

Towards a Semantic Dataspace for Interoperable and Transparent Local Council Decision Data

Leila Feddoul¹, Stefanie Kerkhof², Antje Dietrich³, Hannes Köninger³ and Marianne Mauch¹

¹Heinz Nixdorf Chair for Distributed Information Systems, Friedrich Schiller University Jena, Jena, Germany

²Agency for Home Affairs, Flemish Government, Belgium

³University of Applied Sciences Kehl, School of Public Administration, Kehl, Germany

Abstract

Local decisions form the backbone of governmental action at the municipal level. To strengthen their impact and interoperability, we aim to implement key components of a dataspace for local decision data. In this environment, decisions from different levels of government can interoperate and be linked to thematic datasets. This integration pursues three main objectives: improving the quality of decision data, enabling more efficient and proactive public service delivery, and enhancing data transparency. This paper presents insights from the DECIDE (Data Driven Exploration in Contextual Information on DEcisions) project, which aims to establish a semantic dataspace for local decision data from two countries (Germany and Belgium). The dataspace provides a policy-aware infrastructure for secure and sovereign data exchange, while knowledge graphs ensure semantic harmonization and machine-interpretable data structures. Furthermore, we introduce practical use cases and describe the overall system architecture, including the semantic mapping of heterogeneous data sources to a unified ontology and the AI-based enrichment of the resulting knowledge graph.

Keywords

Dataspace, Knowledge Graph, Information Extraction, E-government, Public Administration

1. Introduction

Local council decisions contain a wealth of relevant information, but are often stored in documents and archives that are difficult to access and have so far only been machine-readable to a limited extent. Against this backdrop, the European research project DECIDE (Data Driven Exploration in Contextual Information on DEcisions)¹ aims to establish an interoperable dataspace for local decisions, where it acts as the governance/interoperability layer for sharing data across organizations, and the developed Knowledge Graph (KG) is seen as a semantic layer for describing, linking, and querying that data [1]. As an initiative within the context of the European Data Space for Smart Communities (DS4SSCC.EU)², DECIDE makes a concrete contribution to the digital transformation of European municipalities by making municipal decisions machine-readable, interoperable, and transparent. Since political decisions are often interdependent and form a complex network of effects, this approach creates the basis for greater transparency and informed decision-making. At the same time, it taps into automation potential and enables cross-sector digital services that support citizens, businesses, and administrations in navigating the legal and political framework. By the end of the project, DECIDE will have developed a system that prepares municipal decisions in a machine-readable format and makes them available in a trusted dataspace. Such a dataspace represents a digital ecosystem in which organizations can securely exchange, link, and reuse data without losing control over their own datasets. The project is coordinated by the Flemish Agency for Home Affairs³. Ghent, Freiburg, and Bamberg serve as pilot cities. The accompanying scientific research is being carried out by the Friedrich Schiller University

Posters, Demos, Blue Sky, and Tutorials at SEMANTiCS 2026, Sep 2026, Ghent, Belgium

✉ leila.feddoul@uni-jena.de (L. Feddoul); stefanie.kerkhof@vlaanderen.be (S. Kerkhof); dietrich@hs-kehl.de (A. Dietrich); koeninger@hs-kehl.de (H. Köninger); marianne.mauch@uni-jena.de (M. Mauch)

✉ 0000-0001-8896-8208 (L. Feddoul); 0000-0003-1478-1867 (M. Mauch)



© 2026 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

¹<https://www.vlaanderen.be/lokaal-bestuur/digitale-transformatie/slimme-lokale-databronnen/project-decide>

²<https://www.ds4sscc.eu/>

³<https://www.vlaanderen.be/agentschap-binnenlands-bestuur>

Jena⁴ and the University of Applied Sciences Kehl⁵. The consortium is complemented by the Urban Software Institute⁶. The main contributions of this paper are as follows:

- Mapping local decision data from multiple heterogeneous formats and languages (Dutch and German) to a unified data model.
- Automatically enriching the resulting KG using Artificial Intelligence (AI) techniques, thereby enabling practical analysis tailored to specific use cases (Section 4).
- Implementing a semantic dataspace infrastructure that provides standardized discovery, controlled access, and interoperable data sharing through a triple store and authorization mechanisms.

