

Towards a Common Reference Infrastructure for Documenting Public Sector Administrative Processes: The Case of “MITOS”

Stylios Bourmpoulis*, Dimitris Zeginis and Konstantinos Tarabanis

University of Macedonia, 156 Egnatia Street, 546 36, Thessaloniki, Greece

Abstract

The Greek government introduced in 2020 the National Registry of Administrative Procedures (“MITOS”) to standardize public sector administrative processes, improve transparency, and provide accurate information to citizens and businesses. While various processes have been documented, the lack of standardized metadata and metaprocesses has led to ambiguity, inconsistent content, and unnecessary complexity. To address these issues, the paper proposes the creation of a common reference infrastructure of standardized metadata and metaprocesses to ensure accuracy and consistency in the final documentation of administrative processes. As a case study, the process of enrolling students in the first grade of high school was chosen. The implementation of the proposed method revealed that using process-oriented metamodels to annotate the underlying legal texts helps to extract crucial information comprehensively. Furthermore, establishing standardized metadata reduces heterogeneity, while outlining the hierarchical structure of the process creates a precise framework for adequately organizing and naming the processes. Lastly, publishing the extracted information in a machine-readable format can pave the way for applications to simplify information access for public servants and citizens.

Keywords

Public Sector, Administrative Processes, Documentation, Metadata, Metaprocesses, MITOS

1. Introduction and motivation

The main objective of Public Administration is to deliver high-quality public services to citizens and businesses. These services are realized through specific administrative processes outlined in the state's legal documents to ensure transparency, accountability, and equal access. However, the implementation of administrative processes often deviates from the prescribed legal requirements, causing confusion among the consumers of public services. As a result, citizens and businesses face obstacles in accessing public services, which in turn harms the state's credibility.

Recognizing the aforementioned issues, the Greek government passed in 2020 a law to establish a National Registry of Administrative Procedures known as “MITOS”[‡]. This registry would serve as the central repository for administrative processes within the Greek public sector. The government aimed for “MITOS” to promote standardization of administrative processes, improve transparency and legal certainty, and provide accurate and reliable information on the operations of public administration.

According to the law, competent public authorities must record all public administrative processes, whether physical or digital, into “MITOS”. Specifically, they must register a process when

Companion Proceedings of the 17th IFIP WG 8.1 Working Conference on the Practice of Enterprise Modeling Forum, M4S, FACETE, AEM, Tools and Demos co-located with PoEM 2024, Stockholm, Sweden, December 3-5, 2024

* Corresponding author.

✉ steliosbour@uom.edu.gr (S. Bourmpoulis); zeginis@uom.edu.gr (D. Zeginis); kat@uom.edu.gr (K. Tarabanis)

🆔 0009-0006-5288-3580 (S. Bourmpoulis); 0000-0002-5426-4031 (D. Zeginis); 0000-0002-4663-2113 (K. Tarabanis);



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

‡ https://en.mitos.gov.gr/index.php/Main_Page

The term “MITOS” originates from ancient Greek mythology. According to the myth, Ariadne provided Theseus with a ball of yarn (referred to as “*mitos*” in Greek), which enabled him to navigate and escape the Labyrinth after defeating the Minotaur. Similarly, the objective of the National Registry of Administrative Procedures is to assist government employees and consumers of public services in navigating through complex administrative procedures.

it is initiated for the first time and document all the essential elements. These elements include the official title of the process, the legal and regulatory framework, the relevant departments, the necessary documents, the steps, the estimated time required to complete the process, as well as any associated fees and charges. Additionally, they must update the processes when they are modified in any way and remove them from the system when they are retired.

Since the introduction of "MITOS", the competent public authorities have been focusing on quickly documenting numerous processes of the Greek public sector. However, this has been accomplished using a limited number of metadata and metaprocesses. The lack of precise definitions for concepts has led to ambiguity, while using different terms for the same concept has caused inconsistency and confusion. Moreover, the absence of commonly accepted models for the hierarchical structure and the process steps has resulted in unnecessary complexity. Lastly, the manual documentation of processes has often led to incorrect or incomplete information. Consequently, the documented processes often exhibit structural heterogeneity, content inconsistency, and deviation from the governing legal framework.

