Transparency and Traceability of Congressional Earmarks: a System for Monitoring and Analyzing Public Data

Fernando Antonio Dantas Gomes Pinto[1], **Jônatas Tavares Pessanha**[1], **Bruno Bondarovsky**[1]
Sérgio Lifschitz[1], **Edward Hermann Haeusler**[1]

[1]Pontifícia Universidade Católica do Rio de Janeiro (PUC-Rio)
Departamento de Informática
Rio de Janeiro – RJ – Brazil

{fpinto, sergio, hermann, jpessanha}@inf.puc-rio.br

bruno.bondarovsky@treetech.inf.puc-rio.br

Abstract. *Transparency in public funds is essential for oversight and social control. However, some current monitoring solutions for* **Congressional Earmarks** *are* ***incomplete****. Official government sources are distributed across multiple endpoints, making data extraction challenging and creating user uncertainties. In addition, other initiatives are limited to specific types of congressional earmarks, failing to provide a comprehensive view of the process. This paper presents the development of a system capable of tracking the entire lifecycle of congressional earmarks, from proposal to payment. The solution was made with a relational database, allowing structured queries to identify authors, recipients, amounts, and congressional earmarks locations (municipal, state, and regional). Additionally, the system integrates Business Intelligence (BI) tools to generate interactive analytical views, facilitating data comprehension and supporting decision-making. With a clear and user-friendly interface, the platform improves transparency and is essential for researchers, journalists, and citizens interested in monitoring budget allocations.*

1. Introduction

Transparency in allocating public funds is essential to ensure accountability and social accountability. One of the main instruments for directing public resources in Brazil is the Legislative Earmarks[1], which allow legislators (deputies and senators) to modify the national budget to allocate funds to specific areas such as states, municipalities, and institutions (private, public, civil society organization). Congressional earmarks were introduced in Brazil's budget legislation with the 1988 Constitution [Brasil 1988] as a mechanism for the legislative to influence the budget allocation and decentralize control over public finances. Since then, this instrument has undergone several regulations and modifications, including introducing the mandatory execution of **congressional earmarks**.

However, existing solutions for monitoring these congressional earmarks are incomplete. Official government sources are distributed across multiple endpoints [da União 2025] [Federal 2025] [Parlamentar 2023], making data extraction challenging and creating uncertainties for users. Furthermore, many available initiatives are

[1]Known as "Emendas Parlamentares" in Portuguese

limited to specific types of congressional earmarks, failing to provide a comprehensive view of the process. The lack of standardization and the dispersion of data hinder transparency and make it difficult to track the stages from proposal to payment.

Recent legal and journalistic developments highlight the urgency of addressing the transparency gap in congressional earmarks. One example is the front-page report from *Folha de São Paulo*, which stated that over BRL 150 billion has been allocated through congressional amendments between 2020 and 2025, often with limited public oversight and fragmented, non-standardized data [de S.Paulo 2025]. Judicial rulings by the Brazilian Supreme Federal Court (STF) have declared certain categories of amendments unconstitutional—such as the Rapporteur's Amendments (RP9)—due to their lack of traceability and the anonymity of their authors [Federal 2023, de Contas da União (TCU) 2021]. Similar issues were identified in Commission Amendments and PIX Amendments, which were suspended in 2024 for failing to comply with constitutional principles of transparency and specificity, as documented in an official letter from the Brazilian Congress to the Supreme Court in response to cases ADPF 854, ADI 7688, ADI 7695, and ADI 7697 [Nacional 2024].

Brazilian news outlets have repeatedly reported on the structural weaknesses and irregularities associated with the amendment system. *Poder360* highlighted that even after a new legislative proposal intended to regulate the allocation of amendments was approved, it preserved loopholes that limit transparency and prevent full traceability [Poder360 2025]. *Estadão* revealed cases in which earmarked funds were directed to companies linked to congressional substitutes and even to a gas station owned by a sitting deputy, illustrating the extent to which these public resources are vulnerable to clientelism and private appropriation [de S. Paulo 2025]. Furthermore, in early 2025, *O Globo* reported that Brazilian civil society organizations submitted a formal complaint to the United Nations denouncing systemic waste, opacity, and corruption linked to the use of parliamentary amendments [Globo 2025]. These developments indicate that existing monitoring tools are not only technically insufficient but also inadequate from a governance and accountability perspective.

