

Using the EU Big Data Test Infrastructure to Publish MITOS Public Service Descriptions as Linked Open Data

Efthimios Tambouris^{1,*}, Dimitris Zeginis¹, Georgios Matziaras¹, Nikolaos Stefanidis¹, Rafail Promikyridis¹, Konstantinos Tarabanis¹, Konstantinos Doutsos Oikonomou², Iraklis Varlamis^{3,4} and Maria Claudia Bodino⁵

¹University of Macedonia, 156 Egnatia Street, 54636, Thessaloniki, Greece

²Hellenic Open University, Greece

³Harokopio University of Athens, 70 Eleftheriou Venizelou Street, 17676, Athens, Greece

⁴GRNET S.A. – National Infrastructures for Research and Technology, 7 Kifisias Street, 11523, Athens, Greece

⁵European Commission, Brussels, Belgium

Abstract

The provision of public services is one of the main goals of public authorities worldwide. In this frame the availability of structured public service descriptions openly is very important since it can improve transparency and trust. The CPSV-AP specification has been proposed as a standard model to publish public service descriptions as Linked Open Data (LOD) and thus enhance their interoperability and integration with other data sources. The aim of this project is publish public service descriptions of the National Registry of Administrative Procedures (MITOS) in Greece as LOD based on CPSV-AP by leveraging the EU Big Data Test Infrastructure (BDTI). The methodology involves understanding the context, modelling public service descriptions, generating LOD, and publishing the data. The project architecture utilizes Apache Airflow for orchestration and Open Link Virtuoso for data storage and retrieval. Through various usage scenarios, the paper highlights the potential benefits of publishing public service descriptions as LOD.

Keywords

BDTI, MITOS, LOD, CPSV-AP

1. Introduction

The provision of public services is one of the main goals of public authorities worldwide. An important part of public service provision is the publishing of structured public service

Proceedings EGOV-CeDEM-ePart conference, September 1-5, 2024, Ghent University and KU Leuven, Ghent/Leuven, Belgium

*Corresponding author.

✉ tambouris@uom.edu.gr (E. Tambouris); zeginis@uom.edu.gr (D. Zeginis); ics20009@uom.edu.gr (G. Matziaras); ics23003@uom.edu.gr (N. Stefanidis); r.promikyridis@uom.edu.gr (R. Promikyridis); kat@uom.edu.gr (K. Tarabanis); std147764@ac.eap.gr (K. D. Oikonomou); varlamis@admin.grnet.gr (I. Varlamis); Mariaclaudia.BODINO@ec.europa.eu (M. C. Bodino)

ORCID 0000-0001-8036-9788 (E. Tambouris); 0000-0002-5426-4031 (D. Zeginis); 0009-0003-7920-9033 (G. Matziaras); 0009-0002-6519-2659 (N. Stefanidis); 0000-0002-5552-2629 (R. Promikyridis); 0000-0002-4663-2113 (K. Tarabanis); 0000-0002-0876-8167 (I. Varlamis); 0009-0001-2407-3760 (M. C. Bodino)



© 2024 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

descriptions through Public Service Description Catalogues (PSDCs) that can empower citizens, contribute to open government, and promote trust between public administrations and citizens.

Public service descriptions are usually based on underlying models [1] that enhance semantic interoperability, transparency, accessibility, and efficiency in public service delivery. A public service model developed by the EU is the Core Public Service Vocabulary Application Profile (CPSV-AP) [2] which aims to be a de facto standard model for publishing public service descriptions as Linked Open Data and thus enhances their interoperability and integration with other data sources (e.g., DBpedia).

In Greece, the government has launched MITOS¹ as the official National Registry of Administrative Procedures, which currently contains descriptions of almost 4000 public services. The data model underlying MITOS is compatible with CPSV-AP. MITOS provides access to its data through an API however its data have not yet been published as Linked Open Data.

The aim of the MITOS_LOD project presented in this paper is to publish MITOS public service descriptions as Linked Open Data. Towards this direction we: i) apply an existing framework for publishing PSDCs as LOD based on CPSV-AP and ii) leverage the EU Big Data Test Infrastructure (BDTI) to pilot the developed solution.

The remainder of this paper is structured as follows. Section 2 provides background information on MITOS, LOD, and BDTI. Section 3 presents the project approach. Section 4 presents the project architecture, section 5 provides a discussion and section 6 the main conclusions along with directions for future research.

