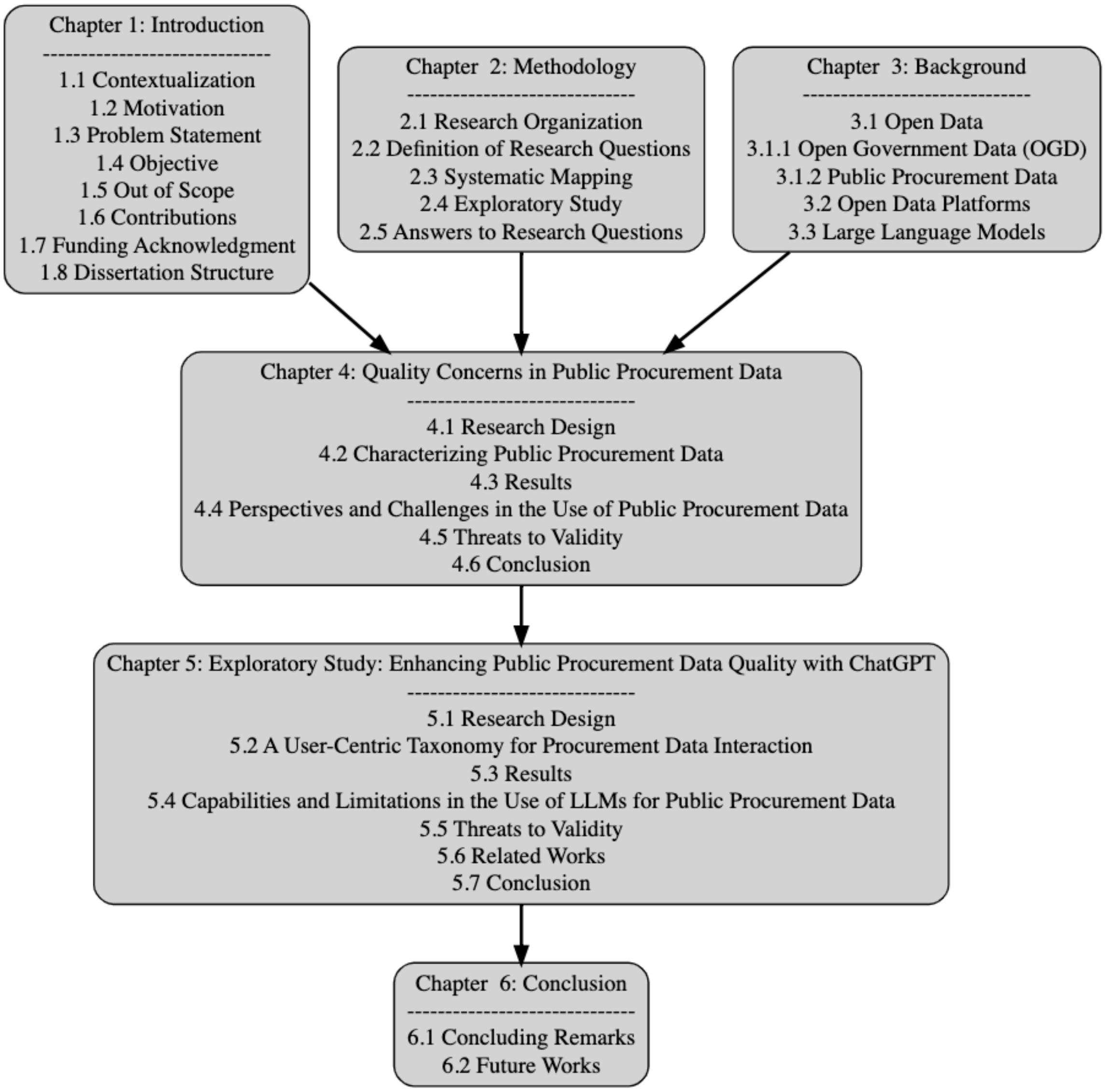


Figure 1 – Structure of the Dissertation: An Overview of Key Chapters and Their Relationships from Introduction to Conclusion and Future Work.

# 2
# Methodology

This chapter outlines the methodological approach adopted in this dissertation, emphasizing the structured and rigorous path followed to investigate and answer the research questions. The overall methodology is visually represented in Figure 2, which encapsulates the step-by-step process from initial research organization, through systematic literature review and an exploratory study, to the concluding answers. Each component of the process plays a crucial role in ensuring a systematic and rigorous exploration of public procurement data concerns, contributing to the findings and contributions of this study. The methodology combines both qualitative and quantitative techniques to provide a comprehensive understanding of public procurement data quality, addressing not only the theoretical aspects but also the practical applications of the research. This approach ensures that the study is well-grounded in the existing literature while also offering innovative solutions to address the identified challenges.

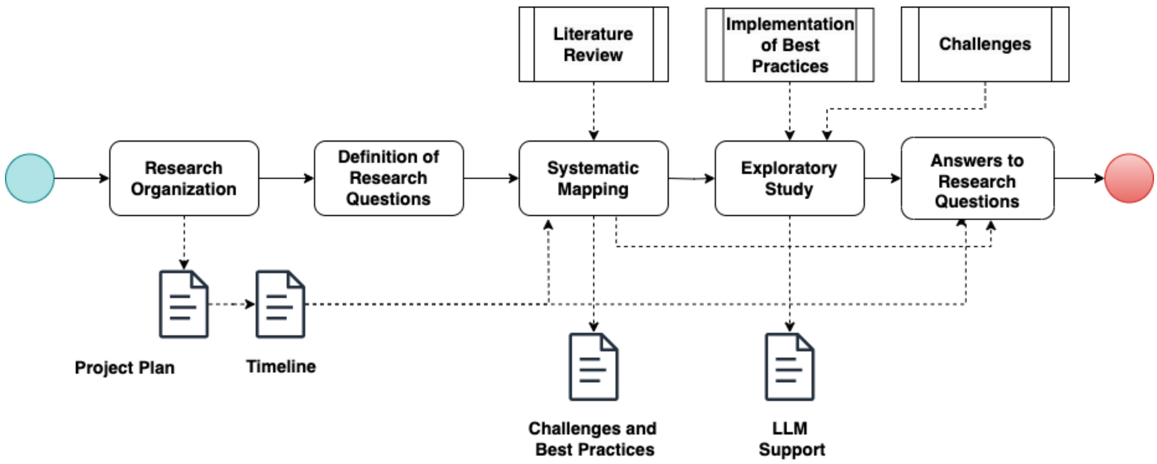


Figure 2 – Phases of the Research Project: Visual Representation of Research Organization, Systematic Mapping, Exploratory Study, and Analysis of Research Questions.

## 2.1 Research Organization

The research employs a twofold protocol: a systematic mapping study to identify and classify existing challenges and solutions, and an exploratory experimental protocol leveraging Large Language Models (LLMs) to assess their capacity to improve data discoberability, usability and accessibility.

This phase is responsible for meticulously planning and organizing all activities necessary for conducting the research. This initial stage is crucial for establishing a clear roadmap for the entire study, ensuring that all subsequent steps are aligned with the overall objectives. During this phase, the structure of the project plan, the schedule of activities, and the relevant research questions were defined to align with the project's objectives. This included setting timelines for each phase, allocating necessary resources, and determining the scope of the investigation to ensure a focused and effective research process.

The systematic mapping, detailed in Section 2.3, involves a comprehensive literature review to identify key concepts, highlight gaps, and establish a theoretical framework for the research. This phase is crucial for identifying challenges and best practices in public procurement data quality. The exploratory study, further detailed in Section 2.4, is designed to offer a real-world understanding of the dynamics at play in public procurement data management and serves as a testing ground for LLM applications. The study uses tailored prompts and interaction modeling to demonstrate how LLMs can assist users in navigating complex procurement datasets, translating technical jargon, and enhancing data discoverability.

## 2.2 Definition of Research Questions

Once the project is organized, the next step is the Definition of Research Questions. This stage is pivotal, as research questions are central to guiding the study, as they help direct the focus toward specific, well-defined issues or knowledge gaps related to public procurement data. The questions posed are informed by preliminary readings and initial assessments of the problem domain, ensuring relevance to both academic literature and practical concerns in public procurement. These questions also frame the subsequent stages of the research, ensuring that each phase contributes toward answering them. The research questions are carefully formulated to address the main challenges in public procurement data quality, focusing on discoverability, accessibility, and usability and are designed to be both academically rigorous and practically relevant, providing clear direction for the entire research process.

## 2.3 Systematic Mapping

The third component, Systematic Mapping, serves as the foundational research phase, in which a comprehensive literature review is conducted. This involves identifying, selecting, and reviewing existing research related to public procurement. A systematic approach ensures that all relevant literature is considered, providing a complete view of the current state of knowledge. This step is essential in identifying Challenges and Best Practices from the literature, which will inform both the exploratory study and the final recommendations. The Literature Review within this phase helps define key concepts, highlight gaps, and establish a theoretical grounding for the research. This phase follows a structured process including planning, execution, and reporting, as outlined by Kitchenham and Charters (KITCHENHAM; CHARTERS, 2007), to ensure a thorough and unbiased review of the existing literature. This systematic approach includes developing a detailed protocol for search terms, defining inclusion and exclusion criteria, and meticulously analyzing the selected studies to extract relevant data and insights.

## 2.4 Exploratory Study

Building on the insights from the systematic mapping, the Exploratory Study is conducted to apply the identified best practices related to public procurement data. The exploratory study phase focuses on the Implementation of Best Practices—applying theoretical frameworks and methodologies to real-world procurement data to assess their validity and effectiveness. This involves putting into practice the knowledge gained from the literature review to see how well these best practices perform in a practical setting. This phase also involves identifying any Challenges encountered during the application of these practices, whether related to data quality, accessibility, or technological constraints. These challenges are documented to provide a realistic view of the practical hurdles involved in improving public procurement data.

The exploratory study provides valuable practical insights and is designed to offer a real-world understanding of the dynamics at play in public procurement data management. It also serves as a testing ground for LLM applications, examining how these AI-driven models can enhance procurement data usability and transparency. Specifically, this study evaluates how Large Language Models (LLMs), such as ChatGPT, can aid in overcoming the challenges identified in the systematic mapping phase, particularly in the areas of data discoverability, accessibility, and usability. This phase helps gather key evidence needed to address the research questions and informs the final discussion.

## 2.5 Answers to Research Questions

Finally, the research culminates in providing Answers to Research Questions. Drawing on the literature review, identified challenges, and exploratory study findings, this phase presents evidence-based responses to the initial research questions posed in Section 2.2. This critical step synthesizes the findings from all previous phases, including both the systematic mapping and exploratory study to provide robust answers to the research questions. This conclusive step not only summarizes the findings but also ties them back to the original objectives of the dissertation, ensuring that the research contributes meaningfully to the discourse on public procurement data. This also ensures that the initial objectives of the study, which focused on identifying and addressing issues of data quality in public procurement, are clearly met.

# 3

# Background

This chapter offers a comprehensive overview of foundational concepts and technologies pertinent to our study. These concepts establish the theoretical basis for the systematic mapping and exploratory study described in the methodology, ensuring that the research questions are firmly grounded in the context of Open Government Data and public procurement practices.We start by examining the principles of Open Data and the FAIR Principles, and introduce Open Government Data concepts. Additionally, we explore the specific context of public procurement data, the concepts of open data portals, and Large Language Models (LLMs).

## 3.1 Open Data

Open Data[1] are information that should be made available freely, allowing them to be used, reused, and redistributed by anyone, provided that credit is attributed to the source and sharing occurs under the same terms. From another perspective, (WILKINSON, 2016) emphasize the importance of empowering both machines and individuals to discover and use data autonomously. The author introduced the FAIR Principles (an acronym for Findable, Accessible, Interoperable, and Reusable), as a set of guidelines aimed at facilitating the discovery, access, interoperability, and reuse of datasets, aligning with these four key characteristics:

1 Findable Principle: Emphasizes the importance of data discovery through the assignment of unique identifiers, the inclusion of detailed metadata, and registration in searchable resources, thus ensuring accessibility and unambiguous referencing.

2 Accessible Principle: Goes beyond discovery by ensuring efficient access to data through identifiers and standardized communication protocols, prioritizing the accessibility of metadata even when the data is not accessible.

[1] https://okfn.org/opendata/

3 Interoperable Principle: Highlights interoperability by requiring a formal and shared language to represent knowledge, the use of FAIR vocabularies, and qualified references to other metadata to ensure consistency in the interpretation of data across different systems.

4 Reusable Principle: Aiming for data reuse, it emphasizes the need for precise attributes, detailed provenance metadata, clear licenses, and compliance with community standards to facilitate understanding by different users.

Figure 3 represents the process from raw data to open data, emphasizing key steps required to make data accessible and usable.

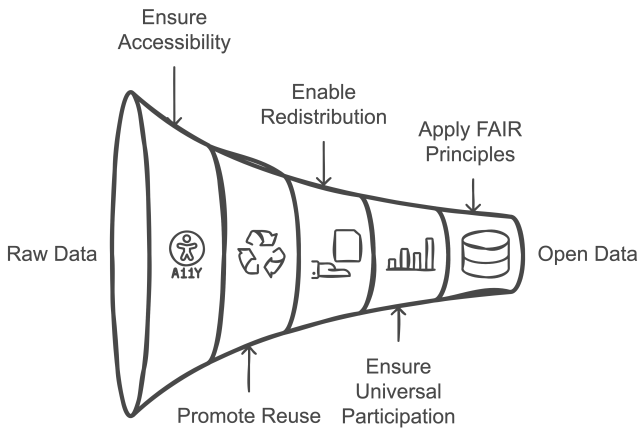


Figure 3 – Lifecycle Stages of Open Data: From Creation to Use.

### 3.1.1 Open Government Data (OGD)

According to the OECD[2], Open Government Data is an approach that seeks to promote transparency and accountability by making public information available to all. For the OPEN GOVERNMENT WORKING GROUP[3], eigth principles establish guidelines for the transparent provision and use of Open Government Data. These principles emphasize that all public data should be disclosed, collected directly from the source, made available to everyone, without the need for registration, as quickly as necessary, structured for automated processing, offered in formats that are not exclusive to a specific entity, and without copyright, patent, trademark, or trade secret regulations. Figure 4 illustrates the core principles of Open Government Data (OGD).

(BACHTIAR, 2020)highlighted that to consolidate trust and transparency in the public sector, it is crucial to implement a specialized platform for disseminating Open Government Data. Nations like India, Brazil, the United Kingdom, and Spain have already incorporated these

[2] https://www.oecd.org/gov/digital-government/open-government-data.htm
[3] https://opengovdata.org/

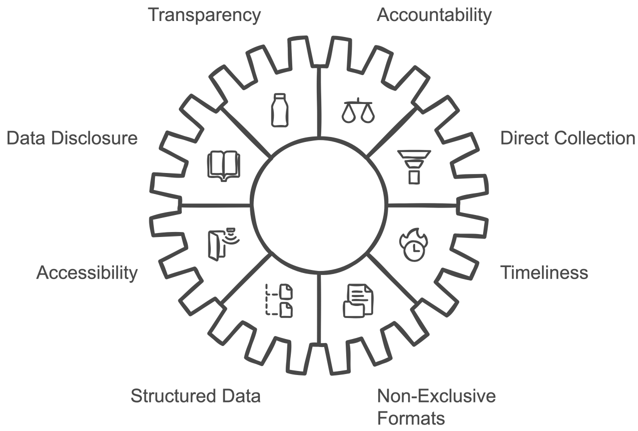


Figure 4 – Core Principles of Open Government Data: Such as Transparency, Accessibility, and Interoperability.

technologies to disclose public data, with the online platform CKAN[4] predominating in this scenario.

### 3.1.2 Public Procurement Data

As defined by (ARROWSMITH, 2010), public procurement involves the acquisition of goods and services by a nation's public sector, encompassing all levels of government. These acquisitions are essential for the functioning of public administration and the delivery of services to the populace in diverse sectors such as education, healthcare, security, energy, and infrastructure (KASHAP, 2004). Public procurement portals, such as USAspending.gov[5] in the United States and the Tenders Electronic Daily (TED)[6] portal in the European Union, serve as key platforms for accessing government procurement data, promoting transparency and enabling public oversight of government spending.

The provision of public procurement data in Brazil is a commitment made by the Brazilian government in the Open Government Partnership[7]. The Public Transparency portal[8], aligned with this initiative, emphasizes the detailed disclosure of government expenditures and the availability of complete information on payments, beneficiaries, and bidding procedures. These guidelines aim to strengthen transparency in public spending and promote research and technological innovation through the implementation of Brazil's Open Data Policy[9].

The National Public Transparency Program (BRASIL, 2023) emphasizes the detailed disclosure of government expenditures and the provision of comprehensive information on

[4] https://ckan.org/
[5] https://www.usaspending.gov/
[6] https://ted.europa.eu/en/
[7] https://www.opengovpartnership.org/
[8] https://portaldatransparencia.gov.br/
[9] https://www.gov.br/conarq/pt-br/legislacao-arquivistica/decretos-federais/decreto-no-8-777-de-11-de-maio-de-2016

payments, beneficiaries, and bidding procedures. These guidelines are aimed at enhancing the transparency of public spending and fostering research and technological innovation by implementing Brazil's Open Data policy[10]. Additionally, there is a digital platform[11] dedicated to the dissemination of information regarding public procurement. This platform is overseen by the Ministry of Economy and is integrated within the Integrated System for General Services Administration (SIASG). The platform is designed to serve entities and public bodies at all levels of government - Federal, State, and Municipal, including all branches of government.

This study examines public procurement data quality issues through the systematic mapping and exploratory study methodology described in Chapter 2, ensuring a comprehensive evaluation of data challenges.

## 3.2 Open Data Platforms

Platforms for Open Government Data are gaining prominence due to their potential to foster innovation in public services, enhance transparency, and contribute to broader societal benefits (DAVIES et al., 2019). These platforms, often referred to as portals, are crucial for data sharing and play a significant role in sustaining open data initiatives such as the Comprehensive Knowledge Archive Network (CKAN)[12], DKAN[13] and Socrata[14]. Open data portals built on these platforms serve as central repositories where data is made accessible, facilitating its reuse and redistribution in accordance with licensing agreements(ALI; ALEXOPOULOS; CHARALABIDIS, 2022).

CKAN is an advanced open-source Data Management System (DMS) designed for the creation and distribution of open data. It supports a variety of platforms at national, international, and federated levels, offering a broad range of features such as APIs, data storage, geospatial capabilities, and customizable themes. CKAN also boasts a variety of extensions that enrich the functionality of open data portals[15]. Widely adopted by governments and enterprises, CKAN supports major open data projects in countries like Brazil[16], the United Kingdom[17], and the United States[18].

DKAN, similar to CKAN, facilitates the establishment of open data environments. It stands out by leveraging PHP[19] and Drupal[20], making it particularly well-suited for organizations

[10] https://dados.gov.br/dados/conjuntos-dados/compras-publicas-do-governo-federal
[11] https://compras.dados.gov.br/docs/home.html
[12] https://ckan.org/
[13] https://dkan.readthedocs.io/en/latest/
[14] https://dev.socrata.com/
[15] https://extensions.ckan.org/
[16] https://compras.dados.gov.br/docs/home.html
[17] https://data.gov.uk/
[18] https://data.gov/open
[19] https://www.php.net/
[20] https://www.drupal.org/

that utilize these technologies for their content management systems (CMS). DKAN comes with a built-in CMS, which sets it apart from CKAN, where users often have to integrate additional CMS systems. Notably, DKAN is utilized by prominent entities such as United States government departments[21] .

Socrata, another key player in the open data landscape, oversees more than a hundred open data catalogs from a variety of organizations around the globe. Socrata's open data is freely available and can be redistributed. It offers a user-friendly interface for data searching and an API known as the SODA[22] API, which simplifies the development of future applications that can deliver more precise results. Examples of platforms powered by Socrata include NASA[23] and the White House[24] data portals.

## 3.3 Large Language Models

Natural Language Processing (NLP) is a field in Computer Science that focuses on computational processing of human language, linked with Artificial Intelligence (AI) and Computational Linguistics. It uses neural network-generated models to represent texts, speech, and non-traditional language specifications.

Large Language Models (LLMs) are expansive language models, usually built on the Transformer architecture (VASWANI et al., 2017), and trained on extensive textual datasets with hundreds of billions of parameters. Among their applications is ChatGPT[25], an advanced model from the GPT series (RADFORD et al., 2018) designed for conversational interactions, capable of generating dialogue that closely mimics human conversation (ZHAO et al., 2023). These models can respond to a variety of user prompts, which can be textual commands or multimedia inputs like images, audio, and videos (SCHULHOFF et al., 2024). A prompt is a structured input that steers the model's output (LIU et al., 2023). Prompt templates, featuring placeholders for specific content, are used to craft prompts suited to a range of tasks (BROWN et al., 2020). Prompts typically contain directives to guide the LLM's response, examples for illustration, formatting instructions for structured outputs, style guidelines for aesthetic control, and role or persona specifications to tailor the response's style (SCHULHOFF et al., 2024).

According to (YüCEL, 2023), Large Language Models (LLMs) can be categorized into generative and extractive models, each serving different purposes (Figure 5). Generative models, like ChatGPT, create fluent and coherent text from scratch, making them ideal for tasks like conversation, programming, and translation. However, they often face issues such as hallucinations, where they produce information that is either inaccurate or fictional, straying

[21] http://www.healthdata.gov/
[22] https://dev.socrata.com/
[23] http://data.nasa.gov
[24] http://open.whitehouse.gov
[25] https://openai.com/blog/chatgpt/

from factual knowledge and possibly delivering responses that are not grounded in the data it was trained on (PERKOVIĆ; DROBNJAK; BOTIČKI, 2024). Extractive models, such as BERT-based models (DEVLIN et al., 2019), retrieve precise information directly from a given text, making them highly useful in scenarios where factual accuracy is paramount, like question-answering and information extraction. While generative models are larger and resource-intensive, extractive models are smaller, less resource-demanding, and better suited for applications requiring textual fidelity and factual correctness. The choice between these models depends on the specific use case and technical needs.

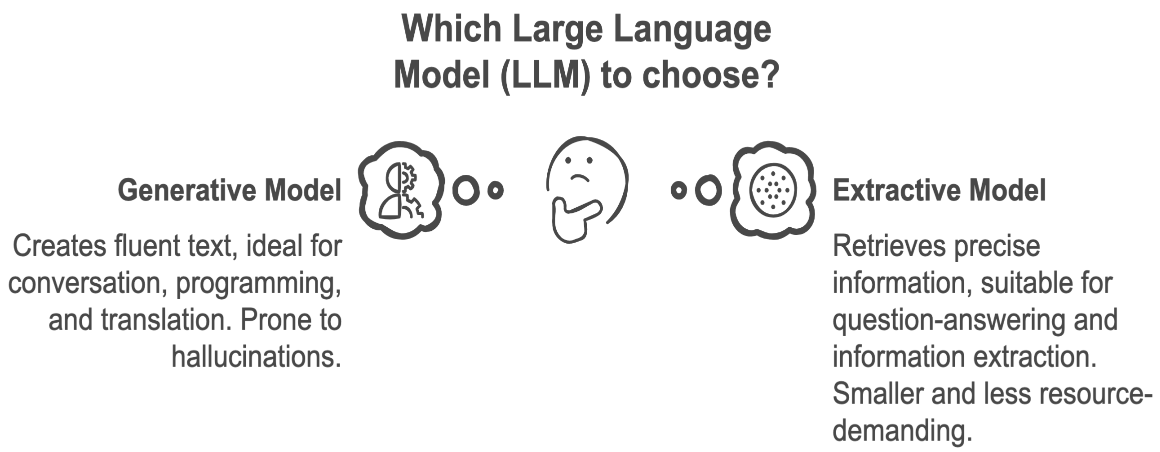


Figure 5 – Categorization of Large Language Models (LLMs): Applications and Use Cases based on (YüCEL, 2023).

Prompt engineering is a technique that enhances the capabilities of LLMs by utilizing task-specific instructions to guide model behavior without altering its core parameters(SAHOO et al., 2024). By providing contextual prompts, this method enables pre-trained models to perform various tasks effectively. It encompasses a range of techniques designed to guide LLMs in generating desired outputs, ranging from basic strategies to more advanced methods.

Effective use of LLMs presents challenges, including cognitive barriers related to query formulation and result interpretation, often arising from limited knowledge or complexity of the topics (WANG et al., 2024). To optimize LLM performance, prompt engineering is essential, creating clear and structured prompts to guide outputs. Users can choose between manual or automated methods to design effective, prompt structures (WHITE et al., 2023), (LIU et al., 2023). Common issues with LLM outputs include vague responses, irrelevant information, and incorrect assumptions about the user's technical knowledge. These can be mitigated by asking precise questions and using clear language, enhancing the relevance and accuracy of responses[26]. Prompting techniques are critical for improving LLM performance, particularly in *few-shot* and *zero-shot* learning within an in-context learning framework, where the model is conditioned on a natural language instruction (BROWN et al., 2020). These methods enable LLMs to perform

[26] https://platform.openai.com/docs/guides/prompt-engineering/six-strategies-for-getting-better-results

tasks with minimal examples (few-shot) or clear instructions (zero-shot) (KOJIMA et al., 2022), (BROWN et al., 2020). *Zero-shot techniques*, especially when paired with *Chain of Thought* (CoT) prompting strategies, engage the LLM in demonstrating its problem-solving approach, which is particularly valuable in complex domains like public procurement (SCHULHOFF et al., 2024).

As discussed in the methodology (Chapter 2), LLMs are central to the exploratory study phase, where their capabilities are assessed in enhancing public procurement data quality.

# 4

# Quality Concerns in Public Procurement Data

This chapter aims to thoroughly examine the quality concerns related to public procurement data, focusing on discoverability, accessibility, and usability. It begins with an outline of the research design and methodology, employing systematic mapping to identify and analyze methods of evaluation, thematic domains, and the aforementioned dimensions of data quality. This structured methodology ensures a comprehensive understanding of quality issues and prepares the ground for practical recommendations. By tackling specific research questions, the chapter characterizes the current state of public procurement data quality and identifies best practices for improvement. This analysis is integral to the dissertation's broader objectives, emphasizing the significance of data quality in improving transparency and efficiency in public procurement processes. The insights gained will underpin the next chapter, which will explore how emerging technologies, especially artificial intelligence, can mitigate these quality concerns.

