Empowering e-services through the Semantic Web

Raffaella Maria Aracri[1,†], Dario Frisardi[1,†], Roberta Radini[1,*,†] and Valerio Santarelli[2,†]

[1] Italian National Institute of Statistics (ISTAT), Italy
[2] OBDA Systems s.r.l., Italy

Abstract
This article illustrates how to enhance data interoperability among Public Administrations (PPAA) by leveraging the publication of e-services based on Semantic Web (SW) technologies such as ontologies, controlled vocabularies, and data schemas, which, through standard languages like OWL [6], RDF[12], and SPARQL [5], ensure harmonization, integrability, and unique semantics for representing administrative data. Furthermore, the potential benefits of implementing semantic e-services through Ontology-based Data Management (OBDM), a data governance methodology that enables data services through ontologies, decoupling their implementation from the physical data sources of PPAA, are discussed. Additionally, the advantages for PPAA of data exchange through semantic e-services utilizing concepts published in the Schema platform, developed by the National Data Catalog project funded through PNRR funds, will be highlighted in the presented use case.

Keywords
PPAA, PNRR, Semantic Web, e-service, ontology, controlled vocabulary, data schema, OBDM

1. Introduction

As part of the investment initiatives delineated within the National Recovery and Resilience Plan (PNRR), which is part of the Next Generation EU (NGEU) program, a measure concerning the digital transition of Public Administrations (PPAA) in data management and interoperability has been designated referred to as the National Digital Data Platform (PDND). This platform acts as the tool to centralize the authentication and authorization methods for data exchange among parties. Therefore, PDND manages the authorization phase of accessing the Application Program Interfaces (APIs), while PPAA set up their automatic connectors to make data accessible and interoperable, promoting their sharing among administrations, as well as between citizens and enterprises. Particularly, this approach avoids citizens from having to provide the same information multiple times to various administrations.

Additionally, investments aimed at enhancing data interoperability within the PDND also encompass the development of Schema, the National Catalog of data for semantic interoperability, which provides semantic clarity to administrative data through an extensive network of Ontologies, Controlled Vocabularies, and data schemas [7]. These are facilitated by standard languages such as OWL, RDF, and SPARQL, ensuring harmonization, integrability, and a unified semantics to represent administrative data information.

The adoption of these standards ensures the accessibility, reusability, and inferential capacities across data originating from various sources, processes, and domains. To fully realize the benefits of these technologies, data ideally should be accessible via standard data access protocols, such as SPARQL [5], the W3C's reference language for querying RDF datasets and OWL ontologies. Among the solutions to ensure adherence to these protocols is the Ontology-based Data Management (OBDM) paradigm [8], which advocates for a virtual approach to data governance, and consequently, data access, through ontologies.

The rest of this paper is structured as follows. In Section 2, we will introduce Schema, the National Data Catalog for semantic interoperability, and explain how the semantic assets published therein can assist PPAA in the implementation of semantic e-services according to data schemas, leading to full data interoperability. Section 3 will illustrate the potential benefits of implementing e-services through OBDM, primarily by decoupling the service layer from the data layer. In Section 4, we will provide an example, through a use case, of the benefits guaranteed to individuals by data interoperability.

Ital-IA 2024: 4th National Conference on Artificial Intelligence, organized by CINI, May 29-30, 2024, Naples, Italy
*Corresponding author.
†These authors contributed equally.
aracri@istat.it (R. M. Aracri); dario.frisardi@istat.it (D. Frisardi); radini@istat.it (R. Radini); santarelli@obdasystems.com (V. Santarelli)
0009-0005-4399-7589 (D. Frisardi)

2. Schema and the semantic e-services

Within the framework of semantic interoperability enhancement, as delineated and overseen by the National Data Catalog (NDC) project [7], the development of the portal, Schema, aims to make the semantic resources of PPAA available. Schema catalogs semantic assets, such as ontologies, controlled vocabularies, and data schemas, making them searchable, reusable, and thus fostering interoperability. The semantic structures of Schema enable the definition of a unified semantics that harmonizes data representation and facilitates information exchange for Italian PPAA. For further information about Schema and the available semantic assets, additional online resources can be found in Appendix A.