To address this issue, this article presents the development of a system capable of tracking the entire lifecycle of congressional earmarks [das Emendas 2025]. The solution employs a relational database, enabling structured queries to identify authors, recipients, amounts, and congressional earmark locations (municipal, state, and regional levels). Additionally, the system integrates Business Intelligence (BI) tools to generate interactive analytical views, facilitating data comprehension and supporting decision-making. With a clear and user-friendly interface, the platform improves transparency and is essential for researchers, journalists, and citizens interested in monitoring budget allocations.

To support the development of this system, we rely on official documents and guidelines, such as the Manual of congressional earmarks published by the Federal Government [do Brasil 2025a], as well as data available on the Portal de Transparência [da União 2025]. These sources provide essential details on the rules, types, and execution of congressional earmarks, ensuring that the proposed solution follows current regulations.

To have a trustful platform, the provenance of the data, and later, knowledge, has to be fully implemented. Provenance in databases and knowledge bases refers to the history and origin of data or information, explaining where it came from and how it was produced; see [Buneman et al. 2001] and [Cheney et al. 2009]. In databases, provenance—also called data lineage—tracks the source of data, the transformations applied to it, and the specific input that led to a result, helping with auditability and explanation of query outputs. In knowledge bases, provenance includes the source of facts, the reasoning or inference steps that derived them, and any metadata related to trust or authorship. It is key in ensuring data integrity, trust, reproducibility, and traceability. The platform we present here has a where-provenance fully implemented and mostly how-provenance elements. In the conclusion, we mention how we will achieve a general provenance implementation.

2. Related Works

Each subsequent subsection describes a platform, or tool, that serves as a platform for analyzing and querying the status of the **congressional earmarks**.

2.1. Portal da Transparência - CGU

Portal da Transparência is the platform developed by the Government Accountability Office, GAO[2]. It is a system made with Microsoft Power BI. It uses technology to promote public accountability and open data, offering real-time access to information on government spending, contracts, and public officials' salaries through an interface designed to be user-friendly, interactive dashboards, and downloadable datasets. The portal ensures data consistency, security, and scalability by adopting open data standards and integrating with federal systems. It supports advanced search features and APIs, enabling developers, researchers, and citizens to analyze and reuse data, thus fostering transparency, civic engagement, and trust in public institutions. That platform is incomplete in some information, e.g., the federative unit of the congressional earmarks author. This is important to distinguish parliamentarians with the same name.

2.2. Siga Brasil

SIGA Brasil is a system created by the Consultancy for Budget, Oversight, and Control of the Congress and Prodasen[3] to allow broad and easy access to data from the Integrated Financial Administration System – SIAFI [do Brasil 2025b]. One of its panels is specific to Congressional Earmarks. Despite being a system with good usability, it does not present a geographic map of the distribution of the Earmarks. This important feature is present in our solution. Furthermore, this platform is incomplete in some information, e.g., the federative unit of the congressional earmarks author. This is important to distinguish parliamentarians with the same name.

2.3. Frente FIT

The Frente Parlamentar Mista de Fiscalização, Integridade e Transparência (FIT) is a non-partisan initiative created in 2023, comprising more than 203 deputies and senators from 18 parties. Its objective is to promote oversight of public acts, integrity, transparency, and

[2]CGU - Controladoria Geral da União in Brazilian Portuguese
[3]The TI company of the Senate

the fight against corruption with the collaboration of experts and civil society organizations [Parlamentar 2023].

The Frente FIT is a tool composed of 3 main panels: Transferências Voluntárias; Emendas PIX [4] e Agenda das Autoridades[5]. We will only present the Emendas PIX panel, our context for this work.

Since it only deals with "Emendas PIX", it is **incomplete**. There are other types of emendas such as: "Emendas de Relator", "Emendas de Bancada", "Emendas de Comissão", "Emendas Individuais - Transferências com Finalidade Definida".They are not contemplated by Frente Fit.

Another important aspect is that the project data is outdated. According to the official website, the last update was made on 30 June 2024.

3. Research Design and Methods

This section details the materials and methods used for collecting, processing, and storing data related to congressional earmarks, legislative history, parliamentarians, party affiliations, Union expenses, and creditors. The platform was developed to structure this information in a relational PostgreSQL database and make it available for analysis through a Business Intelligence (BI) system.

3.1. Data Sources

The developed platform is based on open data provided by different sources from the Brazilian government, including the Câmara dos Deputados, the Federal Senate, the Transparency Portal, and the Brazilian Institute of Geography and Statistics (IBGE). These data are essential for understanding the historical evolution of legislatures, the profile of parliamentarians, their party affiliations, and the allocation of public resources through congressional earmarks. Additionally, detailed information on Union expenses and federal government creditors is included, allowing for a comprehensive analysis of budget execution.

