DCPAPERS

DCMI 2026

Proceedings of the International Conference on Dublin Core and Metadata Applications 2026

ISSN: 1939-1366

OpenDataGOV-AP: A Linked Data Application Profile for parliamentary activities

Tiago Cruz[1], **João Lourenço**[1], **Mariana Curado-Malta**[2], and **João Bispo**[2]

[1] Faculty of Engineering, University of Porto, [2] INESC TEC, Faculty of Engineering, University of Porto

Abstract

This paper presents OpenDataGOV-AP, a Linked Data Application Profile for modelling parliamentary initiatives and related actors in the Portuguese Parliament. Relevant public information exists but is fragmented across heterogeneous XML and PDF sources, limiting its discoverability, interoperability, and reuse by citizens, journalists, and policy-makers. We propose a data-driven profile structured around four modules — Core, Initiatives, Biographical-Profile, and MP-Activities — that reuses established RDF vocabularies and introduces domain-specific terms via the POLIS ontology where gaps exist. Development followed best practices, and the profile is accompanied by a SHACL file for validation. The resulting knowledge graph was validated through SPARQL queries demonstrating cross-party co-authorship analysis and electorate-normalised legislative output. This work provides a semantic foundation for Linked Open Data publishing, semantic search, and improved transparency of parliamentary information.

1. Introduction

Before elections, each political party presents an electoral program that outlines its commitments. Numerous researchers worldwide have examined the fulfilment of these programs and generally agree that most parties do fulfil their promises [1], [2], [3], [4], [5], provided they have the necessary conditions to do so [6]. Despite this academic consensus, the general public often perceives political parties negatively regarding their ability to keep these promises [4]. Serra-Silva et al. [6] argue that many studies compare electoral programs with implemented policies while overlooking legislative initiatives that were not accepted. They suggest that evaluating opposition fulfilment should rely on proposed legislative initiatives rather than just accepted ones, as this approach provides a more comprehensive and accurate estimate of each party's efforts.

The Open Government Partnership (OGP) is an organisation that aims to promote better governance by encouraging transparency, inclusiveness, accountability, and participation[1]. Portugal has been an OGP member since 2017[2], but the Portuguese plan, according to a recent OGP report[3], has problems. The Portuguese government provides numerous websites with open data, including information on legislative initiatives[4], members of parliament[5], public contracts[6] and many other public datasets[7]. However,

Full Paper

Available Aug 01, 2026

Correspondence to
Tiago Cruz
tiagocruz057@gmail.com

1
2
3
4
5
6
7

the data is scarce in many different XML documents, and much information is hidden in PDF documents. The data from the Portuguese Parliament needs to be made accessible to citizens and other stakeholders in a format that breaks the silo of the parliament. A linked dataset already exists [7] however, it is limited to information about members of parliament, leaving broader information needs unaddressed.

Lourenço et al. [8] identify two key societal stakeholder groups. The first comprises "ordinary" citizens who, lacking formal roles in government and the technical skills required, face significant challenges in accessing and interpreting data published on open government platforms. The second group consists of "information brokers"—such as journalists, NGOs, and academic researchers—who possess the expertise and resources needed to effectively navigate and analyse this data.

While information brokers can rely on established strategies to remain informed, ordinary citizens are often left at a disadvantage. Limited time, insufficient technical knowledge, and restricted access to analytical tools hinder their ability to engage critically with available information. This imbalance creates barriers to understanding and increases the risk of misconceptions. Ideally, such a gap should not persist; however, the lack of accessible tools continues to reinforce it, granting a disproportionate advantage to those already equipped with the necessary skills.

As a result, the promise of transparency in open government data remains only partially fulfilled, benefiting a relatively narrow segment of society. This asymmetry becomes particularly problematic in contexts where misinformation and fake news are used to influence democratic processes. Citizens who are unable to independently verify or contextualise information are more vulnerable to manipulation, contributing to a self-reinforcing cycle of distrust and misinformation. Addressing this issue requires the urgent development of inclusive and accessible tools that empower a broader range of users to engage with public data, thereby strengthening informed participation and public trust.

This paper addresses this problem by presenting a linked data application profile, which is a schema to publish linked data in RDF about Portuguese Parliament Initiatives and related actors, relationships between legislative documents, organisations, and policy topics.

This paper is organised as follows. Section 2 provides background on Information Retrieval, semantic search, knowledge graphs, and application profiles, establishing the theoretical context for the work. Section 3 describes the methodology, detailing the data-driven development of the data model across four modules — Core, Initiatives, Biographical-Profile, and MP-Activities — followed by the construction of the constraints matrix and the vocabulary alignment process. Section 4 presents the results, comprising the OpenDataGOV-AP itself and its associated SHACL file, the implementation of the Knowledge Graph through a Python-based parser and RDF converter, and an evaluation through a SPARQL query that demonstrates the profile's applicability and future impact

on the community that will serve. Section 6 concludes the paper with remarks on the contribution and directions for future work.

2. Background

2.1. Information Retrieval

Information Retrieval (IR) is the process by which a user interacts with an IR system by submitting a query. The IR system receives this input query, checks its index of all documents, and returns a ranked list of results, with the most relevant items appearing first [9]. This process can be understood in three parts: document representation, user requirements, and the matching function [9]. The primary goal of IR is to retrieve information that aligns with the user's queries and fulfils their requirements [10].