## 4.1 Research Design and Methodology

Following the guidelines proposed by Kitchenham and Charters (GROUP, 2007), we divided the research design into three phases: planning, execution, and reporting. Table 1 shows the steps for selecting relevant studies, with further details provided in the following sections.

This research design builds on the methodology detailed in Chapter 2, applying systematic mapping techniques to identify key challenges and best practices in public procurement data quality.

Table 1 – Study Selection Process for Systematic Mapping

| | |
|---|---|
| **Step 1** | Search electronic repositories for relevant studies using search strings. |
| **Step 2** | Remove duplicate studies. |
| **Step 3** | Apply exclusion criteria to the list of publications. |
| **Step 4** | Considering the studies not excluded in the previous step, apply inclusion criteria based on the text of the abstract, introduction, and conclusion. |
| **Step 5** | Considering the studies selected from the previous step, apply quality criteria. |

### 4.1.1 Planning

During this phase, we defined a comprehensive review protocol that included Research Questions (RQs), study search strategies, selection and exclusion criteria, and data extraction methods. Table 2 presents the goal of this study following the GQM approach(BASILI; ROMBACH, 1988). The primary Research Question (RQ) of this study is "*How have public procurement data been evaluated from the quality concern perspectives of discoverability, accessibility, and usability?*". Characterizing quality concerns in public procurement data enables those who produce policies, researchers, and stakeholders to learn more about the current state, providing enhancement opportunities and consolidating best practices. In addition, as detailed in Table 3, four Specific Research Questions (SRQs) were formulated to investigate additional aspects of the primary question.

Table 2 – Research Goals Using the Goal Question Metric (GQM) Approach

| | |
|---|---|
| **Analyze** | open government data |
| **for the purpose of** | quality evaluation |
| **with respect to** | discoverability, accessibility and usability |
| **from the point of view of** | researchers, citizens and stakeholders |
| **in the context of** | public procurement |

We used the PICO criteria (STONE, 2002) to build the search strings as presented in Table 4. Table 5 presents the proposed major search terms. These terms served as a basis for constructing the search queries to select studies focusing on public procurement data quality evaluation.

Table 6 shows the search keywords and corresponding alternative terms to be included in the search string. We searched studies in the databases IEEE Xplore Digital Library, Scopus, Web of Science, ACM Digital Library, and Emerald. Table 7 presents the search string we applied and the respective database. We conducted the searches on September 28, 2023, selecting studies published from 2008. The reason to start in 2008 was the landmark Transparency and Open Government (OBAMA, 2009) initiative and its relationship with data transparency and accessibility.

Table 3 – Specific Research Questions and Their Motivations

| **Specific Research Question** | **Motivation** |
|---|---|
| **SRQ1:** What are the main methods and techniques to evaluate public procurement data quality? | Characterizing methods and techniques to evaluate data quality is important to identify strengths and improvement opportunities in public procurement data. |
| **SRQ2:** What are the reported issues related to discoverability, accessibility, and usability in the public procurement data context? | Investigating practical problems users face can provide insights into specific areas to enhance user experience. |
| **SRQ3:** Which thematic domains and government sectors have been reported in public procurement data quality studies? | Examining thematic domains and sectors focused in the selected studies helps identify areas that attract more interest regarding quality issues. This information can lead to improvements in these sectors. |
| **SRQ4:** What are the best practices to promote public procurement data quality? | Recognizing best practices supports benchmarking and public procurement data quality improvement. Quality assurance measures can be applied to identify and evaluate these best practices. |

Table 4 – Construction of Search Strings Based on PICO Criteria

| | |
|---|---|
| (P)opulation | Open Government Data for public procurement |
| (I)ntervention | Quality evaluation of public procurement data |
| (C)omparison | Main methods used to evaluate the quality of public procurement data from the perspective of discoverability, accessibility, and usability |
| (O)utcomes | discoverability, accessibility and usability status of public procurement data and their respective best evaluation methods |

Table 5 – Primary Search Terms for Public Procurement Data Quality Analysis

| | |
|---|---|
| (P)opulation | "public procurement" |
| (I)ntervention | "quality dimensions evaluation" |
| (C)omparison | "methods" |
| (O)utcomes | "discoverability, accessibility and usability" |

Table 8 presents the inclusion, exclusion, and quality criteria. We associate the exclusion criteria with an OR operator, meaning we can exclude a study according to one factor. In contrast, we associate the inclusion criteria with the AND operator, requiring the achievement of all criteria before considering the study. Additionally, we adopted questions based on Dyba and Dingsoyr (DYBÅ; DINGSØYR, 2008) as quality criteria (see Table 8). We evaluated thoroughly all publications that fulfilled the inclusion and exclusion criteria based on these questions.

Table 6 – Alternative Terms and Synonyms for Search String Development

| Terms | Alternative Terms |
|---|---|
| "public procurement " | "government procurement", "public purchasing", "public bidding" |
| "data quality dimensions evaluation" | "quality dimensions assessment" |
| "methods" | "techniques" |
| "discoverability", "accessibility", "usability" | "searchability", "availability", "functionality" |

### 4.1.2 Execution

We used the search string to select peer-reviewed studies from the repositories, considering their titles, abstracts, and keywords. The goal is to improve the retrieval process, reducing occurrences of false-positive results due to unclear or irrelevant keywords.

We carried out the search on September 28, 2023, leading to the identification of 709 articles. In the sequence, we reviewed titles and abstracts to evaluate the relevance of the studies, excluding possible duplicate papers. Following this, we assessed the full text of the studies to meet the inclusion and exclusion criteria. After excluding specific papers during the Eligibility Phase, we obtained the final selection of studies, guided by the quality standards detailed in Table 8.

Figure 6 presents the PRISMA approach (MOHER et al., 2009), illustrating the stages and number of studies in each phase of this systematic mapping. To evaluate the effectiveness of the search string, we selected a minimum set of searched studies before the execution of the whole search analysis. The authors considered the effectiveness acceptable, confirming it later with the results presented in Table 9.

Based on 709 identified studies, the systematic mapping process's search effectiveness is summarized in Table 9. Out of these, 50 studies met the inclusion criteria. Scopus contributed 5 articles, representing a search effectiveness rate of 7.4%. IEEE Xplore Digital Library yielded 7 articles, resulting in a higher search string effectiveness of 12.3%. Web of Science provided 25 articles, with an effectiveness rate of 9.1%. The ACM Digital Library contributed 10 articles, yielding a 5.6% effectiveness rate. Lastly, Emerald contributed 3 articles, resulting in a search effectiveness of 2.4%.

### 4.1.3 Study Selection Process

In the data extraction process from multiple databases, 709 studies were retrieved. Subsequently, after reviewing titles and abstracts and removing duplicates, 107 articles underwent full-text evaluation. Of these, 50 articles met the criteria specified in Table 8. In the following section, we consolidate the collected data and address the predefined SRQs, aligning with the objectives outlined in this study.

Table 7 – Database Sources and Applied Search Strings for Systematic Mapping

| **Repository** | **Search String** |
|---|---|
| Scopus | ( "public procurement" OR "government procurement" OR "public purchasing" OR "public bidding" ) AND<br>( "usability" OR "discoverability" OR "accessibility" OR "searchability" OR "availability" OR ”functionality” ) AND<br>( "quality dimensions evaluation" OR "quality dimensions assessment" OR "methods" OR "techniques" ) |
| IEEE Xplore Digital Library | ("public procurement" OR "government procurement" OR "public purchasing" OR "public bidding" ) AND<br>( ( "quality dimensions evaluation" OR "quality dimensions assessment" OR "methods" OR "techniques" ) OR ( "discoverability" OR "accessibility" OR "usability" OR "searchability" OR "availability" OR ”functionality” ) ) |
| Web of Science | ("public procurement" OR "government procurement" OR "public purchasing" OR “public bidding”) AND<br>("usability" OR "discoverability" OR "accessibility" OR "searchability" OR ”availability” OR ”functionality”) AND<br>(”quality dimensions evaluation” OR ”quality dimensions assessment” OR ”methods” OR ”techniques”) |
| ACM Digital Library | ("public procurement" OR "government procurement" OR "public purchasing" OR “public bidding”) AND<br>("usability" OR "discoverability" OR "accessibility" OR "searchability" OR ”availability” OR ”functionality”) AND<br>(”quality dimensions evaluation” OR ”quality dimensions assessment” OR ”methods” OR ”techniques”) |
| Emerald | ("public procurement" OR "government procurement" OR "public purchasing" OR “public bidding”) AND<br>("usability" OR "discoverability" OR "accessibility" OR "searchability" OR ”availability” OR ”functionality”) AND<br>(”quality dimensions evaluation” OR ”quality dimensions assessment” OR ”methods” OR ”techniques”) |

Table 8 – Inclusion, Exclusion, and Quality Criteria for Study Selection

| **Criterion** | **Type** | **Description** | **Connective or Answer** |
|---|---|---|---|
| Exclusion | E1 | Publications that did not focus on the topic | OR |
| Exclusion | E2 | Publications not written in English or Portuguese. | OR |
| Exclusion | E3 | Publications that are not full papers. | OR |
| Exclusion | E4 | Duplicate publications | OR |
| Exclusion | E5 | Non-primary studies | OR |
| Exclusion | E6 | Studies published earlier 2008 | OR |
| Inclusion | I1 | Publications whose sources were conferences or periodicals | AND |
| Inclusion | I2 | The studies must address the topic of government procurements | AND |
| Quality | Q1 | Were the objectives of the study clearly specified? | YES/NO |
| Quality | Q2 | Was the context of the study clearly defined? | YES/NO |
| Quality | Q3 | Is the way the data is analyzed in line with the study objectives? | YES/NO |

## 4.2 Characterizing Public Procurement Data

This section outlines the dimensions we adopted to characterize the studies as depicted in Figure 7. These dimensions were derived from the systematic mapping methodology described in Chapter 2, ensuring a rigorous and consistent approach to characterizing public procurement data.

We considered four thematic dimensions corresponding to the four SRQs presented in Table 3: Evaluation methods (SRQ1), Discoverability, Accessibility and Usability issues (SRQ2), Thematic Domains and Government Sectors (SRQ3) and Best Practices (SRQ4). These four dimensions aim to comprehensively characterize quality concerns related to public procurement data. The subsequent subsections detail each of these four dimensions.

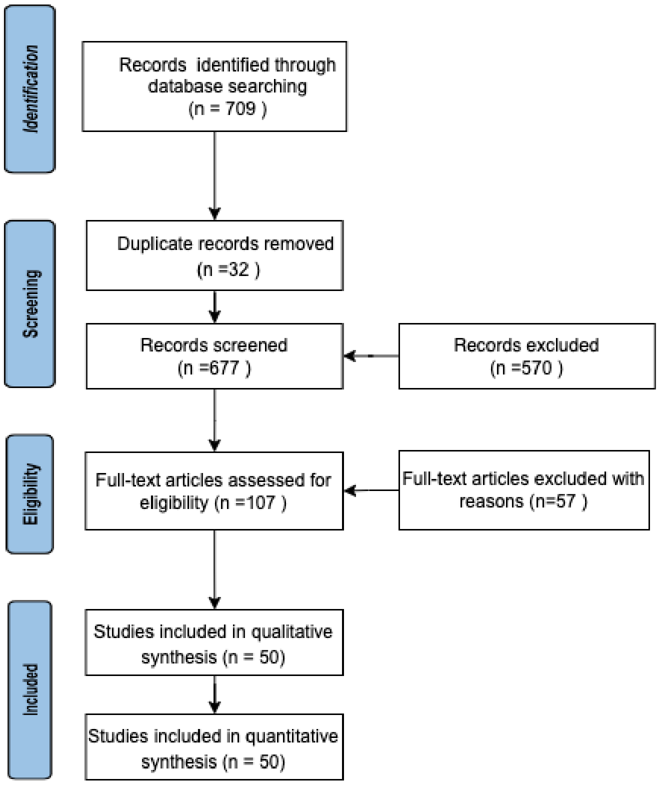


Figure 6 – Systematic Mapping Process: Steps from Initial Study Identification to Final Selection.

Table 9 – Search String Effectiveness Across Database Sources

| **Databases/Search Technique** | **Total** | **Selected** | **Effectiveness** |
|---|---|---|---|
| Scopus | 68 | 5 | 7.4% |
| IEEE Xplore Digital Library | 57 | 7 | 12.3% |
| Web of Science | 276 | 25 | 9.1% |
| ACM Digital Library | 180 | 10 | 5.6% |
| Emerald | 128 | 3 | 2.4% |

### 4.2.1 Evaluation Methods

Examining the possibilities of data quality evaluation approaches is a means to understand their strengths and limitations. The main possibilities for evaluation are Manual (M), Automatic (A), Statistical (S), and Semi-automated (C) methods. Table 10 presents this panorama with the corresponding studies that employed each. A consolidated understanding of evaluation techniques can support an effective selection and adoption of methods to strengthen public procurement data quality standards. Subsection 4.3.1 presents an analysis of evaluation methods applied in each study based on findings from the selected studies, and a complete list can be reached in a public repository[1].

[1] https://doi.org/10.5281/zenodo.11052244

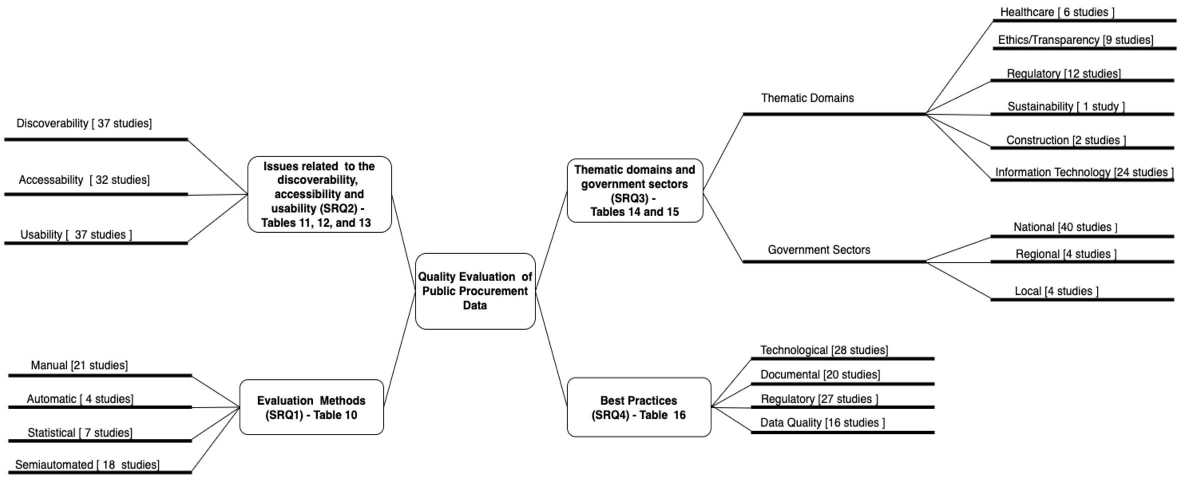


Figure 7 – Characterization of Public Procurement Data Quality Across Four Dimensions.

Table 10 – Evaluation Methods Applied to Public Procurement Data Quality

| Methods | References |
|---|---|
| Manual (21) | (MONDORF; WIMMER; REISER, 2013),(BEHR; ABRAHAMSSON, 2022), (PINTO et al., 2015), (KLUN; SETNIKAR-CANKAR, 2013), (OLIVEIRA et al., 2020), (SIROTKINA; LAZAREVICH, 2023), (MARTINS et al., 2021), (UDUWAGE-DON; HADIWATTAGE; PANUWATWANICH, 2023), (STAKE, 2017), (MELNIKOV; LUKASHENKO, 2019), (ASTBRINK; TIBBEN, 2013), (BJARNASON; PERSSON; RYDENFäLT, 2023), (MONDORF; WIMMER, 2008), (SPACEK; SPACKOVA, 2023), (PUTRI; RULDEVIYANI, 2019), (ARNEY et al., 2014), (LEE; OH; KWON, 2008), (CONCHA; BURR; SUáREZ, 2014), (SILVA et al., 2018), (GAVUROVA; KUBAK, 2021), (CSAKI, 2018) |
| Semiautomated (18) | (ANCARANI et al., 2019), (TOSIN et al., 2016), (NAI et al., 2023), (MODRUŠAN; MRŠIĆ; RABUZIN, 2020), (DAHBI; CHIADMI; LAMHARHAR, 2023), (MENDES; VOIGT, 2022b), (aO et al., 2023), (GONçALVES et al., 2010), (RIIHIAHO et al., 2015), (FERREIRA; AMARAL, 2016), (BALAEVA et al., 2022), (RODRIGUEZ et al., 2019), (VAZQUEZ-ROWE et al., 2021), (LALIĆ et al., 2019), (SOYLU et al., 2022c), (ZHIQIANG et al., 2020), (TYLLINEN et al., 2016) |
| Statistical (7) | (OZYUREK; ERDAL, 2018), (IMAMOGLU; REHAN, 2015), (TAS, 2020), (SANGIL, 2020), (MUHWEZI et al., 2023), (PATRUCCO; AGASISTI; GLAS, 2021), (MELON; SPRUK, 2020) |
| Automatic (4) | (VELASCO et al., 2021),(SOYLU et al., 2022b), (MONTEIRO; CORREIA, 2023), (ALMEIDA et al., 2018) |

### 4.2.2 Discoverability, Accessibility, and Usability issues

Discoverability evaluates to which extent users must be able to discover data relevant to their needs to utilize it (DONKER; LOENEN, 2017). This attribute strongly depends on the quality of descriptive metadata provided along with the data itself (ATTARD et al., 2015). Accessibility is also a relevant quality attribute to a well-functioning open data ecosystem, not only from a technical perspective but also from a legal perspective. For this reason, the importance of policies to define the legal context and standards to facilitate data interoperability of open government data ecosystems usually created by governments to facilitate access to sharing and (re)using of government data (DONKER; LOENEN, 2017). Usability can be expressed as how easily can the published data be used (ATTARD et al., 2015). It depends on other quality dimensions related to the degree of how data is accessible, open, interoperable, complete, and discoverable (ATTARD et al., 2015). We performed a detailed analysis of issues related to discoverability, accessibility, and usability from evidence obtained from the literature in Subsection 4.3.2.

Tables 11 through 13 present an overview of studies that discuss these issues, classifying them according to the technological, documented, regulatory, and data quality perspectives.

Table 11 – Discoverability Issues and Perspectives

| **Related Issues** | **Perspective** | **References** |
|---|---|---|
| Discoverability | Technological | (VELASCO et al., 2021), (MONDORF; WIMMER; REISER, 2013), (BEHR; ABRAHAMSSON, 2022), (PINTO et al., 2015), (IMAMOGLU; REHAN, 2015), (SOYLU et al., 2022b), (TAS, 2020), (TAS, 2020), (SIROTKINA; LAZAREVICH, 2023), (MARTINS et al., 2021), (UDUWAGE-DON; HADIWATTAGE; PANUWATWANICH, 2023), (NAI et al., 2023), (ASTBRINK; TIBBEN, 2013), (ALMEIDA et al., 2018), (MONDORF; WIMMER, 2008), (SPACEK; SPACKOVA, 2023), (DAHBI; CHIADMI; LAMHARHAR, 2023), (NURMANDI; KIM, 2015), (MENDES; VOIGT, 2022b), (aO et al., 2023), (PUTRI; RULDEVIYANI, 2019), (FERREIRA; AMARAL, 2016), (BALAEVA et al., 2022), (RODRIGUEZ et al., 2019), (ARNEY et al., 2014), (SOYLU et al., 2022c) |
| | Documental | (MONDORF; WIMMER; REISER, 2013), (TOSIN et al., 2016), (SOYLU et al., 2022b), (OLIVEIRA et al., 2020), (TAS, 2020), (ASTBRINK; TIBBEN, 2013), (MODRUŠAN; MRŠIĆ; RABUZIN, 2020), (MONDORF; WIMMER, 2008), (SPACEK; SPACKOVA, 2023), (DAHBI; CHIADMI; LAMHARHAR, 2023), (MENDES; VOIGT, 2022b), (aO et al., 2023), (RIIHIAHO et al., 2015), (BALAEVA et al., 2022), (LEE; OH; KWON, 2008), (GAVUROVA; KUBAK, 2021), (CSAKI, 2018) |
| | Regulatory | (SOYLU et al., 2022b),(MONTEIRO; CORREIA, 2023), (MAVIDIS; FOLINAS, 2022), (SIROTKINA; LAZAREVICH, 2023), (MODRUŠAN; MRŠIĆ; RABUZIN, 2020), (MONDORF; WIMMER, 2008), (NURMANDI; KIM, 2015), (MENDES; VOIGT, 2022b), (aO et al., 2023), (RIIHIAHO et al., 2015), (FERREIRA; AMARAL, 2016), (RODRIGUEZ et al., 2019), (VAZQUEZ-ROWE et al., 2021), (ZHIQIANG et al., 2020) |
| | Data quality | (OZYUREK; ERDAL, 2018), (MONTEIRO; CORREIA, 2023), (UDUWAGE-DON; HADIWATTAGE; PANUWATWANICH, 2023), (ASTBRINK; TIBBEN, 2013), (MODRUŠAN; MRŠIĆ; RABUZIN, 2020), (DAHBI; CHIADMI; LAMHARHAR, 2023), (BALAEVA et al., 2022), (RODRIGUEZ et al., 2019), (ARNEY et al., 2014), (CONCHA; BURR; SUáREZ, 2014), (SOYLU et al., 2022c), (CSAKI, 2018) |

Table 12 – Accessibility Issues and Perspectives

| Related Issues | Perspective | References |
|---|---|---|
| Accessibility | Technological | (VELASCO et al., 2021), (MONDORF; WIMMER; REISER, 2013), (SOYLU et al., 2022b), (UDUWAGE-DON; HADIWATTAGE; PANUWATWANICH, 2023), (NAI et al., 2023), (ASTBRINK; TIBBEN, 2013), (ALMEIDA et al., 2018), (MODRUŠAN; MRŠIĆ; RABUZIN, 2020), (MONDORF; WIMMER, 2008), (DAHBI; CHIADMI; LAMHARHAR, 2023), (NURMANDI; KIM, 2015), (MENDES; VOIGT, 2022b), (aO et al., 2023), (GONçALVES et al., 2010), (PUTRI; RULDEVIYANI, 2019), (BALAEVA et al., 2022), (RODRIGUEZ et al., 2019), (VAZQUEZ-ROWE et al., 2021) |
| | Documental | (SOYLU et al., 2022b), (SIROTKINA; LAZAREVICH, 2023), (UDUWAGE-DON; HADIWATTAGE; PANUWATWANICH, 2023), (ALMEIDA et al., 2018), (aO et al., 2023), (FERREIRA; AMARAL, 2016), (LEE; OH; KWON, 2008), (CSAKI, 2018) |
| | Regulatory | (ANCARANI et al., 2019), (MONDORF; WIMMER; REISER, 2013), (KLUN; SETNIKAR-CANKAR, 2013), (SOYLU et al., 2022b), (OLIVEIRA et al., 2020), (STAKE, 2017), (SANGIL, 2020), (SPACEK; SPACKOVA, 2023), (MENDES; VOIGT, 2022b), (GONçALVES et al., 2010), (RIIHIAHO et al., 2015), (CONCHA; BURR; SUáREZ, 2014), (GAVUROVA; KUBAK, 2021) |
| | Data quality | (SOYLU et al., 2022b),(MONTEIRO; CORREIA, 2023), (SIROTKINA; LAZAREVICH, 2023), (MONDORF; WIMMER, 2008), (DAHBI; CHIADMI; LAMHARHAR, 2023), (aO et al., 2023), (SOYLU et al., 2022c), (CSAKI, 2018) |

### 4.2.3 Thematic Domains and Government Sectors

Identifying thematic areas and government sectors that focus on public procurement data quality issues provides insights into the characterizing of public procurement data quality concerns. Understanding why thematic areas and government sectors must analyze quality issues can also reveal the importance of transparency, accountability, and participatory governance in public procurement data practices. Tables 14 and 15 respectively present thematic domains and government sectors most mentioned in the selected studies. Table 14 classifies papers based on specific thematic areas, e.g., information technology, regulatory issues, ethics and transparency, healthcare, sustainability, and construction. Table 15 organizes the papers by government sectors, encompassing national, regional, and local levels. We present more details of these findings in Subsection 4.3.3.

### 4.2.4 Best Practices

Best practices applied to promote public procurement data quality provide appropriate conditions to improve the discoverability, accessibility, and usability characteristics of these data. Table 16 organizes studies according to the best practices identified in the literature, categorized into technological, documental, regulatory, and data quality, with references for each category. We discuss these findings in detail in Subsection 4.3.4. Additional details are accessible in the Zenodo URL previously provided.

Table 13 – Usability Issues and Perspectives

| Related Issues | Perspective | References |
|---|---|---|
| Usability | Technological | (VELASCO et al., 2021), (MONDORF; WIMMER; REISER, 2013), (TOSIN et al., 2016), (PINTO et al., 2015), (IMAMOGLU; REHAN, 2015), (TAS, 2020), (SIROTKINA; LAZAREVICH, 2023), (MARTINS et al., 2021), (UDUWAGE-DON; HADIWATTAGE; PANUWATWANICH, 2023), (MUHWEZI et al., 2023), (MONDORF; WIMMER, 2008), (SPACEK; SPACKOVA, 2023), (DAHBI; CHIADMI; LAMHARHAR, 2023), (aO et al., 2023), (BALAEVA et al., 2022), (RODRIGUEZ et al., 2019), (CONCHA; BURR; SUáREZ, 2014), (TYLLINEN et al., 2016) |
| | Documental | (SOYLU et al., 2022b), (UDUWAGE-DON; HADIWATTAGE; PANUWATWANICH, 2023), (SANGIL, 2020), (NURMANDI; KIM, 2015), (BALAEVA et al., 2022), (CSAKI, 2018) |
| | Regulatory | (TOSIN et al., 2016), (PINTO et al., 2015), (MONTEIRO; CORREIA, 2023), (TAS, 2020), (UDUWAGE-DON; HADIWATTAGE; PANUWATWANICH, 2023), (ASTBRINK; TIBBEN, 2013), (MUHWEZI et al., 2023), (BJARNASON; PERSSON; RYDENFäLT, 2023), (DAHBI; CHIADMI; LAMHARHAR, 2023), (RIIHIAHO et al., 2015), (FERREIRA; AMARAL, 2016), (BALAEVA et al., 2022), (ARNEY et al., 2014), (LEE; OH; KWON, 2008), (SILVA et al., 2018), (LALIĆ et al., 2019), (TYLLINEN et al., 2016) |
| | Data quality | (MONDORF; WIMMER; REISER, 2013), (TOSIN et al., 2016), (OZYUREK; ERDAL, 2018), (SOYLU et al., 2022b), (UDUWAGE-DON; HADIWATTAGE; PANUWATWANICH, 2023), (NAI et al., 2023), (SANGIL, 2020), (ALMEIDA et al., 2018), (DAHBI; CHIADMI; LAMHARHAR, 2023), (MENDES; VOIGT, 2022b), (aO et al., 2023), (PUTRI; RULDEVIYANI, 2019), (RODRIGUEZ et al., 2019), (SOYLU et al., 2022c) |

## 4.3 Results

This section discusses the findings of each specific research question (SRQ) as previously presented in Table 3. This discussion is a reference to answer the primary research question (RQ) already presented in the Subsection 4.1.1. It aims to provide a comprehensive panorama of the challenges in assessing and improving public procurement data quality by examining evaluation methods, challenges, thematic domains, and best practices.