Data are obtained through `REST APIs`, `XML` files, and `CSV` file format downloads, then processed and stored in a PostgreSQL [Douglas and Douglas 2003] database. The extraction and organization of this data enable the creation of interactive dashboards and analytical queries that facilitate data-driven decision-making.

The main data sources used include the following endpoints:

3.2. Extract, Transform and Load process.

Extract, transform, load (ETL) is a foundational process in data engineering that underpins every data, analytics, and AI workload [Kimball and Ross 2002]. The process was implemented to ensure the collection, standardization, and proper storage of data in the PostgreSQL database.

[4] The official name is Emendas Individuais - Transferências Especiais.

[5] Each one of these names refers to a specific type of congressional earmarks, each of them with a respective set of norms

Entity	Resource Link
Legislaturas	`dadosabertos.camara.leg.br/api/v2/legislaturas`
Parliamentarian data (deputies)	`dadosabertos.camara.leg.br/api/v2/deputados`
Parliamentarian data (senator)	`legis.senado.leg.br/dadosabertos/senador/lista/atual`
Mandates	`legis.senado.leg.br/dadosabertos/senador/01/mandatos`
Congressional earmarks	`portaldatransparencia.gov.br/download-de-dados/emendas-parlamentares`
Congressional earmarks documents	`portaldatransparencia.gov.br/download-de-dados/emendas-parlamentares-documentos/`
Demographic data	`sidra.ibge.gov.br/`
Business registration (CNPJ)	`portaldatransparencia.gov.br/download-de-dados/favorecidos-pj`

Table 1. Entities and Resource Links used in the project.

As already presented, data extraction was performed from `REST APIs`, `XML` files, and `CSV` downloads provided by government institutions. API access was managed using the request library, while XML files were processed with `xml.etree.ElementTree`. This process is automated to ensure that the database remains continuously updated.

During the transformation phase, the data undergo a series of treatments for normalization and cleaning. Techniques are applied to ensure information consistency by removing redundancies, filling in missing values. In addition, the encoding of `CSV` files is converted to maintain compatibility between different data sources. Data standardization is also performed, such as unifying the names of congressional earmarks authors. In some sources, names appeared with variations like "SEM INFORMACAO"[6] or "Sem Informação". These values were adjusted to a standardized format, such as "S/I" to maintain uniformity.

Finally, in the load phase, the processed data is inserted into the PostgreSQL database using `psycopg2`[7] version 3, enabling efficient transaction handling. Indexes are implemented to optimize query speed and ensure better performance in subsequent analyses. Data integrity is maintained through the definition of primary and foreign keys, ensuring correct relationships between tables. This ETL workflow guarantees the reliability and quality of stored data, allowing detailed analyses of budget execution and the allocation of public resources through congressional earmarks.

3.3. Data Enrichment with Provenance

Data Provenance refers to the tracking and documentation of the origin, transformation, and manipulation of data throughout its lifecycle. Provenance allows understanding of how, when, and by whom data was generated, extracted, and processed, ensuring that information is traceable, reliable, and transparent. According to [Simmhan et al. 2005], data provenance is the "record of the sources, processes, and transformations that produce or modify the data". The traceability of these data is crucial for auditing, verification, and validation, especially in contexts such as government data, where accuracy and integrity are essential.

Enriching data with provenance is a key practice to ensure traceability, transparency, and reliability of the information used in our system. To ensure that each data entry is clearly documented, specific metadata are added during the extraction and load-

[6]No Information
[7]PostgreSQL database adapter for Python [Psycopg 2025].

ing process. These metadata allow tracking the origin and history of each piece of data, facilitating continuous auditing and validation. Following are the main elements of data en provenance. provenance.

Extraction Date: The date when the data was extracted from the original source. This information helps track the temporal aspects of the data, identifying when the information was collected.
Load Date and Time: The exact date and time when the data was loaded into our database. This ensures that any analysis is based on up-to-date data and that the data-loading process is accurately documented.
Data Source URL: The direct link to the original data source from which the data was extracted. This ensures transparency, allowing any user to access the original source and verify the accuracy and integrity of the data.
Endpoint Name: The name of the endpoint used to access the data from the original source. This information is crucial to understand the method of accessing the data and ensuring the data was retrieved as expected.
Extraction Parameters Notes: A detailed description of the parameters used during the extraction process. This includes filters, limits, and any other criteria that were applied to select the data for extraction. These parameters are vital to ensure the reproducibility of the extraction process and to understand how the data was selected.