Current traditional IR systems struggle to comprehend the meaning of user queries, as phrases can have different interpretations based on context [9], [11], [12], [13]. Additionally, traditional IR systems cannot effectively retrieve results when words, despite having different forms, are synonyms [14], [15]. While using a thesaurus could help address the synonym issue[8], this method is costly due to the need for periodic updates with new terms [14], which complicates scalability. By integrating semantic web technologies, we tackle this challenge and bridge the existing semantic gap [10], [13]. Semantic IR focuses on delivering documents based on matching meanings rather than just matching literal words [9], [11], [12], [13], [15], [16].

2.2. Semantic Search

Combining keyword-based interfaces with semantic technologies is considered one of the most challenging aspects of semantic search. In this context, ontologies are increasingly used to develop applications for specific domains. They allow users to identify and extract semantic components from documents, effectively addressing the limitations of keyword-based search methods. Ontologies are a primary tool for knowledge representation since they formalise concepts and relationships within explicit, often hierarchical, conceptual structures [17], [18], [19], [20]. To achieve this, ontologies define a shared vocabulary that specifies concepts, their attributes, and their interrelations [17]. They are grounded in formal logic, particularly description logics, which provide a rigorous semantic basis for representation and automated reasoning [21], [22].

With an ontology, a search engine can deduce concepts from the user's query, which can then be used in techniques like query expansion. Query expansion involves adding terms to a user's query to enhance the number of relevant documents returned or to improve search accuracy. This technique is recognised for its positive impact on IR systems.

2.3. Knowledge Graphs and Application Profiles

To operationalise a semantic layer, the extracted concepts and relations can be represented in a knowledge graph (KG).

[8]

A KG is a graph-structured representation of knowledge in which entities are nodes and semantic relations are edges, typically organised according to a schema so that machines can interpret and reason over the data [23], [24]. The schema is called an Application Profile (AP), which defines how generic RDF vocabularies[9] are used, constrained and combined for a specific domain and set of use cases. This follows the seminal perspective introduced by Heery and Patel [25], where application profiles are used to ``mix and match'' metadata terms while preserving semantic clarity and interoperability [25]. In our case, the AP is designed for parliamentary open data, where users need to discover initiatives, trace actors, and follow relationships between legislative documents, organisations, and policy topics. Rather than creating an entirely new RDF vocabulary, an AP reuses established RDF vocabularies and clarifies how they should be applied consistently and interoperably. This is particularly important in public-sector settings, where data quality and comparability directly affect transparency and accountability [7], [8].

The next section presents the methodology used to develop a Linked Data Application Profile OpenDataGOV, which serves as the operational bridge between source systems and KG construction, enabling consistent ingestion, enrichment, and semantic search over parliamentary data.

3. Methodology

The development of the OpenGOV Linked Data Application Profile (OpenGOV-AP) was informed by previous work [26], [27], [28]. In this context, we followed the deployment guidelines proposed by Tim Berners-Lee [29], aiming to transform the three-star data in the XML files and one-star data in the PDF files into five-star data. Additionally, our design was guided by the FAIR principles[10], which emphasise Findability, Accessibility, Interoperability, and Reusability to promote effective management and long-term usability of the data.

During development, we began by defining the data model. Next, we established the constraints matrix, defining for each term its domain and range, its cardinality and any encoding scheme (value or syntax) for the values, to reduce ambiguity and enhance downstream retrieval. Finally, we proceeded with vocabulary alignment, ultimately selecting the most appropriate RDF vocabulary terms. The classes and terms were aligned with ontology-engineering principles to make the semantics explicit and machine-processable [17], [18], [19].

The following sections present the development process and its milestones, including the techniques and methods used.

9

10

3.1. Data Model

The development process is data-driven, using the existing XML files published by the Portuguese Parliament and the PDF files. The process focused on data profiling, which involved analysing the actual XML structures and the content of the PDF files. The data model was iteratively shaped based on the empirical structure, constraints, and semantics observed in the XML files [7], [8], [29].

3.1.1. *Core*

The first set of entities and their relations, which we call "Core", are[11]:

- *Legislature*: defined as the duration between elections and the dissolution of parliament, typically lasting four years and divided into annual legislative sessions. To accurately represent this structure in the data model, a dedicated Legislative Session Concept is introduced to capture session-specific information. Additionally, a relationship between the Legislative Session and the Legislature is necessary to maintain this hierarchical structure.
- *LegislativeSession*: defined as a yearly period within a legislature during which the Parliament is in session.
- *Person* has three categories: *MemberOfParliament*, *Guest*, and *MemberOfGovernment*. *MemberOfParliament* is distinguished by two types of attributes: general attributes, which remain consistent across different legislatures, and legislature-specific attributes, which may vary (for example, the name of the parliamentary body). To address this, a dedicated entity is introduced, *Mandate*.
- *Mandate*: establishes a relationship between a *Person* and a *Legislature*, serving as a concept for information that is dependent on the *Legislature*. This information includes attributes that change throughout a legislative session, such as the situation (e.g., a substitute becoming permanent), posts (e.g., holding a specific office during a session), and parliamentary group membership (e.g., changes in party affiliation).
- *Situation*: the status of a *MemberOfParliament* within the Parliament, e.g whether a suspension or a resignation, among other possibilities.
- *Post*: a role or office held by a *MemberOfParliament* within the Parliament (e.g., President, Vice-President), of a *MemberOfGovernment* in the Government or of a Guest to the Parliament.
- *Membership*: the affiliation of a *MemberOfParliament* to a *Group* or *OrganisationalBody* over a specific time interval. Therefore, a *Mandate* may include multiple memberships, each associated with one of the previous concepts and a defined temporal scope.
- A political affiliation is represented through two distinct entities:

[11]