In order to tackle these challenges, the paper proposes an approach for documenting public sector administrative processes based on metadata and metaprocesses standardization. Specifically, it suggests creating a common reference infrastructure of standardized metadata and metaprocesses to ensure accuracy and consistency in the final documentation of administrative processes. The infrastructure is planned to be established as a key component of MITOS, aiming to serve as the conceptual basis for developing systems and applications within the Greek Public Administration. To apply the suggested approach and pinpoint any areas requiring improvement, the paper examines the administrative process of enrolling students in high school as a case study.

The rest of the paper is structured as follows: Section 2 develops the background and related work of the study. Section 3 analyses the scope, goal, requirements, and method of developing the proposed infrastructure. Section 4 applies our approach to the administrative process of student enrollment. Finally, section 5 presents conclusions, study limitations, research gaps, and recommendations for future work.

2. Background and related work

In all democratic societies, public services and the corresponding administrative processes are based on law. However, according to the literature, problems of interpretation and implementation often arise from the adoption and publication of legal regulations to their practical application [1]. Even within the same state, the services and processes published on governmental sites and portals often differ in their name, structure, and execution, showing deviations from the law [2]. Different interpretations by the large number of stakeholders and the absence of interoperable descriptions of legal regulations lead to inconsistency and complexity [3]. Additionally, the absence of data standards exacerbates the problem and makes data collection, processing, and analysis a constant challenge.

The literature suggests methods to address different interpretations of the law [1] and achieve an end-to-end digitization of public services [3]. Researchers also emphasize the importance of standardizing both data and processes to create a unified view of public administration [2]. A common example is the use of controlled vocabularies, which are defined as "organized sets of controlled terminology values" [4]. Controlled vocabularies are widely used for indexing, categorization, and information retrieval. They include controlled lists, synonym rings, taxonomies, and thesauri. For instance, the Common Education Data Standards (CEDS) in the USA has developed a common vocabulary, data models, tools, and metadata to help stakeholders understand and use education data consistently [5]. Similarly, [6] proposes a technical infrastructure to integrate educational data from various sources.

In the literature, there are also process frameworks with a hierarchical organization, such as the APQC's Process Classification Framework (PCF)[7]. These frameworks can serve as a guide and can be customized for specific application areas. They help in organizing an organization's processes,

which are often numerous. Each metaprocess includes more specific processes. This hierarchical structure can be continued at several levels, culminating in a few metaprocesses at the first level that describe the organization's functions in a generalized, abstract way (e.g., 5–12 for an organization). Moving from higher levels of the hierarchy to lower ones, we transition from defining "what" work is accomplished to "how" a particular organization accomplishes its work.

Researchers [2] recommend using ontologies to create ontology-based business process models, which have the advantage of being both human and machine-interpretable. Ontologies make knowledge about the modelling language (e.g. BPMN) and the application domain explicit, thus ensuring a precise shared interpretation of the information encapsulated in the process model to both humans and machines. This can be done either by annotating process model elements with classes from the modelling language and domain ontology and exporting them as instances of ontology classes ("semantic lifting") or by creating directly process model elements as instances in the modelling language and annotating them with domain knowledge from the application domain ontology ("semantic metamodeling") [8]. Since the semantics can be made formal through an ontology language, like RDF(S), ontology-based process models have the potential to allow automated analysis, decision making and digitalization.

3. Towards a common reference infrastructure for documenting processes

3.1. Scope and goal

The paper scopes the process of documenting and recording administrative processes in "MITOS". This process involves studying the legal text, gathering essential information about the administrative process (such as involved parties and steps), and entering this information into the system. The paper proposes an approach for documenting and registering administrative processes by creating a common reference infrastructure of standardized metadata and metaprocesses. This infrastructure will be a reference for all public servants involved, providing precise definitions of terms and relationships, outlining the hierarchical structure of processes, and defining standard steps for process execution. It is expected that it will improve the consistency and quality of the final documentation of administrative processes in the Greek public sector. The type of metadata and metaprocesses contained in the proposed infrastructure concerns the operational (conceptual) level and not the technical elements related to the physical representation of data and processes in information systems.