By leveraging ontology semantics, coherence and consistency in data are ensured, significantly enhancing interoperability. Within ontologies, there is uniformity in the description and referencing of entities present, supported by cardinality constraints that govern the relationships among them.

Data schemas become an essential tool for PPAA to expose their data and facilitate communication through shared channels, exploiting the fundamental principle of a unified semantics. The strength of data schemas lies in their ability to expose the data structure and enforce type and format constraints, ensuring not only syntactic but also semantic interoperability, targeting entities and concepts defined within ontologies and controlled vocabularies [10].

In the context of e-services, the accurate definition of data schemas plays a fundamental role in ensuring data coherence and complete interoperability. Establishing a shared ontological semantics, therefore, becomes crucial to guarantee a uniform interpretation of data by all involved parties. Through a proper implementation of data schemas, e-services become essential tools for promoting the effective utilization of semantic technologies within the realm of public administration.

The data schema format for the e-service must adhere to the specifications of a YAML file [2] (if version 3.0 of OpenAPI is utilized, the YAML file should be named with an extension oas3.yaml). Within the e-service data schema, the main components of the service in question must be defined, taking into account the semantic references of ontologies and controlled vocabularies. An example of how the data schema should be structured to define the concept of "Person" is provided below, showcasing some of the key commands:

```
Person:
 type: object
 description: https://w3id.org/
     italia/onto/CPV/Person
 x-jsonld-context:
  ...
  tax_code: https://w3id.org/italia/
     onto/CPV/taxCode
  date_of_birth: https://w3id.org/
     italia/onto/CPV/dateOfBirth
  family_name: https://w3id.org/
     italia/onto/CPV/familyName
  ...
 properties:
  tax_code:
   $ref: "#/components/schemas/
      TaxCode"
  date_of_birth:
   format: date
   type: string
   pattern:
      [0-9]{4}-[0-1][0-9]-[0-3][0-9]
  family_name:
   type: string
...
TaxCode:
 type: string
 description: https://w3id.org/
     italia/onto/CPV/taxCode
...
```

where the `type` is reported, as well as the URI pointing to the semantic resource is included in `description`. All the `properties`, whether required or necessary for the service, are inserted after specifying the corresponding semantic reference. Thus, it can be observed how data schemas incorporate format and type constraints, as well as references to other components (e.g., `tax_code`), which will then reference a semantic resource defining its concept. Currently, JSON-LD standard [13] does not allow to define in a unique way the context for primitive values, i.e. `string`, and in such cases it is necessary to adopt an *ad-hoc* shared strategy. More detailed data schemas are available in Appendix A.

3. Implementation of e-services with OBDM

The Ontology-based Data Management (OBDM) [4, 8, 9] is a paradigm introduced and promoted by the Department of Computer, Control, and Management Engineering "Antonio Ruberti" (DIAG) at

Sapienza University of Rome and by OBDA Systems[1]. Its aim is the integration and governance of data stored in an organization's information system through an ontology. The purpose of this approach is to create a single conceptual access point to the organization's information assets, enabling the conceptual realization of all data governance services within a complex system.

Indeed, OBDM can be viewed as a form of virtual data integration. However, it is based on the notion of replacing the global schema, which represents the unified view of the domain, with a conceptual and formal representation formulated through an ontology expressed in a logical language. This choice ensures that the integrated view offered by an OBDM system is not limited to a structure accommodating data from sources but constitutes a semantically rich description of the relevant concepts within the domain of interest and the relationships between them. Similarly to how it occurs in data integration systems, conceptual relationships, or mappings, are utilized to establish semantic correspondences between the global schema and the data in the sources.