- *ParliamentaryGroup*: a group of *MemberOfParliament*s who are elected from the same party list in a given election. It exists only for the duration of a legislature or until it no longer has active members (for example, a *MemberOfParliament* can leave the party and become independent in the middle of a Legislature).
- *Party*: a political party that exists independently of the parliamentary structure. A *ParliamentaryGroup* is affiliated with a political party.
- *OrganisationalBody*: a collegiate body constituted within the scope of the parliament. It can be a *Commission* (that can have sub-commissions), an *ARWorkingGroup*, or an *WorkingGroup*. Instances of this entities correspond to formally established internal structures of the parliamentary institution, tasked with specific legislative, supervisory, or deliberative functions.
- *Government*: a *MemberOfGovernment* belongs to a *Government* of the country.
- *Constituency*: an administrative division of the country used for electoral purposes, representing the territory and population that a *MemberOfParliament* is mandated to represent in parliament.

3.1.2. *Initiatives*

The second set of entities and their relations (which we call "Initiative") are[12]:

- *Initiative*: an initiative that is proposed at the Parliament, it is catalogued within a taxonomy. Initiatives can lead to other initiatives and can grow out of existing ones. An initiative can also result from a *EuropeanInitiative*. An initiative is also related to a *Legislature* and a *LegislativeSession*. An initiative is proposed and authored. An authorship encompasses only the persons who effectively wrote the initiative, while a proposal differs from the authorship in the way that it encompasses the group who jointly proposed it to the Parliament. The authors' names are revealed only when the proponent is a *ParliamentaryGroup* or a *MemberOfParliament*. Several entities can propose an *Initiative*: one or more *ParliamentaryGroup*s or *MemberOfParliament*s, and other entities such as:
- *Commission*: a specific Commission of the Parliment.
- *Government*: the government.
- *CitizenGroup*: Groups of citizens that collectively bring an initiative to parliament. But details about these groups are not included because such inclusion could raise privacy concerns.
- *RegionalLegislativeAssembly*: A local parliament.
- *Petition*: By bringing a specific call for legislative change to the attention of the legislature. A petition is related to the *Legislature* and to the *LegislativeSession*.

[12]

- *ChangeProposal*: An initiative can be changed. The change has authorship (a *MemberOfParliament* or *ParliamentaryGroup*) and is published in the Diário da República (DARPublication)[13].
- A DARPublication is a sub-entity of *Publication*. They are related to the *Legislature* and to the *LegislativeSession*.
- *Document*: an initiative has content and annexes.
- *Event*: an initiative is related to a set of events. An event relates to activities that occur during the discussions of the initiative. An *Event* may :
- have one or more *Annex*,;
- result in one or more *DARPublication*
- have the appeals of *ParliamentaryGroup*s or *MemberOfParliament*s;
- result in an approved *Diploma*
- happen jointly with other *Initiatives*;
- have *Petition*s jointly discussed;
- have one or more associated activities.
- An event also tracks the discussions (*Discussion*) of the work by means of interventions (*Intervention*) that can be made by *MemberOfParliament*s and *MemberOfGoverment*s, and also by *Guest*s.
- an intervention is also published in a *DARPublication*.
- an event tracks the voting of the *MemberOfGoverment* and *ParliamentaryGroup* (*Vote*) - votes in favour, against and abstentions. It also records those absent from the vote.
- When an event is related to a *Commission*, it contains extra information. This is called a *CommissionEvent*. It:
- tracks votes.
- has one or more *Rapporteur*s and a rapporteur is a *MemberOfParliament*.
- may have one or more *Transmission*s
- may produce one or more *Document*s
- may have hearings and auditions of related activity (*RelatedActivity*)
- may have publications and reports in the *DARPublication*.
- is attributed to *Commission*s and may also have sub-commissions.
- is part of a *LegislativeSession* and of a *Legislature*.

[13]

3.1.3. *Biographical-Profile*

The third set of entities and their relations (which we call "Biographical-Profile") are[14] :

- *Award*: The awards a *MemberOfParliament* has received throughout their life, such as medals, badges, and formal official appraisals, for example.
- *PublishedWork*: Works that a *MemberOfParliament* has published throughout their life.
- *Habilitation*: Literary habilitations of a *MemberOfParliament*. Contains information regarding their formal academic degree, along with extra information such as the status of completion (complete or incomplete) and the level of the degree, ranging from 9 (primary education) to 16 (post-graduate).
- *Role*: The Role entity represents roles performed by a *MemberOfParliament* within or outside the Parliament, such as being president of a municipality or simply being a member of parliament in a certain legislature.
- *Title*: The Title entity represents formal titles of the *MemberOfParliament* throughout their lifetime.

3.1.4. *MP-Activities*

The fourth set of entities and their relations (which we call "MP-Activities") are[15] :

- *ParliamentaryActivity*: a formally recorded action undertaken by an *MemberOfParliament* within the Parliament. Each activity is assigned a unique identifier and classified according to a defined activity type, including plenary debates, government questions, motions, votes, and ceremonies, among others. Each record also specifies the corresponding legislature, legislative session, entry date, debate date, and the subject matter of the activity.
- *Requisition*: a parliamentary instrument that allows *MemberOfParliament* to exercise their individual or collective rights regarding the Assembly's activities. There are two types of *Requisition*: Requests and Inquiries. This tool is essential for parliamentary oversight, enabling *MemberOfParliament*s to examine various levels of public administration and hold them accountable.
- *PublicAccount*: pertains to the state's public finances, where the *MemberOfParliament* served as a rapporteur, overseeing the overall management of government funds. A *PublicAccount* is related to a *Legislature*.
- *YouthParliamentaryActivity*: refers to activities, typically meetings, conducted by the *MemberOfParliament* within the context of the Youth Parliament. This meetings are held at *School* in a specific *Constituency* area, and are part of a Legislature.
- *Membership* in *Group*:
- *FriendshipParliamentaryGroup*: a multi-party organisations within the Portuguese Parliament that focus on fostering dialogue and cooperation with the parliaments

14

15

of allied countries. These groups serve as key instruments of parliamentary diplomacy, promoting inter-parliamentary relations through knowledge exchange, information sharing, and mutual consultation. Their broader goal is to coordinate positions in international forums and advance shared political, economic, social, and cultural interests.