With this metadata, we ensure that all data loaded into our database is fully traceable and that its origin is documented. This not only facilitates the validation and auditing of data but also enhances confidence in the analyses conducted, as the sources and criteria used are clearly and accessibly documented. Hence, we provided a complete where-provenance to our platform.

3.4. Data Visualization

For data visualization and analysis, Business Intelligence (BI) tools were employed to create interactive dashboards and analytical reports. Microsoft Power BI was the chosen tool for its significant advantages, notably in development speed and flexibility. Its intuitive interface and robust features enable rapid dashboard creation, accelerating project timelines. Additionally, Power BI's adaptability allows for seamless integration with various data sources, providing the flexibility to meet our project requirements.

4. Results

The results of this work can be separated into two parts: data model and data visualization.

4.1. The Data Model

The data model is the cornerstone of any Power BI project. Describes how data are organized, related, and calculated, directly impacting the accuracy and efficiency of the analyzes. Currently, our model data is composed of 5 entities: "Mandatos", "Emendas", "Documentos", "Dados IBGE Documentos" and "Dados IBGE Emendas". The correct relationships between tables ensure that data is aggregated and filtered accurately. An efficient model reduces data loading and processing time and can also adapt to new analysis needs and data growth.

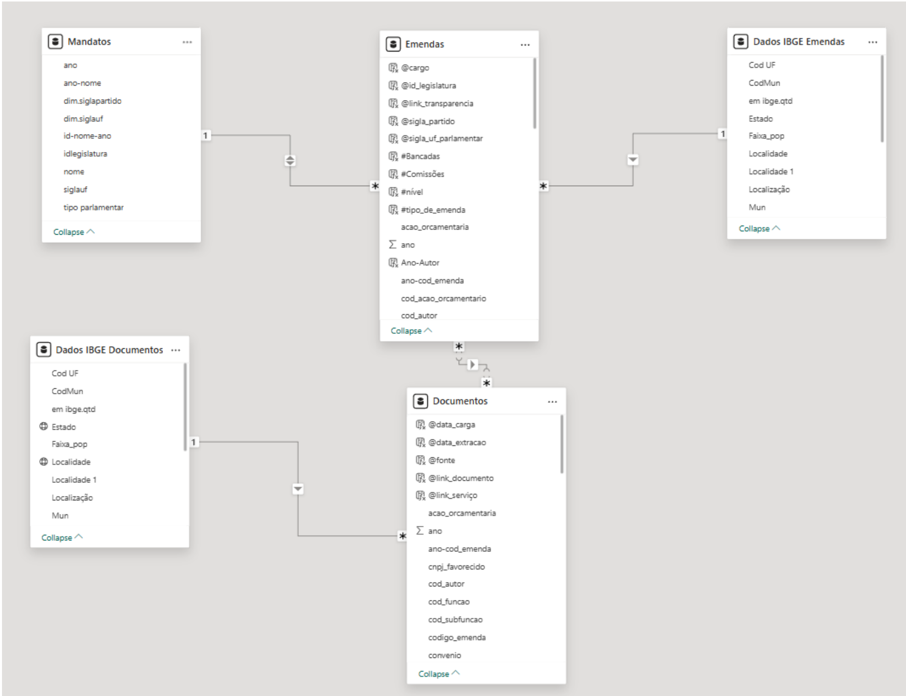


Figure 1. The Power BI Data Model

The core of our data model is the ”Emendas” entity, from which all Earmarks analysis logic stems. The ”Mandatos” entity holds annual mandate information of deputies and senators, including their parties. ”Documentos” contains all documents related to a specific Earmark. ”IBGE” data tables provide the necessary information to group data based on location, as well as population data per location.

This configuration does not mirror the database, but is rather a manipulation of the raw data present in our database, enabling us to answer the desired questions and present them visually on the Power BI platform.

4.2. Data Visualization

In terms of visualization, after data manipulation and the design of the ideal model, our goal was to create visuals that enable detailed viewing of how this distribution of Earmarks occurs, thus facilitating the analysis of these public data.

To facilitate a comprehensive analysis, a dashboard was designed with a wide variety of visuals and indicators, employing a top-down approach. This structure allows users to initially explore high-level Earmarks data and subsequently delve into related Documents.

Initially, we present the spatial distribution of committed Earmark values at the municipal level, as illustrated in Figure 2. The annual distribution of Earmark types was analyzed, showcasing the percentage of committed value for each type over the years, as depicted in Figure 3. The amount of values committed by government function is also presented in Figure 4, highlighting the significant allocation to areas such as health and national defense. Additionally, we included the distribution of Earmarks by Bench, Committees, Parliamentarians, and Party. In Figure 5, is presented a selection of Earmark committed values by "Bancada" and the respective data provenance of "Bancada do Mato Grosso". A comparison of the evolution of the committed and paid values of parliamentary Earmarks over time is presented chart in Figure 6. Finally, we arrive at the document level in Figure 7, where, depending on the filters previously applied in the dashboard, it is possible to view the documents related to the selected set of earmarks.