2. Background Information

This section presents relevant background information on i) Public Service Description Catalogues (PSDCs) including MITOS, i.e. the official Greek PSDC, ii) Linked Open Data (LOD) and CPSV-AP that provide principles and vocabularies for publishing and connecting structured public service data and iii) the EU Big Data Test Infrastructure (BDTI) which provides cloud-based services for big data (including LOD) cases.

2.1. Public Service Description Catalogues and MITOS

Public Service Description Catalogues (PSDCs) are comprehensive repositories of standardized descriptions of public services provided by governments or public agencies. These catalogues contain detailed information about the services offered, including their scope, objectives, procedures, eligibility criteria, and any associated costs or fees. The purpose of PSDCs is to enhance transparency, accessibility, and efficiency in public service delivery by providing consistent information about available services. PSDCs enable users to easily find and understand the services they need and facilitating interactions with government agencies.

MITOS is the Greek official National Registry of Administrative Procedures [3]. It allows the modeling of existing administrative procedures of public services, as well as documenting the required supporting documents in collaboration with the competent units of the entities. It incorporates a mechanism for continuous updating the procedures to support the legislative

¹<https://mitos.gov.gr/>

goal of continuous compliance and exclusive information on changes after any simplification. The data model underlying MITOS is compatible with CPSV-AP, and most of the main entities have URIs, however, the data are not yet published according to the LOD principles. MITOS also offers a publicly available API to provide data in JSON format.

2.2. Linked Open Data and CPSV-AP

Linked Open Data (LOD)[4] is an approach to publishing and connecting structured data on the web, making it interlinked and easily accessible. It follows principles such as using URIs to identify entities, employing RDF (Resource Description Framework) for representing data, and utilizing HTTP for accessing data and metadata. LOD aims to enable data interoperability, integration, and reuse across various domains and applications [5].

CPSV-AP (Core Public Service Vocabulary Application Profile) is a specification developed by the European Commission for describing public services in a standardized manner as LOD. LOD and CPSV-AP offer a powerful framework for representing, sharing, and integrating public service-related data in a linked and interoperable manner. By leveraging LOD principles and adhering to the CPSV-AP specification, organizations can enhance the discoverability, accessibility, and usability of public service information [6].

In this frame, the literature has proposed a relevant framework [7] for publishing public service descriptions as LOD using CPSV-AP. The framework was piloted based on data from the Region of Epirus in Greece. A similar approach applied at the European level [8] transformed information from European government portals into LOD. They mapped data from various portals using a single model but faced challenges with manual effort and limitations of poorly formatted data. Their experience highlights the importance of designing scalable solutions for data transformation that prioritize interoperability across systems.

2.3. The EU Big Data Test Infrastructure (BDTI)

The Big Data Test Infrastructure (BDTI)² is a ready-to-use and free analytics cloud platform which aims to help European public administrations to experiment with data and derive insights to support them in their decision-making. To achieve this goal it provides open-source data science tools, completely deployed in the cloud, to support public administrations in their data journey, from data collection, to data visualization through data analysis and data orchestration.

More concretely, public administrations can easily apply for the initiative and benefit of a free of charge cloud-based analytics infrastructure to experiment and to prototype solutions before deploying them in the production environment on their own premises. The test environment provided by the BDTI consists of several integrated open-source solutions, and the required cloud infrastructure that includes virtual machines, analytics clusters, storage and networking facilities. Among other BDTI provides services related to LOD management (namely Virtuoso) and automation (namely Airflow). Many pilots have successfully helped national, regional and local administrations to optimize their processes [9], to design better services and increase transparency, improving the availability, quality and usability of public sector information.