- *Delegation*: a formally constituted group of representatives created by an institution to act on its behalf in a specific context. There are two types of delegations: *TemporaryDelegation*: a group formed for a specific purpose, mandate, or event, with a clearly defined duration. It is typically created to address a particular task, such as participating in an international meeting, conducting a visit, or tackling a defined issue. The delegation ceases to exist once its objective is achieved. Its composition, scope, and timeframe are explicitly defined at the time of its creation, and it has no ongoing institutional responsibilities beyond its stated mandate. *PermanentDelegation*: a standing body that is established on a continual basis to represent an institution in ongoing or long-term contexts, such as participation in international organisations or interparliamentary assemblies. Unlike temporary delegations, it has stable membership (although individual members may rotate), a continuous mandate, and institutional continuity. It performs regular duties, maintains relationships over time, and functions within the formal structure rather than being created for a single, isolated purpose.
- *CommissionActivity*: activities done by the *MemberOfParliament*, related to or that involve a *Commission*. There are 4 types of commission activities:
- *Audition*: a formally recorded activity in which a parliamentary committee summons or invites an external entity, such as a public official, nominee, or expert, to testify or provide an account of their actions on a matter under scrutiny. It primarily serves as a tool of oversight and accountability, identifying the external entity involved, the committee responsible, and the subject under examination.
- *Hearing*: a formally recorded activity in which a parliamentary committee or dedicated working group receives civil society organisations, associations, or other stakeholders to gather their positions and input on a given matter. It is consultative in nature, and identifies the external entity involved, the committee or working group responsible, and the topic under discussion.
- *Visit*: The Visit class represents formal visits conducted by a *MemberOfParliament*, encompassing both national and international engagements. It captures the destination of the visit, its start and end dates, the subject matter discussed, the associated commission's name and acronym, and details about the relevant legislature and legislative session.
- *Event*: a formal event the *MemberOfParliament* attends, mostly either ceremonies, debates, and conferences.

3.2. Constraints Matrix and Vocabulary Alignment

A constraints matrix defines each property identified in the data model. Per entity, it presents: a label and description, the RDF vocabulary term to be used with that property, the range, the cardinality, and whether a controlled vocabulary is required. The constraints matrix is available in the following link[16]: https://politrackapp.com/ontologies/OpenDataGOV-AP/[17]. To develop this matrix, an analysis of the RDF vocabularies that exist in the context of the Parliaments was conducted (Table A1 at Appendix A displays the ontologies that model parliamentary activities). The data model consists of numerous concepts that need to be mapped to RDF vocabulary terms. For each concept in the data model and for every property related to that concept, we identified potential candidate terms from RDF vocabularies. [26] call this the vocabulary alignment. All of these decisions were made amid the tension between interoperability and semantic precision, two opposing forces that tend to repel rather than attract.

The POLIS ontology was developed to serve the OpenDataGOV-AP[18]. It comprises four modules. This development was based on the identification of classes and properties that lacked suitable candidate terms.

4. Results, Implementation and Validation

4.1. Results: A OpenDataGOV Linked Data Application Profile

The OpenDataGOV-AP is presented in https://politrackapp.com/ontologies/OpenDataGOV-AP/ and the SHACL file in https://politrackapp.com/ontologies/OpenDataGOV-AP/SHACL.ttl.

The profile is organised around the four modules introduced in the data model — Core, Initiatives, Biographical-Profile, and MP-Activities — and covers several dozen classes and over a hundred properties. For each property, the constraints matrix specifies the RDF vocabulary term, domain, range, cardinality, and any required controlled vocabulary or encoding scheme, yielding a tightly constrained, machine-processable specification. Where existing RDF vocabularies offered suitable terms, these were reused; where gaps were identified, new terms were introduced through the POLIS ontology, which was developed in parallel to serve the OpenDataGOV-AP. The SHACL file operationalises these constraints and can be used directly to validate any RDF dataset that claims conformance with the profile, providing a concrete mechanism for quality assurance in downstream applications.

4.2. Implementation

The construction of the Knowledge Graph (KG) was made possible through the development of a Python-based parser and RDF converter. Initially, XML data were collected from various sources on the Parliament website, including fundamental information,

16

17

18

initiatives, Members of Parliament's activities, and the biographical registry. After retrieving this content, it was parsed and temporarily stored in memory using Python dataclasses[19]. Due to the XML files' inherently non-linked structure, we introduced an intermediate population step to replace static attributes with appropriate class-based representations. This approach enabled us to establish semantic relationships. Whenever possible, we integrated this population process into the parsing phase. Finally, we performed a comprehensive iteration over all instantiated data classes to transform the structured data into RDF triples using the Python library *rdflib*[20]. The resulting KG was then validated against the developed SHACL to ensure compliance with OpenDataGOV-AP.