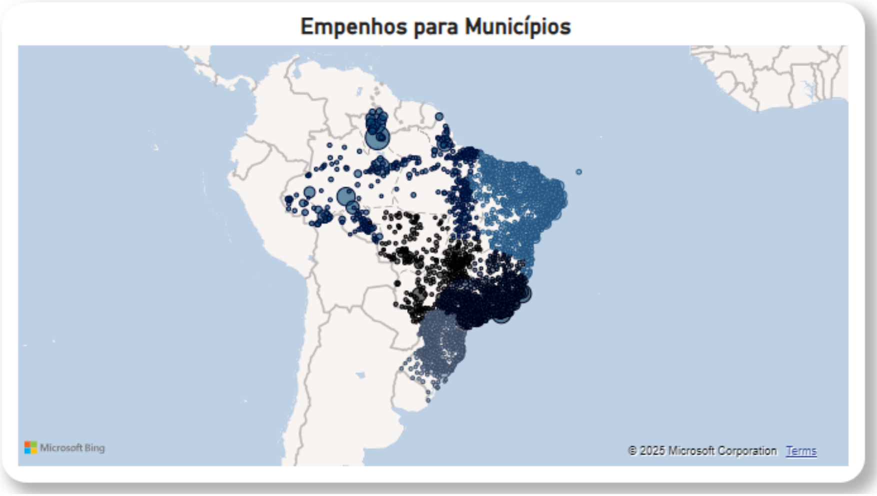


Figure 2. Amount of values committed by MUNICIPIO.

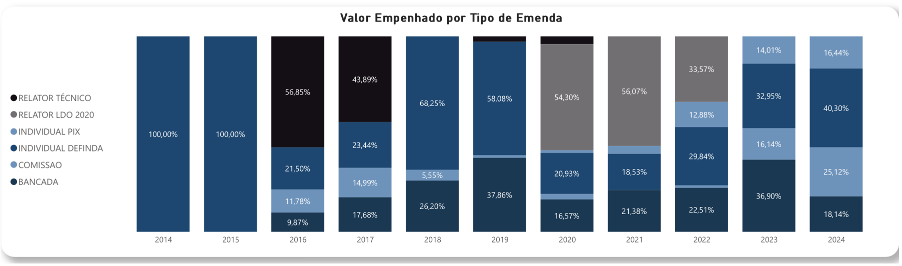


Figure 3. Annual Variation in the Types of EMENDAS PARLAMENTARES.

5. Conclusion and future works

The enrichment of the data with its provenance was essential to ensure traceability and transparency in the process of building and making the data available in our system. The origin history of the data was carefully preserved, which allowed for a clear and reliable view of its trajectory from collection, processing, and availability in the BI. This process ensured that the data could be audited and verified at all stages. Additionally, cross-validation (conducted by experts) between different data sources was carried out to ensure the consistency and integrity of the information. The quality of the data was monitored through periodic feedback from this expert, enabling the detection and correction of inconsistencies. This continuous monitoring helped ensure that the analyses were based on accurate, up-to-date data with full visibility of its origin. From the user perspective (journalists or researchers), this provenance metadata provides a temporal context of the data, ensuring greater reliability of our solution. An important distinction of our dashboard model—unlike the structure adopted by the Frente FIT—is that it includes not only the traditional types of congressional earmarks, but also incorporates other categories defined