3.2. Requirements

Before developing our approach, specific requirements were established. The requirements were derived from the vision of Society 5.0 [9], which is expected to shape the operations of public administrations in the future. Its objective is to establish a highly intelligent society where individuals are at the center of a network consisting of interconnected services, data, information, and knowledge. In other words, its purpose is to create a human-centered, knowledge-intensive and data-driven society. Moreover, since the proposed infrastructure mainly relates to the Greek registry of administrative procedures, the requirements need to align with the main objectives of "MITOS" as described in the relevant law. These objectives include standardizing administrative processes, increasing transparency and legal certainty, and providing accurate and reliable information.

Therefore, the proposed infrastructure:

- *Requirement 1:* must ensure that the information entered in "MITOS" is accurate and aligns with the legal text without any distortions. This will enhance transparency and help citizens understand how to access public services.

- *Requirement 2:* must ensure that the structure and the content of administrative processes are clearly defined enabling consistency and understanding among all involved parties. Civil servants should use the same documentation toolkit to avoid inconsistencies and complexity.
- *Requirement 3:* should allow the publication of documented information in a machine-readable format to enable processing by advanced systems like artificial intelligence. The processed information can then be used to improve people's lives in the real world.

3.3. Methodology

The development of the proposed infrastructure is at the core of several activities and stages, beginning with drafting the legal text and concluding with the publication of the documented information in "MITOS" (Figure 1). Precisely, the process consists of seven stages as follows:

1. *Stage 1: Select text(s).* In this stage the appropriate legal text (e.g. law, ministerial decision) defining the legal framework for an administrative process is selected. It is crucial to select the current version of the legal text, as there are often amendments or even complete repeals of law articles.
2. *Stage 2: Annotate text(s).* Stage 2 involves annotating texts to ensure full understanding of their content. Three methods of annotation are proposed. The first method involves a top-down approach using a conceptual model, such as the TOGAF Core Metamodel customized for process modeling requirements [10]. The text is annotated and its information is structured according to the key elements of the model (e.g. actors, roles, steps). This can be done manually or automated through a language model, such as ChatGPT. The second method is bottom-up, where domain experts study the text to identify structural elements. The third method combines the top-down and bottom-up approaches, comparing and complementing their results.
3. *Stage 3: Standardize metadata.* At this stage, standardized metadata is created using controlled vocabularies. More specifically, a list of terms is generated from the legal text, which is the simplest form of metadata. This list is then cleaned up, organized, and its terms are correlated to produce more complex metadata in the form of a controlled vocabulary. The controlled list will serve as a basis for developing other types of controlled vocabulary, such as synonym ring lists, taxonomies, and thesauri.
4. *Stage 4: Standardize metaprocesses.* At this stage we use the standardized metadata created in Stage 3 to define metaprocesses. These metaprocesses describe a set of related processes at a high level and, through hierarchical specialization, reach the level of the executed processes. This creates a hierarchical tree of an organization's processes. At this stage, a process model structure can be developed with standard steps/phases in a sequence. The steps/phases will have standard names using terms from controlled metadata vocabulary.
5. *Stage 5: Create process model.* At this stage, the standardized metadata and metaprocesses created in stages 3 and 4 are utilized to develop the administrative process model. This model will comprehensively represent various elements of the process, including actors, steps, and necessary documentation.
6. *Stage 6: Ontologize model.* During this stage, reference ontologies are utilized to structure and publish the information from the process model in both human-readable and machine-readable formats. This is crucial for making the information easily discoverable, as it allows for quick retrieval of needed information through queries. This will enable automated extraction of useful answers to questions such as "What documents do I need to provide to the public administration to access a service?" and "What actions do I need to take as a public official in the specific administrative process, and when do these actions need to be implemented?".