2. Related work

While existing initiatives tackle important aspects of interoperability and public-sector data integration, none fully integrate dataspace principles, KG-based semantic harmonization, and AI-driven enrichment within a unified architecture. LBLD (Local Decisions as Linked Open Data in Flanders) [2] is a Flemish initiative that publishes local government decisions as Linked Open Data, making municipal decision-making more transparent and interoperable using KGs. Open Standards for Linked Organisations (OSLO) [3] is a Flemish government initiative that develops semantic standards and shared data models to enable interoperable and consistent exchange of public-sector data across Flemish administrations. At the European level, the ISA² programme [4] (Interoperability solutions for public administrations, businesses and citizens) delivers a portfolio of semantic standards and reusable digital building blocks (e.g., core vocabularies) for public administrations across Europe. However, these initiatives focus primarily on structured publication rather than dynamic AI enrichment or policy-aware data sharing within a dataspace. OParl (Open Parliament Interface) is a German open standard that defines a RESTful API for publishing structured, machine-readable parliamentary and municipal decision data. Unlike previous projects, it uses JSON instead of semantic formats, providing only syntactic interoperability. DATEN:RAUM:FREIBURG [5] is a smart city initiative that develops a distributed architecture integrating urban data sources (not just municipal decisions) and applications through a central data hub. Thus, it focuses on integration rather than ontology-driven semantic harmonization. [6] proposes a standardized reference architecture for urban data spaces and provides architectural guidance for organizing the technical components and services required for secure data sharing between municipalities. However, it remains at the conceptual architecture level, and neither develops a semantic representation model for municipal decision data nor implements a semantic dataspace. In [7], an empirical/conceptual study is conducted to identify success factors for urban data spaces based on literature and stakeholder perspectives. The research gap lies in the lack of an integrated approach combining federated dataspaces, KGs, and AI-based enrichment to transform heterogeneous municipal decision data into semantically interoperable, policy-aware assets. By going beyond conceptual reference architectures and empirical studies, DECIDE addresses this gap by developing and implementing such an integrated solution.

3. System Overview

Figure 1 depicts the overall architecture of the proposed system. We consider three data sources originating from different data providers and represented in heterogeneous formats. The first data source is provided by the Flemish Government and contains local decisions of the city of Ghent in the RDFa format following the OSLO Standard. The second data source originates from the city council of Freiburg and provides local decision data in the OParl API format⁷ (JSON). The third data source is provided by the city council of Bamberg and delivers decisions as PDF/HTML files. Given the