The first evidence of this challenge is the variety of stakeholders' profiles that deal with public procurement data. Each stakeholder brings a specific perception and set of needs and concerns, emphasizing the importance of data providers for being aware of multiple viewpoints when planning, implementing, assessing, and utilizing public procurement data infrastructure (WANG et al., 2023). Table 17 highlights how diverse stakeholders rely on procurement data for varying purposes, demonstrating the multifaceted importance of transparency, accessibility, and accuracy in public procurement. Meanwhile, Table 18 presents the profiles of these six different stakeholders cited in the selected studies. The characteristics of these profiles should drive public procurement policies and practices through research and policy efforts, including researchers, academics, and non-profit organizations (BEHR; ABRAHAMSSON, 2022)(TAS, 2020), along with governmental institutions and regulatory bodies (TAS, 2020)(MARTINS et al., 2021)(ASTBRINK; TIBBEN, 2013). The varied backgrounds of stakeholders (e.g., healthcare (GAVUROVA; KUBAK, 2021), sustainability (SILVA et al., 2018) and ICT accessibility (MAR-

Table 14 – Thematic Domains of Public Procurement Data Quality Research

| Thematic Domains | References |
|---|---|
| Information Technology (24) | (MONDORF; WIMMER; REISER, 2013), (TOSIN et al., 2016), (BEHR; ABRAHAMSSON, 2022), (PINTO et al., 2015), (IMAMOGLU; REHAN, 2015), (MONTEIRO; CORREIA, 2023), (OLIVEIRA et al., 2020), (SIROTKINA; LAZAREVICH, 2023), (UDUWAGE-DON; HADIWATTAGE; PANUWATWANICH, 2023), (MELNIKOV; LUKASHENKO, 2019), (NAI et al., 2023), (MODRUŠAN; MRŠIĆ; RABUZIN, 2020), (MONDORF; WIMMER, 2008), (DAHBI; CHIADMI; LAMHARHAR, 2023), (aO et al., 2023), (GONçALVES et al., 2010), (PUTRI; RULDEVIYANI, 2019), (RIIHIAHO et al., 2015), (RODRIGUEZ et al., 2019), (LEE; OH; KWON, 2008), (PATRUCCO; AGASISTI; GLAS, 2021), (SOYLU et al., 2022c), (CSAKI, 2018), (TYLLINEN et al., 2016) |
| Regulatory (12) | (KLUN; SETNIKAR-CANKAR, 2013), (TAS, 2020), (MARTINS et al., 2021), (MELNIKOV; LUKASHENKO, 2019), (ASTBRINK; TIBBEN, 2013), (BJARNASON; PERSSON; RYDENFäLT, 2023), (NURMANDI; KIM, 2015), (FERREIRA; AMARAL, 2016), (BALAEVA et al., 2022), (RODRIGUEZ et al., 2019), (LEE; OH; KWON, 2008), (PATRUCCO; AGASISTI; GLAS, 2021) |
| Ethics/Transparency (9) | (VELASCO et al., 2021),(BEHR; ABRAHAMSSON, 2022), (PINTO et al., 2015), (STAKE, 2017), (ALMEIDA et al., 2018), (DAHBI; CHIADMI; LAMHARHAR, 2023), (MENDES; VOIGT, 2022b), (aO et al., 2023), (CSAKI, 2018) |
| Healthcare (6) | (MUHWEZI et al., 2023), (BJARNASON; PERSSON; RYDENFäLT, 2023), (MENDES; VOIGT, 2022b), (ARNEY et al., 2014), (GAVUROVA; KUBAK, 2021), (TYLLINEN et al., 2016) |
| Construction Industry (2) | (UDUWAGE-DON; HADIWATTAGE; PANUWATWANICH, 2023), (VAZQUEZ-ROWE et al., 2021) |
| Sustainability (1) | (SILVA et al., 2018) |

TINS et al., 2021)(ASTBRINK; TIBBEN, 2013)) justifies the multiple interests and concerns addressed in the selected studies, from quality management practices (LALIĆ et al., 2019) to the impact of e-procurement (MELON; SPRUK, 2020) and the promotion of accessibility and inclusivity (ASTBRINK; TIBBEN, 2013).

### 4.3.1 Specific Research Question 1 (SRQ1)

Specific Research Question 1 (SRQ1) investigates the main methods and techniques to evaluate public procurement data quality. Based on the evidence obtained from the selected studies, we classified the methods and techniques into four categories: manual, automatic,

Table 15 – Government Sectors Addressed in Selected Public Procurement Studies

| **Government Sectors** | **References** |
|---|---|
| National (40) | (ANCARANI et al., 2019), (VELASCO et al., 2021), (MONDORF; WIMMER; REISER, 2013), (TOSIN et al., 2016), (PINTO et al., 2015), (IMAMOGLU; REHAN, 2015), (SOYLU et al., 2022b), (OLIVEIRA et al., 2020), (TAS, 2020), (SIROTKINA; LAZAREVICH, 2023), (MARTINS et al., 2021), (UDUWAGEDON; HADIWATTAGE; PANUWATWANICH, 2023), (STAKE, 2017), (MELNIKOV; LUKASHENKO, 2019), (NAI et al., 2023), (NAI et al., 2023), (ASTBRINK; TIBBEN, 2013), (ALMEIDA et al., 2018), (MUHWEZI et al., 2023), (MODRUŠAN; MRŠIĆ; RABUZIN, 2020), (MONDORF; WIMMER, 2008), (SPACEK; SPACKOVA, 2023), (DAHBI; CHIADMI; LAMHARHAR, 2023), (GONçALVES et al., 2010), (PUTRI; RULDEVIYANI, 2019), (RIIHIAHO et al., 2015), (FERREIRA; AMARAL, 2016), (BALAEVA et al., 2022), (RODRIGUEZ et al., 2019), (ARNEY et al., 2014), (LEE; OH; KWON, 2008), (CONCHA; BURR; SUáREZ, 2014), (VAZQUEZ-ROWE et al., 2021), (SILVA et al., 2018), (GAVUROVA; KUBAK, 2021), (MELON; SPRUK, 2020), (LALIĆ et al., 2019), (SOYLU et al., 2022c), (CSAKI, 2018), (ZHIQIANG et al., 2020), (TYLLINEN et al., 2016) |
| Regional (4) | (VELASCO et al., 2021), (SANGIL, 2020), (MENDES; VOIGT, 2022b), (aO et al., 2023) |
| Local (4) | (KLUN; SETNIKAR-CANKAR, 2013), (BJARNASON; PERSSON; RYDENFäLT, 2023), (NURMANDI; KIM, 2015), (PATRUCCO; AGASISTI; GLAS, 2021) |

statistical, and semi-automated. Table 10 summarizes the reported evaluation methods.

**Manual Methods**, the most identified methods in the selected studies, primarily rely on human intervention and expertise. Within this category, methods such as literature reviews (SPACEK; SPACKOVA, 2023) (CONCHA; BURR; SUáREZ, 2014) (ARNEY et al., 2014), secondary data analyses (BEHR; ABRAHAMSSON, 2022) and structured data collection (KLUN; SETNIKAR-CANKAR, 2013) (OLIVEIRA et al., 2020) play pivotal roles. Semi-structured literature reviews coupled with interviews (ARNEY et al., 2014) and general research methodologies (KLUN; SETNIKAR-CANKAR, 2013) also contribute significantly. Methods such as questionnaire surveys (SPACEK; SPACKOVA, 2023)(PUTRI; RULDEVIYANI, 2019), evaluations based on the most economically advantageous tender (MEAT) (STAKE, 2017), case studies (ASTBRINK; TIBBEN, 2013) and focus groups (ASTBRINK; TIBBEN, 2013) (CONCHA; BURR; SUáREZ, 2014) have also been cited. The use of semi-structured interviews (PINTO et al., 2015) and field research via observation methods (PINTO et al., 2015) also shows other possibilities of evaluation. In the context of analysis, compliance, and evaluation, certain methods hold particular relevance, namely conformance testing (MONDORF; WIM-

Table 16 – Best Practices for Improving Public Procurement Data Quality

| **Practices** | **References** |
|---|---|
| Technological (28) | (ANCARANI et al., 2019), (VELASCO et al., 2021), (MONDORF; WIMMER; REISER, 2013), (TOSIN et al., 2016), (KLUN; SETNIKAR-CANKAR, 2013), (IMAMOGLU; REHAN, 2015), (IMAMOGLU; REHAN, 2015), (SOYLU et al., 2022b), (MONTEIRO; CORREIA, 2023), (MARTINS et al., 2021), (UDUWAGE-DON; HADIWATTAGE; PANUWATWANICH, 2023), (MELNIKOV; LUKASHENKO, 2019), (NAI et al., 2023), (ALMEIDA et al., 2018), (MODRUŠAN; MRŠIĆ; RABUZIN, 2020), (MONDORF; WIMMER, 2008), (DAHBI; CHIADMI; LAMHARHAR, 2023), (NURMANDI; KIM, 2015), (GONçALVES et al., 2010), (FERREIRA; AMARAL, 2016), (RODRIGUEZ et al., 2019), (LEE; OH; KWON, 2008), (PATRUCCO; AGASISTI; GLAS, 2021), (CONCHA; BURR; SUáREZ, 2014), (SILVA et al., 2018), (LALIĆ et al., 2019), (SOYLU et al., 2022c), (CSAKI, 2018), (ZHIQIANG et al., 2020) |
| Documental (20) | (ANCARANI et al., 2019), (VELASCO et al., 2021), (MONDORF; WIMMER; REISER, 2013), (TOSIN et al., 2016), (OZYUREK; ERDAL, 2018), (PINTO et al., 2015), (KLUN; SETNIKAR-CANKAR, 2013), (MONTEIRO; CORREIA, 2023), (OLIVEIRA et al., 2020), (ASTBRINK; TIBBEN, 2013), (ALMEIDA et al., 2018), (BJARNASON; PERSSON; RYDENFäLT, 2023), (SPACEK; SPACKOVA, 2023), (RIIHIAHO et al., 2015), (FERREIRA; AMARAL, 2016), (RODRIGUEZ et al., 2019), (SILVA et al., 2018), (LALIĆ et al., 2019), (SOYLU et al., 2022c), (CSAKI, 2018), (ZHIQIANG et al., 2020) |
| Regulatory (27) | (ANCARANI et al., 2019), (VELASCO et al., 2021), (MONDORF; WIMMER; REISER, 2013), (OZYUREK; ERDAL, 2018), (KLUN; SETNIKAR-CANKAR, 2013), (MONTEIRO; CORREIA, 2023), (OLIVEIRA et al., 2020), (TAS, 2020), (SIROTKINA; LAZAREVICH, 2023), (MARTINS et al., 2021), (UDUWAGE-DON; HADIWATTAGE; PANUWATWANICH, 2023), (STAKE, 2017), (SANGIL, 2020), (ASTBRINK; TIBBEN, 2013), (ALMEIDA et al., 2018), (MODRUŠAN; MRŠIĆ; RABUZIN, 2020), (MONDORF; WIMMER, 2008), (NURMANDI; KIM, 2015), (MENDES; VOIGT, 2022b), (PUTRI; RULDEVIYANI, 2019), (RIIHIAHO et al., 2015), (FERREIRA; AMARAL, 2016), (RODRIGUEZ et al., 2019), (VAZQUEZ-ROWE et al., 2021), (SILVA et al., 2018), (GAVUROVA; KUBAK, 2021), (ZHIQIANG et al., 2020) |
| Data quality (16) | (VELASCO et al., 2021), (MONDORF; WIMMER; REISER, 2013), (OZYUREK; ERDAL, 2018), (KLUN; SETNIKAR-CANKAR, 2013), (SOYLU et al., 2022b), (MARTINS et al., 2021), (NAI et al., 2023), (SANGIL, 2020), (ALMEIDA et al., 2018), (MENDES; VOIGT, 2022b), (FERREIRA; AMARAL, 2016), (RODRIGUEZ et al., 2019), (GAVUROVA; KUBAK, 2021), (SOYLU et al., 2022c), (CSAKI, 2018), (ZHIQIANG et al., 2020) |

MER; REISER, 2013), compliance testing (MONDORF; WIMMER; REISER, 2013), content analysis (BEHR; ABRAHAMSSON, 2022), qualitative content analysis (BJARNASON; PERSSON; RYDENFäLT, 2023) and usability evaluations (BJARNASON; PERSSON; RYDENFäLT, 2023). These methods are usually used in conjunction to achieve an effective data evaluation. According to Table 10, 21 studies reported using manual methods and techniques to evaluate public procurement data quality. This indicates that there is still room for other methods to improve the evaluation's effectiveness and scale, such as automated or semiautomated methods.

**Automatic Methods** employ algorithms and computational processes to analyze data. This involves techniques like automatic pattern identification, information extraction through text mining (ALMEIDA et al., 2018), and anomaly detection for irregularities (SOYLU et al.,

Table 17 – Stakeholder Groups and Their Needs in Public Procurement

| Stakeholders | Needs and Objectives |
|---|---|
| **Suppliers** | Access to data to identify business opportunities, respond to government demands, and ensure competitive proposals. Essential for understanding market needs, fulfilling contractual obligations, and fostering fair competition. |
| **Buyers** | Require transparent information about suppliers, contract history, and pricing to ensure efficient procurement decisions, optimize resource allocation, and comply with regulations. |
| **Researchers** | Need access to procurement data for analyzing trends, evaluating public spending, and producing studies that inform public policy and improve procurement processes. |
| **Policy Makers** | Use procurement data to craft evidence-based policies, assess the impact of government spending, and ensure alignment with economic and social objectives. |
| **Citizens** | Seek open and transparent data to monitor government procurement activities, ensure accountability, and contribute to public discourse and oversight. |
| **Governmental Institutions** | Require comprehensive data access to streamline internal processes, enhance collaboration between agencies, and ensure efficiency and transparency in public procurement. |

2022b). Beyond processing, automatic methods also utilize tools for data extraction to uncover hidden relationships and insights within the data. For example, one can cite tasks such as fraud detection (anomaly detection and pattern recognition in the data, with techniques like data clustering, transaction analysis, and identifying extreme values(SOYLU et al., 2022b) and risk ranking (VELASCO et al., 2021), linguistic term conversion (text mining techniques for reclassifying descriptive texts) (ALMEIDA et al., 2018), and electronic auction algorithms (blockchain technology) (ALMEIDA et al., 2018). According to Table 10, only four studies reported using automatic methods to evaluate public procurement data quality. This indicates that despite the effectiveness of these methods, only a few studies have reported their use.

**Statistical Methods** employ statistical techniques for analysis, like regression analysis and life cycle costing, to explore relationships between variables (IMAMOGLU; REHAN, 2015) (TAS, 2020) (SANGIL, 2020). Data comparison and evaluation techniques, including multiple regression and cluster analysis, employ statistical tools to compare datasets, identify patterns, and assess data quality (MELON; SPRUK, 2020). These methods offer a robust toolkit for quantitative data exploration and evaluation. Only seven studies reported using statistical methods to evaluate public procurement data quality.

**Semiautomated Methods** integrate multiple approaches, including combining manual, automatic, and statistical methods for a more comprehensive analysis. For example, in the evaluation of Open Government Data quality frameworks, combined methods utilize both automated algorithms and manual assessments to ensure thorough evaluations. Comparative

Table 18 – Stakeholder Representation Across Selected Studies

| **Stakeholders** | **References** |
|---|---|
| Suppliers | (ANCARANI et al., 2019), (MONDORF; WIMMER; REISER, 2013), (PINTO et al., 2015), (IMAMOGLU; REHAN, 2015), (STAKE, 2017), (SANGIL, 2020), (CONCHA; BURR; SUáREZ, 2014), (GAVUROVA; KUBAK, 2021), (SOYLU et al., 2022c) |
| Buyers | (IMAMOGLU; REHAN, 2015), (STAKE, 2017), (MODRUŠAN; MRŠIĆ; RABUZIN, 2020), (SILVA et al., 2018), (BALAEVA et al., 2022), (PATRUCCO; AGASISTI; GLAS, 2021), (GAVUROVA; KUBAK, 2021), (SOYLU et al., 2022c) |
| Researchers | (BEHR; ABRAHAMSSON, 2022), (TAS, 2020), (STAKE, 2017), (ASTBRINK; TIBBEN, 2013), (SILVA et al., 2018), (ZHIQIANG et al., 2020) |
| Policy makers | (VELASCO et al., 2021), (SOYLU et al., 2022b), (TAS, 2020), (MARTINS et al., 2021), (STAKE, 2017), (MELNIKOV; LUKASHENKO, 2019), (ASTBRINK; TIBBEN, 2013), (MONDORF; WIMMER, 2008), (MENDES; VOIGT, 2022b), (aO et al., 2023), (RODRIGUEZ et al., 2019), (ARNEY et al., 2014), (PATRUCCO; AGASISTI; GLAS, 2021), (SILVA et al., 2018), (MELON; SPRUK, 2020), (CSAKI, 2018), (ZHIQIANG et al., 2020) |
| Citizens | (TOSIN et al., 2016), (SOYLU et al., 2022b), (MENDES; VOIGT, 2022b), (PUTRI; RULDEVIYANI, 2019), (SOYLU et al., 2022c), (ZHIQIANG et al., 2020), (TYLLINEN et al., 2016) |
| Governmental institutions | (OZYUREK; ERDAL, 2018), (KLUN; SETNIKAR-CANKAR, 2013), (IMAMOGLU; REHAN, 2015), (MONTEIRO; CORREIA, 2023), (OLIVEIRA et al., 2020), (SIROTKINA; LAZAREVICH, 2023), (STAKE, 2017), (MELNIKOV; LUKASHENKO, 2019), (NAI et al., 2023), (ASTBRINK; TIBBEN, 2013), (ALMEIDA et al., 2018), (MUHWEZI et al., 2023), (MONDORF; WIMMER, 2008), (SPACEK; SPACKOVA, 2023), (DAHBI; CHIADMI; LAMHARHAR, 2023), (NURMANDI; KIM, 2015), (aO et al., 2023), (GONçALVES et al., 2010), (RIIHIAHO et al., 2015), (RODRIGUEZ et al., 2019), (LEE; OH; KWON, 2008), (VAZQUEZ-ROWE et al., 2021), (SILVA et al., 2018), (LALIĆ et al., 2019) |

summative usability evaluations combine automated usability tests with manual evaluations to provide comprehensive insights into user experiences (BALAEVA et al., 2022). Business intelligence and massive data processing techniques combine automated algorithms with manual interpretations to extract actionable insights from large datasets (VAZQUEZ-ROWE et al., 2021). Online surveys combine automated data collection with manual descriptive statistics analysis, providing an understanding of survey responses (RODRIGUEZ et al., 2019). BIM (Building Information Modeling) integrates automated modeling techniques with manual assessments to evaluate life cycle sustainability and make informed decisions in construction projects (TOSIN et al., 2016). REBUS-PLS combines automated effects' sizes calculation with manual permutation tests to assess the quality of structural equation models, ensuring robust and reliable analyses (TYLLINEN et al., 2016). Eighteen studies reported using semiautomated methods to evaluate public procurement data quality.

The evaluation approach methods indicate challenges and opportunities in public procurement data activities. This enriches the understanding of the presented evaluation methods related to SRQ1 and provides insights for effective improvement strategies.

### 4.3.2 Specific Research Question 2 (SRQ2)

This subsection addresses SRQ2 by examining the challenges related to discoverability, accessibility, and usability within public procurement data. The analysis categorizes these challenges into four issues, as mentioned in Subsection 4.2.2: Technological, Documental, Regulatory, and Data Quality. This discussion also offers insights for stakeholders navigating the complexities of managing and leveraging public procurement data.

**Discoverability Issues**. From the *technological* perspective, the discoverability of data landscapes is not trivial. Various data sources often use different formats and structures, which makes it difficult to integrate their requirements into a single strategy (PUTRI; RULDEVIYANI, 2019). The *TheyBuyForYou* platform (SOYLU et al., 2022c) and knowledge graph (DAHBI; CHIADMI; LAMHARHAR, 2023) are attempts to deal with these challenges, expanding horizons in public procurement through the use of open linked data. Heterogeneity in data formats requires robust search systems to handle diverse datasets, yet such systems often fall short, leading to inefficient discoverability (ALMEIDA et al., 2018). One possibility to tackle this situation is through mining techniques (ALMEIDA et al., 2018). Additionally, the need for interoperability further highlights these challenges (MONDORF; WIMMER, 2008). The lack of advanced tools and methodologies indicates the need for researchers and practitioners to focus on this problem (BEHR; ABRAHAMSSON, 2022). Addressing data relevance also poses significant discoverability issues. Technological barriers often prevent systems from efficiently filtering and presenting the most relevant data to users (VELASCO et al., 2021). For example, ineffective search algorithms and insufficient metadata tagging can result in irrelevant data being presented, complicating the user experience. This issue is compounded by interoperability hurdles between different systems and platforms, which inhibit seamless data sharing and retrieval (MONDORF; WIMMER; REISER, 2013) (MENDES; VOIGT, 2022b) (MAVIDIS; FOLINAS, 2022). These concerns, beyond impacting accessibility and usability, hinder users' ability to efficiently find and locate the data they need without having to sift through irrelevant or unrelated information.

From the *documental* perspective, information scarcity is a critical issue in discoverability. Many datasets are incomplete or poorly structured, leading to gaps in discovering available information (MONDORF; WIMMER; REISER, 2013). For instance, studies focusing on discoverability requirements for mobile apps highlight these deficiencies (OLIVEIRA et al., 2020). The lack of comprehensive data discoverability and collection practices can affect public procurement procedures, competition, and cost-effectiveness (ASTBRINK; TIBBEN, 2013) (TAS, 2020). Some studies argue that inadequate documentation protocols contribute to these gaps, as seen in the electronic public procurement case report (SIROTKINA; LAZAREVICH, 2023) and the broader challenges in public e-procurement (FERREIRA; AMARAL, 2016). While these issues can also affect accessibility and usability, these documental deficiencies impact the ability to discover information. Validation shortcomings of ontologies used to classify and organize data can hinder discoverability. If ontologies are not appropriately validated, they may fail to represent the data landscape accurately, causing difficult data categorization and searchability (TOSIN et al., 2016). This makes it difficult for users to find the needed data (LEE; OH; KWON, 2008).

*Regulatory* complexities impact the discoverability of data. Various regulations governing data access and usage create barriers to discovering useful data (MONTEIRO; CORREIA, 2023) (MONDORF; WIMMER, 2008). Different regions and sectors have varying compliance requirements, making it challenging to navigate and locate data that adheres to all relevant regulations (KLUN; SETNIKAR-CANKAR, 2013). This complexity is compounded by the lack of unified agreements and guidelines across jurisdictions, creating further obstacles to discovering data (MONTEIRO; CORREIA, 2023) (aO et al., 2023). Additionally, the absence of comprehensive guidelines for data interoperability among public authorities limits data discoverability (MAVIDIS; FOLINAS, 2022). When governmental bodies lack standardized protocols for data sharing, it results in fragmented data ecosystems where data is not easily accessible or discoverable (SOYLU et al., 2022b).

From the *data quality* perspective, discoverability issues within public procurement data can be attributed to shortcomings in metadata management and data quality. Inconsistent, incomplete, or inaccurate metadata hinders effective search and retrieval (BALAEVA et al., 2022). Furthermore, poorly structured or erroneous data within the databases themselves impedes discoverability (MODRUŠAN; MRŠIĆ; RABUZIN, 2020). Textual data stored in CSV files is susceptible to discoverability challenges due to improper handling, resulting in missing content, inconsistencies, and data loss (CSAKI, 2018).

**Accessibility Issues**. From the *technological* perspective, accessibility issues often stem from data integration complexities. Integrating diverse data sources into a coherent and accessible system requires overcoming significant technical challenges (NURMANDI; KIM, 2015). For instance, differences in data formats and structures necessitate sophisticated integration solutions to ensure that all relevant data is accessible and user-friendly (NAI et al., 2023) (RODRIGUEZ et al., 2019). Furthermore, the scarcity of tools and methodologies to facilitate this integration process hampers accessibility (MODRUŠAN; MRŠIĆ; RABUZIN, 2020) (VELASCO et al., 2021). Data accessibility is further influenced by authentication processes. Overly complex procedures can act as a deterrent, impeding user access to necessary data (ASTBRINK; TIBBEN, 2013).

From the *documental* perspective, data accessibility hinges on the quality of information structure and completeness. Well-organized and comprehensive datasets empower users to readily access and effectively utilize the data (SOYLU et al., 2022b) (aO et al., 2023). Structural deficiencies and lack of comprehensiveness in data sets can severely hinder accessibility. In such scenarios, users find it difficult to locate relevant information and effectively utilize the data due to its fragmented or poorly organized nature. (UDUWAGE-DON; HADIWATTAGE; PANUWATWANICH, 2023) (ALMEIDA et al., 2018).

*Regulatory* complexities significantly affect data accessibility, being an hindrances to

SMEs' involvement in public procurement, such as administrative requirements and a shortage of resources (ANCARANI et al., 2019). Continual changes in legislation is another major barrier to municipalities, highlighting accessibility issues due to evolving regulatory requirements (KLUN; SETNIKAR-CANKAR, 2013). Interoperability problems between different public authorities' profiles limit accessibility to information (SOYLU et al., 2022b). Moreover, the lack of agreements and guidelines on data sharing and accessibility further complicates the scenario (MONDORF; WIMMER; REISER, 2013).