When seeking information, a query is expressed on the ontology (rather than on the information system's databases), and the correspondences established between the data and the ontology's concepts enable the ontological reasoning engine [11, 3] to derive the response. This relieves the user from the necessity of understanding the technical aspects of data storage and the specifics of where and how data is storage. Similarly, when carrying out a data governance task (such as quality assessment, re-engineering, data cleaning, etc.), direct access to the informational sources is bypassed, and the appropriate functions are performed through the domain ontology.

The latter aspect not only formally describes the enterprise's information model but also serves as a means to embrace a declarative approach to data governance. Through the explicit delineation of the domain representation, knowledge reusability is achieved, a feat not achieved when the global schema merely provides a unified description of the underlying sources.

OBDM systems generally have a common structure divided into three layers or tiers: ontology, mappings, and data sources. The distinction between ontology and data sources reflects the separation between the conceptual or semantic level, which is presented to users, and the logical and physical level of the information system, stored in the sources, with mappings serving the role of reconciliation between these levels. In this scenario, the ontology and corresponding mappings to the sources provide not only a tool for data access but also a common basis for documenting an enterprise's information assets. This approach brings significant benefits for governance and management of the information system.

E-services represent an opportunity to leverage the wealth of shared ontological models and controlled vocabularies in Schema, not only as tools for conceptual sharing, but also as a means of semantically accessing data according to standardized and shared models. The value added by implementing e-services through OBDM techniques is two-fold.

Firstly, it resides in the capability to decouple the implementation of e-services from the physical sources of repositories that host the data of PPAA. In this scenario, the realization of e-services could potentially be accomplished solely through the articulation of requirements or queries on the ontological models published on Schema by the data-owning PPAA. This approach delegates the task of mediating with the data structures of individual PPAA to the mappings, while the ontological reasoning engine is entrusted with leveraging ontologies and mappings to translate the ontological requirement into queries on the physical data.

Secondly, OBDM provide the possibility to access data using SPARQL [5], the W3C standard language for querying ontologies and RDF datasets, without the implementation of data transformation and migration processes from their sources, typically consisting relational DBMS, to triple stores for formatted RDF data. Given the complexity and the volume of such data commonly managed by the information systems of PPAA, such processes naturally require significant efforts, including infrastructural ones. OBDM, on the other hand, offers a solution distinct from this scenario, favoring a virtual approach to data access, where queries expressed in SPARQL on the ontology are transformed into SQL queries on the physical sources at query time.

The decision to employ OBDM presents clear advantages both in terms of e-service implementation, utilizing standard Semantic Web languages for querying ontologies, i.e. SPARQL, and in terms of their maintenance and evolution.

The decoupling from the physical layer would allow isolating the e-service layer from the usual dynamics of reorganization, restructuring, distribution, or replication typically encountered by databases in information systems, particularly in the case of large organizations. In light of these developments,

[1] www.obdasystems.com

adjustments to ensure the consistency of data sharing services would be confined to the mapping layer, thus modifying the assertions that express the correspondences between elements of the data layer and those of the semantic layer.

4. A use case of semantic interoperability

Based on what has been introduced in the previous sections, we aim to introduce an illustrative use case of data interoperability among PPAA and a valuable aid to citizens by implementing the principle of once only [1]. It can be hypothesized that the Ministry of Culture, MiC, intends to offer a promotional service for the enjoyment of cultural assets (such as certain museums) by high-level education students from the local area, i.e., their municipality of residence.

This use case involves an exchange of information between e-services exposed by the relevant PPAA. For its modeling, as introduced in Section 2, a portion of the Schema semantic network is utilized, particularly the core Location (CLV), People (CPV), and Organization (COV) ontologies, along with domain-specific ontologies, such as the Italian Learning Ontology (Learning), Resident Population Ontology (RPO), and the Cultural Heritage Ontology (CulturalHeritage). All the reported ontologies are available in Appendix A, while in Figure 1 there is a portion of the semantic network.