⁴<https://www.opendva.de>

⁵<https://www.hs-kehl.de/>

⁶<https://www.ui.city/en/>

⁷<https://oparl.org/>

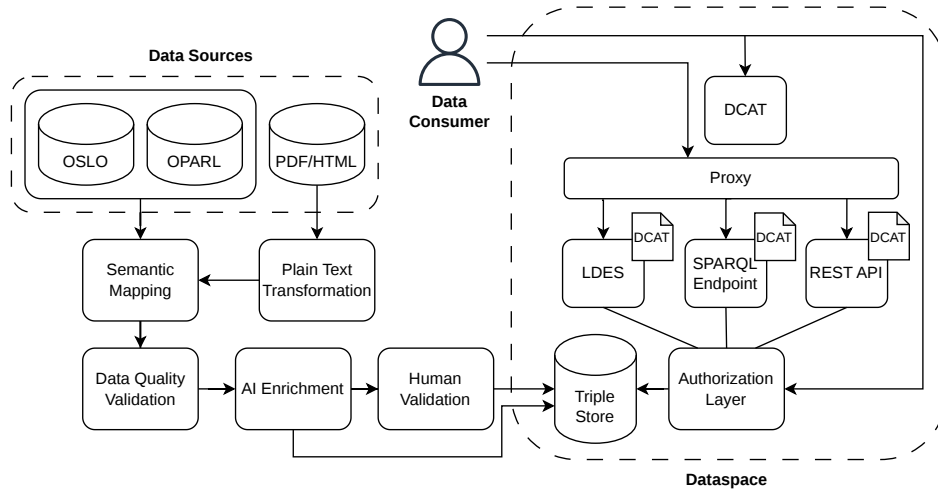


Figure 1: System architecture overview

heterogeneity of the considered data sources and their underlying semantic models, a prerequisite for integration was the definition of a shared ontology tailored to the target domain. We selected and reused the existing ontologies ELI (the “European Legislation Identifier” ontology)⁸ and ELI-DL (the “European Legislation Identifier” ontology for draft legislation)⁹ as the common reference models. These ontologies provide the necessary concepts and relationships to harmonize the different representations of local decisions across providers. It subsequently served as the semantic foundation for mapping, enrichment, and validation within the dataspace. The first step is to map the data in different formats (OParL API JSON, OSLO, PDF/HTML) to the reference ontologies (*semantic mapping*). Semi-structured and unstructured data are first preprocessed into an intermediate format that facilitates semantic mapping (*plain text transformation*). Given the limited structural information available in these sources, we rely heavily on AI-based enrichment techniques to extract, disambiguate, and contextualize relevant entities before integrating the instances into the KG. Next, a *data quality validation* is performed using a SHACL-based¹⁰ component that verifies the completeness and model compliance of the generated ELI and ELI-DL data. The generated linked data is subsequently enriched through AI-based pipelines (*AI Enrichment*) covering two main scenarios: (1) Entities are automatically extracted from unstructured sources such as HTML pages and PDF documents using transformer-based models (xlm-roberta-base, RoBERTa, and legal-bert-base-uncased), and mapped to the shared ontologies (entity linking) using AI agents and a Model Context Protocol (MCP) Server that provides tools for querying SPARQL endpoints, searching locations, performing web searches, and utilizing a vector-based knowledge base. (2) The existing KG is further enriched with complementary information to support specific use cases, such as linking local decisions to higher-level policy goals (e.g., the Sustainable Development Goals) in order to enable policy impact analysis and reporting (Subsection 4.1). The code for the entity recognition and linking services is available on GitHub [8, 9, 10]. Afterward, enrichments are either stored directly in the triple store or first presented to a human expert, who can approve, reject, or modify the AI-generated annotations (*human validation*). The latter are also semantically described using the Web Annotation Vocabulary¹¹. The resulting KG is exposed via SPARQL, REST interfaces, and as a Linked Data Event Stream (LDES)¹², enabling incremental access and transparent version management. To access the data, a data consumer first consults the well-known location of the DCAT (Data Catalog Vocabulary)

⁸<https://op.europa.eu/web/eu-vocabularies/dataset/-/resource?uri=http://publications.europa.eu/resource/dataset/eli>

⁹<https://op.europa.eu/web/eu-vocabularies/dataset/-/resource?uri=http://publications.europa.eu/resource/dataset/eli-dl>

¹⁰<https://www.w3.org/TR/shacl/>

¹¹<https://www.w3.org/TR/annotation-vocab/>

¹²<https://interoperable-europe.ec.europa.eu/collection/semic-support-centre/linked-data-event-streams-ldes>

catalog¹³ to discover the available dataset distributions and their corresponding access endpoints. The DCAT description provides information about dataset content, access endpoints, licensing conditions, and provenance. Each pilot city's data sources are described using a DCAT-AP v3-compliant¹⁴ catalog. Based on this metadata, the data consumer can determine whether authentication is required and, if necessary, complete the authentication process (*authorization layer*). Access to endpoints is mediated by a proxy layer that acts as an intermediary between the consumer and the underlying services, enforcing security policies and routing requests accordingly. Once authenticated, the consumer can interact with the relevant endpoints via the proxy using the specified protocol to retrieve or query the datasets of interest. Beyond dataset discovery, the architecture also supports governance-related metadata by publishing Open Digital Rights Language (ODRL)¹⁵ policies. While DCAT describes datasets, data services, and their access endpoints, ODRL expresses machine-readable access rights and usage policies. The architecture also supports participant trust through the Universal Trust Data Registry (VC), which extends the DCAT discovery layer by providing W3C Decentralized Identifiers (DIDs)¹⁶ and Verifiable Credentials (VCs)¹⁷, enabling authentication and trust verification of data space participants.