²<https://big-data-test-infrastructure.ec.europa.eu/>

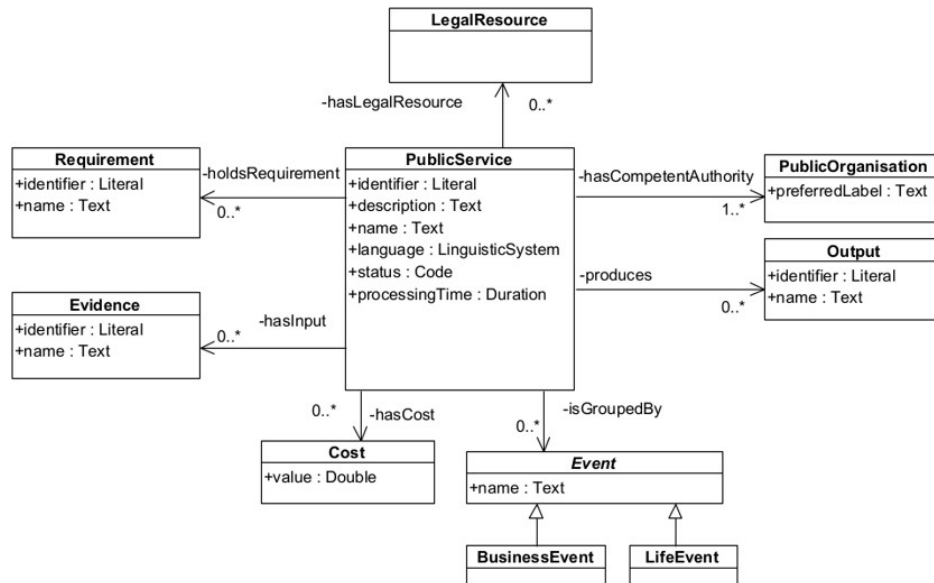


Figure 1: Subset of CPSV-AP AP that is used for pilot purposes

3. The MITOS_LOD project Approach

The aim of the MITOS_LOD project is to publish MITOS public service descriptions as LOD using CPSV-AP. The process followed is adopted by [7] and comprises the following steps.

Step 1: Understanding the context, preparation, and Public Service (PS) selection. This step involves grasping the context including geographical scope and current PS models. This information has been presented in section 2. In addition, in the project there was no PS selection; instead it was decided that all PS in MITOS would be transformed to LOD.

Step 2: Reuse or definition of a URI design policy. In this regard, the EU advocates adopting the following URI structure: `http://domain/type/concept/reference`. In the frame of this project we used the domain "mitos.gov.gr:8890" while an example full URI is `https://mitos.gov.gr:8890/PublicServices/id/evidence/evidence14`

Step 3: PS descriptions modelling. In this step the Public Service descriptions of MITOS are mapped to CPSV-AP concepts. Figure 1 below shows the subset of CPSV-AP that is used for pilot purposes.

Step 4: Linked data generation (and interlinking). This step involves converting CPSV-AP compliant Public Service descriptions into linked data. The linked data generation is performed by collecting MITOS data through an API and transforming them to RDF through a mapping implemented in Python. More details are provided in section 4. An excerpt of the result is depicted in Figure 2.

Step 5: Linked data validation. During this stage, the generated linked data undergo validation to identify any syntax errors and ensure alignment with the CPSV-AP schema.

Step 6: Linked data publication (and interlinking). The linked dataset, such as the RDF/Turtle file, is then uploaded to an RDF store (Virtuoso), making it accessible as an RDF graph via a

```

<https://mitos.gov.gr:8890/id/ps/963466> a cpsv:PublicService ;
dct:identifier "5fd82410-1a65-4a08-a35f-6ec32601e944" ;
dct:title "Άδεια εργασίας προσωπικού ασφαλείας Ιδιωτικών Επιχειρήσεων Παροχής Υπηρεσιών Ασφαλείας (Ι.Ε.Π.Υ.Α.)" ;
skos:altLabel "Άδεια εργασίας προσωπικού για ιδιωτικές υπηρεσίες ασφάλειας" ;
dct:description "Η διαδικασία αφορά στην έκδοση άδειας εργασίας σε προσωπικό ασφαλείας Ιδιωτικών Επιχειρήσεων
Παροχής Υπηρεσιών Ασφαλείας (Ι.Ε.Π.Υ.Α.)." ;
cv:hasCost <https://mitos.gov.gr:8890/PublicServices/id/cost/cost963466> ;
cv:holdsRequirement <https://mitos.gov.gr:8890/PublicServices/id/requirement/requirement963466> ;
cv:hasInput <https://mitos.gov.gr:8890/PublicServices/id/evidence/evidence963466> ;
cv:hasLegalResource <https://mitos.gov.gr:8890/PublicServices/id/rule/rule963466> ;
cv:isGroupedBy<https://mitos.gov.gr:8890/PublicServices/id/event/event8> ;
cv:isGroupedBy<https://mitos.gov.gr:8890/PublicServices/id/event/event9> ;
cpsv:produces <https://mitos.gov.gr:8890/PublicServices/id/praxis/praxis963466>.

```

Figure 2: Excerpt of the produced RDF result (in Greek)

SPARQL query	Description
<pre> prefix cv: <http://data.europa.eu/m8g/> select ?name, (count(*) as ?total) where { ?x cv:isGroupedBy ?event. ?event cv:name ?name } GROUP BY ?name ORDER BY DESC (?total) </pre>	Get the number of Public Services per life event (e.g., child birth).
<pre> prefix cv: <http://data.europa.eu/m8g/> select ?name, (count(*) as ?total) where { ?x cv:hasInput ?input. ?input cv:name ?name } GROUP BY ?name ORDER BY DESC (?total) </pre>	Get the number of public services per evidence that is required as input.