From the *data quality* perspective, heterogeneous data characteristics impact accessibility. When data exhibits significant variability in terms of format, quality, and structure, it is a challenge to create systems that can universally access and process data (SOYLU et al., 2022b) (MONDORF; WIMMER, 2008) (CSAKI, 2018).

**Usability Issues**. From the *technological* perspective, there is a significant impact to usability of data systems. Interoperability issues between different systems and platforms can result in fragmented data, making it difficult for users to effectively utilize the data (IMAMOGLU; REHAN, 2015) (SIROTKINA; LAZAREVICH, 2023). When systems cannot seamlessly communicate, it leads to inconsistent data presentation and user experiences, hampering the overall usability (MONDORF; WIMMER, 2008). Furthermore, data integration complexities also play a critical role in affecting usability. Users often face difficulties when trying to work with integrated data from diverse sources due to inconsistencies and incompatibilities (TOSIN et al., 2016) (STAKE, 2017). Interface concerns are another critical issue affecting usability. Poorly designed user interfaces can make data systems difficult to navigate and use efficiently (PINTO et al., 2015) (SPACEK; SPACKOVA, 2023) (TYLLINEN et al., 2016). Interfaces that do not cater to user needs or are overly complex can significantly diminish the usability of a data system (SIROTKINA; LAZAREVICH, 2023).

From the *documental* perspective, issues such as structured and complete data deficiencies also impact usability. When data sets are incomplete or lack proper organization, it becomes challenging for users to make informed decisions based on the data (UDUWAGE-DON; HADIWATTAGE; PANUWATWANICH, 2023) (SANGIL, 2020) (CSAKI, 2018). Heterogeneous data characteristics further complicate usability (SOYLU et al., 2022b).

From the *regulatory* perspective, complex environments often lead to systems that are difficult to navigate and use effectively due to the need to comply with various legal requirements (UDUWAGE-DON; HADIWATTAGE; PANUWATWANICH, 2023) (BALAEVA et al., 2022). This complexity can result in interfaces that are not intuitive or user-friendly, further impacting usability (TAS, 2020) (SILVA et al., 2018). Different standardized protocols for data sharing and usability, result in inconsistent user experiences and difficulties in data utilization (LEE; OH; KWON, 2008). When public procurement systems lack clear usability guidelines, it results in interfaces that are not user-friendly, thereby restricting access to the data (PINTO et al., 2015) (MONTEIRO; CORREIA, 2023) (TYLLINEN et al., 2016).

From the *data quality* perspective, usability can be directly impacted. Issues like missing or inadequate data elements (SOYLU et al., 2022b), inconsistencies (UDUWAGE-DON; HADIWATTAGE; PANUWATWANICH, 2023) (DAHBI; CHIADMI; LAMHARHAR, 2023), and lack of validation checks (UDUWAGE-DON; HADIWATTAGE; PANUWATWANICH, 2023) are just some examples. Reported challenges and interpretability issues further hinder usability (NAI et al., 2023) (SOYLU et al., 2022c).

Figure 8 and Tables 19 through 22 present a detailed categorization of studies related to Technological, Documental, Regulatory, and Data Quality Issues impacting the Discoverability, Accessibility, and Usability of public procurement data. These tables summarize the key challenges identified across different research areas and provide references to the relevant studies that have addressed these issues. Each table is organized by specific characteristics of the issues, allowing for a clearer understanding of how these challenges affect various aspects of public procurement systems.


Figure 8 – Multidimensional Analysis of Challenges in Public Procurement Data: Technological, Regulatory, Documental, and Data Quality Perspectives Intrinsically Connected.

### 4.3.3 Specific Research Question 3 (SRQ3)

Evidence to answer Specific Research Question 3 (SRQ3) from the selected studies is categorized in (Table 14) as thematic domains and in (Table 15) as government sectors.

*Information Technology* stands out among thematic domains according to the evidence from 24 studies. They cover topics from interoperability testing in pan-European public service provision (MONDORF; WIMMER; REISER, 2013) to data integration within government procurement systems in Brazil(TOSIN et al., 2016). There is also a focus on AI governance in public procurement (BEHR; ABRAHAMSSON, 2022) and the discussion of the Decentralized Autonomous Organizations (DAOs) related to public procurement data (MONTEIRO; CORREIA, 2023). Interoperability in e-tendering across European states is examined (MONDORF; WIMMER, 2008), whereas there is room also for discussing intelligent monitoring systems in specific regions (MODRUŠAN; MRŠIĆ; RABUZIN, 2020). There is also the discussion of web accessibility concerns within electronic procurement platforms in Portugal (GONçALVES et al., 2010). This evidence illustrates the importance of Information Technology (IT) in fostering transparency, efficiency, and efficacy within public procurement. They delve into critical issues of interoperability and governance in public procurement systems, emphasizing the transformative potential of IT in streamlining and optimizing procurement practices.

On the *regulatory* front, the Spanish Public Sector Contracting Platform is the subject of analysis, shedding light on regulatory procedures (RODRIGUEZ et al., 2019). Additionally, another study discussed the quality assessment of the EU's Tenders Electronic Daily (TED) dataset (CSAKI, 2018), highlighting the significance of regulatory evaluations in ensuring efficiency and standards compliance in public procurement practices.

Regarding the *ethics* and *transparency* thematic domains, we found evidence of the examination of AI governance and ethics in public procurement processes (BEHR; ABRAHAMSSON, 2022), and also an investigation of fraud detection and assessment of transparency aspects of the public procurement procedures and portal in Brazil (PINTO et al., 2015)(ALMEIDA et al., 2018). A discussion about ethics and transparency in public procurement also takes place in the Sweden scenario, including minimizing bias in awarding contracts that impact small and medium-sized enterprises (SMEs), particularly when considering the ethical implications of MEAT (Most Economically Advantageous Tender) versus solely focusing on the lowest price (STAKE, 2017). We also found evidence of discussion regarding transparency and efficiency in e-government procurement in Malaysia, Thailand, and China (LEE; OH; KWON, 2008). The study underscores the significance of incorporating ethical considerations and transparency mechanisms in e-government procurement initiatives to uphold public confidence in the procurement process.

Six selected studies focus on the *healthcare* sector exploring different facets. Procurement activities related to COVID-19 in Germany were examined, shedding light on the strategies implemented during the pandemic (MENDES; VOIGT, 2022b). The efficiency evaluation within the healthcare sector in the Slovak Republic was scrutinized, highlighting the importance of optimizing processes for enhanced performance (GAVUROVA; KUBAK, 2021). We also found a study focusing on procuring health commodities in sub-Saharan Africa, emphasizing the challenges and strategies (ARNEY et al., 2014) and a study discussed a large-scale public IT system procurement in healthcare and social welfare domains, showcasing the complexities and considerations involved in such endeavors (TYLLINEN et al., 2016). These studies collectively provide valuable insights into the challenges, strategies, and dynamics of procurement practices within the healthcare domain across different regions, reflecting the diverse and critical nature of procurement activities in the healthcare sector.

*Sustainability* is also a thematic domain in one of the selected studies, discussing sustainable procurement practices, notably in the Federal Public Institution in Brazil (SILVA et al., 2018). We identified two studies focusing on the *construction* domain. The first discusses the enhancement of bids within Public-Private Partnerships in the Sri Lankan context (UDUWAGE-DON; HADIWATTAGE; PANUWATWANICH, 2023), and the second discusses seismic retrofitting of schools in Lima, Peru (VAZQUEZ-ROWE et al., 2021). These studies shed light on various aspects impacting different sectors and regions within the realm of public procurement.

**Government Sectors**: We identified 40 studies that delve extensively into national contexts, with a particular focus on specific countries such as Brazil (7 studies), Portugal (3 studies), Russia (3 studies), Italy (3 studies), European Union (EU) (5 studies), OECD Countries (1 study), Slovenia (2 studies), Hungary (2 studies), Slovak Republic (2 studies), Sweden (2 studies), and Indonesia (2 studies) (ANCARANI et al., 2019) (VELASCO et al., 2021)(TOSIN et al., 2016)(PINTO et al., 2015)(OLIVEIRA et al., 2020) (IMAMOGLU; REHAN, 2015) (SIROTKINA; LAZAREVICH, 2023)(MELNIKOV; LUKASHENKO, 2019)(BALAEVA et al., 2022) (STAKE, 2017)(BJARNASON; PERSSON; RYDENFäLT, 2023) (NAI et al., 2023) (ASTBRINK; TIBBEN, 2013) (SPACEK; SPACKOVA, 2023) (DAHBI; CHIADMI; LAMHARHAR, 2023) (MELON; SPRUK, 2020) (CSAKI, 2018). They provide in-depth insights into the different public procurement landscapes, practices, and challenges encountered within the context of each specific country, contributing valuable knowledge to the broader understanding of procurement processes across diverse national settings. Figure 9 illustrates the distribution of public procurement selected studies by country and region.

Specific *regional* studies provide valuable insights into public procurement practices such as public procurement data in Brazilian states (VELASCO et al., 2021)(aO et al., 2023), the province of Albay in the Philippines (SANGIL, 2020), and the state of North Rhine-Westphalia (NRW) in Germany (MENDES; VOIGT, 2022b). In a *local* context, we found a study that delves into administrative challenges in Slovenian municipalities (VELASCO et al., 2021), while another study provides insights from 92 out of 290 Swedish municipalities (SANGIL, 2020). We also found a study exploring e-procurement in Indonesian cities such as Yogyakarta, Tangerang, and Kutaikartanegara (MENDES; VOIGT, 2022b) and another paper that combines data from Italian and American municipalities to gain insights into various procurement variables (aO et al., 2023).

The analysis of thematic domains and government sectors addresses SRQ3 by providing a structured overview of the research landscape. It offers insights for future exploration and decision-making by incorporating various topics and specific case studies from different countries.

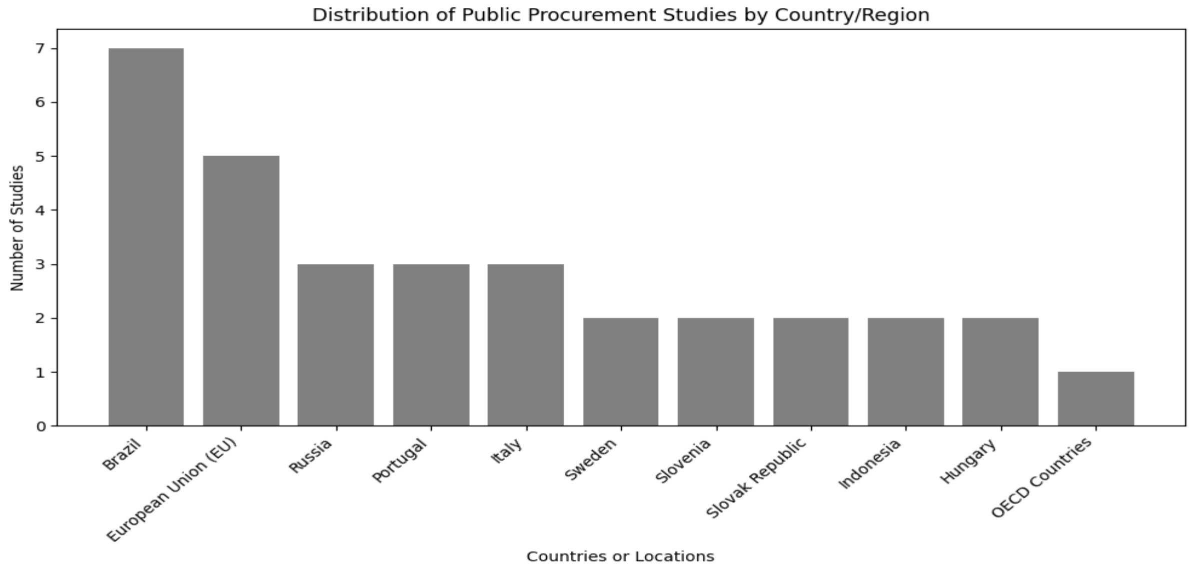


Figure 9 – Geographic Distribution of Studies on Public Procurement: Countries and Regions.

This approach answers RQ and emphasizes the global significance and applicability of the findings, being a reference for policymakers, practitioners, and researchers aiming to improve government procurement practices worldwide.

### 4.3.4 Specific Research Question 4 (SRQ4)

Specific Research Question 4 (SRQ4) outlines the best practices identified in selected studies for enhancing public procurement data quality.

From the **technological** perspective, we found 28 studies discussing best practices for public procurement data (Table 16). Based on evidence from these studies, it is possible to conclude that adopting technological solutions is a powerful tool for promoting transparency and competition within the procurement landscape (ZHIQIANG et al., 2020) and prioritizing technological interventions based on effectiveness can be achieved through *Importance-Performance Analysis* (PUTRI; RULDEVIYANI, 2019). Investing in information and communication technologies (ICT) and providing technical support can further modernize the public marketplace (ANCARANI et al., 2019). Implementing decision support systems and automating data mapping based on relational databases can further streamline the process and improve efficiency (VELASCO et al., 2021) (TOSIN et al., 2016). Technological advancements like electronic submissions and standardized data formats can also significantly streamline the process (KLUN; SETNIKAR-CANKAR, 2013) (TOSIN et al., 2016). Implementing interoperable systems specifically designed for e-procurement is another recommendation (IMAMOGLU; REHAN, 2015). However, addressing technological interoperability challenges remains vital for smooth operations (MONDORF; WIMMER, 2008).

Transparency and data quality are essential for trustworthy public procurement. Data analysis techniques like information retrieval and graph analysis facilitate efficient data integration (NAI et al., 2023). Building audit trails and reference price databases effectively deter fraud (aO et al., 2023). Text mining and knowledge graphs help standardize data and identify potential frauds (ALMEIDA et al., 2018) (DAHBI; CHIADMI; LAMHARHAR, 2023). Leveraging semantic technologies can integrate disparate open data sources, leading to a more comprehensive and coherent data landscape (SOYLU et al., 2022c).

Public procurement systems must also be accessible and secure. Using web accessibility evaluation tools ensures compliance with accessibility standards, making the system inclusive for all (GONçALVES et al., 2010). A clear and transparent public procurement system with well-designed interfaces is essential for accessibility (PINTO et al., 2015). Role-Based Access Control (RBAC) protocols can help ensure that only authorized users have access to data and systems within Decentralized Autonomous Organizations (DAOs) used in procurement (MONTEIRO; CORREIA, 2023).

Usability is another important factor. Focusing on the various aspects of transparency can lead to a more user-friendly experience (PINTO et al., 2015). Making public procurement data user-friendly and accessible, along with adopting Decentralized Autonomous Organizations (DAOs) and electronic auctions, are all strategies that can further enhance the efficiency, transparency and fairness of public procurement processes (SOYLU et al., 2022b) (MELNIKOV; LUKASHENKO, 2019) (MONTEIRO; CORREIA, 2023) (ZHIQIANG et al., 2020). Incorporating ergonomics into the design of digital systems used in procurement ensures a user-friendly experience for everyone (BJARNASON; PERSSON; RYDENFäLT, 2023). User-centric design should also be a priority throughout the procurement process, as evidenced by the importance of clearly defined user scenarios in evaluations (TYLLINEN et al., 2016).

Beyond the strategies mentioned above, several additional considerations are important. Establishing a central Electronic Public Procurement Platform (PeP) can improve service delivery and overall efficiency (IMAMOGLU; REHAN, 2015). Leveraging interoperability testing across European e-government services ensures seamless integration (MONDORF; WIMMER; REISER, 2013). Integrating ethical considerations into artificial intelligence (AI) in procurement practices is crucial for responsible decision-making (BEHR; ABRAHAMSSON, 2022).

The **documental** perspective. Reducing entry costs, such as financial requirements and excessive contract bundling, can make the process more accessible for a wider range of participants processes (ANCARANI et al., 2019). Simplifying the procurement process goes hand-in-hand with clear and well-organized documentation. Establishing clear and consistent documentation practices is crucial. This includes creating well-defined artifacts and application profiles, implementing organized tender evaluation practices, and standardizing instructions for tender dossier preparation (MONDORF; WIMMER; REISER, 2013)(OZYUREK; ERDAL, 2018) (KLUN; SETNIKAR-CANKAR, 2013).

Transparency and data quality are fundamental to ensuring the integrity of procurement documentation. Emphasis on transparency policies, data standardization, and accessibility is paramount (ASTBRINK; TIBBEN, 2013). Utilizing techniques like developing customized code for classification in transparency portals and data triangulation for comprehensive analysis further contribute to organized and reliable information flow (ALMEIDA et al., 2018) (SILVA et al., 2018). Furthermore, adhering to quality management standards like ISO 9001:2008 and addressing technical issues in open data generation ensure the overall quality and trustworthiness of the documentation (LALIĆ et al., 2019) (CSAKI, 2018). Finally, developing robust information governance frameworks provides a strong foundation for maintaining documental coherence throughout the procurement process (ZHIQIANG et al., 2020).

The **regulatory** perspective. Ensuring fair and open competition is a cornerstone of effective public procurement. This can be achieved by reducing unnecessary pre-qualification requirements and focusing on post-contractual arrangements that ensure quality and delivery (ANCARANI et al., 2019). Guidelines and recommendations specifically designed for SMEs can further level the playing field (STAKE, 2017). Furthermore, improved regulation that promotes competitive and cost-effective procurement practices fosters fairness for all participants (TAS, 2020).

Clear regulations and consistent compliance are essential for a well-functioning procurement system. Monitoring legislative changes and ensuring good governance of ontologies, the standardized categories used to classify information, contribute to regulatory clarity (MONDORF; WIMMER; REISER, 2013). Additionally, promoting transparency through accessibility standards and collaboration with stakeholders helps ensure adherence to regulatory norms (OLIVEIRA et al., 2020). Compliance with national legislation specifically focused on transparency is also crucial (FERREIRA; AMARAL, 2016). Regular review and potential reform of national procurement legislation can further strengthen the regulatory framework (BALAEVA et al., 2022).

Accessibility and environmental considerations are increasingly important aspects of public procurement regulations. Emphasis on transparent and effective monitoring regimes ensures that accessibility standards are met (ASTBRINK; TIBBEN, 2013). This focus on accessibility should be driven by a strong ethical and moral obligation to create a fair and inclusive system (OLIVEIRA et al., 2020). Similarly, mandatory adoption of accessibility criteria when procuring information and communication technologies (ICT) promotes inclusivity for users with disabilities (ASTBRINK; TIBBEN, 2013). Additionally, providing guidelines and training opportunities for people with disabilities is crucial for ensuring they can fully participate (GONçALVES et al., 2010). Finally, establishing common standards and guidelines for implementing environmental policies strengthens the overall regulatory framework for sustainable public procurement practices (VAZQUEZ-ROWE et al., 2021).

The **data quality** perspective. Reliable and accurate data is essential for effective public procurement. A focus on improving the overall quality of public spending data through techniques like data mining can also lead to better quality procurement data (VELASCO et al., 2021).

Developing tools specifically designed to ensure the quality of ontologies, the standardized categories used to classify information further promote data consistency and reliability (MONDORF; WIMMER; REISER, 2013). Compliance with established data quality standards, such as EN 301 549[2], and addressing any identified data quality issues are also important steps (MARTINS et al., 2021).

Standardization methods also play a key role, as focusing on these methods helps ensure reliable data (ALMEIDA et al., 2018). Data quality improvement is ongoing, and continuous efforts to improve the quality of disclosed procurement data are essential (SANGIL, 2020). Emphasizing transparency policies, data standardization, and accessibility creates an environment that fosters overall data quality (MENDES; VOIGT, 2022b).

Beyond quality, ensuring comprehensive data collection and building trust in the system are crucial. Techniques like web scraping to gather relevant information can contribute to a more complete data set (NAI et al., 2023). Trust in the central body responsible for system security is also paramount, as it helps ensure the overall quality and reliability of the data (SPACEK; SPACKOVA, 2023).

While the strategies mentioned above are key, there are other factors to consider. Enhancing procurement knowledge to reduce issues with bid responsiveness can indirectly improve data quality (UDUWAGE-DON; HADIWATTAGE; PANUWATWANICH, 2023). Understanding the relationships between quality management practices and time-based performance also contributes to overall data quality (LALIĆ et al., 2019).

**Other Practices Recommendations**: Public procurement can be more inclusive by fostering collaboration among small and medium-sized enterprises (SMEs) and established industry players (ANCARANI et al., 2019). Accessibility is another key aspect of inclusivity. Collaboration and clear communication are essential to ensure accessibility requirements are effectively integrated into the procurement process (OLIVEIRA et al., 2020).

Informed decision-making is essential for successful public procurement. This requires careful consideration of internal and external factors that can affect the use of information and communication technologies (ICT) within companies (CONCHA; BURR; SUáREZ, 2014). A comprehensive approach also considers domestic, international, and political-economic factors that can influence e-procurement (MELON; SPRUK, 2020). Recognizing the benefits of SME participation in e-procurement framework agreements can encourage greater engagement from this important sector (CONCHA; BURR; SUáREZ, 2014).

Examination of best practices for enhancing public procurement data quality addresses SRQ4 by providing a thorough overview of recommendations across technological, regulatory, and other domains. Specific strategies such as technological advancements, regulatory clarity,

2 https://www.etsi.org/deliver/etsi$_e n/301500_301599/301549/03.02.01_60/en_301549v030201p.pdf$

data quality enhancement, and usability improvements are explored, offering actionable insights and emphasizing data quality in procurement processes. The focus on collaboration, transparency, and ethical considerations underscores the holistic approach to cultivating an environment conducive to reliable and transparent procurement practices. This analysis serves as a valuable resource for stakeholders seeking to improve the integrity and efficacy of public procurement systems worldwide .

### 4.3.5 Addressing the Primary Research Question (RQ): Evaluation of the Quality of Government Procurement Data

We found evidence in selected studies regarding public procurement data quality, addressing the primary Research Question (RQ): "How have public procurement data been evaluated from the quality perspectives of discoverability, accessibility, and usability?". The analysis of secondary research questions (SRQs) provided valuable insights into the effectiveness of answering the primary RQ. This section synthesizes the findings from the specific research questions (SRQs) to provide a comprehensive answer to the main research question (RQ).

We identified and categorized the evaluation methods employed by selected studies to analyze the quality of open government data in public procurement processes. These categorizations include manual, automatic, statistical, and semi-automated techniques and highlight the data quality assessment. This categorization allows for a clear understanding of the different approaches used to assess data quality, showcasing a variety of methods employed in the literature.

The challenges related to discoverability, accessibility, and usability are categorized into technological, documental, regulatory, and data quality issues, underscoring the complexities involved in ensuring reliable and usable public procurement data. This categorization reveals the multifaceted nature of the problems affecting public procurement data, which are not solely technical but also involve issues of documentation, regulation, and inherent data quality.

We discussed diverse perspectives and areas of interest within public procurement data quality evaluation by examining government sectors and thematic domains such as information technology, regulation, and transparency. We identified strategies across technology, regulation, documentation, and data quality domains to improve discoverability, accessibility, and usability of public procurement data. These findings highlight the importance of a multidisciplinary approach to improving data quality, emphasizing that solutions must address technical, regulatory, and organizational aspects to achieve effective results. By evaluating these key aspects addressed in the secondary research questions (SRQs) and aligning them with the primary Research Question (RQ) on public procurement data quality assessment, the study provides a comprehensive understanding of addressing quality concerns in public procurement data. This alignment of the SRQs with the RQ demonstrates how each specific question contributes to the broader understanding of public procurement data quality.

## 4.4 Perspectives and Challenges in the Use of Public Procurement Data

Based on the discussion from Section 4.3, this section presents the challenges impacting public procurement data quality, giving its importance to government transparency and efficiency. It highlights the relevance of discoverability, accessibility, and usability in data by exploring these perspectives.

***Challenges Related to Data Evaluation Methods***: Public procurement data analysis employs a diverse toolkit of evaluation methods, each with its limitations. Manual methods, while insightful, can be time-consuming and subjective. Conversely, automated techniques can be fast but overlook subtle data patterns and require careful setup. Statistical methods offer quantifiable results but may struggle with complex data. Finally, semi-automated approaches combine these strengths but require smooth integration. To navigate these challenges, researchers must adopt a balanced approach, acknowledging the limitations and strengths of each method while aligning them with the specific research goals.

***Perspectives Related to Data Evaluation Methods***: To address challenges reported in evaluation methods in public procurement analysis, embracing technological advancements streamlines data analysis and overcomes limitations posed by manual and automatic methods. Investment in innovative tools and methodologies enhances evaluation processes' speed, accuracy, and scalability, unlocking new opportunities for insights and decision-making. Implementing a strategy encompassing technological advancement, collaboration, data quality assurance, and regulatory adherence is imperative for unlocking the full potential of evaluation methods.

***Challenges Related to Discoverability, Accessibility, and Usability Issues***: The complexity of public procurement data landscapes poses challenges to efficient analysis and processing. It arises from the wide variation in data sources regarding format, structure, and accessibility. The heterogeneity complicates efforts to streamline data analysis and reduces the effectiveness of processing methods. The sheer volume of data also exacerbates challenges, impeding stakeholders' ability to extract meaningful insights promptly. The challenges in procurement systems include difficulty in identifying relevant data among vast information, exacerbated by the lack of standardized protocols for data classification and prioritization. This impedes the extraction of actionable insights from procurement datasets. Interoperability challenges within procurement systems hinder effective data sharing and management. Fragmentation often leads to data silos, restricting access and collaboration among stakeholders. This lack of interoperability results in inefficiencies and impedes information exchange. Moreover, complexities in data integration exacerbate these challenges, demanding significant time and resources for reconciling disparate data sources. For these reasons, adopting a framework can help deal with these challenges.

Studies have highlighted challenges in efficiently accessing necessary data within procurement systems, primarily due to limitations in tools and methodologies. Scarce resources and outdated technology impede progress in data acquisition and analysis, thereby restricting the efficacy of procurement processes. The absence of adoption of standardized tools and methodologies exacerbates these difficulties, hindering stakeholders from effectively utilizing available data. CKAN is one framework widely used to implement open government data portals. [3]

The complexities inherent in procurement systems pose significant hurdles, namely challenges with search functionality, authentication processes, and interface design. Ineffective search algorithms and authentication procedures can hinder access to vital data, impeding decision-making processes and reducing productivity. Interface issues, such as poor user experience and system inefficiencies, compound usability challenges, making it challenging for stakeholders to navigate procurement systems efficiently.

***Perspectives Related to Discoverability, Accessibility, and Usability Issues***: Addressing technological challenges in procurement systems offers transformative opportunities to enhance system efficiency and effectiveness.