4. Use Cases

4.1. Policy Impact Report

In the following, we present practical use cases that illustrate the benefits of semantically describing local council decisions and making them accessible through a dataspace.

In this use case, the goal is to showcase the impact of local council decisions on policy goals with a *Policy impact report*. To achieve that, each decision needs to be linked to one or more relevant policy goals (e.g., the Sustainable Development Goals (SDGs)¹⁸ or locally created code lists). For example, the following decision *The city will expand social housing units by 15% by 2030* can be linked to SDG 11 (Sustainable Cities and Communities) in a positive sense because it contributes to the achievement of this goal, and *The new zoning plan permits industrial expansion into wetland areas* can be linked to SDG 15 (Life on Land) but in a negative sense because it is expected to have an adverse effect on this goal. This is achieved in the AI enrichment step (Section 3) in two phases: (1) Cold start: this is the initial phase where no annotated data exists. A Large Language Model (LLM) (Mistral, OpenAI, Anthropic, or Ollama using Langchain) performs classification solely based on the codelist's concept definitions, providing immediate responses while human validators gradually create the annotated dataset required for the second phase. (2) Supervised classifier: When sufficient validated annotations are available, the system trains a local supervised classifier (distilbert-base-uncased) on the new ground truth and can replace the LLM, lowering API costs. Next, a new annotation instance (oa:Annotation, based on the Web Annotation Vocabulary) is stored in the KG, and a link between the actual decision and the policy goal is created. Finally, each (decision, policy label) pair is sent to the LLM with the task of inferring the policy impact direction (positive, negative, or uncertain) and confidence level (low, medium, or high). The impact is also stored in the KG. The code for mapping decisions to policy goals is publicly available on GitHub [10].

4.2. Mapping Local Decisions on Restricted Mobility Zones to Geo-locations

This use case focuses on making local government decisions visible and accessible through geo-portals. By linking decisions to specific geographic locations, citizens, businesses, and city officials can better understand the impact of policies. The approach involves enriching decision data with location metadata,

¹³<https://www.w3.org/TR/vocab-dcat-3/>

¹⁴<https://semiceu.github.io/DCAT-AP/releases/3.0.0/>

¹⁵<https://www.w3.org/TR/odrl-model/>

¹⁶<https://www.w3.org/TR/did-1.1/>

¹⁷<https://www.w3.org/TR/vc-data-model/>

¹⁸<https://sdgs.un.org/goals>

integrating it into city mapping systems, and ensuring compatibility with Geographical Information Systems (GIS). Practical applications include visualizing restricted and car-free mobility zones on interactive city maps. This improves public awareness, simplifies compliance, and supports data-driven urban planning, for example, by creating more space for bicycles and pedestrians. To achieve that, we practically map local decisions about restricted mobility zones to geo-coordinates and integrate them into city portals or as a standalone viewer for citizens. This is done in the AI enrichment step (Section 3) by first detecting elements of locations in the decision text (e.g., streets, cities, or postcodes) using transformer-based models (xlm-roberta-base, RoBBERT, and legal-bert-base-uncased), then combining the elements into candidate locations and searching for their geolocation in the map (e.g., using Nominatim¹⁹ for geocoding), and incorporating the newly derived information as semantically structured entities and relations within the KG. The code for the geo-location extraction service is publicly available on GitHub [8]. Preliminary evaluation results for the extraction of locations and other entity types are available on Hugging Face²⁰.

4.3. Support Access for Subsidies for Private Owners

This use case facilitates access to subsidies for property owners undertaking climate-friendly renovations. This is achieved by developing a Retrieval Augmented Generation (RAG)-based search system that enables search and question answering over existing decisions on subsidies. This system takes the user's question as input and performs a semantic search to retrieve the top-k most relevant decisions. These decisions are then provided, together with the user's query, to an LLM (Mistral, OpenAI, or Ollama), which generates the final response in the form of a list of relevant decisions. The code for the question answering service is publicly available on GitHub [11].