Figure 3: Exploration SPARQL queries

SPARQL endpoint. The Virtuoso RDF store is provided through BDTI.

Step 7: Linked data exploitation. The published linked dataset serves multiple purposes. It aids in acquiring valuable insights into essential public service characteristics or aggregated statistics related to public services. Some example exploration SPARQL queries are presented in Figure 3.

4. Architecture

The project is developed based on a distributed architecture (Figure 4) that uses a multitude of technologies for of data acquisition, transformation, storage and publishing. The main architecture components are as follows:

Big Data Test Infrastructure (BDTI): The BDTI plays a crucial role in this project by providing a cloud-based environment and resources specifically designed for experimenting with open-source big data tools and technologies, including Airflow and Virtuoso. This readily available infrastructure eliminates the need for extensive setup and configuration, allowing for rapid development and testing.

Apache Airflow Orchestration: An instance of Apache Airflow is deployed within the BDTI environment in order to orchestrate the entire data processing pipeline. It offers a user-friendly web interface for workflow visualization, allowing for real-time monitoring of task progress.

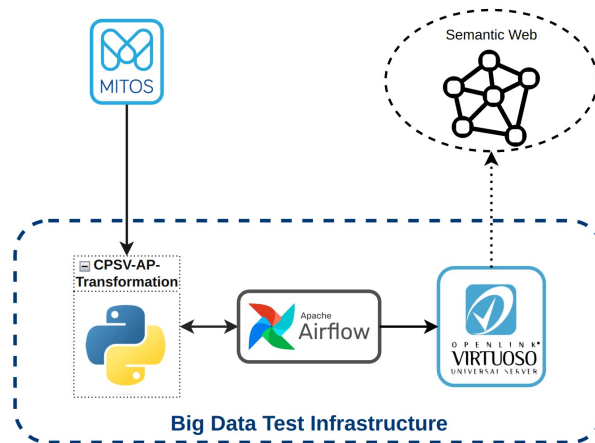


Figure 4: MITOS_LOD architecture

MITOS: The data acquisition process commences with interaction with the MITOS through a call to its open API. The API of MITOS provides easy access to the available data in a structured machine-readable way.

CPSV-AP-Transformation to LOD: The core of the project is a series of Python scripts that run through Airflow to transform the data retrieved through the API, based on a mapping, to a format compliant with CPSV-AP. The transformation process results in RDF data using the Turtle syntax.

Open Link Virtuoso: The LOD data created are stored in the deployed Open Link Virtuoso instance on the BDTI that offers also a publicly available SPARQL endpoint.

The pipeline for producing and publishing linked data contains the following steps:

1. **Airflow orchestrator:** The process begins with Apache Airflow orchestrating the entire workflow by first calling the services responsible for data retrieval.
2. **Retrieve Public Services IDs:** The Python scripts send asynchronous HTTP GET requests to the Mitos.gov.gr API to retrieve a list of IDs of the available public service.
3. **Retrieve Public Services Descriptions:** Using the retrieved IDs the Python script requests through the MITOS API the full details of all public services using a separate request for each ID.
4. **Processor function:** As the details are received, they are processed, structured, and stored in CSV files organized by groups of informations i.e., the classes of CPSV-AP. For example the groups include General information, Evidences, Requirements etc..
5. **Transformation:** The data read from the CSV files are converted into RDF format, based on a mapping to CPSV-AP, using the Turtle syntax.
6. **Importing to Virtuoso:** With the data ready and correctly formatted, Apache Airflow is setup to use the available Virtuoso API to import the data onto its RDF store.
7. **Publish LOD:** The data is published as Linked Open Data (LOD) for use on the web.