7. *Stage 7: Publishing information.* In the final stage, the critical information that comprises the identity of the administrative procedure (e.g. competent authorities, steps, evidence) is published in "MITOS" in both human-readable and machine-readable formats.

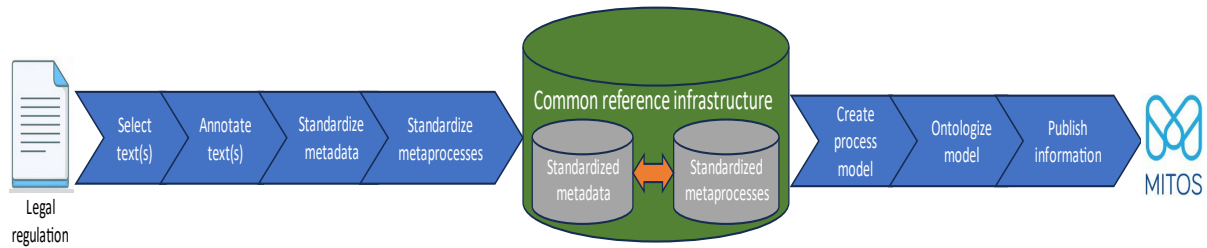


Figure 1: The common reference infrastructure and its surrounding activities.

4. Case study: The process of “Student Enrollment in High School”

In order to implement the proposed approach, we searched "MITOS" to find processes already recorded in the system that demonstrate the issues mentioned in the introduction. Due to the large number of processes, we focused on processes related to the field of education, which is our area of expertise. By analyzing the legal texts of the processes and drawing on our experience as domain experts, we identified candidate processes that demonstrated documentation problems, such as ambiguity in the official title and content, as well as inaccurate information. Ultimately, we chose the "Student Enrollment in High School" process from the candidate processes we identified. We chose this particular process because it encompasses a significant number of the documentation issues mentioned above and involves a large number of users, including student guardians and students.

4.1. The current process documentation in “MITOS”

Before implementing our method, we decided to review the "Student Enrollment in High School" process in "MITOS". First, we studied the relevant legal text, namely a ministerial decision of 2019 amended in 2022. Subsequently, we compared the information contained in the ministerial decision with the information entered for this process in "MITOS". We then categorized the documentation issues into key process elements, including the official title, the competent authorities, the required evidence, and the process steps (Table 1).

More precisely, the process' title ("Student Enrollment in High School") was found to encompass a number of processes, each differing in several aspects such as the participants involved and the steps followed. For instance, the process for enrolling in the first grade of high school differs significantly from enrolling in the second and third grades. The issue arises from the lack of a hierarchical model for organizing these processes, where the second-to-last level should group the different versions of the process (e.g., enrolling in a secondary school), and the last level should detail the specific versions of the process (e.g., enrolling in the first grade of secondary school). The absence of this structure causes confusion for users, as they encounter information and steps that do not apply to all versions of the high school enrollment process.

The current documentation is also causing confusion as it fails to clearly outline the steps and required documents for the enrollment process. It is unclear which steps pertain to enrolling in first grade and which are for other grades. Additionally, specific processes (e.g., "Carrying out overdue registrations") have been entered incorrectly in the system as steps, which seems to be a result of misunderstanding. Moreover, the responsibility for carrying out the steps is vaguely attributed to the "Director of the Educational Structure" instead of specific roles such as "Head teacher of the High School." There is also inconsistency in the terminology used, with different terms like "School", "School unit" and "Educational structure" used interchangeably for the same concept, causing further

confusion. Lastly, the description of the process is unclear, leading to confusion among those involved in the process.

Table 1

Documentation issues identified in the process

Key elements of process	Documentation issue
Official title	The title does not clearly specify which version of the process is being referred to.
Competent authority	There is ambiguity regarding who is responsible for carrying out the process activities.
Required evidence	The particular evidence does not indicate which version of the service they correspond to.
Process steps	There is inconsistency in the terminology used, with different terms used interchangeably for the same concept. The specific steps do not specify which version of the service they correspond to.