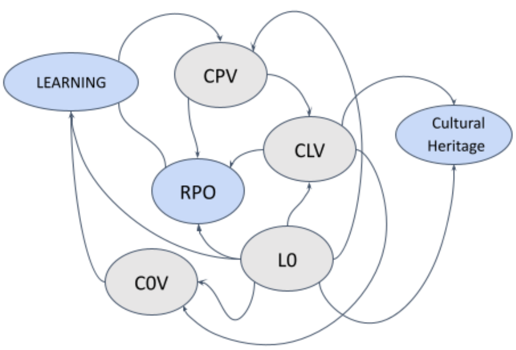

Figure 1: Extract of Schema's semantic network.

To achieve this, the Ministry of Culture, MiC, can provide an e-service, *S12*, to a cultural institution (e.g., a museum), which, given the unique identifier of a person (tax code) and the municipality where the cultural asset subject to promotion is located, retrieves:

i. an e-service, *S1*, exposed by the National Register of the Resident Population (ANPR) system, which returns a confirmation if the person's current municipality of residence matches that of the cultural asset;
ii. an e-service, *S2*, provided by the Ministry for Universities and Research (MUR), confirming that the person is a student enrolled in a High-Level Education course.

The e-service *S12*, after retrieving the validations returned by the e-services *S1* and *S2*, returns to the requesting institution the authorization for the promotion for the individual.

An example of the YAML code configuration for the e-service *S1* is provided below, with the key commands:

```
components:
 schemas:
  RegisteredResidentPerson:
   x-jsonld-context:
    RPO: https://w3id.org/italia/
        onto/RPO/
    tax_code: https://w3id.org/
        italia/onto/CPV/taxCode
    currently_registered_residence:
      "@id": "RPO:
          currentlyHasRegistered
          ResidenceIn"
      "@context":
       "@base": "https://w3id.org/
           italia/controlled-
           vocabulary/territorial-
           classifications/cities"
   type: object
   description: https://w3id.org/
      italia/onto/RPO/
      RegisteredResidentPerson
   required:
   - tax_code
   properties:
    taxCode:
      $ref: "#/components/schemas/
          TaxCode"
    currently_registered_residence:
      type: string
      enum: [...]
      example: '058103-(1871-01-15)'
  TaxCode:
    type: string
    description: https://w3id.org/
        italia/onto/CPV/taxCode
```

The addressing to the tax code concept is ensured by the URI identifying the concept (https://w3id.org/italia/onto/CPV/taxCode), similarly to the use of the controlled vocabulary of cities (https://w3id.org/italia/controlled-vocabulary/territorial-classifications/cities) for current registered residence. Therefore, unique semantics emerge as a fundamental tool that, in the reported use case, enables interoperability among the 3 e-services.

In Figure 2 there is an illustration depicting the main components, and their connections through e-services, in the use case.

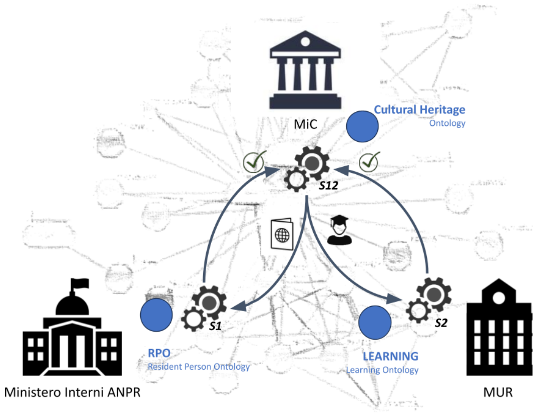

Figure 2: Use case representation.

5. Conclusions

As previously outlined, the use of e-services represents a pivotal element in automating data exchange among PPAA. Their implementation can play a crucial role as the driving force for achieving data interoperability, facilitated by a common semantics represented by the data schemas provided by the Schema platform.