From a data relevance perspective, fostering collaboration and knowledge-sharing among stakeholders presents an opportunity to refine data prioritization and classification. Establishing cross-functional teams and collaborative platforms enables procurement professionals to collectively pinpoint pertinent data points and organize information in alignment with strategic objectives. Advocating for data literacy initiatives and training programs empowers stakeholders to make informed decisions and extract meaningful insights from procurement datasets, ultimately enhancing data relevance and usability.

Interoperability challenges in procurement systems present an opportunity to advocate for standardized data formats and protocols. Procurement professionals can facilitate seamless information exchange and collaboration by promoting interoperable data standards and fostering data-sharing agreements among stakeholders. Investing in interoperability infrastructure and governance frameworks can establish a more cohesive and integrated procurement ecosystem, enhancing data accessibility and utilization.

To overcome data access limitations, it is essential to prioritize investments in data management technologies and capacity-building initiatives. This involves enhancing data infrastructure and modernizing procurement systems to streamline data acquisition processes and improve accessibility. Promoting data-sharing initiatives and open data policies enhances transparency and availability, empowering stakeholders to make informed decisions and foster innovation within procurement systems.

From a user experience standpoint, there is a chance to enhance stakeholder engagement and satisfaction by prioritizing usability and interface design improvements. Incorporating user

[3] https://www.ckan.org/

feedback and user-centered design principles into procurement system development can lead to intuitive interfaces, boosting productivity and workflow efficiency. Investing in user training and support programs can empower stakeholders to navigate procurement systems confidently, fostering greater adoption and utilization of these essential tools.

Leveraging advanced technologies and innovative solutions can overcome data landscape obstacles. Investing in data integration platforms and analytics tools streamlines data processing workflows, enabling stakeholders to extract actionable insights more efficiently. Additionally, adopting emerging technologies such as artificial intelligence and machine learning automates tasks and enhances decision-making processes within procurement systems.

***Perspectives and Challenges Related to Thematic Domains Analysis***: In exploring the perspectives and challenges of thematic domain analysis in public procurement data quality, a broad spectrum of insights have surfaced from studies across diverse sectors.

Thematic analysis of public procurement research reveals information technology (IT) as a dominant theme. Studies delve into specific areas like interoperability testing, AI ethics, and web accessibility for e-procurement platforms. This focus highlights the transformative power of IT in enhancing transparency and efficiency. From an IT perspective, digital tools and investments in infrastructure, data integration, and cybersecurity are crucial for streamlining processes, improving data quality and accessibility, mitigating risks, and unlocking the full potential of digitalization in procurement.

Public procurement regulations require scrutiny to ensure data accuracy and reliability. Regulatory revisions are proposed to address inconsistencies that hinder compliance and efficiency. Stakeholders advocate for standardized protocols, clear guidelines, and oversight mechanisms to strengthen compliance and uphold regulatory standards.

Ethics and transparency represent a critical thematic domain in public procurement, with studies focusing on AI governance, fraud detection, and transparency in procurement systems across different nations. These investigations highlight the importance of accountability and integrity in procurement practices, emphasizing the need for transparency measures to prevent corruption and promote fair competition.

Emerging research highlights the critical role of public procurement in healthcare, emphasizing its impact on securing essential resources and ensuring their effectiveness (especially during crises). Furthermore, sustainability is gaining traction, with studies exploring how procurement practices can promote environmental and social responsibility through long-term goals.

***Perspectives and Challenges Related to Govern Sectors Analysis***: Across all levels of government, several common challenges in public procurement emerge. These include ensuring transparency, combating fraud, and maintaining high data quality.

National public procurement policies prioritize overarching frameworks but exhibit variations. While there is a need to prioritize SME inclusion through simplified procedures, it is also relevant to emphasize sustainability by integrating environmental and social considerations. A key challenge lies in balancing robust regulations with flexibility, exemplified by the need for evolving AI governance frameworks that address ethical concerns without stifling innovation.

Regional public procurement practices act as a crucial bridge, translating national policies into specific local needs, especially in areas with diverse economic and social landscapes. However, ensuring consistency and interoperability across jurisdictions remains a challenge. Initiatives like the Pan-European Public Procurement Online exemplify efforts to standardize and facilitate seamless public procurement across regions.

Public procurement by local governments presents a valuable opportunity for targeted community development and economic stimulation. However, achieving efficiency and transparency can be challenging. While some municipalities succeed through centralization, standardization, and digitalization, others struggle to implement national policies, risking corruption and mismanagement. This highlights the need for adaptable public procurement approaches that acknowledge national frameworks and local contexts for optimal outcomes.

Despite the global reach of public procurement research, encompassing studies from Europe, America, Asia, Africa, and Oceania, some regions like Africa, Central America, parts of South America, Japan, Australia, and emerging economies (India, South Africa) present a relative scarcity of in-depth analyses in published studies. This gap necessitates further research to illuminate public procurement practices in these regions and countries.

***Perspectives and Challenges Surrounding Best Practices Related to Discoverability, Accessibility, and Usability in Public Procurement*** : In public procurement, improving discoverability, accessibility, and usability practices is pivotal for fostering inclusivity, transparency, and efficiency.

Technological advancements offer significant potential to improve public procurement discoverability. Investments in ICT infrastructure, interoperability testing, and standardized data formats can streamline processes, reduce costs, and promote SME participation. However, challenges remain in navigating the complexities of integrating diverse regional and sectoral procurement systems alongside the substantial investment required.

Public procurement data accessibility requires prioritizing user-friendly interfaces and robust data models. Web accessibility tools, public-private partnerships, and automation all contribute by improving platform usability and streamlining operations. Ethical considerations and inclusivity mandates further emphasize ensuring accessibility for all stakeholders, even in the face of potential resource constraints and resistance to change. Overcoming these hurdles necessitates comprehensive stakeholder engagement and awareness campaigns.

Public procurement systems can achieve optimal usability through clear instructions, intuitive interface design, and accessibility considerations. This user-centric approach achieved through transparency, ergonomics, and collaboration, fosters efficient navigation, inclusivity, and stakeholder satisfaction. However, aligning stakeholder needs and balancing usability with other objectives requires iterative testing and continuous improvement cycles.

## 4.5 Threats to Validity in Systematic Mapping

In this section, we discuss the threats to the validity based on the guidelines proposed by Kitchenham and Charters (GROUP, 2007). In the next paragraphs, we consider the following types of threats: construct validity, internal validity, and external validity (ZHOU et al., 2016).

*Threats to Construct Validity.* Using different terminologies for public procurement data-related concepts can lead to misclassification. We included well-known terms from the theme (REJEB et al., 2023) (PURNOMO et al., 2021) to mitigate this possible threat. We selected and peer-reviewed the studies iteratively, considering the steps presented in Table 1 and the criteria listed in Table 8.

*Threats to Internal Validity*. We identify that one of the major issues is the risk of missing relevant studies. For this reason, we adopted the guidelines proposed by Kitchenham and Charters (GROUP, 2007) to define and validate our search string and selected five digital libraries to execute it. Possible biases in applying the inclusion and exclusion criteria are another threat to the validity. We mitigated this threat by examining each selected paper by at least two co-authors.

*Threats to External validity*. There is a risk of not identifying and extracting the relevant information from the selected studies or inaccurate interpretation of the extracted data. A possible consequence would be a mistaken mapping of some selected studies. We mitigated this threat by examining each selected paper by at least two co-authors and dealing with eventual discrepancies in a meeting discussion.

## 4.6 Chapter Overview

Public procurement data plays an important role in the context of open government data due to its contribution to transparency, participatory governance, accountability, competition among government suppliers, administrative efficiency, and corruption control (ATTARD et al., 2015) (FAZEKAS; CZIBIK, 2021) (BAUHR et al., 2020). The availability of government data and its procedures allows for monitoring by those outside (MEIJER, 2013). The quality of public spending, especially those related to public procurement processes and outcomes, can be evaluated through specific and reliable measures that require data quality awareness (FAZEKAS; CZIBIK, 2021).

In this study, we investigated quality concerns of public procurement data based on evidence from the literature regarding discoverability, accessibility, and usability issues, presenting a comprehensive categorization of these issues from the technological, regulatory, documental, and data quality perspectives. Discussing and identifying gaps in previous research highlights the importance of addressing the challenges to enhance transparency and efficiency in government procurement processes. Stakeholders play an important role in this scenario due to the possibility of improving data quality according to their needs. The findings confirm the role of collaboration, transparency, and ethical considerations in fostering reliable global procurement practices by means of evaluation of data quality dimensions, examination of challenges related to discoverability, accessibility, and usability, and analysis of evaluation methods, stakeholder diversity, and best practices.

Future research could explore the impact of implementing the recommended best practices on government procurement data quality through longitudinal studies. Investigation into the integration of these practices with existing data management systems could be valuable. It would also be potentially useful to explore the impact of emerging technologies such as blockchain and artificial intelligence, specifically large language models, on the characterization of quality concerns of public procurement data. We plan to conduct a longitudinal study to track the evolution of data quality concerns over time and undertake comparative analyses of public procurement policies and practices across regions.

In summary, this chapter critically examines the importance of data quality in public procurement, focusing on discoverability, accessibility, and usability through a systematic mapping. It explores various evaluation methods and emphasizes the significance of data quality across different government sectors and thematic domains, addressing the challenges stakeholders face when accessing and utilizing procurement data. Best practices are outlined with an emphasis on technological advancements, document management, regulatory frameworks, and data quality considerations, all aimed at improving transparency and operational efficiency in procurement processes. It stabilishes a foundational understanding of the current state of public procurement data quality, setting the stage for exploring innovative solutions in subsequent chapter.

Table 19 – Studies Addressing Technological Issues to Procurement Data Quality

| **Issues** | **Discoverability** | **Accessibility** | **Usability** |
|---|---|---|---|
| Diverse Data Landscapes | (ASTBRINK; TIBBEN, 2013), (SOYLU et al., 2022c) | | (VELASCO et al., 2021), (TAS, 2020) |
| Addressing Data Relevance | (VELASCO et al., 2021), (NAI et al., 2023), (TAS, 2020), (DAHBI; CHIADMI; LAMHARHAR, 2023), (RODRIGUEZ et al., 2019) | (VELASCO et al., 2021) | |
| Interoperability Hurdles | (MONDORF; WIMMER; REISER, 2013), (MENDES; VOIGT, 2022b), (FERREIRA; AMARAL, 2016) | (MONDORF; WIMMER; REISER, 2013), (UDUWAGE-DON; HADIWATTAGE; PANUWATWANICH, 2023), (MONDORF; WIMMER, 2008), (BALAEVA et al., 2022) | (MONDORF; WIMMER; REISER, 2013), (IMAMOGLU; REHAN, 2015), (SIROTKINA; LAZAREVICH, 2023) |
| Data Integration Complexities | (NAI et al., 2023), (SPACEK; SPACKOVA, 2023), (NURMANDI; KIM, 2015) | (NAI et al., 2023), (MONDORF; WIMMER, 2008), (NURMANDI; KIM, 2015), (RODRIGUEZ et al., 2019) | (TOSIN et al., 2016), (MUHWEZI et al., 2023) |
| Scarce Tools and Methodologies | (BEHR; ABRAHAMSSON, 2022) | (MODRUŠAN; MRŠIĆ; RABUZIN, 2020), (VAZQUEZ-ROWE et al., 2021) | |
| Search Challenges for PP Systems | (PINTO et al., 2015), (UDUWAGE-DON; HADIWATTAGE; PANUWATWANICH, 2023), (aO et al., 2023), (PUTRI; RULDEVIYANI, 2019), (BALAEVA et al., 2022) | (DAHBI; CHIADMI; LAMHARHAR, 2023), (MENDES; VOIGT, 2022b), (RODRIGUEZ et al., 2019) | |
| Authentication Processes | | (PINTO et al., 2015), (aO et al., 2023) | (ASTBRINK; TIBBEN, 2013) |
| Interface Concerns | (SIROTKINA; LAZAREVICH, 2023), (MARTINS et al., 2021) | (MARTINS et al., 2021), (ASTBRINK; TIBBEN, 2013), (GONçALVES et al., 2010), (PUTRI; RULDEVIYANI, 2019) | (PINTO et al., 2015), (MARTINS et al., 2021), (UDUWAGE-DON; HADIWATTAGE; PANUWATWANICH, 2023), (SPACEK; SPACKOVA, 2023), (DAHBI; CHIADMI; LAMHARHAR, 2023), (aO et al., 2023), (BALAEVA et al., 2022), (RODRIGUEZ et al., 2019), (TYLLINEN et al., 2016) |
| Fragmented PP Systems | (IMAMOGLU; REHAN, 2015), (aO et al., 2023), (ARNEY et al., 2014) | (aO et al., 2023) | (MONDORF; WIMMER, 2008) |
| Heterogeneous Data Sources | (SOYLU et al., 2022b) | | |
| Diverse Data Formats | (ALMEIDA et al., 2018), (MONDORF; WIMMER, 2008), (SOYLU et al., 2022c) | (SOYLU et al., 2022b), (ALMEIDA et al., 2018), (aO et al., 2023), (PUTRI; RULDEVIYANI, 2019) | (CONCHA; BURR; SUáREZ, 2014), (CONCHA; BURR; SUáREZ, 2014) |

Table 20 – Studies Addressing Documental Issues to Procurement Data Quality

| **Issues** | **Discoverability** | **Accessibility** | **Usability** |
|---|---|---|---|
| Information Scarcity | (MONDORF; WIMMER; REISER, 2013), (OLIVEIRA et al., 2020), (TAS, 2020), (ASTBRINK; TIBBEN, 2013), (SPACEK; SPACKOVA, 2023), (MENDES; VOIGT, 2022b), (RIIHIAHO et al., 2015), (RIIHIAHO et al., 2015), (GAVUROVA; KUBAK, 2021) | (SIROTKINA; LAZAREVICH, 2023), (FERREIRA; AMARAL, 2016) | |
| Ontology Validation Shortcomings | (TOSIN et al., 2016) | (LEE; OH; KWON, 2008) | |
| Structured and Complete Data Deficiency | (SOYLU et al., 2022b), (MODRUŠAN; MRŠIĆ; RABUZIN, 2020), (DAHBI; CHIADMI; LAMHARHAR, 2023), (MENDES; VOIGT, 2022b), (aO et al., 2023), (BALAEVA et al., 2022) | (UDUWAGE-DON; HADIWATTAGE; PANUWATWANICH, 2023), (ALMEIDA et al., 2018) | (UDUWAGE-DON; HADIWATTAGE; PANUWATWANICH, 2023), (SANGIL, 2020), (CSAKI, 2018) |
| Heterogeneous Data Characteristics | (MONDORF; WIMMER, 2008), (LEE; OH; KWON, 2008), (CSAKI, 2018) | (SOYLU et al., 2022b), (aO et al., 2023), (CSAKI, 2018) | (SOYLU et al., 2022b), (NURMANDI; KIM, 2015), (BALAEVA et al., 2022), (CSAKI, 2018) |

Table 21 – Studies Addressing Regulatory Issues to Procurement Data Quality

| **Issues** | **Discoverability** | **Accessibility** | **Usability** |
|---|---|---|---|
| PP Regulatory Complexity | (MONTEIRO; CORREIA, 2023), (SIROTKINA; LAZAREVICH, 2023), (MONDORF; WIMMER, 2008) | (ANCARANI et al., 2019), (KLUN; SETNIKAR-CANKAR, 2013), (STAKE, 2017), (RIIHIAHO et al., 2015), (CONCHA; BURR; SUáREZ, 2014) | (UDUWAGE-DON; HADIWATTAGE; PANUWATWANICH, 2023), (BALAEVA et al., 2022) |
| Lack of Agreements and Guidelines | (MONTEIRO; CORREIA, 2023), (MODRUŠAN; MRŠIĆ; RABUZIN, 2020), (MENDES; VOIGT, 2022b), (aO et al., 2023), (RIIHIAHO et al., 2015), (RODRIGUEZ et al., 2019), (VAZQUEZ-ROWE et al., 2021), (ZHIQIANG et al., 2020) | (MONDORF; WIMMER; REISER, 2013), (OLIVEIRA et al., 2020), (SANGIL, 2020), (MENDES; VOIGT, 2022b), (GONçALVES et al., 2010), (GAVUROVA; KUBAK, 2021) | (TAS, 2020), (MUHWEZI et al., 2023), (BJARNASON; PERSSON; RYDENFäLT, 2023), (DAHBI; CHIADMI; LAMHARHAR, 2023), (ARNEY et al., 2014), (SILVA et al., 2018), (LALIĆ et al., 2019) |
| Usability Rule Gaps in PP System Interfaces | (NURMANDI; KIM, 2015) | (SPACEK; SPACKOVA, 2023) | (PINTO et al., 2015), (MONTEIRO; CORREIA, 2023), (RIIHIAHO et al., 2015), (FERREIRA; AMARAL, 2016), (BALAEVA et al., 2022), (TYLLINEN et al., 2016) |
| Interoperability Issues among Public Authorities | (FERREIRA; AMARAL, 2016) | (SOYLU et al., 2022b) | (LEE; OH; KWON, 2008) |

Table 22 – Studies Addressing Data Quality Issues in Public Procurement

| **Issues** | **Discoverability** | **Accessibility** | **Usability** |
|---|---|---|---|
| Validation of Document Instances | (MONTEIRO; CORREIA, 2023), (ARNEY et al., 2014) | (SIROTKINA; LAZAREVICH, 2023) | (MONDORF; WIMMER; REISER, 2013) |
| Semantic Annotation Implementation | (UDUWAGE-DON; HADIWATTAGE; PANUWATWANICH, 2023) | (MONTEIRO; CORREIA, 2023), (MONDORF; WIMMER, 2008) | (TOSIN et al., 2016), (NAI et al., 2023) |
| Dataset Fragmentation and Granularity Deficits | (OZYUREK; ERDAL, 2018) | (aO et al., 2023) | (OZYUREK; ERDAL, 2018) |
| Concerns Regarding Poor Data Quality | (ASTBRINK; TIBBEN, 2013), (MODRUŠAN; MRŠIĆ; RABUZIN, 2020), (DAHBI; CHIADMI; LAMHARHAR, 2023), (BALAEVA et al., 2022), (RODRIGUEZ et al., 2019), (CONCHA; BURR; SUáREZ, 2014), (SOYLU et al., 2022c), (CSAKI, 2018) | (DAHBI; CHIADMI; LAMHARHAR, 2023), (SOYLU et al., 2022c), (CSAKI, 2018) | (SOYLU et al., 2022b), (UDUWAGE-DON; HADIWATTAGE; PANUWATWANICH, 2023), (NAI et al., 2023), (SANGIL, 2020), (ALMEIDA et al., 2018), (DAHBI; CHIADMI; LAMHARHAR, 2023), (MENDES; VOIGT, 2022b), (aO et al., 2023), (PUTRI; RULDEVIYANI, 2019), (RODRIGUEZ et al., 2019), (SOYLU et al., 2022c) |

# 5

# Enhancing Discoverability, Accessibility, and Usability of Public Procurement Data Using ChatGPT

This chapter presents an exploratory study focused on using ChatGPT to enhance the quality of public procurement data. It addresses a research gap concerning the application of Large Language Models (LLMs) in this domain, emphasizing the need for methods to improve data discoverability, accessibility, and usability for suppliers. The study employs a user-centric taxonomy, detailed in Section 5.2 to categorize key elements of user interaction with procurement data. It tackles four specific research questions regarding how ChatGPT can assist stakeholders in overcoming procurement data challenges. The findings offer insights into the practical application of LLMs, highlighting their strengths and weaknesses. Building on quality concerns from Chapter 4, this study aligns with the dissertation's goal to demonstrate improvements in public procurement data quality through AI technologies. The concluding sections will outline best practices for leveraging ChatGPT effectively and set the stage for future research on integrating AI into public procurement systems. Through the systematic application of best practices derived from a systematic mapping, the exploratory study rigorously evaluates the potential of LLMs like ChatGPT to improve data quality, ensuring practical and actionable outcomes.

Table 23 – Exploratory Study Goal according to GQM

| | |
|---|---|
| **Analyze** | LLM support effectiveness |
| **for the purpose of** | characterization |
| **with respect to** | discoverability, accessibility and usability |
| **from the point of view of** | government suppliers |
| **in the context of** | public procurement data |

## 5.1 Exploratory Study Design

This section introduces the methodology of our exploratory study, detailing how we investigate the potential of Large Language Models (LLMs) to improve public procurement data quality, with a particular emphasis on the perspective of government suppliers. Our focus on government suppliers as key stakeholders is justified because they are directly impacted by the quality of procurement data, which is essential for identifying business opportunities, preparing competitive bids, and fulfilling contractual obligations.

The research design detailed in this section employs the Goal-Question-Metric (GQM) methodology (CALDIERA; ROMBACH, 1994), as summarized in Table 23. Utilizing the GQM approach ensures a systematic framework for our investigation, with clearly articulated research goals, precisely formulated questions, and the application of measurable metrics to assess the efficacy of LLMs within the public procurement data landscape. The GQM methodology is directly linked with the usercentric taxonomy detailed in Section 5.2. The taxonomy helps operationalize the goals and questions defined by the GQM by providing a framework for identifying the relevant dimensions of analysis and interaction within the procurement process, aligning the specific research questions (SRQs), and ensuring a targeted evaluation of LLM support from the perspective of government suppliers.

The primary research question (RQ) guiding this study is:: "How can LLMs, such as the ChatGPT model, support government suppliers in improving the discoverability, accessibility, and usability of public procurement data?" To address this question comprehensively, we developed four Specific Research Questions (SRQs), each designed to explore different facets of the primary RQ, as detailed in Table 24.

We consider four key dimensions to characterize how Large Language Models (LLMs) can support stakeholders in the public procurement domain, as depicted in Figure 10. Each dimension is explored in detail below.

**Quality Assessment of LLM Responses (SRQ1)**. This dimension focuses on evaluating the quality of outputs generated by LLMs in response to procurement-related queries, corresponding to SRQ1. The evaluation process involves qualitative analysis, as depicted in Figure 10, to ensure that the information generated by LLMs is clear, consistent, comprehensive, and

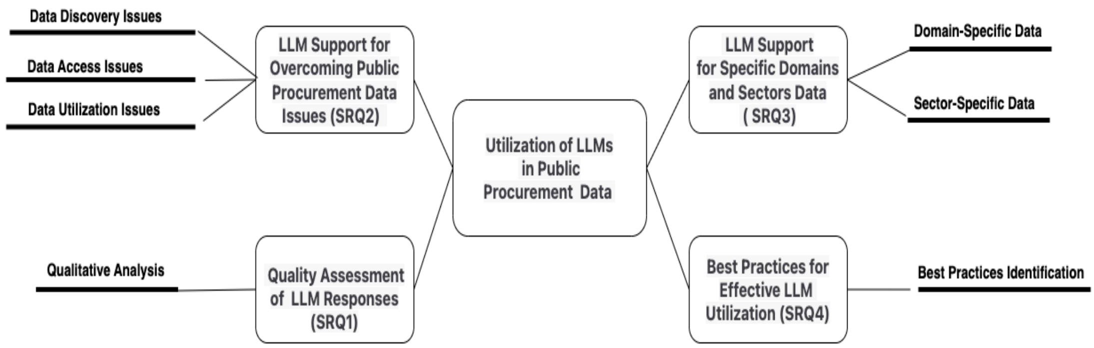


Figure 10 – LLM Support Areas Public Procurement Data Quality Across Four Dimensions.

Table 24 – Research Questions on Supplier Engagement with LLMs in Public Procurement

| **Specific Research Questions** | **Motivation** |
|---|---|
| **SRQ1:** How can suppliers assess the quality of LLM responses in addressing their informational requirements? | This is about suppliers need to ensure that the information they receive from LLMs is both reliable and pertinent to their specific procurement needs. |
| **SRQ2:** How can LLM support suppliers overcome barriers to discovering, accessing and using public procurement data? | This investigates how LLM can assist suppliers in effectively searching for and utilizing relevant procurement data by addressing associated issues. |
| **SRQ3:** How can LLM support suppliers within specific domains and government sector procurement data? | This explores using advanced AI to enhance procurement processes in specialized sectors, particularly in national contexts. |
| **SRQ4:** What best practices can be learned from the results to effectively use LLM in public procurement? | This aims to develop strategies and methods to maximize the benefits of using LLM in public procurement. |

aligned with the needs of stakeholders. Key evaluation criteria include cross-checking information against reliable sources, assessing contextual relevance, ensuring clarity of communication, verifying the completeness of responses, and maintaining consistency across similar queries.

**LLM Support for Overcoming Public Procurement Data Issues (SRQ2)**. This dimension investigates how LLMs can address common challenges in public procurement data, such as barriers to discoverability, accessibility issues, and usability constraints. Figure 10 outlines these challenges in relation to SRQ2. The analysis explores the potential of LLMs to improve search and retrieval processes, enhance user interaction with data, and facilitate more effective analytical reporting.

**LLM Support for Specific Domains and Governement Sectors (SRQ3)**. Addressing SRQ3, as illustrated in Figure 10, this dimension examines how LLMs can assist suppliers in navigating and excelling within specific government sector procurement systems. The focus is on utilizing domain-specific data to tailor LLM support, ensuring that suppliers receive relevant and specialized information that is most pertinent to their sector.

**Best Practices for Effective LLM Utilization (SRQ4)**. This dimension proposes a set of best practices aimed at optimizing the use of LLMs in procurement data management, directly addressing SRQ4. These practices are derived from insights gained through LLM interactions and include strategies such as effective prompt design, iterative refinement techniques, and the integration of feedback to enhance response quality concerning the discovery, access, and utilization of public procurement data.

## 5.2 A User-Centric Taxonomy for Procurement Data Interaction

This section introduces a taxonomy designed to foster a user-centric interaction with procurement data, streamlining the procurement processes. Figures 11, 12, and 13 provide visual representations of this structured taxonomy and the supplier-centric interaction model, aiming to enhance comprehension of the proposed framework and its practical implications.

Figure 11 showcases a detailed taxonomy that focuses on user interaction with procurement data, offering a structured overview of the core elements in public procurement from the user's standpoint and emphasizing stakeholder roles alongside the infrastructure that aids in data discoverability, accessibility, and usability. The user-centric taxonomy, developed through an iterative process, concentrates on the distinct needs and viewpoints of government suppliers. It is designed to reflect the essential stages and components of suppliers' engagement with procurement data, from identifying their roles to the adept use of available resources.