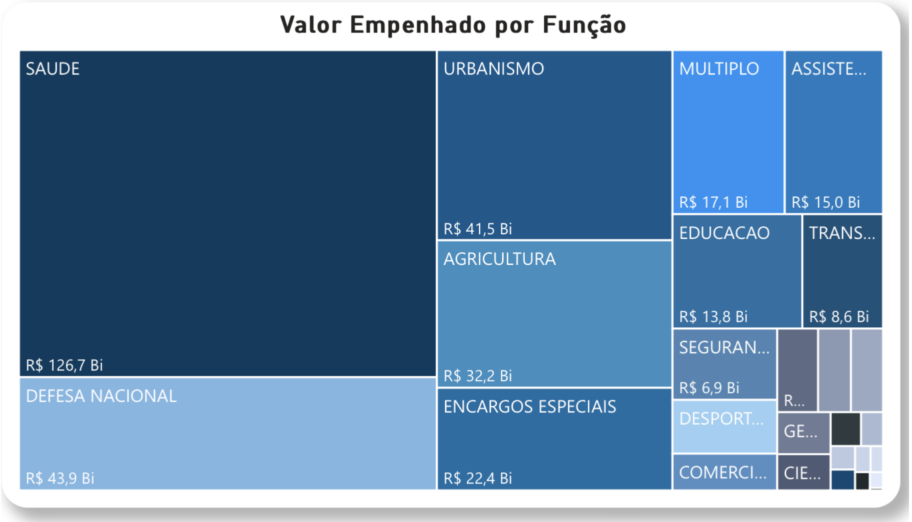


Figure 4. Amount of values committed by government function.

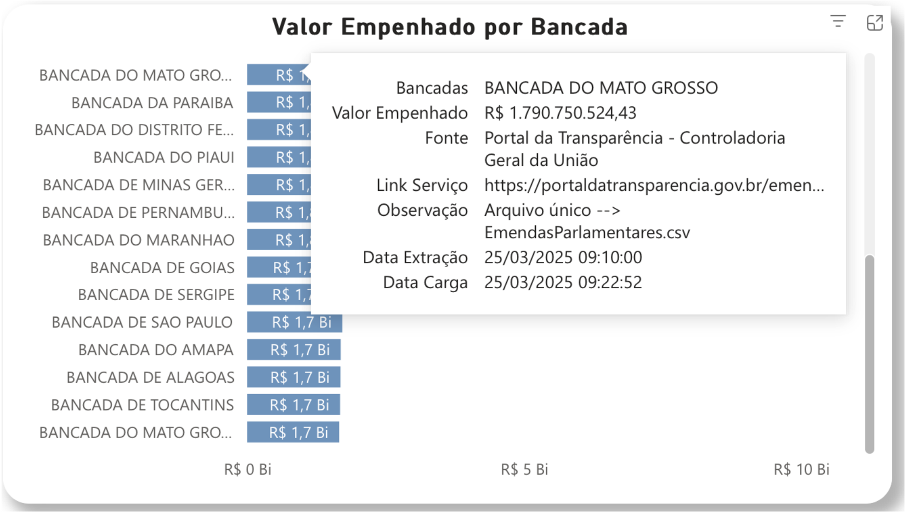


Figure 5. Data provenance for Bancada do Mato Grosso.

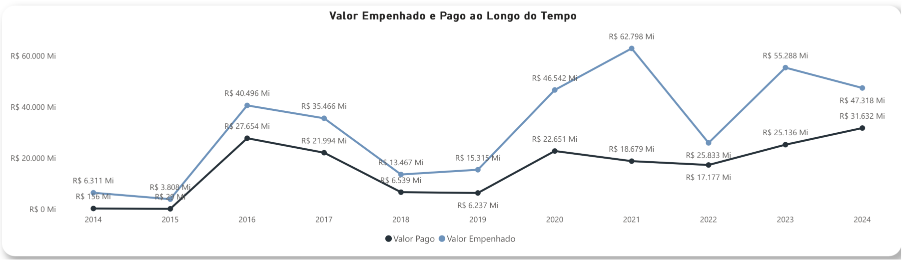
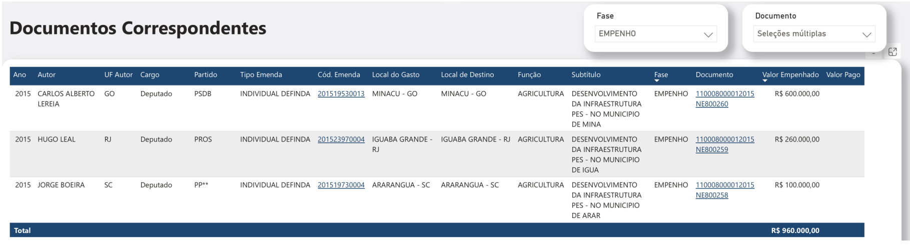