Managing data access through the OBDM technique is an area worth investing in the short term, as it allows for the decoupling of the physical data layer from the semantic one and enables the application of these innovative techniques even on existing data structures without the need for re-conversion. Furthermore, based on a logical language, these techniques also provide Artificial Intelligence services that infer knowledge from data and facilitate quality management.

The proposed solutions move in this direction, highlighting that much work still needs to be done. Only the commitment of PPAA to implement semantic interconnection of concepts, and consequently data, can fully achieve the objective of "digitalization, innovation and security in the Public Administration." This could be the prompting for the digitalization of the entire Italian country.

References

[1] AgID. *Piano Triennale per l'informatica nella Pubblica Amministrazione 2024 - 2026*. Agenzia per l'Italia Digitale, Rome, Italy, 2023. Available online at: https://docs.italia.it/italia/piano-triennale-ict/pianotriennale-ict-doc/it/2024-2026/index.html.

[2] Oren Ben-Kiki, Clark Evans, and Ingy döt Net. YAML ain't markup language version 1.2. 2009. Available on: http://yaml.org/spec/1.2/spec.html.

[3] Diego Calvanese, Giuseppe De Giacomo, Domenico Lembo, Maurizio Lenzerini, Antonella Poggi, Mariano Rodriguez-Muro, Riccardo Rosati, Marco Ruzzi, and Domenico Fabio Savo. The MASTRO system for ontology-based data access. *Semantic Web*, 2(1):43–53, 2011.

[4] Diego Calvanese, Giuseppe De Giacomo, Domenico Lembo, Maurizio Lenzerini, and Riccardo Rosati. Tractable reasoning and efficient query answering in description logics: The DL-Lite family. *Journal of Automated reasoning*, 39:385–429, 2007.

[5] Steve Harris and Andy Seaborne. SPARQL 1.1 Query Language, 2013. Available online at: https://www.w3.org/TR/sparql11-query/.

[6] Pascal Hitzler, Markus Krötzsch, Bijan Parsia, Peter F. Patel-Schneider, Sebastian Rudolph, et al. OWL 2 web Ontology Language Primer. *W3C recommendation*, 27(1):123, 2009. Available online at: https://www.w3.org/TR/owl2-primer/.

[7] Istat. Trasformazione digitale della pubblica amministrazione. Metodi per l'interoperabilità per lo sviluppo di *e-service*. In *Letture statistiche – Metodi*. Istituto nazionale di statistica, 2024. Available online at: https://www.istat.it/it/archivio/293230.

[8] Maurizio Lenzerini. Ontology-based data management. In *Proceedings of the 20th ACM international conference on Information and knowledge management*, pages 5–6, 2011.

[9] Maurizio Lenzerini. Managing data through the lens of an ontology. *AI Magazine*, 39(2):65–74, 2018.

[10] Barry Nouwt. Tight integration of web APIs

with Semantic Web. In *SEMANTiCS (Workshops)*, 2017.
[11] Antonella Poggi, Domenico Lembo, Diego Calvanese, Giuseppe De Giacomo, Maurizio Lenzerini, and Riccardo Rosati. Linking data to ontologies. In *Journal on data semantics X*, pages 133–173. Springer, 2008.
[12] Guus Schreiber, Yves Raimond, Frank Manola, Eric Miller, and Brian McBride. RDF 1.1 Primer. *World-Wide Web Consortium*, 2014. Available online at: https://www.w3.org/TR/rdf11-primer/.
[13] Manu Sporny, Dave Longley, Gregg Kellogg, Markus Lanthaler, and Niklas Lindström. JSON-LD 1.1. *W3C Recommendation, Jul*, 2020. Available on: https://www.w3.org/TR/json-ld11/.

A. Online Resources

The National Data Catalog (NDC) platform:

- Schema

The sources for data schemas in Schema:

- INAIL
- INPS

The mentioned ontologies in Section 4:

- CLV
- COV
- CPV
- CulturalHeritage
- L0
- Learning
- RPO