By adopting a user-centric approach, the intention is to utilize the Language Learning Model (LLM) to address the challenges suppliers face, ensuring that the LLM is a supportive tool rather than the taxonomy's main subject. This strategy aligns with the Goal-Question-Metric (GQM) framework, as described in Section 3, which offers a robust basis for evaluating the effects of LLMs on suppliers. The LLM thus acts as an enabler to help suppliers overcome their challenges, maintaining its role secondary to the taxonomy's focus. The synergy with the GQM framework, as thoroughly explained in Section 5.1, guarantees that the taxonomy provides a sturdy framework for measuring the impact of LLMs on suppliers.

This taxonomy, illustrated in Figure 11 provides a foundational framework for navigating the complex landscape of public procurement data. It clearly defines the roles of key stakeholders—such as suppliers, public agents, and citizens—and integrates Open Government Data (OGD) frameworks such as CKAN[1], DKAN[2], and Socrata[3], which are designed to enhance data accessibility across various thematic domains and government sectors. Thematic domains

[1] https://ckan.org/
[2] https://dkan.readthedocs.io/en/latest/
[3] https://dev.socrata.com/

encompass broad categories like healthcare and information technology, while government sectors refer to administrative levels—national, regional, or local—where procurement data is managed.

Additionally, Figure 11 incorporates the use of Large Language Models (LLMs), including free models like ChatGPT free version, AmberChat[4], Open assistant[5], as well as commercial models like ChatGPT, Gemini[6], Claude[7]. It also outlines strategies for interacting with LLMs, such as Thought Generation and in-context learning, which are crucial for effective text processing and generation.

**Implementing the Proposed Taxonomy for Government Suppliers** The evolution of a user-centric taxonomy from a conceptual model to a practical tool designed to enhance user engagement is illustrated in Figure 12. It maps the structured journey of government suppliers within the taxonomy, encompassing interaction routes, engagement support, and data quality improvements. It delineates the systematic progression of a government supplier through the taxonomy, categorized into pathways for interaction,facilitators for engagement, and enhancements for data quality.

*User Interaction Pathways* focus on guiding the government supplier through essential decisions that shape their interaction with procurement data. The process begins with selecting the stakeholder role, specifically identifying the user as a government supplier. Following this, the user selects an appropriate OGD framework, such as CKAN, which is instrumental in managing procurement data. The next steps involve choosing a thematic domain, like Information Technology, and selecting a relevant government sector, such as National. Finally, the user selects an LLM (e.g., the free version of ChatGPT) and applies specific prompt strategies like Thought Generation and In-Context Learning, as illustrated in Table 25. These choices form the foundation of the user's interaction with the procurement data.

*User Engagement Facilitators* are activated as the user moves through these interaction pathways. These facilitators refine and enhance the interaction by focusing on role identification, ensuring the user's role as a government supplier is clearly defined and understood. Framework alignment ensures that the selected OGD framework is properly integrated with the user's objectives. Contextual relevance ensures that the choices made in the thematic domain and sector selection are appropriate and meaningful for the user's needs. LLM engagement focuses on maximizing the effectiveness of the chosen LLM, ensuring that it is used to its full potential. Finally, prompt engagement refines prompt strategies, ensuring they are applied correctly and effectively.

*Data Quality Improvement* is the ultimate goal of this process, achieved through the

[4] https://huggingface.co/LLM360/AmberChat
[5] https://huggingface.co/OpenAssistant
[6] https://gemini.google.com/
[7] https://claude.ai/

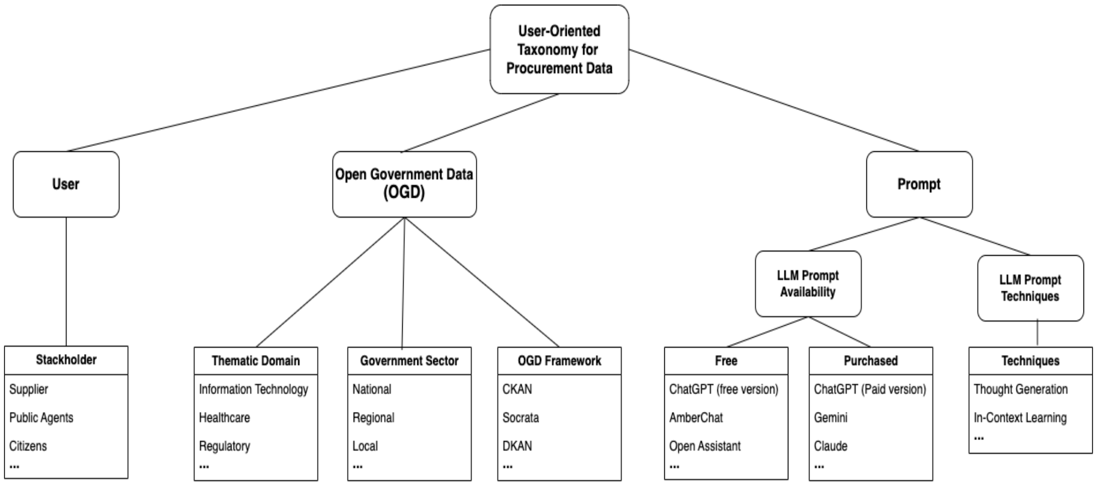


Figure 11 – User-Centric Taxonomy for Public Procurement Data Interactions: Enhancing Discoverability, Accessibility, and Usability.

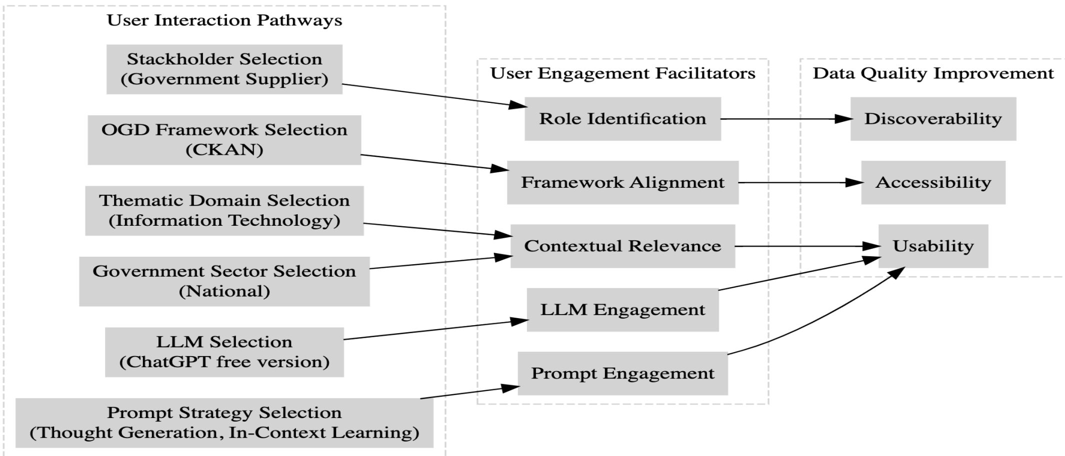


Figure 12 – Application of User-Centric Taxonomy: Streamlining Procurement for Government Suppliers.

facilitators. Role identification enhances the discoverability of relevant data by ensuring the user's role is clear and well-defined. Framework alignment improves accessibility by aligning the selected framework with the user's needs, making it easier to access the data. Contextual relevance, LLM engagement, and prompt engagement collectively contribute to improved usability, ensuring that the data is not only accessible but also meaningful and easy to interact with.

Figure 13 provides a visual representation that encapsulates the user-centric approach for interacting with procurement data, focusing on a government supplier's journey through various selection processes.

To further enrich the understanding of the interaction between government suppliers and LLMs, Figure 13 can be enhanced with the interaction flow presented in Figure 14. This figure provides a detailed, step-by-step perspective on how a supplier specifically interacts with an LLM, such as ChatGPT. The flow begins with the identification of procurement data needs,

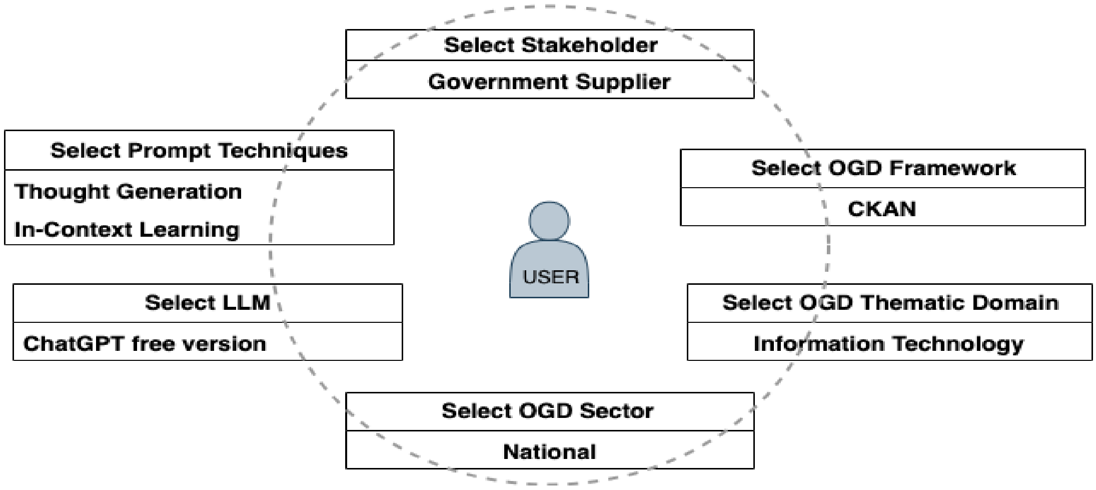


Figure 13 – Supplier-Centric Interaction Model: Leveraging LLMs to Enhance Usability and Data Navigation.

Table 25 – Comparative Analysis of Prompt Engineering Techniques

| Prompt Strategy | Characteristics |
|---|---|
| **General Inquiry:** ’*What are the important rules to get the best answer to the questions asked to you?*’ | **Technique:** Zero-shot learning (ZHAO et al., 2023). **Details Provided:** Minimal context. **Purpose:** To evaluate how well ChatGPT can generate comprehensive responses based solely on the question. |
| **Context-Specific Inquiry:** ’*Based on these tips, create a prompt for a government supplier who wants to discover websites with open government procurement data.*’ | **Technique:** Few-shot learning (ZHAO et al., 2023). **Details Provided:** Specific examples and context. **Purpose:** To examine how additional context improves the relevance and correctness of the response. |
| **Thought Generation:** ’*Can you provide a step-by-step guide on how to find these portals, access the data, and utilize it effectively for my procurement analysis?*’ | **Technique:** Zero-shot thought generation (ZHAO et al., 2023). **Details Provided:** Specific task request. **Purpose:** To evaluate the model’s ability to generate detailed and practical guidance. |

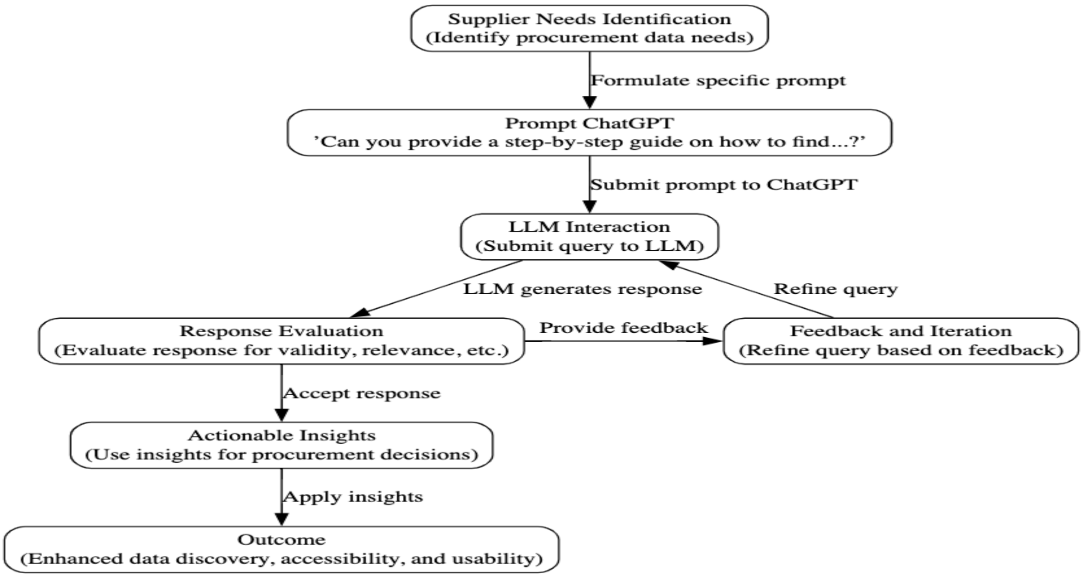


Figure 14 – Supplier Interaction Flow with LLM for Public Procurement Data Acquisition.

progresses to the formulation of a specific prompt (e.g., 'Can you provide a step-by-step guide on how to find...'), the submission of this prompt to the LLM, and then the evaluation of the LLM's response for validity and relevance. The process is iterative, with feedback leading to the refinement of queries. Ultimately, the supplier applies the actionable insights gained from the LLM to enhance data discovery, accessibility, and usability. Therefore, Figure 14 complements Figure 13 by focusing on the practical, iterative interaction with an LLM, highlighting the steps a supplier takes to effectively use LLMs in navigating public procurement data rather than the broader selection process.

## 5.3 Exploratory Study Results

In this section, we will address four Specific Research Questions (SRQs) that are integral to understanding the role of Large Language Models (LLMs) in enhancing public procurement data quality from the supplier perspective. Each SRQ is designed to explore different dimensions of the overarching research question: "How can LLMs, such as the ChatGPT model, support government suppliers in improving the discoverability, accessibility, and usability of public procurement data?". By systematically examining these SRQs, we aim to provide comprehensive insights into the effectiveness of LLMs in addressing the challenges faced by suppliers in navigating public procurement processes.The complete set of queries used in this study, along with the corresponding responses generated by ChatGPT, are available in a public repository[8]. This repository provides access to all the examples discussed below, offering a transparent view of the interactions with ChatGPT and the data-driven insights derived from these interactions.

[8] https://doi.org/10.5281/zenodo.13748278

### 5.3.1 Analysis of Specific Research Question 1 (SRQ1)

Given the exploratory nature of this study and its focus on user experience with Large Language Models (LLMs), a qualitative approach was determined to be the most appropriate. This method effectively captures the intricate details of interactions and the subjective quality of information, aligning with the study's emphasis on the dynamics between suppliers and LLMs. The subsequent section offers a qualitative analysis of the responses generated by LLMs.

Addressing SRQ1: "How can suppliers assess the quality of LLM responses in addressing their informational requirements?", we evaluate ChatGPT's responses across several key quality dimensions: validity, relevance, clarity, comprehensiveness, consistency, and helpfulness.

The chosen attributes for evaluating ChatGPT's performance are grounded in widely accepted criteria for measuring information quality and the effectiveness of information systems. These dimensions—validity, relevance, clarity, comprehensiveness, consistency, and helpfulness—are fundamental in determining whether the information delivered by a system is both suitable and dependable for the user. They are especially pertinent in Large Language Models (LLMs), where ensuring high-quality information is essential for users to make well-informed and secure decisions. Additionally, these attributes have been employed in various studies that evaluate the responses of LLMs across different domains, including healthcare, thereby substantiating their relevance for application in this research (CADAMURO et al., 2023).

It is important to mention that LLMs, such as ChatGPT, may exhibit the phenomenon of 'hallucination,' where the model generates information that seems plausible but is not true or based on facts. To mitigate this problem, we adopted cross-referencing of ChatGPT responses with authoritative sources, such as government data websites. Additionally, the model's temperature, a parameter that influences the randomness of responses, was kept constant during testing to ensure the consistency of the results. According to OpenAI's documentation[9], the default temperature setting is 1.0. Although this research uses a free version of ChatGPT, the validation method and the maintenance of parameters such as temperature contribute to increasing the reliability of the conclusions.

As detailed below, each quality dimension—validity, relevance, clarity, comprehensiveness, consistency, and helpfulness—is assessed through specific examples, demonstrating ChatGPT's performance in various procurement scenarios. This evaluation aligns directly with the usercentric taxonomy presented in Section Section **??**. The selected attributes for evaluation are integral components of the taxonomy, ensuring that the information assists government suppliers effectively. For example, the focus on improving data quality within the taxonomy is reflected in our strategies for assessing LLM responses, which include verifying against authoritative sources to ensure validity and evaluating the contextual relevance of the information provided. These strategies are directly related to the user engagement facilitators as defined

9 https://platform.openai.com/docs/api-reference/chat/object

within the taxonomy.

**Cross-Referencing ChatGPT Responses for Validity**. To assess the validity and correcteness of ChatGPT's responses, cross-referencing with authoritative sources is essential, especially considering the model's knowledge cutoff date (AZARIA; AZOULAY; RECHES, 2024). For instance, one evaluated query was: '*What are the important rules to get the best answer to the questions asked to you?*'. This prompt used a zero-shot technique, , where the model is asked to generate a response without prior examples.The response was cross-referenced against best practices from OpenAI's documentation, confirming the validity of the generated information. Another example is: "*Where can I independently verify the details provided about public procurement procedures?*". This prompt also employs a zero-shot technique, as it directly asks the model to identify authoritative sources without providing prior examples or context. The response provides a curated list of authoritative sources for verifying public procurement procedures, ensuring users have the means to authenticate information and thus increase the trustworthiness of their understanding of procurement data. Finally, to mitigate potential inaccuracies, suppliers should validate both the correctness of content and its alignment with current, evidence-based practices in public procurement.

**Ensuring Relevance through Context-Specific Responses**. To ensure the relevance of ChatGPT's responses to specific procurement inquiries, it is important to use context-specific prompting strategies. For example, the prompt '*Based on these tips, create a prompt for a government supplier who wants to discover websites with open government procurement data.*' employed a few-shot technique to generate tailored guidance. The resulting response was highly relevant, providing actionable advice that directly met the user's need for identifying procurement data sources. This highlights the effectiveness of context-specific prompts in eliciting relevant, user-focused outputs from LLMs. Another question that aligns with the example provided is: "*What are the first steps I need to take to find public procurement opportunities?*". This query mirrors the specific, scenario-focused method seen in the example, aiming for pertinent, actionable advice. A suitable response details practical steps like market identification, registration on relevant platforms, alert configurations, and research strategies, thereby assisting suppliers in effectively locating procurement data sources. These examples demonstrate how using context-specific prompting strategies can significantly enhance the relevance of responses, ensuring they are directly applicable to the user's situation.

**Evaluating Clarity and Ease of Understanding**. Clarity in ChatGPT's responses is essential for ensuring that suppliers can easily follow and implement the provided information. This can be evaluated by assessing whether the responses are structured in a coherent and easily understandable manner. For instance, these prompts employed the strategy of a thought generation inquiry using a zero-shot technique to request detailed procedural instructions: ''*Can you provide a step-by-step guide to find these portals, access the data, and utilize it effectively for my procurement analysis?*' and '*Now provide detailed step-by-step instructions on how a* *government supplier can access data from a site that uses CKAN as a data catalog through its API.*'. The responses were structured in a clear, step-by-step manner, with specific examples that facilitated ease of understanding and following the instructions. It guided the user through the process of identifying CKAN or CKAN-like portals, accessing data, and utilizing it effectively for procurement analysis, ensuring that the instructions were comprehensible and actionable for the intended audience. The structured nature of these responses facilitates ease of understanding and practical application, ensuring that suppliers can effectively use the information provided.

**Assessing Comprehensive Responses and Helpfulness**. To evaluate the comprehensiveness and helpfulness of ChatGPT's responses, suppliers should assess whether the responses cover all relevant aspects of the inquiry and provide actionable insights. For example, this prompt employed a thought generation inquiry strategy using a few-shot technique to assess the thoroughness and utility of the model's responses: '*Now use this refined prompt to obtain the answers I need, asking for the information step-by-step from each section of the refined prompt, so you can provide me with more specific and useful information about public procurement portals with open tenders for supplying a specific material from my portfolio.*' The response comprehensively covered all aspects of the inquiry by addressing each part of the refined prompt in detail. It guided the supplier from identifying relevant portals to navigating them and accessing specific data, providing detailed examples of how to find, understand, and fetch data via API. The thoroughness with which ChatGPT addressed each aspect of the prompt demonstrates its capability to offer comprehensive, actionable advice, supporting the broader research goal of enhancing data usability for suppliers.

**Maintaining Consistent and Reliable Output**. Consistency in ChatGPT's responses is essential for ensuring reliability over time, particularly in dynamic procurement environments. Suppliers can evaluate consistency by comparing responses to similar questions posed at different times, ensuring that the information remains aligned and reliable. For instance, the prompts '*How can I be sure that these websites you listed use CKAN to provide their open data?*' and '*What specific features of these websites should I look for to identify CKAN usage?*' were used to assess whether the model provides consistent, coherent guidance. The responses to both questions were consistent, offering aligned information and step-by-step instructions on identifying CKAN usage features. This consistency reinforces the reliability of the responses and underscores the model's consistent output quality. Regular checks for consistency across different prompts ensure that the information provided by ChatGPT remains dependable and trustworthy over time.

In conclusion, the findings indicate that ChatGPT effectively supports government suppliers by providing highquality responses that are contextually relevant. These responses span key quality dimensions, including validity, relevance, clarity, comprehensiveness, and consistency. ChatGPT's strong performance in these areas demonstrates its potential as a valuable tool for managing the intricacies of public procurement data. Consequently, the research suggests that ChatGPT is a capable aid for suppliers in evaluating the quality of responses from Large Language


Figure 15 – Snapshot of a Typical Interaction with ChatGPT.

Models (LLMs), thus addressing the primary research question concerning the assessment of LLM response quality by suppliers.

### 5.3.2 Analysis of Specific Research Question 2 (SRQ2)

This section investigates SRQ2: "How can LLM support suppliers overcome barriers to discovering, accessing and using public procurement data?" The focus is on evaluating how ChatGPT can assist suppliers in overcoming common obstacles encountered in discovering, accessing, and using public procurement data. Each dimension will be addressed below with examples of how LLMs can facilitate the public procurement process for suppliers.

To better illustrate the practical application of Large Language Models (LLMs) in addressing these challenges, Figure 15 provides a snapshot of a typical interaction with ChatGPT. The figure presents a simplified chat interface featuring a dedicated space for the user's query. For instance, the query ”Can you provide a step-by-step guide on how to find these portals, access the data, and utilize it effectively for my procurement analysis?” serves as a thought-generation prompt designed to solicit detailed procedural instructions. Accompanying this prompt, an excerpt of ChatGPT's response is presented, highlighting the model's ability to offer structured, stepby-step guidance. This visual example demonstrates how the model processes information and delivers it to the user, thereby improving the user's understanding of the practical uses of LLMs in public procurement.

Improving public procurement processes requires overcoming challenges related to data discovery, accessibility, and usability. The following sections detail how LLMs can improve each of these aspects.

**Enhancing Data Discovery**. In public procurement, effective data discovery ensures that all relevant information is identified and utilized efficiently. Suppliers often struggle with crafting effective search queries. For instance, the prompt, '*What are the important rules to get* *the best answer to the questions asked to you?* ', elicits a response from ChatGPT that outlines a structured approach to formulating inquiries. By generating comprehensive guidelines, this prompt helps suppliers in understanding the significance of specificity and context within their queries, which can substantially enhance the discovery process to successfully locate relevant procurement data. Similarly, the response to the prompt, '*As a government supplier looking to access public procurement data, what are the key pieces of information I need to know?*' effectively identifies reliable sources of procurement data and provides a clear starting point for suppliers to locate relevant information, addressing the initial barrier to discovery.

**Improving Data Accessibility**. Once data discovery challenges are addressed, suppliers must access the data they've located. Complex procurement portals present both technical and regulatory challenges. For instance, the prompt, '*Based on these tips, create a prompt for a government supplier who wants to discover websites with open government procurement data*', provides context-specific advice to guide suppliers in finding relevant data portals, and assists suppliers in identifying and accessing procurement data portals that are pertinent to their needs. Additionally, the prompt, '*Now provide detailed step-by-step instructions on how a government supplier can access data from a site that uses CKAN as a data catalog through its API.*', breaks down the process into manageable steps, simplifying navigation through complex portals.

**Addressing Data Usability Issues**. Effective use of procurement data is critical for decision-making. Suppliers may struggle with data analysis or resource utilization post-access. The prompt, '*Can you provide a step-by-step guide on how to find these portals, access the data, and utilize it effectively for my procurement analysis?*', shows how ChatGPT can guide suppliers through the entire process, from discovery to analysis. This not only simplifies the process but also ensures that suppliers can make informed decisions based on the data they access. It was designed to generate ideas or suggestions based on a provided context, and requires the model to engage in critical thinking and deliver structured responses. By producing a step-by-step guide, the prompt assists suppliers in effectively employing accessed data, thereby improving their capacity to analyze and apply procurement data to their business operations. Other examples include the prompt, '*Now use this refined prompt to obtain the answers I need, asking step-by-step for information in each section of the refined prompt, so you can provide me with more specific and useful information about public procurement portals with open tenders for supplying a specific material from my portfolio.*' This approach, characterized by interactive querying through iterative prompting, allows for detailed, context-specific follow-up questions. It ensures that the supplier receives highly relevant, tailored information that meets their specific needs, thereby facilitating the better utilization of procurement data. Similarly, prompts like, '*How can I be sure that these websites you listed indeed use CKAN to provide their open data?*' and *"What specific features of these websites should I look for to identify CKAN usage?"* The both responses highlight specific features to look for, which helps suppliers quickly determine whether a site uses CKAN, thereby simplifying the usability of the information and reducing the time spent on verifying data sources.

In conclusion, the evidence suggests that LLMs such as ChatGPT can play a practical role in enhancing the discoverability, accessibility and usability of public procurement data, thus answering our second research question by offering clear, actionable guidance for suppliers.

### 5.3.3 Analysis of Specific Research Question 3 (SRQ3)

This section investigates SRQ3: "How can LLM support suppliers within specific domains and government sector procurement data?". The analysis below examines how LLMs offer domain-specific and sector-specific guidance, showcasing its capacity to tailor assistance to the specific needs of suppliers within different government procurement systems, providing context and analysis to directly answer SRQ3.

**Addressing Domain-Specific Data**. Suppliers need to understand procurement data within their specific domains. For example, the prompt, '*As a supplier of IT services, what procurement data should I focus on to find relevant opportunities?*' provides IT suppliers with targeted strategies for identifying relevant procurement opportunities. This demonstrates LLMs' ability to offer industry-specific guidance, helping suppliers discover and access relevant data more efficiently.