4.3. Validation

To assess the applicability of the OpenDataGOV-AP schema we have developed, we seek answers to the following questions regarding the knowledge graph (KG):

- What are the most common co-author pairs across different parties?
- How many initiatives have originated from each Member of Parliament (MP) in their respective constituencies, adjusted for electorate size to facilitate comparison of representational output across various regions?

  Due to space constraints, the queries and their results are presented in Appendix B.

5. Discussion on Impact of the work

The results on Table 1 reveal cross-party collaboration patterns that aren't visible from voting records alone. Specifically:

- For journalists:
- Pairs with high co-authorship despite opposing parties signal pragmatic alliances worth investigating — is there a deal, shared constituency interest, or personal relationship driving it?
- Common ideologies surfaced from Wikidata can expose when parties market themselves as distinct but share underlying political DNA — useful for "parties are more alike than they admit" stories.
- Outliers (e.g., far-left and far-right MPs co-authoring) are inherently newsworthy.
- For policy-makers:
- High co-authorship pairs identify natural coalition-building opportunities for getting legislation passed.
- If certain ideological overlaps consistently produce more collaborative output, that's a signal for which parties to negotiate with on specific issue areas.

19

20

- The exclusion of PSD&CDS (the parties at the moment in the Government) means the analysis focuses on opposition dynamics — useful for understanding how minority parties coordinate.

  The results on Table 2 show that the combination of all five metrics in one table is what makes it powerful — each metric alone is interesting, but together we can ask whether underrepresented constituencies are also underserved legislatively. The data connects representational input (voters, seats) with legislative output (initiatives), which opens several angles:

- For journalists:
- Constituencies where MPs submit far fewer initiatives per 100,000 voters than average are underperforming — a straightforward accountability story.
- The "voters per seat" metric exposes malapportionment: some voters effectively have more representation than others, which is a structural fairness story.
- Comparing high-electorate vs. low-electorate constituencies reveals whether smaller districts punch above their weight legislatively.
- For policy makers:
- Persistent low initiative output from certain constituencies may indicate MPs are overloaded, under-resourced, or disengaged — a case for institutional support or electoral reform.
- The voters-per-seat ratio feeds directly into debates about redrawing constituency boundaries or reforming the electoral system.
- Normalising by 100,000 voters makes it defensible as a KPI for representational equity across very differently sized districts.
- For the general public:
- People can look up their own constituency and see how active their MPs are relative to others — a rare, concrete accountability tool.
- The voters-per-seat figure makes abstract concepts like "vote weight" tangible: your vote may literally count less than someone else's depending on where you live.
- It demystifies parliament by translating legislative activity into a familiar unit (per 100,000 people), similar to how health or crime statistics are typically presented.

  When legislative data is open, semantised, and machine-readable, it becomes more than a record — it becomes a tool. Anyone can query it, connect it to other knowledge sources, and tell accurate, evidence-based stories about how power is exercised. That kind of radical transparency is one of democracy's best defences against misinformation and populism.

6. Conclusion and Future Work

This paper addressed the challenge of making Portuguese parliamentary data more accessible and reusable through a Linked Data Application Profile (OpenDataGOV-AP). Starting from heterogeneous XML and PDF sources, we defined a semantic model (a Linked Data Application Profile) that structures parliamentary initiatives, actors, institutional relations, and policy-related metadata within an RDF framework.

The resulting Linked Data Application Profile establishes a foundation for publishing and integrating parliamentary data beyond isolated systems, and for supporting downstream applications such as semantic search. By explicitly modelling entities, properties, and relationships relevant to legislative activity, this work helps reduce technical barriers to accessing public information and strengthen transparency-oriented data infrastructures. modelling As future work, we plan to complete and validate the full profile implementation with real publishing pipelines, enrich the model with additional parliamentary and electoral sources, and evaluate retrieval effectiveness in a semantic search prototype.

Acknowledgments

This work is funded by the Universidade do Porto - Faculdade de Engenharia.

During the preparation of this manuscript, the authors used Grammarly.com to improve the readability, grammar, and overall structure of the text. The authors reviewed, edited, and validated the resulting content and take full responsibility for the final version of the manuscript.

Appendix A. Parliament-related ontologies

Table A1. *Parliament-related ontologies and their URIs*

Name	URI
Map for Scrutiny	http://purl.org/map4scrutiny-core#
LEX ontology	Archived
Popolo	404 error
Ontologia della Camera dei deputati	http://dati.camera.it/ocd/
Parliament	https://ukparliament.github.io/ontologies/
POLIS – Parliamentary Core	http://purl.org/polis/ar/core#
POLIS – Parliamentary Initiatives	http://purl.org/polis/ar/initiatives#
POLIS – Biographical data of the Members of Parliament	http://purl.org/polis/ar/biographical-profile#
POLIS – Activities of the Members of the Parliament	http://purl.org/polis/ar/mp-activities#

The "Archived" LEX ontology refers to https://interoperable-europe.ec.europa.eu/collection/eu-semantic-interoperability-catalogue/solution/cen-metalex/distribution/metalex-owl-ontology – accessed in April 2026. The "404 error" Popolo entry refers to https://www.w3.org/ns/opengov – accessed in April 2026.

Appendix B. SPARQL Queries and Query Results

This Appendix presents the queries to address the questions in Section 4.3.