5. Conclusion and future work

In this paper, we present an approach for constructing a unified KG of local council decisions from heterogeneous, bilingual data sources. We address heterogeneous structured and unstructured data by mapping them to a domain-relevant ontology. The resulting KG is enriched using a range of AI-based techniques, including information extraction and entity linking, which enhances the completeness and analytical value of the graph, enabling its use in practical applications. The KG is embedded within a dataspace architecture that provides a secure data exchange, thereby combining semantic interoperability with governance and access control mechanisms. The system is still under active development. Future work includes finalizing the dataspace setup, extending and evaluating AI-based enrichment and validation, and conducting user evaluations.

Acknowledgments

The research project DECIDE is funded by DS4SSCC within the DS4SSCC-DEP program, which is funded by the European Union Digital Europe Work Programme 2021-2022 under Grant Agreement No. 101123342.

Declaration on Generative AI

The author(s) have not employed any Generative AI tools.

¹⁹<https://nominatim.org/>

²⁰<https://huggingface.co/svercoutere/longformer-classifier-refinement-abb#evaluation>

References

- [1] J. Theissen-Lipp, M. Kocher, C. Lange, S. Decker, A. Paulus, A. Pomp, E. Curry, Semantics in dataspace: Origin and future directions, in: Companion Proceedings of the ACM Web Conference 2023, WWW '23 Companion, Association for Computing Machinery, New York, NY, USA, 2023, p. 1504–1507. URL: <https://doi.org/10.1145/3543873.3587689>. doi:10.1145/3543873.3587689.
- [2] R. Buyle, P. Colpaert, M. Van Compernelle, P. Mechant, V. Volders, R. Verborgh, E. Mannens, Local Council Decisions as Linked Data: a Proof of Concept., in: ISWC (Posters & Demos), 2016.
- [3] R. Buyle, L. De Vocht, M. Van Compernelle, D. De Paepe, R. Verborgh, Z. Vanlischout, B. De Vidts, P. Mechant, E. Mannens, OSLO: open standards for linked organizations, in: Proceedings of the International Conference on Electronic Governance and Open Society: Challenges in Eurasia, EGOSE '16, Association for Computing Machinery, New York, NY, USA, 2016, p. 126–134. URL: <https://doi.org/10.1145/3014087.3014096>. doi:10.1145/3014087.3014096.
- [4] V. Kalogirou, Y. Charalabidis, The European union landscape on interoperability standardisation: status of European and national interoperability frameworks, in: Enterprise Interoperability VIII: Smart Services and Business Impact of Enterprise Interoperability, Springer, 2019, pp. 359–368.
- [5] D. K. Tscheulin, P. M. Mähner, A. Gelantia, Smart City Initiatives and the Common Welfare in Modern Societies, in: Lebenslagen und Gemeinwohl, 2024, pp. 331–340.
- [6] S. Cuno, L. Bruns, N. Tcholtchev, P. Lämmel, I. Schieferdecker, Data governance and sovereignty in urban data spaces based on standardized ict reference architectures, Data 4 (2019). URL: <https://www.mdpi.com/2306-5729/4/1/16>. doi:10.3390/data4010016.
- [7] J. Haberkern, T. Schäffer, Towards data ecosystems in smart cities: Success factors of urban data spaces., in: DATA, 2023, pp. 548–558.
- [8] Pedro Antonio Jeuris and others, Geocoding and NER Service, <https://github.com/semantic-ai/decide-geocoding-service>, 2026. Last commit: 1e7b52bce62c1d77bc70d1e1af07bff2bbdb19e5, accessed 2026-05-22.
- [9] Stefaan Vercoutere and others, Entity Linking Backend MCP Server, <https://github.com/semantic-ai/entity-linking-backend>, 2026. Last commit: 733b9679f4436b5893c25866e3c4a12150125de7, accessed 2026-05-22.
- [10] Ward De Muer and others, codelist-labeling-service, <https://github.com/semantic-ai/codelist-labeling-service>, 2026. Last commit: 5009c3f48cc5e6c78bdca421ee87a40e348f67d5, accessed 2026-05-22.
- [11] Pedro Antonio Jeuris and others, UC2 Subsidies RAG System, <https://github.com/semantic-ai/decide-question-answering>, 2026. Last commit: 27c62e430e44071ed08fb0707b4f61669fac6448, accessed 2026-05-22.