**Addressing Sector-Specific Data**. After addressing domain-specific data, it is equally important Understanding sector-specific data for aligning offerings with government needs. For instance, the prompt, '*What specific opportunities exist for IT suppliers in national or federal public procurement?*' demonstrate the ability of LLMs to identify key opportunities within specific sectors. The response highlighted contracts for cybersecurity assessments and cloud migration services—critical areas for IT suppliers to focus on. This illustrates the model's ability to provide actionable insights that are directly relevant to the supplier's sector. Additionally, the prompt, '*As a government supplier aiming to access public procurement data in the IT and Healthcare sectors, focusing on national processes in Brazil, the United States, the United Kingdom, China, and Switzerland, what are the key pieces of information I need to know?* leads to insights about CKAN data portals and available datasets, emphasizing the importance of understanding data types and relevant portals for each country.

These examples demonstrate LLMs' ability to deliver tailored guidance across domain and sector-specific contexts within government procurement. By doing so, LLMs enhance suppliers' ability to navigate and succeed in the complex procurement landscape, providing a concrete answer to our third research question.

### 5.3.4 Analysis of Specific Research Question 4 (SRQ4)

This section examines SRQ4: "What best practices can be learned from the results to effectively use LLM in public procurement?" The analysis below examines a variety of prompts and their alignment with essential criteria, including targeted questioning, clarity of phrasing, thoughtful prompt design, iterative refinement, and the incorporation of insights. This process enables the identification of best practices for effectively utilizing Large Language Models (LLMs). By concentrating on these criteria, we can discern strategies that significantly improve the discovery, accessibility, and utilization of public procurement data.

**Crafting Targeted Prompts**. One of the most effective strategies in prompt design is to formulate specific questions that target particular aspects (MONDAL; BAPPON; ROY, 2024), in this case, those of public procurement. For instance, the prompt, '*As a government supplier looking to access public procurement data, what are the key pieces of information I need to know?*' targets a broad but specific aspect of procurement. The response is structured into clear key points, ensuring comprehensibility. Another example is '*What specific opportunities exist for IT suppliers in national or federal public procurement?*' provides a focused response detailing relevant opportunities, helping suppliers understand the landscape. These prompt's design ensure clarity and comprehensiveness, which are crucial for effective communication and understanding.

**Enhancing Response Quality**. While precision in initial inquiries is crucial, refining prompts based on previous interactions improves response quality (CHEN et al., 2023). For example, the prompt, '*Based on these tips, create a prompt for a government supplier who wants to discover websites with open government procurement data*', refines initial tips into a specific prompt tailored to a government supplier's needs. Similarly, the prompt, '*Can you provide a step-by-step guide on how to find these portals, access the data, and utilize it effectively for my procurement analysis?*' builds upon initial responses to deliver more detailed insights. The structured format, which includes an introduction and specific request, directs the LLM to deliver focused and useful information. This demonstrates how refining initial inputs can lead to more precise and actionable responses.

**Ensuring Accurate Responses**. Clarity in phrasing is essential for obtaining accurate and useful responses (MU et al., 2024). For example, the prompt, '*How can I be sure that these websites you listed indeed use CKAN to provide their open data?*' results in a detailed, actionable response. The steps are clearly articulated, making them easy to follow. The response leverages detailed insights about CKAN features to provide practical verification steps, illustrating the importance of integrating specific knowledge and insights into the prompts. Another prompt, '*What are the first steps I need to take to find public procurement opportunities?*' results in a structured response outlining essential actions, including identifying the target market and registering on government portals. This clarity assists the LLM in generating precise answers and ensures that the responses are directly relevant to the suppliers' needs.

**Maximizing Insight Generation**. Prompts that encourage comprehensive responses can yield more detailed insights (LOU; ZHANG; YIN, 2023). For instance, the prompt '*What are some common challenges suppliers face when accessing public procurement data and how can they be overcome?*' not only identifies challenges but also provides solutions, thereby equipping suppliers with actionable strategies. Another example is the prompt '*Is there any way to automate the process of discovering which open data portals are available?*'. This prompt elicits a response that offers actionable insights into automating the discovery of open data portals. By focusing on specific requests, these prompts aligns with the best practices for effective querying, ensuring that LLM responses are both relevant and useful for the user's needs.

In summary, best practices for using LLMs in public procurement include crafting precise inquiries, refining prompts iteratively, ensuring clarity in phrasing, and encouraging comprehensive responses. These practices, derived from the analysis of the fourth research question's prompts, significantly enhance the discoverability, accessibility, and usability of procurement data.

## 5.4 Capabilities and Limitations in the Use of LLMs for Public Procurement Data

Large Language Model such as ChatGPT can revolutionize the handling of public procurement data, offering a spectrum of capabilities alongside some significant challenges.

On the capabilities side, LLMs can significantly enhance the discoverability, accessibility, and usability of public procurement data. As demonstrated in Section 5.3.2, ChatGPT showed a strong capability in guiding suppliers to discover relevant data, but this was highly dependent on the specificity of the prompt as noted in Section 5.3.4. They efficiently process vast volumes of information, providing tailored responses to specific inquiries that can aid suppliers in navigating complex datasets, identifying relevant opportunities, and accessing detailed procedural guidance. Through advanced natural language processing, LLMs deliver comprehensive, contextually relevant responses that can precisely meet users' informational needs, thereby simplifying the procurement process. Their adaptability to various public sector domains can offer sector-specific insights and automate repetitive tasks, reducing cognitive load and improving efficiency.

Despite these advantages, several limitations are associated with LLMs in public procurement. A key concern is the correctness and reliability of the information these models provide, which is critical in decision-making processes within public procurement. As highlighted in Section 5.3.1, the validity of ChatGPT's responses must be cross-referenced with authoritative sources, and suppliers should validate the correctness of content, aligning with best practices. The quality of outputs heavily depends on the quality of the prompt, underlying data, and the model's interpretation capabilities, potentially leading to incorrect, generic, or shallow responses, particularly in complex scenarios requiring domain-specific expertise. LLMs may also struggle with the discoverability of niche or highly specialized procurement information, potentially resulting in incomplete or misleading guidance. While iterative refinement of prompts can mitigate some issues, it requires user expertise and engagement that may not always be feasible, a factor explored in Section 5.3.4.

In conclusion, integrating LLMs into public procurement workflows must be approached with care, ensuring that the technology enhances, rather than substitutes, human judgment and domain expertise.

## 5.5 Threats to Validity in the Exploratory Study

This section addresses the threats to the validity of our exploratory study, acknowledging the limitations that might affect the generalizability and robustness of our findings. It is crucial to understand these constraints to contextualize the study's results and highlight areas for future research. We have identified several key factors that may influence the validity of our conclusions.

**Selected Data Sources**. The study primarily focuses on CKAN data frameworks for public procurement. Although CKAN is widely used, this focus may limit the applicability of our findings to other data management systems or repositories. Consequently, the generalizability of the results to different contexts or frameworks within public procurement could be affected. Moreover, public procurement data is just one category of open government data, and this focus may limit the study's relevance to other operational environments where different types of data might present unique challenges.

**Selected Specific Domain and Sector**. The study concentrates on public procurement data within the Information Technology domain and the national government sector. While this focus enables an in-depth analysis, it may restrict the applicability of the findings to other sectors or domains within public procurement where distinct challenges and data characteristics may prevail.

**Selected LLM**. This study is based on a free version of ChatGPT, which lacks some advanced capabilities available in the paid or enterprise versions, such as enhanced processing power and more sophisticated language understanding. This limitation may lead to less accurate or comprehensive responses, potentially downplaying the full potential of LLMs in handling complex public procurement data. Thus, findings may not fully represent the capabilities of more advanced LLMs in real-world applications. Additionally, another threat to validity in the use of LLMs may arise from dated versions (CHENG et al., 2024). While an LLM may have a general cutoff date announced, its knowledge of certain subjects or features may be significantly outdated, leading to inaccurate or incorrect responses.

**Selected Prompts**. The effectiveness of ChatGPT in this study is significantly influenced by the quality and specificity of the prompts used. Although various prompting strategies have been explored, variations in prompt formulation can lead to significant differences in the quality of responses. The specific formulations may not encompass the entire spectrum of possible interactions with the LLM. This variability may affect the consistency of the findings, suggesting that the results are contingent upon the user's expertise in crafting effective prompts.

While this exploratory study offers valuable insights into the use of LLMs in public procurement, these threats to validity emphasize the necessity for future research to broaden the scope of the investigation.

## 5.6 Related Works

The development of large language models (LLMs) has significantly advanced the field of natural language processing, enabling new methods for extracting and interpreting information procurement, including healthcare, education, law, and finance (ZHAO et al., 2023). Despite this range of applications, research into the potential of LLMs within the context of Open Government Data (OGD) remains underexplored (LOUKIS et al., 2023). The present study seeks to address this gap by specifically investigating the use of ChatGPT in public procurement, a critical area that has not yet received significant attention regarding LLM utilization.

Recent efforts in OGD have focused on integrating LLMs like GPT-3.5 to enhance data access through natural language interfaces (MAMALIS et al., 2023). For instance, ongoing research aims to incorporate ChatGPT into existing frameworks to improve 12 the interpretability of government data (BARCELLOS et al., 2024). These efforts align with our study's goal of demonstrating how LLMs can enhance the discoverability, accessibility, and usability of public procurement data.

Large Language Models are also valuable tools for improving efficiency and analysis in government oversight processes such as supreme audit institutions and internal audit functions (UGALE; HALL, 2024). For instance, the Brazilian Federal Court of Accounts (TCU) has implemented LLM systems to automate critical tasks, including case analysis and adjudication recommendations, thereby significantly improving operational efficiency (PEREIRA et al., 2024). The TCU has developed ChatTCU (SILVA et al., 2024), a customized LLM that integrates internal data to provide users with access to jurisprudence and document summaries. While these advancements highlight the potential of LLMs in public sector applications, our study diverges by focusing on how individual suppliers can leverage these technologies within public procurement.

Despite the recognized benefits of artificial intelligence in procurement processes (WASEEM et al., 2023), to our knowledge, research specifically exploring the impact of ChatGPT on public procurement is still incipient. This study provides a preliminary exploration of this domain by presenting a structured taxonomy and interaction flow tailored to suppliers. This approach aims to empower suppliers to effectively utilize LLMs in analyzing public procurement data, enhancing their ability to navigate complex procurement landscapes.

## 5.7 Chapter Overview

This chapter explored the potential of Large Language Models (LLMs), represented by ChatGPT, to improve the discoverability, accessibility, and usability of public procurement data. Through a detailed analysis, we examined how ChatGPT can support government suppliers in navigating complex procurement processes, enhancing their ability to participate effectively in public procurement. The research provides a structured approach for suppliers to interact with LLMs, highlighting the importance of precise inquiries, iterative prompt refinement, and clear phrasing to achieve more effective results.

The findings indicate that ChatGPT can significantly enhance suppliers' ability to discover, access, and utilize procurement data by providing contextually relevant, clear, and actionable guidance. This capability helps bridge existing gaps in the procurement process, promoting a more informed and transparent participation of suppliers.

The study demonstrates that LLMs can assist suppliers in overcoming common obstacles in discovering, accessing, and utilizing public procurement data, offering clear, actionable guidance that improves their ability to make datadriven decisions. Specifically, it emphasizes using LLMs to navigate open data portals, extract relevant information, and tailor search strategies to specific domains and government sectors.

Additionally, by offering a structured taxonomy and interaction flow, the study provides a foundation for the broader adoption of LLMs in public procurement and open government data initiatives. This framework, detailed in Section 5.2, guides suppliers through the critical stages of data interaction, ensuring that they can effectively leverage the power of LLMs to enhance their participation in public procurement.

Furthermore, the study provides practical insights into crafting targeted prompts, iteratively refining queries, and ensuring the clarity of questions, which significantly enhances the quality of LLM responses and increases suppliers' ability to understand, access, and use procurement data. Using techniques such as few-shot learning and thought generation enables LLMs to provide more context-specific and relevant answers, thereby improving the overall usability of public procurement data. The findings also underscore the need for suppliers to validate the information provided by LLMs against authoritative sources, ensuring that their understanding of procurement data is trustworthy and dependable.

This study contributes to the existing body of knowledge by addressing the gap in research concerning the application of LLMs in public procurement, specifically from the supplier's perspective. It also recognizes the limitations of using a free ChatGPT and focusing on the Information Technology sector, suggesting the need for broader research across different sectors and with more advanced LLM models.

Future research should also explore the impact of LLMs in diverse cultural and legal contexts and address ethical and security implications. This encompasses the examination of diverse data at sources, sectors, and more advanced LLM models. It also requires exploring various prompt strategies to develop a comprehensive understanding of the role of LLMs in improving public procurement processes and ensuring that LLMs can be used to promote a more transparent and efficient process.

In conclusion, this exploratory study establishes a foundation for understanding and applying LLMs to improve public procurement data, fostering a more transparent, efficient, and inclusive marketplace for government suppliers. The practical implications of this research highlight the potential for LLMs to revolutionize public procurement processes, making them more accessible and manageable for suppliers of all sizes.

# 6
# Conclusion

This dissertation investigated the challenges associated with public procurement data quality, focusing on three key dimensions: discoverability, accessibility, and usability. The research adopted a dual-method approach consisting of a systematic mapping study to identify existing challenges and solutions and an exploratory study to evaluate the potential of Large Language Models (LLMs), such as ChatGPT, in addressing these issues. By integrating computational and user-centric methodologies, this study sought to bridge the gap between complex procurement datasets and their practical usability by government suppliers and other stakeholders. The findings provide a structured analysis of procurement data limitations and present an AI-driven approach to mitigating them, contributing to ongoing efforts to enhance public procurement transparency and efficiency.

## 6.1 Final Remarks

This study explored methods for assessing the quality of public procurement data, identifying four primary approaches: manual, automatic, statistical, and semi-automatic. Manual methods, though effective in capturing nuanced insights through expert evaluation, are resource-intensive and prone to human bias. Automatic techniques leverage artificial intelligence to efficiently analyze large datasets, excelling in detecting trends and anomalies but lacking qualitative depth. Statistical methods offer robust quantitative analysis but require advanced expertise and may overlook contextual details. Semi-automatic approaches combine the strengths of manual and automated methods, ensuring comprehensive evaluation but demanding complex implementation and significant computational resources. A hybrid approach integrating multiple methodologies is recommended for optimal procurement data assessment.

The research also examined challenges in procurement data discoverability, accessibility, and usability from technological, documental, regulatory, and data quality perspectives. Technological barriers include data format heterogeneity and interoperability issues across platforms, complicating integration and analysis. Documental challenges arise from incomplete or poorly structured datasets and inadequately validated ontologies. Regulatory constraints, such as inconsistent data access policies across jurisdictions, further hinder usability. Finally, data quality issues—including missing, inconsistent, or inaccurate records—undermine trust and usability. Addressing these challenges requires a comprehensive strategy that incorporates technological advancements, improved documentation, regulatory harmonization, and continuous data quality enhancement.

An analysis of thematic areas and government sectors in procurement research highlighted a strong focus on national-level studies, particularly in Information Technology (IT), regulatory frameworks, ethics, and transparency. Studies on IT procurement cover interoperability, data integration, and AI applications. Research on regulatory frameworks emphasizes the role of legal structures in procurement efficiency, as seen in studies on the Spanish Public Sector Contracting Platform and EU's TED dataset. Ethics-focused studies explore AI governance, fraud detection, and bias mitigation in contract awards. Additionally, healthcare procurement research examines COVID-19 procurement strategies and IT implementations in social welfare. Sustainability and construction procurement are also explored, with emphasis on public-private partnerships and infrastructure resilience projects.

Best practices for improving procurement data quality involve technological, documental, and regulatory measures. Adopting ICT solutions, decision support systems, and automated data mapping enhances transparency and efficiency. Standardizing data formats and ensuring interoperability are critical for seamless e-procurement operations. Techniques such as information retrieval, graph analysis, and text mining facilitate data standardization and fraud detection. Secure procurement systems require accessibility evaluation tools and Role-Based Access Control (RBAC) mechanisms, especially for Decentralized Autonomous Organizations (DAOs). Regulatory consistency and clear documentation reduce entry barriers and foster competitive procurement environments. Ensuring data reliability through compliance with quality standards and fostering collaboration among SMEs and larger suppliers can further improve accessibility and inclusivity in procurement.

The exploratory study demonstrated the potential of Large Language Models (LLMs), such as ChatGPT, in enhancing procurement data usability. By enabling natural language queries, LLMs improve data discoverability, making procurement information more accessible to non-experts. Their ability to generate concise, user-friendly responses facilitates navigation of complex procurement data, reducing barriers to participation. However, limitations such as reliance on prompt quality, the need for domain-specific fine-tuning, and risks of inaccurate responses underscore the necessity of further research. The study suggests refining prompt engineering techniques to enhance response precision, ensuring structured and iterative query refinement for improved data retrieval.

## 6.2 Future Works

Future research should focus on enhancing LLM capabilities for procurement applications through domain-specific fine-tuning and comparative evaluations of different LLM architectures. Expanding studies to regional and local procurement systems can provide insights into jurisdictional challenges and regulatory variations. Ethical considerations, including bias mitigation, misinformation risks, and legal compliance, must be systematically addressed. Integrating LLMs into procurement dashboards via APIs can improve real-time data retrieval and decision-making. Additionally, AI-driven procurement training programs can empower suppliers to optimize their engagement with procurement platforms. Further exploration of LLM applications in fraud detection, risk assessment, and contract anomaly identification could contribute to more transparent procurement processes.

# Bibliography

ABDELRAHMAN, O. H. Open government data: Development, practice, and challenges. In: *Open Data*. [S.l.]: IntechOpen, 2021. CC BY 3.0 License. Citado na página 14.

ALI, M.; ALEXOPOULOS, C.; CHARALABIDIS, Y. A comprehensive review of open data platforms, prevalent technologies, and functionalities. In: *Proceedings of the 15th International Conference on Theory and Practice of Electronic Governance*. [S.l.: s.n.], 2022. p. 203–214. Citado 2 vezes nas páginas 13 and 26.

ALMEIDA, G. et al. Improvement of transparency through mining techniques for reclassification of texts: The case of brazilian transparency portal. In: *Proceedings of the 19th Annual International Conference on Digital Government Research: Governance in the Data Age*. New York, USA: ACM, 2018. ISBN 9781450365260. Citado 18 vezes nas páginas 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 50, 53, 54, 55, 63, 64, and 65.

ANCARANI, A. et al. A comparative analysis of sme friendly public procurement: Results from canada, hungary and italy. *International Journal of Public Administration*, ROUTLEDGE JOURNALS, TAYLOR & FRANCIS LTD, 2-4 PARK SQUARE, MILTON PARK, ABINGDON OX14 4RN, OXON, ENGLAND, v. 42, n. 13, p. 1106–1121, 2019. ISSN 0190-0692. Citado 12 vezes nas páginas 37, 39, 42, 43, 45, 48, 51, 52, 53, 54, 55, and 64.

aO, M. A. B. et al. Plus: A semi-automated pipeline for fraud detection in public bids. *Digit. Gov.: Res. Pract.*, Association for Computing Machinery, New York, NY, USA, sep 2023. ISSN 2691-199X. Citado 13 vezes nas páginas 37, 38, 39, 40, 41, 42, 45, 47, 51, 53, 63, 64, and 65.

ARNEY, L. et al. Strategic contracting practices to improve procurement of health commodities. *Global Health Science and Practice*, Johns Hopkins University Press, v. 2, n. 3, p. 295 – 306, 2014. ISSN 2169575X. Cited by: 28; All Open Access, Gold Open Access, Green Open Access. Citado 10 vezes nas páginas 37, 38, 40, 41, 42, 45, 50, 63, 64, and 65.

ARROWSMITH, S. Public procurement: basic concepts and the coverage of procurement rules. In: ARROWSMITH, S.; TREUMER, S.; FEJØ, J. (Ed.). *Public Procurement Regulation: An Introduction*. Nottingham, UK: The EU Asia Inter University Network for Teaching and Research in Public Procurement Regulation, 2010. Citado na página 25.

ASTBRINK, G.; TIBBEN, W. Ict accessibility criteria in public procurement in oecd countries–the current situation. In: SPRINGER. *7th International Conference, UAHCI 2013, Las Vegas, USA*. [S.l.], 2013. p. 155–164. Citado 15 vezes nas páginas 37, 38, 39, 40, 41, 42, 43, 45, 46, 47, 51, 54, 63, 64, and 65.

ATTARD, J. et al. A systematic review of open government data initiatives. *Government Information Quarterly*, v. 32, n. 4, p. 399–418, 2015. ISSN 0740-624X. Citado 2 vezes nas páginas 38 and 61.

AZARIA, A.; AZOULAY, R.; RECHES, S. Chatgpt is a remarkable tool—for experts. *Data Intelligence*, MIT Press One Rogers Street, Cambridge, MA 02142-1209, USA journals-info . . . , v. 6, n. 1, p. 240–296, 2024. Citado na página 75.

BACHTIAR, A. e. a. Literature review of open government data. In: *2020 International Conference on Information Technology Systems and Innovation (ICITSI)*. [S.l.]: IEEE, 2020. p. 329–334. Citado na página 24.

BALAEVA, O. et al. Public procurement efficiency as perceived by market participants: The case of russia. *INTERNATIONAL JOURNAL OF PUBLIC ADMINISTRATION*, ROUTLEDGE JOURNALS, TAYLOR & FRANCIS LTD, 2-4 PARK SQUARE, MILTON PARK, ABINGDON OX14 4RN, OXON, ENGLAND, v. 45, n. 16, p. 1156–1167, DEC 10 2022. ISSN 0190-0692. Citado 14 vezes nas páginas 37, 38, 39, 40, 41, 42, 45, 47, 48, 51, 54, 63, 64, and 65.

BARCELLOS, R. et al. Exploring interpretability in open government data with chatgpt. In: *Proceedings of the 25th Annual International Conference on Digital Government Research*. [S.l.: s.n.], 2024. p. 186–195. Citado na página 83.

BASILI, V. R.; ROMBACH, H. D. The tame project: Towards improvement-oriented software environments. *IEEE Transactions on software engineering*, IEEE, v. 14, n. 6, p. 758–773, 1988. Citado na página 31.

BAUHR, M. et al. Lights on the shadows of public procurement: Transparency as an antidote to corruption. *Governance*, Wiley Online Library, v. 33, n. 3, p. 495–523, 2020. Citado na página 61.

BEHR, T. V.; ABRAHAMSSON, P. Ai governance and ethics in public procurement: Bridging the gap between theory and practice. In: *2022 IEEE 28th ICE/ITMC  31st IAMOT Joint Conference*. [S.l.: s.n.], 2022. p. 1–7. Citado 11 vezes nas páginas 37, 38, 40, 41, 42, 43, 45, 46, 50, 53, and 63.

BJARNASON, E.; PERSSON, J.; RYDENFäLT, C. Initial case study findings for requirements on work-related health aspects. In: *2023 IEEE 31st REW*. [S.l.: s.n.], 2023. p. 388–396. Citado 8 vezes nas páginas 37, 40, 41, 42, 43, 51, 53, and 64.

BRASIL. *Cartilha Acesso à Informação na Prática: Orientações para cidadãos, gestores públicos e Tribunais de Contas*. [S.l.: s.n.], 2023. Citado na página 25.

BROWN, T. et al. Language models are few-shot learners. *Advances in neural information processing systems*, v. 33, p. 1877–1901, 2020. Citado 3 vezes nas páginas 27, 28, and 29.

CADAMURO, J. et al. Potentials and pitfalls of chatgpt and natural-language artificial intelligence models for the understanding of laboratory medicine test results. an assessment by the european federation of clinical chemistry and laboratory medicine (eflm) working group on artificial intelligence (wg-ai). *Clinical Chemistry and Laboratory Medicine (CCLM)*, De Gruyter, v. 61, n. 7, p. 1158–1166, 2023. Citado na página 74.

CALDIERA, V. R. B. G.; ROMBACH, H. D. The goal question metric approach. *Encyclopedia of software engineering*, p. 528–532, 1994. Citado na página 67.

CARDOSO, R. A. L.; CARNEIRO, G. de F.; MENESES, J. E. X. de. Dados abertos governamentais no contexto de políticas públicas de saúde e sistemas prisionais: Realidade ou utopia? *arXiv preprint*, arXiv:2103.00541, 2021. Citado na página 14.

CARNEIRO, G. de F. et al. Perspectives and challenges in the analysis of prison systems data: A systematic mapping. *arXiv preprint*, arXiv:2101.03539, 2021. Citado na página 14.

CHEN, Y. et al. Iterative prompt refinement for mining gene relationships from chatgpt. *bioRxiv*, Cold Spring Harbor Laboratory Preprints, 2023. Citado na página 80.

CHENG, J. et al. Dated data: Tracing knowledge cutoffs in large language models. *arXiv preprint arXiv:2403.12958*, 2024. Citado na página 82.

CONCHA, G.; BURR, G.; SUáREZ, P. Study on the public market platform's impact on sme in chile. In: *Proceedings of the 8th International Conference on Theory and Practice of Electronic Governance*. New York, USA: ACM, 2014. p. 435–441. ISBN 9781605586113. Citado 11 vezes nas páginas 37, 38, 39, 40, 42, 43, 45, 55, 63, 64, and 65.

CSAKI, C. Towards open data quality improvements based on root cause analysis of quality issues. In: PARYCEK, P. et al. (Ed.). *ELECTRONIC GOVERNMENT (EGOV 2018)*. [S.l.]: SPRINGER INTERNATIONAL PUBLISHING AG, 2018. (Lecture Notes in Computer Science, v. 11020), p. 208–220. ISBN 978-3-319-98690-6; 978-3-319-98689-0. ISSN 0302-9743. Citado 15 vezes nas páginas 37, 38, 39, 40, 41, 42, 43, 45, 47, 48, 50, 51, 54, 64, and 65.

DAHBI, K. Y.; CHIADMI, D.; LAMHARHAR, H. Knowledge graph based representation to extract value from open governement data an application to the public procurement domain. *International Journal of Advanced Computer Science and Applications*, Science and Information Organization, v. 14, n. 3, p. 255 – 262, 2023. ISSN 2158107X. Citado 15 vezes nas páginas 37, 38, 39, 40, 41, 42, 43, 45, 46, 49, 51, 53, 63, 64, and 65.