B.1. Query 1: Cross-Party Co-Authorship and Ideological Common Ground

Query 1 presented in Listing 1 retrieves the most frequent co-author pairs among different parties across Legislatures 13 to 17 and enhances each pair with the ideological common ground shared by their respective parties. The ideological common ground is retrieved from Wikidata. The results are presented in Table 1. For each pair of Members of Parliament (MP1 and MP2), the table shows their respective parties (Party 1 and Party 2), any common ideology according to Wikidata, and the number of co-authored initiatives.

```
PREFIX m4s: <http://purl.org/map4scrutiny-core#>
PREFIX polis-ini: <http://purl.org/polis/ar/initiatives#>
PREFIX polis-core: <http://purl.org/polis/ar/core#>
PREFIX polis-graph: <http://purl.org/polis/ar/graph#>
PREFIX ocd: <http://dati.camera.it/ocd/>
PREFIX foaf: <http://xmlns.com/foaf/0.1/>
PREFIX qudt: <http://qudt.org/schema/qudt/>
PREFIX dbo: <http://dbpedia.org/ontology/>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX wdt: <http://www.wikidata.org/prop/direct/>

SELECT ?legislatureNumber ?mp1ParliamentaryName ?mp2ParliamentaryName ?
party1Acronym ?party2Acronym (COALESCE(GROUP_CONCAT(DISTINCT ?
commonIdeology; separator=", "), "") AS ?commonIdeologies) ?
coauthoredInitiatives
WHERE {
  {
    SELECT ?legislatureNumber ?mp1ParliamentaryName ?
mp2ParliamentaryName ?party1Acronym ?party2Acronym ?party1 ?party2
(COUNT(DISTINCT ?initiative) AS ?coauthoredInitiatives)
    WHERE {
      ?initiative polis-ini:hasAuthor ?mp1, ?mp2 ;
                  polis-core:belongsToLegislature ?legislature .
      FILTER (STR(?mp1) < STR(?mp2))

      ?mp1 a m4s:Parliamentarian ; foaf:name ?mp1Name ;
           ocd:rif_mandatoCamera ?mandate1 .
      ?mp2 a m4s:Parliamentarian ; foaf:name ?mp2Name ;
           ocd:rif_mandatoCamera ?mandate2 .

      ?mandate1 polis-core:belongsToLegislature ?legislature ;
                polis-core:parliamentaryName ?mp1ParliamentaryName ;
                polis-core:hasMembership [ polis-core:isMemberOf ?
party1 ] .
      ?mandate2 polis-core:belongsToLegislature ?legislature ;
                polis-core:parliamentaryName ?mp2ParliamentaryName ;
                polis-core:hasMembership [ polis-core:isMemberOf ?
party2 ] .

      ?legislature dbo:number ?legislatureNumber .
      ?party1 qudt:acronym ?party1Acronym .
      ?party2 qudt:acronym ?party2Acronym .

      FILTER (?party1 != ?party2)
      FILTER (?party1Acronym != "PSD" && ?party2Acronym != "PSD")
    }
    GROUP BY ?legislatureNumber ?mp1ParliamentaryName ?
mp2ParliamentaryName ?party1Acronym ?party2Acronym ?party1 ?party2
    ORDER BY DESC(?coauthoredInitiatives)
    LIMIT 20
  }

  OPTIONAL {
    ?party1 owl:sameAs ?wd1 .
    ?party2 owl:sameAs ?wd2 .

    SERVICE <https://query.wikidata.org/sparql> {
      ?wd1 wdt:P1142 ?ideologyNode .
      ?wd2 wdt:P1142 ?ideologyNode .
      ?ideologyNode rdfs:label ?commonIdeology .
      FILTER(LANG(?commonIdeology) = "pt")
    }
  }
}
GROUP BY ?legislatureNumber ?mp1ParliamentaryName ?mp2ParliamentaryName ?
party1Acronym ?party2Acronym ?party1 ?party2 ?coauthoredInitiatives
ORDER BY DESC(?coauthoredInitiatives)
```

Listing B1. *Cross-Party Co-Authorship and Ideological Common Ground*

Table B1. *Top MP Pairs by Co-authored Legislative Initiatives, with Party Affiliation and Shared Ideological Positions - excluding the parties in the Government*

MP 1	MP 2	Party 1	Party 2	Common Ideologies	Co-authored
Joana Mortágua	Rui Tavares	BE	L	democratic socialism	5
Rui Tavares	Alfredo Maia	L	PCP		5
André Silva	Isabel Pires	PAN	BE		4
Duarte Alves	Mariana Silva	PCP	PEV		4
Eurídice Pereira	Joana Mortágua	PS	BE		4
Eurídice Pereira	Pedro dos Santos Frazão	PS	CH		4
Eurídice Pereira	Rui Rocha	PS	IL	pro-Europeanism	4
Eurídice Pereira	Rui Tavares	PS	L	pro-Europeanism	4
Joana Mortágua	Alfredo Maia	BE	PCP		4
Joana Mortágua	Pedro dos Santos Frazão	BE	CH		4