Figure 6. Committed and paid values over the years



Documentos Correspondentes

Fase: EMPENHO

Documento: Seleções múltiplas

Ano	Autor	UF Autor	Cargo	Partido	Tipo Emenda	Cód. Emenda	Local do Gasto	Local de Destino	Função	Subtítulo	Fase	Documento	Valor Empenhado	Valor Pago
2015	CARLOS ALBERTO LEREIA	GO	Deputado	PSDB	INDIVIDUAL DEFINDA	201519530013	MINACU - GO	MINACU - GO	AGRICULTURA	DESENVOLVIMENTO DA INFRAESTRUTURA PES - NO MUNICIPIO DE MINA	EMPENHO	110008000012015 NE800260	R$ 600.000,00	
2015	HUGO LEAL	RJ	Deputado	PROS	INDIVIDUAL DEFINDA	201523970004	IGUABA GRANDE - RJ	IGUABA GRANDE - RJ	AGRICULTURA	DESENVOLVIMENTO DA INFRAESTRUTURA PES - NO MUNICIPIO DE IGUA	EMPENHO	110008000012015 NE800259	R$ 260.000,00	
2015	JORGE BOEIRA	SC	Deputado	PP**	INDIVIDUAL DEFINDA	201519730004	ARARANGUA - SC	ARARANGUA - SC	AGRICULTURA	DESENVOLVIMENTO DA INFRAESTRUTURA PES - NO MUNICIPIO DE ARAR	EMPENHO	110008000012015 NE800258	R$ 100.000,00	
Total													R$ 960.000,00	

Figure 7. Detailed breakdown of documents related to the selected earmarks.

by the legislation in effect at the time of execution. One example is the use of Rapporteur's Amendments (Emendas de Relator) during 2020, 2021, and 2022. Although the term was also used in previous years to describe technical budget adjustments, our dashboard distinguishes between these two usages: "Emendas de Relator LDO 2020" refers to the political amendments introduced under the 2020 Budget Guidelines Law, while "Emendas de Relator Técnico" refers to technical adjustments made prior to Law 13.898 of 2019 [Brasil 2019].

Future Work, continuing the development of the platform, we are working on a version that will enable advanced queries in the context of a semantic web. To achieve this, we plan to create a specific ontology in the domain of congressional earmarks, which will serve as the foundation for the semantic modeling of the data.

After defining the ontology, we will proceed with converting the current relational data model into a graph-based data model. This process will involve transforming the data into RDF triples, which will be stored and managed in a compatible database, such as AllegroGraph [AllegroGraph 2019].

A small experiment has already been conducted as a proof of concept, and its results are presented in Figure 8, which shows a congressional earmark in `RDF/XML` format. This initiative demonstrated the feasibility of the approach and provided insights for adjustments in the data conversion and storage model.

```
<?xml version="1.0" encoding="utf-8"?>
<rdf:RDF
  xmlns:emendas="http://example.org/EmendasOnto#"
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#">
  <emendas:Emenda rdf:about="http://example.org/resource/emenda/202291910008_2022">
    <emendas:codigoEmenda>202291910008</emendas:codigoEmenda>
    <emendas:ano rdf:datatype="http://www.w3.org/2001/XMLSchema#integer">2022</emendas:ano>
    <emendas:tipoEmenda>EMENDA INDIVIDUAL - TRANSFERENCIAS</emendas:tipoEmenda>
    <emendas:temAutor>
      <emendas:Autor rdf:about="http://example.org/resource/autor/9191">
        <emendas:codigoAutor>9191</emendas:codigoAutor>
        <emendas:nomeAutor>ZE NETO</emendas:nomeAutor>
      </emendas:Autor>
    </emendas:temAutor>
  </emendas:Emenda>
```

Figure 8. Piece of RDF/XML file.

Additionally, we recognize that the `SPARQL` query language is not yet widely known, especially among our target audience. To make queries more accessible and intuitive, we plan to develop a language model based on LLM (Large Language Model) that translates natural language queries into SPARQL. This model will be integrated into the platform, allowing users to perform searches more intuitively and efficiently without requiring in-depth technical knowledge of SPARQL syntax.

With these advancements, our aim is to enhance the accessibility and efficiency of the system, making the exploration of congressional earmarks data more dynamic and flexible for different user profiles.

Concerning the combined use of relational databases and KGs, or even OWL-DL ontologies as a need for representing knowledge on the National Budget *Norms and Laws*, we foresee the incorporation of more sophisticated provenance techniques, such as the semi-rings-based provenance annotations.