5. Discussion

This section describes some usage scenarios of the produced LOD for MITOS that highlight the prospective benefits. The usage scenarios include:

1. Publishing PS descriptions as LOD provide citizens and businesses with structured and standardized information about PS offerings, potentially improving discoverability and streamlining the application process. This could ultimately enhance user experience and promote the utilization of PS catalogs, leading to resource optimization (e.g., time, cost).
2. Utilizing LOD, service providers could create personalized recommendations for relevant PSs based on individual user profiles and needs, leading to a more efficient and targeted service experience.
3. LOD representations of PS descriptions empower policymakers to conduct in-depth analyses and evaluate the potential impact of PS redesign or improvement opportunities. This enables them to address complex questions such as the effects of abolishing specific document requirements on a PS.
4. Publishing PS descriptions as LOD fosters transparency and accountability by making PS-related information publicly accessible, open, shareable, and reusable. This approach also facilitates more effective and efficient PS portfolio management, allowing for better resource allocation and service optimization. Furthermore, it enables public organizations to participate more readily in PS description federation, thereby enhancing interoperability and collaboration with other public entities.
5. By leveraging LOD, public organizations can track and measure the performance of PSs, enabling them to identify areas for improvement and optimize service delivery based on datadriven insights.
6. The availability of LOD datasets containing PS descriptions presents an opportunity for the IT industry to develop new value-added services and applications that take advantage of this data.

6. Conclusions and Future Work

This paper demonstrates the feasibility and advantages of using the EU Big Data Test Infrastructure to publish Public Service Descriptions from MITOS as Linked Open Data leveraging the CPSV-AP specification. It can serve as a valuable reference for other public entities by showcasing the practical application of BDTI for hosting public projects. Additionally, our experience with BDTI also fostered a collaborative relationship, allowing us to provide valuable feedback on the platform, ensuring its continuous improvement for future users.

The proposed approach (including the model, process and architecture) to create LOD from MITOS can easily be adapted and adopted by other public administrations that want to transform and publish their PSDCs as LOD.

Going forward, an important area for future development is linking MITOS linked data with other data in the LOD cloud as well as creating user-friendly visualization tools for these data. These tools can transform complex LOD datasets into readily understandable formats, empowering a wider range of users to interact with and leverage public service data. Additionally,

customizable dashboards could be developed to provide policymakers with real-time insights into service usage and performance metrics for data-driven decision-making. Another area for future development is optimizing the LOD update process. Currently, a notable performance improvement could be achieved by minimizing the number of calls to the PSD source. This would reduce the load placed on their systems and enhance overall efficiency. One potential approach involves implementing an incremental update strategy. This strategy would focus on identifying and updating only the portions of the LOD dataset that have actually changed, rather than retrieving and reprocessing the entire dataset on each update cycle.

Acknowledgments

The ready-to-use, free of charge analytics cloud stack infrastructure used to implement the pilot has been provisioned by the Big Data Test Infrastructure Support Team from the European Commission (Directorate General for Digital services, Unit B1, Data, Artificial Intelligence & Web) that provided guidance and assistance throughout the duration of the project. The Big Data Test Infrastructure programme is part of the Digital Europe Programme (DEP). This work has been partially supported by the GR digiGOV-innoHUB project (Grant Agreement no. 101083646), which is co-funded by the European Union under Competitiveness Programme (ESPA 2021-2027).

References

- [1] A. Gerontas, V. Peristeras, E. Tambouris, E. Kaliva, I. Magnisalis, K. Tarabanis, Public service models: a systematic literature review and synthesis, *IEEE Transactions on Emerging Topics in Computing* 9 (2019) 637–648.
- [2] E. Commission, Core public service vocabulary application profile 3.2.0, 2024. URL: <https://github.com/SEMICEu/CPSV-AP/tree/master/releases/3.2.0>.
- [3] OPSI, Mitos – the national registry of administrative procedures. observatory of public sector innovation., 2024. URL: <https://oecd-opsi.org/innovations/mitos-the-national-registry-of-administrative-procedures/>.
- [4] C. Bizer, T. Heath, T. Berners-Lee, Linked data: The story so far, in: *Semantic services, interoperability and web applications: emerging concepts*, IGI global, 2011, pp. 205–227.
- [5] L. Ding, V. Peristeras, M. Hausenblas, Linked open government data [guest editors' introduction], *IEEE Intelligent systems* 27 (2012) 11–15.
- [6] D. Zeginis, K. Tarabanis, An event-centric knowledge graph approach for public administration as an enabler for data analytics, *Computers* 13 (2024) 17.
- [7] A. Gerontas, E. Tambouris, K. Lazopoulou, K. Tarabanis, On using cpsv-ap to publish public service descriptions as linked open data, *Service Oriented Computing and Applications* 16 (2022) 231–261.
- [8] P. Fragkou, L. Maglaras, Transforming points of single contact data into linked data, *Computers* 11 (2022) 122.
- [9] BDTI, Bdti success stories, 2024. URL: https://big-data-test-infrastructure.ec.europa.eu/success-stories_en.