B.2. Query 2: Electorate-Normalised Legislative Initiative Rate

Query 2 presented in Listing 2 retrieves the number of persons that vote in a Constituency (Electorate), the number of seats that from each constituency across Legislatures 13 to 17. To facilitate comparison of representational output among constituencies, the results are normalized based on electorate size. The findings are presented in Table 2. This table includes details for each constituency, such as the number of voters in the electorate, the number of seats elected from that electorate, the number of voters needed to elect a seat, the total number of initiatives submitted by the constituency, and the number of initiatives per 100,000 voters.

```
PREFIX polis-core: <http://purl.org/polis/ar/core#>
PREFIX polis-graph: <http://purl.org/polis/ar/graph#>
PREFIX polis-ini: <http://purl.org/polis/ar/initiatives#>
PREFIX m4s: <http://purl.org/map4scrutiny-core#>
PREFIX ocd: <http://dati.camera.it/ocd/>
PREFIX pe: <https://ukparliament.github.io/ontologies/election/election-
ontology.ttl#>
PREFIX foaf: <http://xmlns.com/foaf/0.1/>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX wd: <http://www.wikidata.org/entity/>
PREFIX wdt: <http://www.wikidata.org/prop/direct/>

SELECT ?constituencyName
  ?electorate
  ?seats
  (ROUND((?electorate / ?seats) * 10) / 10 AS ?electoratePerSeat)
  (?initiatives AS ?initiativeCount)
  (ROUND((?initiatives / ?electorate) * 100000 * 10) / 10 AS ?
initiativesPer100kElectors)
WHERE {
  ?constituency a pe:ConstituencyArea ;
                foaf:name ?constituencyName ;
                owl:sameAs ?wd .

  OPTIONAL {
    SELECT ?constituency (COUNT(DISTINCT ?initiative) AS ?
initiativeCountRaw)
    WHERE {
      ?initiative a polis-ini:Initiative ;
                  polis-ini:hasAuthor ?mp ;
                  polis-core:belongsToLegislature ?legislature .

      ?mp a m4s:Parliamentarian ;
          ocd:rif_mandatoCamera ?mandate .

      ?mandate polis-core:belongsToLegislature ?legislature ;
               polis-core:inConstituency ?constituency .
    }
    GROUP BY ?constituency
  }

  BIND(COALESCE(?initiativeCountRaw, 0) AS ?initiatives)

  SERVICE <https://query.wikidata.org/sparql> {
        OPTIONAL { ?wd wdt:P1831 ?electorate } .
        OPTIONAL { ?wd wdt:P1410 ?seats } .
  }
}
ORDER BY DESC(?electoratePerSeat)
```

Listing B2. *Electorate-Normalized Legislative Initiative Rate*

Table B2. *Electoral Size, Parliamentary Representation, and Legislative Initiative Rates Across Portuguese Constituencies - in the last two columns, in bold, the highest values.*

Constituency	Electorate	Seats	Electors per Seat	Initiatives	Initiatives per Seat	Initiatives per 100k electors
Açores	228 975	5	45 795.0	618	123.6	269.9
Aveiro	645 747	16	40 359.2	2 151	134.4	333.1
Beja	123 032	3	41 010.7	1 130	**376.7**	**918.5**
Braga	778 359	19	40 966.3	2 412	126.9	309.9
Bragança	141 587	3	47 195.7	355	118.3	250.7
Castelo Branco	170 152	4	42 538.0	614	153.5	360.9
Coimbra	380 064	9	42 229.3	1 188	132.0	312.6
Europa	895 515	2	447 757.5	330	165.0	36.9
Évora	136 725	3	45 575.0	929	309.7	79.5
Faro	376 882	9	41 875.8	1 880	208.9	498.8
Fora da Europa	570 435	2	285 217.5	428	214.0	75.0
Guarda	151 557	3	50 519.0	547	182.3	360.9
Leiria	415 359	10	41 535.9	1 651	165.1	397.5
Lisboa	1 921 189	48	40 024.8	6 581	137.1	342.5
Madeira	257 897	6	42 982.8	753	125.5	292.0
Portalegre	9 6425	2	48 212.5	390	195.0	404.5
Porto	1 595 205	40	39 880.1	4 918	123.0	308.3
Santarém	380 976	9	42 330.7	2 066	229.6	542.3
Setúbal	737 285	18	40 960.3	4 400	244.4	596.8
Viana do Castelo	240 942	6	40 157.0	762	127.0	316.3
Vila Real	219 112	5	43 822.4	706	141.2	322.2
Viseu	348 016	8	43 502.0	788	98.5	226.4

References

[1] A. M. Belchior, C. Moury, L. Hanenberg, and P. Nunes, "O cumprimento dos programas eleitorais em Portugal: Tendências, reformas políticas e o papel dos veto players (1995-2019)," *Análise Social*, 2022, [Online]. Available: https://scielo.pt/scielo.php?script=sci_arttext&pid=S0003-25732022000100160&nrm=iso

[2] J. Artés, "Do spanish politicians keep their promises?," *Party Politics*, vol. 19, 2013.

[3] E. Naurin, "Is a promise a promise? Election pledge fulfilment in comparative perspective using sweden as an example," *West European Politics*, vol. 37, 2014, doi: 10.1080/01402382.2013.863518.

[4] C. Moury, "Minority governments and pledge fulfilment: Evidence from portugal," *Government and Opposition*, 2016, doi: 10.1017/gov.2016.14.

[5] R. Thomson, "Citizens' evaluations of the fulfilment of election pledges: Evidence from ireland," *The Journal of Politics*, vol. 73, 2011, doi: 10.1017/S0022381610000952.

[6] S. Serra-Silva and A. M. Belchior, "Understanding the pledge fulfilment of opposition parties using evidence from portugal," *European Politics and Society*, vol. 21, 2020, doi: 10.1080/23745118.2019.1605961.

[7] L. Inês, B. Ana Alice, and A. Óscar Atanázio, *MapAScrutiny – A Linked Open Data Solution for Politicians Interest Registers*. in Proceedings of the International Conference on Dublin Core and Metadata Applications. Dublin Core Metadata Initiative, 2023. doi: 10.23106/DCMI.953176634.

[8] R. P. Lourenço, "Data disclosure and transparency for accountability: A strategy and case analysis," *Information Polity*, vol. 18, 2013, doi: 10.3233/IP-130310.