Semi-ring provenance, [Green et al. 2007], is a mathematically grounded framework that provides a unified approach to modeling data provenance. It leverages algebraic structures known as semi-rings to annotate data and propagate these annotations through query evaluation. Unlike traditional methods, semi-ring provenance is notably expressive, compositional, and extensible, supporting diverse provenance types such as lineage, why-provenance, and how-provenance. It is particularly useful in hybrid data systems that integrate relational and knowledge graph (KG) data, ensuring consistent and interpretable provenance across different data models. A semi-ring, defined as $(K, \oplus, \otimes, 0, 1)$, uses associative operations that adhere to distributive and identity laws. In provenance contexts, database tuples are tagged with semi-ring elements, and query operations manipulate these annotations using the semi-ring's operators. Various semi-rings serve different purposes: the Boolean semi-ring tracks the presence or absence of data for lineage; provenance polynomials represent how-provenance by expressing derivation paths; security labels manage access control; and natural numbers count the number of derivations. For example, a provenance annotation as $(d_1 \otimes d_2) \oplus d_3$, is the same as $(d_1 \otimes d_3) \oplus (d_2 \otimes d_3)$, although they are not the same structurally. The calculi of these algebraic identities can be assisted by an AI agent that learned how to deal with these identities on top of an equational theory of semi-rings. Compared to traditional techniques, semi-ring provenance offers several advantages. It provides a uniform formalism for multiple provenance types, aligns naturally with relational algebra operations due to its compositionality, and offers high expressiveness by accommodating uncertainty, multiplicity, and dependencies. It also allows for efficient computation through optimization and factorization strategies, and its extensibility makes it adaptable to domain-specific annotations like trust levels or probabilistic measures. However, we believe that such a level of sophistication will not be necessary in the context of a tool built on the principles of a regulated and transparent set of data and knowledge bases. On the other side, $\oplus$ terms, i.e., those of the form $D_1 \oplus D_2$ for provenance, may be quite useful, for in many cases the data can be provided from many different sources, in almost equivalent ways, taking associativity and commutativity and distributivity into account. This is a matter of future work.

When it comes to integrating relational data with knowledge graphs, semi-ring provenance plays a key role in unifying their structurally and semantically diverse models. It allows consistent annotation and tracking across both types of data, supports reasoning in knowledge graphs by recording inference steps, handles hybrid queries combining SPARQL and SQL, and enables explanations for query results that involve both structured and semantic content.

References

AllegroGraph (2019). Allegrograph - the enterprise knowledge graph. Accessed: 2019-11-12.

Brasil (1988). *Constituição da República Federativa do Brasil de 1988*. Presidência da República. Accessed: 2025-03-24.

Brasil (2019). Lei nº 13.898. diretrizes para a lei orçamentária de 2020. *Diário Oficial da República Federativa do Brasil*.

Buneman, P., Khanna, S., and Tan, W.-C. (2001). Why and where: A characterization of data provenance. In *International Conference on Database Theory*, pages 316–330. Springer.

Cheney, J., Chiticariu, L., and Tan, W.-C. (2009). Provenance in databases: Why, how, and where. *Foundations and Trends in Databases*, 1(4):379–474.

da União, C.-G. (2025). Portal da transparência. Accessed: 2025-03-29.

das Emendas, C. (2025). Central das emendas.

de Contas da União (TCU), T. (2021). Contas do governo 2021 – emendas de relator geral.

de S. Paulo, O. E. (2025). Emendas de parlamentares vão parar em empresas de suplentes e até em posto de gasolina de deputado. *O Estado de S. Paulo*.

de S.Paulo, F. (2025). Emendas movimentam r$ 150 bi em 5 anos. *Folha de S.Paulo*.

do Brasil, G. F. (2025a). Manual de emendas parlamentares. Accessed: 2025-03-24.

do Brasil, T. N. (2025b). Manual do siafi. Accessed: 2025-03-24.

Douglas, K. and Douglas, S. (2003). *PostgreSQL*. New Riders Publishing, USA.

Federal, S. (2025). Portal siga brasil. Accessed: 2024-10-11.

Federal, S. T. (2023). Decisões do stf reforçam transparência na destinação de emendas parlamentares. *Diário da Justiça*.

Globo, O. (2025). Ongs brasileiras citam desperdício e corrupção ao denunciar falta de transparência nas emendas parlamentares à onu. *O Globo*.

Green, T. J., Karvounarakis, G., and Tannen, V. (2007). Provenance semirings. In *PODS '07: Proceedings of the 26th ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems*, pages 31–40, Beijing, China. ACM.

Kimball, R. and Ross, M. (2002). *The Data Warehouse Toolkit: The Complete Guide to Dimensional Modeling*. Wiley, 2 edition.

Nacional, C. (2024). Ofício ao stf sobre adpf 854, adi 7688, adi 7695 e adi 7697.

Parlamentar, F. (2023). Frente fit. Accessed: 2025-02-02.

Poder360 (2025). Projeto que padroniza emendas é aprovado com brecha na transparência. *Poder360*.

Psycopg (2025). Psycopg3 documentation. Accessed: 2025-02-03.

Simmhan, Y. L., Plale, B., and Gannon, D. (2005). A survey of data provenance techniques.