DAVIES, T. et al. *The state of open data: Histories and horizons*. [S.l.]: African Minds, 2019. Citado na página 26.

DEVLIN, J. et al. Bert: Pre-training of deep bidirectional transformers for language understanding. In: *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*. Association for Computational Linguistics, 2019. p. 4171–4186. Disponível em: <https://aclanthology.org/N19-1423>. Citado na página 28.

DONKER, F. W.; LOENEN, B. V. How to assess the success of the open data ecosystem? *International Journal of Digital Earth*, Taylor & Francis, v. 10, n. 3, p. 284–306, 2017. Citado na página 38.

DYBÅ, T.; DINGSØYR, T. Empirical studies of agile software development: A systematic review. *Information and software technology*, Elsevier, v. 50, n. 9-10, p. 833–859, 2008. Citado na página 32.

FAZEKAS, M.; CZIBIK, Á. Measuring regional quality of government: the public spending quality index based on government contracting data. *Regional Studies*, Taylor & Francis, v. 55, n. 8, p. 1459–1472, 2021. Citado na página 61.

FERREIRA, I.; AMARAL, L. A. Public e-procurement: Advantages, limitations and technological "pitfalls". In: *Proceedings of the 9th International Conference on Theory and Practice of Electronic Governance*. New York, NY, USA: ACM, 2016. p. 9–12. ISBN 9781450336406. Citado 11 vezes nas páginas 37, 38, 39, 40, 41, 42, 43, 46, 54, 63, and 64.

GAVUROVA, B.; KUBAK, M. The efficiency evaluation of public procurement of medical equipment. *ENTREPRENEURSHIP AND SUSTAINABILITY ISSUES*, ENTERPRENEURSHIP & SUSTAINABILITY CENTER, M K CIURLIONIO STR 86A, VILNUS, 03100, LITHUANIA,

v. 8, n. 3, p. 193–209, MAR 2021. ISSN 2345-0282. Citado 10 vezes nas páginas 37, 38, 39, 40, 41, 42, 43, 45, 50, and 64.

GONçALVES, R. et al. Portuguese web accessibility in electronic public procurement platforms. In: *5th Iberian Conference on Information Systems and Technologies*. [S.l.: s.n.], 2010. p. 1–5. Citado 11 vezes nas páginas 37, 39, 41, 42, 43, 45, 50, 53, 54, 63, and 64.

GROUP, S. E. *Guidelines for performing systematic literature reviews in software engineering*. [S.l.], 2007. Citado 2 vezes nas páginas 30 and 61.

IMAMOGLU, M. Y.; REHAN, M. Cost anaylsis of interoperable based e-procurement system in the republic of turkey. In: *Proceedings of the The International Conference on Engineering amp; MIS 2015*. New York, USA: ACM, 2015. (ICEMIS '15). ISBN 9781450334181. Citado 13 vezes nas páginas 37, 38, 40, 41, 42, 43, 44, 45, 48, 51, 52, 53, and 63.

KASHAP, S. Public procurement as a social, economic and political policy. In: *International Public Procurement Conference*. Fort Lauderdale, Florida, USA: [s.n.], 2004. p. 133–147. Citado na página 25.

KITCHENHAM, B.; CHARTERS, S. *Guidelines for Performing Systematic Literature Reviews in Software Engineering*. [S.l.], 2007. Citado na página 21.

KLUN, M.; SETNIKAR-CANKAR, S. Better regulation and public procurement in slovenian municipalities. *TRANSYLVANIAN REVIEW OF ADMINISTRATIVE SCIENCES*, BABES-BOLYAI UNIV, FACULTY POLITICAL, ADMINSTRATIVE & COMMUNICATION SCIENCES, PUBLIC ADMIN DEPT, CENTER RES PUBLIC ADM, CLUJ-NAPOCA, 400132, ROMANIA, n. 38E, p. 96–105, FEB 2013. ISSN 1842-2845. Citado 11 vezes nas páginas 37, 39, 41, 42, 43, 45, 47, 48, 52, 53, and 64.

KOJIMA, T. et al. Large language models are zero-shot reasoners. *Advances in neural information processing systems*, v. 35, p. 22199–22213, 2022. Citado na página 29.

LALIĆ, B. et al. The impact of quality management purchasing practices on purchasing performance in transitional economies. *Tehnički vjesnik*, Strojarski fakultet u Slavonskom Brodu; Fakultet elektrotehnike, računarstva ..., v. 26, n. 3, p. 815–822, 2019. Citado 9 vezes nas páginas 37, 40, 41, 42, 43, 45, 54, 55, and 64.

LEE, J.; OH, K.-T.; KWON, H. Y. Striving for transparency and efficiency in e-government: procurement reform through e-procurement. In: *Proceedings of the 2nd international conference on Theory and practice of electronic governance*. [S.l.: s.n.], 2008. p. 183–190. Citado 12 vezes nas páginas 37, 38, 39, 40, 41, 42, 43, 45, 47, 48, 50, and 64.

LIU, P. et al. Pre-train, prompt, and predict: A systematic survey of prompting methods in natural language processing. *ACM Computing Surveys*, ACM New York, NY, v. 55, n. 9, p. 1–35, 2023. Citado 2 vezes nas páginas 27 and 28.

LOU, R.; ZHANG, K.; YIN, W. Is prompt all you need? no. a comprehensive and broader view of instruction learning. *arXiv preprint arXiv:2303.10475*, 2023. Citado na página 80.

LOUKIS, E. et al. Chatgpt application vis-a-vis open government data (ogd): Capabilities, public values, issues and a research agenda. In: SPRINGER. *International Conference on Electronic Government*. [S.l.], 2023. p. 95–110. Citado na página 83.

MAMALIS, M. E. et al. Can large language models revolutionalize open government data portals? a case of using chatgpt in statistics. gov. scot. In: *Proceedings of the 27th Pan-Hellenic Conference on Progress in Computing and Informatics*. [S.l.: s.n.], 2023. p. 53–59. Citado na página 83.

MARTINS, M. et al. En 301 549: Parameter decoding, evolution and translation into european portuguese - steps towards the creation of a tool for the application of ict accessibility requirements in public procurement. In: *8th International Conference on ICT Accessibility (ICTA)*. [S.l.: s.n.], 2021. p. 1–6. Citado 9 vezes nas páginas 37, 38, 40, 41, 42, 43, 45, 55, and 63.

MAVIDIS, A.; FOLINAS, D. From public e-procurement 3.0 to e-procurement 4.0; a critical literature review. *Sustainability*, MDPI, v. 14, n. 18, p. 11252, 2022. Citado 3 vezes nas páginas 38, 46, and 47.

MEIJER, A. Understanding the complex dynamics of transparency. *Public administration review*, Wiley Online Library, v. 73, n. 3, p. 429–439, 2013. Citado na página 61.

MELNIKOV, V. V.; LUKASHENKO, O. A. Evolution of public procurement auctions in russia. *Terra Economicus*, Southern Federal University, v. 17, n. 1, p. 150 – 173, 2019. ISSN 20736606. Citado 7 vezes nas páginas 37, 41, 42, 43, 45, 51, and 53.

MELON, L.; SPRUK, R. The impact of e-procurement on institutional quality. *JOURNAL OF PUBLIC PROCUREMENT*, EMERALD GROUP PUBLISHING LTD, HOWARD HOUSE, WAGON LANE, BINGLEY BD16 1WA, W YORKSHIRE, ENGLAND, v. 20, n. 4, p. 333–375, OCT 27 2020. ISSN 1535-0118. Citado 7 vezes nas páginas 37, 41, 42, 44, 45, 51, and 55.

MENDES, M.; VOIGT, J. Open government data as a panacea against corruption and mismanagement: An analysis of open public data on covid-19-related procurement at the sub-national level. In: *Proceedings of the 15th International Conference on Theory and Practice of Electronic Governance*. [S.l.: s.n.], 2022. p. 180–186. Citado na página 15.

MENDES, M.; VOIGT, J. Open government data as a panacea against corruption and mismanagement: An analysis of open public data on covid-19-related procurement at the sub-national level. In: *Proceedings of the 15th International Conference on Theory and Practice of Electronic Governance*. New York, USA: ACM, 2022. (ICEGOV '22), p. 180–186. ISBN 9781450396356. Citado 15 vezes nas páginas 37, 38, 39, 40, 41, 42, 43, 45, 46, 50, 51, 55, 63, 64, and 65.

MODRUŠAN, N.; MRŠIĆ, L.; RABUZIN, K. Intelligent public procurement monitoring system powered by text mining and balanced indicators. In: SPRINGER. *International Conference on Data Management Technologies and Applications*. [S.l.], 2020. p. 115–133. Citado 12 vezes nas páginas 37, 38, 39, 41, 42, 43, 45, 47, 50, 63, 64, and 65.

MOHER, D. et al. Preferred reporting items for systematic reviews and meta-analyses: the prisma statement. *Annals of internal medicine*, American College of Physicians, v. 151, n. 4, p. 264–269, 2009. Citado na página 33.

MONDAL, S.; BAPPON, S. D.; ROY, C. K. Enhancing user interaction in chatgpt: Characterizing and consolidating multiple prompts for issue resolution. In: IEEE. *2024 IEEE/ACM 21st International Conference on Mining Software Repositories (MSR)*. [S.l.], 2024. p. 222–226. Citado na página 80.

MONDORF, A.; WIMMER, M. Interoperability in e-tendering: the case of the virtual company dossier. In: *Proceedings of the 2nd international conference on Theory and practice of electronic governance*. [S.l.: s.n.], 2008. p. 110–116. Citado 16 vezes nas páginas 37, 38, 39, 40, 41, 42, 43, 45, 46, 47, 48, 50, 52, 63, 64, and 65.

MONDORF, A.; WIMMER, M. A.; REISER, D. A framework for interoperability testing in pan-european public service provision. In: SPRINGER. *Electronic Government: 12th IFIP WG 8.5 International Conference, EGOV 2013*. [S.l.], 2013. p. 188–199. Citado 17 vezes nas páginas 37, 38, 39, 40, 41, 42, 43, 45, 46, 48, 50, 53, 54, 55, 63, 64, and 65.

MONTEIRO, F.; CORREIA, M. Decentralised autonomous organisations for public procurement. In: *Proceedings of the 27th International Conference on Evaluation and Assessment in Software Engineering*. [S.l.: s.n.], 2023. p. 378–385. Citado 13 vezes nas páginas 37, 38, 39, 40, 41, 43, 45, 47, 48, 50, 53, 64, and 65.

MU, F. et al. Clarifygpt: A framework for enhancing llm-based code generation via requirements clarification. *Proceedings of the ACM on Software Engineering*, ACM New York, NY, USA, v. 1, n. FSE, p. 2332–2354, 2024. Citado na página 80.

MUHWEZI, M. et al. Information integration, procurement internal controls, material and purchasing procedure standardization and procurement performance in humanitarian organizations. *Journal of Humanitarian Logistics and Supply Chain Management*, Emerald Publishing Limited, jan. 2023. Citado 7 vezes nas páginas 37, 40, 41, 42, 45, 63, and 64.

NAI, R. et al. Exploitation and merge of information sources for public procurement improvement. In: *Machine Learning and Principles and Practice of Knowledge Discovery in Databases, ECML PKDD*. [S.l.]: SPRINGER, 2023. v. 1752, p. 89–102. ISBN 978-3-031-23617-4; 978-3-031-23618-1. ISSN 1865-0929. Citado 15 vezes nas páginas 37, 38, 39, 40, 41, 42, 43, 45, 47, 49, 51, 53, 55, 63, and 65.

NUNES, L. dos S.; MORENO, E. D.; CARNEIRO, G. de F. A review of covid-19 data usage in correctional institutions from developing countries. *IEEE Transactions on Computational Social Systems*, 2023. Citado na página 14.

NURMANDI, A.; KIM, S. Making e-procurement work in a decentralized procurement system. *International Journal of Public Sector Management*, Emerald Group Publishing Limited, v. 28, n. 3, p. 198–220, jan. 2015. Citado 10 vezes nas páginas 38, 39, 40, 41, 42, 43, 45, 47, 63, and 64.

OBAMA, B. Transparency and open government, memorandum for the heads of executive departments and agencies. *http://www. whitehouse. gov-/the_press_office/TransparencyandOpenGovernment*, 2009. Citado na página 31.

OLIVEIRA, A. C. et al. Do brazilian federal agencies specify accessibility requirements for the development of their mobile apps? In: *Proceedings of the XVI Brazilian Symposium on Information Systems*. New York, USA: ACM, 2020. ISBN 9781450388733. Citado 12 vezes nas páginas 37, 38, 39, 41, 42, 43, 45, 46, 51, 54, 55, and 64.

OpenDataSoft. *What is Open Data? A Practical Guide*. 2024. Accessed: 2024-10-16. Disponível em: <https://www.opendatasoft.com/en/what-is-open-data-practical-guide/>. Citado na página 13.

OZYUREK, I.; ERDAL, M. Assessment of qualification criteria described in public procurement law code 4734 in construction works by analytic hierarchy process (ahp). *GAZI UNIVERSITY JOURNAL OF SCIENCE*, GAZI UNIV, CENTER CAMPUS TECHNOLOGY FAC B BLOCK EK BINA, 2ND FL, ANKARA, 06500, TURKEY, v. 31, n. 2, p. 437–454, 2018. ISSN 2147-1762. Citado 7 vezes nas páginas 37, 38, 40, 43, 45, 53, and 65.

PATRUCCO, A. S.; AGASISTI, T.; GLAS, A. H. Structuring public procurement in local governments: The effect of centralization, standardization and digitalization on performance. *PUBLIC PERFORMANCE & MANAGEMENT REVIEW*, ROUTLEDGE JOURNALS, TAYLOR & FRANCIS LTD, 2-4 PARK SQUARE, MILTON PARK, ABINGDON OX14 4RN, OXON, ENGLAND, v. 44, n. 3, p. 630–656, MAY 4 2021. ISSN 1530-9576. Citado 5 vezes nas páginas 37, 41, 42, 43, and 45.

PEREIRA, J. et al. Inacia: Integrating large language models in brazilian audit courts: Opportunities and challenges. *Digital Government: Research and Practice*, ACM New York, NY, 2024. Citado na página 83.

PERKOVIĆ, G.; DROBNJAK, A.; BOTIČKI, I. Hallucinations in llms: Understanding and addressing challenges. In: IEEE. *2024 47th MIPRO ICT and Electronics Convention (MIPRO)*. [S.l.], 2024. p. 2084–2088. Citado na página 28.

PINTO, F. C. et al. Avaliação da transparência do sistema de compras do governo brasiliero: Uma visão baseada na perspectiva da usabilidade. In: *2015 Latin American Computing Conference (CLEI)*. [S.l.: s.n.], 2015. p. 1–11. Citado 13 vezes nas páginas 37, 38, 40, 41, 42, 43, 45, 48, 50, 51, 53, 63, and 64.

PURNOMO, A. et al. E-procurement research mapping: Lesson from bibliometric approach (1975-2019). In: IEEE. *2021 ICIMTech*. [S.l.], 2021. v. 1, p. 801–806. Citado na página 61.

PUTRI, M. E.; RULDEVIYANI, Y. Prioritization strategy for government's website information quality. case study: Indonesia national public procurement agency. In: *2019 ITCC*. [S.l.]: ACM, 2019. p. 37–44. ISBN 978-1-4503-7228-2. Citado 12 vezes nas páginas 37, 38, 39, 40, 41, 42, 43, 45, 46, 52, 63, and 65.

RADFORD, A. et al. Improving language understanding by generative pre-training. San Francisco, CA, USA, 2018. Citado na página 27.

REJEB, A. et al. The landscape of public procurement research: a bibliometric analysis and topic modelling based on scopus. *Journal of Public Procurement*, Emerald Publishing Limited, v. 23, n. 2, p. 145–178, jan. 2023. Citado na página 61.

RIBEIRO, C. G. e. a. Unveiling the public procurement market in brazil: A methodological tool to measure its size and potential. *Development Policy Review*, v. 36, p. O360–O377, 2018. Citado na página 14.

RIIHIAHO, S. et al. Procuring usability: Experiences of usability testing in tender evaluation. In: . [S.l.]: SPRINGER-VERLAG BERLIN, 2015. v. 223, p. 108–120. ISBN 978-3-319-21783-3; 978-3-319-21782-6. ISSN 1865-1348. 6th Scandinavian Conference on Information Systems, FINLAND. Citado 9 vezes nas páginas 37, 38, 39, 40, 41, 42, 43, 45, and 64.

RODRÍGUEZ, M. J. G. et al. Spanish public procurement: legislation, open data source and extracting valuable information of procurement announcements. *Procedia Computer Science*, Elsevier, v. 164, p. 441–448, 2019. Citado na página 15.

RODRIGUEZ, M. J. G. et al. Spanish public procurement: legislation, open data source and extracting valuable information of procurement announcements. In: *CENTERIS2019, SYSTEMS/PROJMAN2019,HCIST2019*. [S.l.]: ELSEVIER, 2019. v. 164, p. 441–448. ISSN 1877-0509. Citado 13 vezes nas páginas 37, 38, 39, 40, 41, 42, 43, 45, 47, 50, 63, 64, and 65.

SAHOO, P. et al. A systematic survey of prompt engineering in large language models: Techniques and applications. *arXiv preprint arXiv:2402.07927*, 2024. Citado na página 28.

SANGIL, M. J. Exploratory data analysis of government procurement data to influence bidding decision and strategy in albay, philippines. In: *2020 IEEE International Symposium on Technology and Society (ISTAS)*. [S.l.: s.n.], 2020. p. 235–245. Citado 12 vezes nas páginas 37, 39, 40, 42, 43, 44, 45, 48, 51, 55, 64, and 65.

SCHULHOFF, S. et al. The prompt report: A systematic survey of prompting techniques. *arXiv preprint arXiv:2406.06608*, 2024. Citado 2 vezes nas páginas 27 and 29.

SILVA, E. H. M. da et al. Chattcu. *Revista do TCU*, v. 153, p. 19–45, 2024. Citado na página 83.

SILVA, R. C. D. et al. Sustainable public procurement: the federal public institution's shared system. *Revista de Gestão*, Emerald Publishing Limited, v. 25, n. 1, p. 9–24, jan. 2018. Citado 10 vezes nas páginas 37, 40, 41, 42, 43, 45, 48, 51, 54, and 64.

SIROTKINA, N.; LAZAREVICH, S. Electronic public procurement: Case of russia. *INTERNATIONAL JOURNAL OF PUBLIC ADMINISTRATION*, ROUTLEDGE JOURNALS, TAYLOR & FRANCIS LTD, 2-4 PARK SQUARE, MILTON PARK, ABINGDON OX14 4RN, OXON, ENGLAND, v. 46, n. 11, p. 783–794, AUG 18 2023. ISSN 0190-0692. Citado 14 vezes nas páginas 37, 38, 39, 40, 41, 42, 43, 45, 46, 48, 51, 63, 64, and 65.

SOYLU, A. et al. Data quality barriers for transparency in public procurement. *Information*, MDPI, v. 13, n. 2, p. 99, 2022. Citado na página 15.

SOYLU, A. et al. Data quality barriers for transparency in public procurement. *Information*, MDPI, ST ALBAN-ANLAGE 66, CH-4052 BASEL, SWITZERLAND, v. 13, n. 2, FEB 2022. Citado 15 vezes nas páginas 37, 38, 39, 40, 42, 43, 44, 45, 47, 48, 49, 53, 63, 64, and 65.

SOYLU, A. et al. Theybuyforyou platform and knowledge graph: Expanding horizons in public procurement with open linked data. *SEMANTIC WEB*, IOS PRESS, NIEUWE HEMWEG 6B, 1013 BG AMSTERDAM, NETHERLANDS, v. 13, n. 2, p. 265–291, 2022. ISSN 1570-0844. Citado 13 vezes nas páginas 37, 38, 39, 40, 41, 42, 43, 45, 46, 49, 53, 63, and 65.

SPACEK, D.; SPACKOVA, Z. Issues of e-government services quality in the digital-by-default era - the case of the national e-procurement platform in czechia. *JOURNAL OF PUBLIC PROCUREMENT*, EMERALD GROUP PUBLISHING LTD, HOWARD HOUSE, WAGON LANE, BINGLEY BD16 1WA, W YORKSHIRE, ENGLAND, v. 23, n. 1, p. 1–34, FEB 24 2023. ISSN 1535-0118. Citado 12 vezes nas páginas 37, 38, 39, 40, 42, 43, 45, 48, 51, 55, 63, and 64.

STAKE, J. Evaluating quality or lowest price: consequences for small and medium-sized enterprises in public procurement. *JOURNAL OF TECHNOLOGY TRANSFER*, SPRINGER, ONE NEW YORK PLAZA, SUITE 4600, NEW YORK, NY, UNITED STATES, v. 42, n. 5, SI, p. 1143–1169, OCT 2017. ISSN 0892-9912. Citado 11 vezes nas páginas 37, 39, 41, 42, 43, 45, 48, 50, 51, 54, and 64.

STONE, P. W. Popping the (pico) question in research and evidence-based practice. *Applied nursing research: ANR*, v. 15, n. 3, p. 197–198, 2002. Citado na página 31.

TANG, R.; JIANG, J. Characteristics of open government data (ogd) around the world: A country-based comparative meta-analysis. In: *2020 ASIS&T Asia-Pacific Regional Conference (Virtual Conference)*. [s.n.], 2020. Disponível em: <https://doi.org/10.2478/dim-2020-0026>. Citado na página 14.

TAS, B. K. O. Effect of public procurement regulation on competition and cost-effectiveness. *Journal of Regulatory Economics*, SPRINGER, VAN GODEWIJCKSTRAAT 30, 3311 GZ DORDRECHT, NETHERLANDS, v. 58, n. 1, p. 59–77, AUG 2020. ISSN 0922-680X. Citado 13 vezes nas páginas 37, 38, 40, 41, 42, 43, 44, 45, 46, 48, 54, 63, and 64.

TOSIN, T. et al. A model for data integration in open and linked databases with the use of ontologies. In: *2016 35th International Conference of the Chilean Computer Science Society (SCCC)*. [S.l.: s.n.], 2016. p. 1–9. Citado 15 vezes nas páginas 37, 38, 40, 41, 42, 43, 45, 47, 48, 50, 51, 52, 63, 64, and 65.

TYLLINEN, M. et al. We need numbers! - heuristic evaluation during demonstrations (hed) for measuring usability in it system procurement. In: . [S.l.]: ACM, 2016. p. 4129–4141. ISBN 978-1-4503-3362-7. 34th Annual CHI Conference on Human Factors in Computing Systems (CHI4GOOD). Citado 10 vezes nas páginas 37, 40, 41, 42, 45, 48, 51, 53, 63, and 64.

UDUWAGE-DON, N. L. S.; HADIWATTAGE, C.; PANUWATWANICH, K. Enhancing the responsiveness of bids in public procurement: The case of sri lanka. *PUBLIC WORKS MANAGEMENT & POLICY*, SAGE PUBLICATIONS INC, 2455 TELLER RD, THOUSAND OAKS, CA 91320 USA, 2023 JUL 27 2023. ISSN 1087-724X. Citado 15 vezes nas páginas 37, 38, 39, 40, 41, 42, 43, 47, 48, 49, 51, 55, 63, 64, and 65.

UGALE, G.; HALL, C. Generative ai for anti-corruption and integrity in government: Taking stock of promise, perils and practice. OECD, 2024. Citado na página 83.

VASWANI, A. et al. Attention is all you need. *Advances in neural information processing systems*, v. 30, 2017. Citado na página 27.

VAZQUEZ-ROWE, I. et al. A method to include life cycle assessment results in choosing by advantage (cba) multicriteria decision analysis. a case study for seismic retrofit in peruvian primary schools. *SUSTAINABILITY*, MDPI, ST ALBAN-ANLAGE 66, CH-4052 BASEL, SWITZERLAND, v. 13, n. 15, AUG 2021. Citado 11 vezes nas páginas 37, 38, 39, 41, 42, 43, 45, 51, 54, 63, and 64.

VELASCO, R. B. et al. A decision support system for fraud detection in public procurement. *International Transactions in Operational Research*, Blackwell Publishing Ltd, v. 28, n. 1, p. 27 – 47, 2021. ISSN 09696016. Citado 15 vezes nas páginas 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 51, 52, 55, and 63.

WANG, B. et al. Task supportive and personalized human-large language model interaction: A user study. In: *Proceedings of the 2024 Conference on Human Information Interaction and Retrieval*. [S.l.: s.n.], 2024. p. 370–375. Citado na página 28.

WANG, F. et al. Paths to open government data reuse: A three-dimensional framework of information need, data and government preparation. *Information & Management*, Elsevier, v. 60, n. 8, p. 103879, 2023. Citado na página 40.

WASEEM, M. et al. Artificial intelligence procurement assistant: Enhancing bid evaluation. In: SPRINGER NATURE SWITZERLAND CHAM. *International Conference on Software Business*. [S.l.], 2023. p. 108–114. Citado na página 83.

WHITE, J. et al. A prompt pattern catalog to enhance prompt engineering with chatgpt. *arXiv preprint arXiv:2302.11382*, 2023. Citado na página 28.

WILKINSON, M. D. e. a. The fair guiding principles for scientific data management and stewardship. *Scientific Data*, v. 3, n. 1, p. 1–9, 2016. Citado na página 23.

YüCEL, B. *Generative vs. Extractive Language Models: Which one is right for your use case?* 2023. <https://haystack.deepset.ai/blog/generative-vs-extractive-models>. Accessed: 2024-10-15. Citado 3 vezes nas páginas 7, 27, and 28.

ZHAO, W. X. et al. A survey of large language models. *arXiv preprint arXiv:2303.18223*, 2023. Citado 3 vezes nas páginas 27, 72, and 83.

ZHIQIANG, M. et al. Unobserved heterogeneity in public procurement governance and value for money. *Lex Localis - Journal of Local Self-Government*, INST LOCAL SELF-GOVERNMENT MARIBOR, SMETANOVA ULICA 30, MARIBOR, 2000, SLOVENIA, v. 18, n. 1, p. 95–121, JAN 2020. ISSN 1581-5374. Citado 9 vezes nas páginas 37, 38, 42, 43, 45, 52, 53, 54, and 64.

ZHOU, X. et al. A map of threats to validity of systematic literature reviews in software engineering. In: IEEE. *2016 23rd APSEC*. [S.l.], 2016. p. 153–160. Citado na página 61.