[9] R. Chauhan, R. Goudar, R. Sharma, and A. Chauhan, *Domain ontology based semantic search for efficient information retrieval through automatic query expansion*. in 2013 International Conference on Intelligent Systems and Signal Processing (ISSP). 2013. doi: 10.1109/ISSP.2013.6526942.

[10] S. Neji, T. Chenaina, A. M. Shoeb, and L. B. Ayed, *SemApp: A Semantic Approach to Enhance Information Retrieval*. in Computational Science and Its Applications – ICCSA 2021. Springer International Publishing, 2021.

[11] M. Tang, J. Chen, and H. Chen, *SemOIR: An ontology-based semantic information retrieval system*. in 2020 IEEE 20th International Conference on Software Quality, Reliability and Security Companion (QRS-C). 2020. doi: 10.1109/QRS-C51114.2020.00042.

[12] S. Redekar *et al.*, "Web Search Engine Using Ontology Learning," *International Journal of Innovative Research in Computer and Communication Engineering*, vol. 5, 2017.

[13] G. D and A. Rastogi, *Integrating Web 3.0 Technologies for Advanced Information Retrieval Systems in the Educational Sector*. in 2023 International Conference on Communication, Security and Artificial Intelligence (ICCSAI). 2023. doi: 10.1109/ICCSAI59793.2023.10421149.

[14] S. Maya, "GDRMA: Graph Neural Networks for Document Retrievals With Mean Aggregation," *IEEE Access*, vol. 12, 2024, doi: 10.1109/ACCESS.2024.3512597.

[15] S. Jiang, T. F. Hagelien, M. Natvig, and J. Li, *Ontology-based semantic search for open government data*. in 2019 IEEE 13th International Conference on Semantic Computing (ICSC). 2019. doi: 10.1109/ICOSC.2019.8665522.

[16] S. Sharma, A. Kumar, and V. Rana, *Ontology Based Informational Retrieval System on the Semantic Web: Semantic Web Mining*. in 2017 International Conference on Next Generation Computing and Information Systems (ICNGCIS). 2017. doi: 10.1109/ICNGCIS.2017.21.

[17] T. R. Gruber, "A translation approach to portable ontology specifications," *Knowledge Acquisition*, vol. 5, 1993, doi: 10.1006/knac.1993.1008.

[18] R. Studer, V. R. Benjamins, and D. Fensel, "Knowledge engineering: Principles and methods," *Data & Knowledge Engineering*, vol. 25, 1998, doi: 10.1016/S0169-023X(97)00056-6.

[19] K. I. Kotis, G. A. Vouros, and D. Spiliotopoulos, "Ontology engineering methodologies for the evolution of living and reused ontologies: Status, trends, findings and recommendations," *The Knowledge Engineering Review*, vol. 35, 2020, doi: 10.1017/S0269888920000065.

[20] L.-D. Ibáñez, E. Simperl, R. Palma, J. Domingue, and P. Casanovas, "Crossing the chasm between ontology engineering and application development: A survey," *Journal of Web Semantics*, vol. 70, 2021, doi: 10.1016/j.websem.2021.100655.

[21] F. Baader, D. Calvanese, D. L. McGuinness, D. Nardi, and P. F. Patel-Schneider, Eds., *The Description Logic Handbook: Theory, Implementation, and Applications*. Cambridge University Press, 2003.

[22] I. Horrocks, P. F. Patel-Schneider, and F. van Harmelen, "From SHIQ and RDF to OWL: The making of a web ontology language," *Journal of Web Semantics*, vol. 1, 2003, doi: 10.1016/S1570-8268(03)00030-8.

[23] S. T. Sukumar, C.-H. Lung, and M. Zaman, *Knowledge Graph Generation for Unstructured Data Using Data Processing Pipeline*. in 2023 IEEE 47th Annual Computers, Software, and Applications Conference (COMPSAC). 2023. doi: 10.1109/COMPSAC57700.2023.00068.

[24] A. Oarga, M. Hart, A. M. Bran, M. Lederbauer, and P. Schwaller, *Scientific knowledge graph and ontology generation using open large language models*. in AI for Accelerated Materials Design-NeurIPS 2024. 2024. [Online]. Available: https://openreview.net/forum?id=wMMhffCxXZ

[25] R. Heery, M. Patel, and et al., "Application profiles: mixing and matching metadata schemas," *Ariadne*, vol. 25, 2000.

[26] M. C. Malta, "Contributo metodológico para o desenvolvimento de perfis de aplicação no contexto da Web Semântica," 2014.

[27] M. C. Malta and A. A. Baptista, *A method for the development of dublin core application profiles (me4dcap v0.1): a description*. in Proceedings of the International Conference on Dublin Core and Metadata Applications. Dublin Core Metadata Initiative, 2013.

[28] M. Curado Malta, P. Centenera, and E. González-Blanco, "Using Reverse Engineering to Define a Domain Model: The Case of the Development of a Metadata Application Profile for European Poetry," in *Developing Metadata Application Profiles*, M. Curado Malta, A. A. Baptista, and P. Walk, Eds., IGI Global, 2017. doi: 10.4018/978-1-5225-2221-8.ch007.

[29] T. Berners-Lee, "Linked data - design issues." [Online]. Available: https://www.w3.org/DesignIssues/LinkedData.html

[30] A. M. Fedotov *et al.*, "Using the thesaurus to develop it inquiry systems," *Journal of Theoretical and Applied Information Technology*, vol. 86, 2016, [Online]. Available: https://www.jatit.org/volumes/Vol86No1/6Vol86No1.pdf