4.2. The proposed approach for documenting the process

After reviewing the current process documentation, we implemented our approach to test its effectiveness in addressing the previously mentioned issues. Since our review revealed that the high school enrollment process consists of several specific processes (e.g., "Enrollment in the first grade of high school", "Enrollment in the second grade of high school") that differ from each other, we decided to focus on one of these. We chose to focus on the enrollment process in the first grade of high school, as it involves a larger number of participants and steps compared to the others.

4.2.1. Select text

In the Greek education system, a ministerial decision of 2019 outlines the processes for enrolling students in different grades and types of secondary schools. This decision was amended in 2022, and the revised version is available in PDF format on the website of the National Printing Office of the Greek government, ensuring easy access for all stakeholders. The ministerial decision was chosen as the reference text for the documentation of the enrolment process in the first grade of high school.

4.2.2. Annotate text

We started by annotating the ministerial decision using a semi-automated method. At first, we manually annotated the ministerial decision by implementing a top-down approach. More precisely, our annotation was guided by a set of questions (Table 2) that were specifically designed to match each component of the TOGAF Core Metamodel, tailored to meet the requirements for process modeling. This metamodel was chosen because it allows for comprehensive process modeling and the documentation of process artifacts.

After tabulating the information, we automated the annotation process using the free edition of the large language model (LLM) ChatGPT. More precisely, we have uploaded the ministerial decision, which outlines the specific administrative process, in PDF format. This ensures that the responses we receive are based on the official institutional text of the Greek government rather than relying on general knowledge that may be available. Since "MITOS" is intended for both citizens and public administration executives, we created separate lists of questions for each category of users (Table 3). The answers we received were recorded in a separate table. Finally, we combined the manually and automatically retrieved information and structured it according to the TOGAF Core Metamodel components in a single table to define the process identity (Table 4).

Table 2

Questions for annotating the legal text

TOGAF metamodel component	Questions
“Process”	<i>“What processes are described in the legal text?”</i>
“Function”	<i>“What functions are realized by the processes?”</i>
“Actor”	<i>“Who are the actors participating in the processes?”</i>
“Role”	<i>“Which roles are assumed by the actors?”</i>
“Organization Unit”	<i>“To which organizational units do they belong?”</i>
“Event”	<i>“Which events are generated by the processes?”</i>
“Product”	<i>“What products are produced by the processes?”</i>
“Control”	<i>“Which controls are guided by the processes?”</i>
“Business service”	<i>“Which public services are realized by the processes?”</i>

Table 3

Questions asked to ChatGPT

Viewpoint	Questions
Student guardian	<i>“What do I need to enroll my child in first grade of high school?”</i> <i>“Which documents should I provide, and where and when?”</i>
Head teacher, Director of Directorate of Education	<i>“What is the process for enrolling a student in the first grade of high school?”</i> <i>“What actions should be taken by the head teacher/Director of the Directorate of Education, and when should these actions be implemented?”</i>

Table 4

The identity of “Student enrolment in the first grade of high school”

TOGAF metamodel component	Information
“Organization unit”	Directorate of Primary Education, Directorate of Secondary Education, Primary School, High School
“Actor”	Student, Citizen, Teacher
“Role”	Student guardian, Head teacher of primary school, Head teacher of high school, Director of Directorate of Secondary Education
“Function”	Student enrollment
“Process”	1. The Director of the Directorate of Secondary Education issues an invitation. 2. In December of the current school year, the student guardian submits a formal declaration of permanent residence, along with proof of permanent residence, to the head teacher of the primary school the student attends. 3. The head teacher of the primary school creates a list of the students' addresses. 4. The head teacher of the primary school sends the list to the Directorate of Secondary Education.

5. In January, the Director of the Directorate of Secondary Education makes a decision on the designation of the secondary schools for the students' enrollment.
6. The Director of the Directorate of Secondary Education sends the lists to the Directorate of Primary Education.
7. The Directorate of Primary Education forwards the lists of the secondary schools where the pupils are enrolled to the primary schools.
8. The head teacher of primary school informs the students' guardians.
9. The head teacher of primary school sends the school-leaving certificates to the high schools of enrollment.
10. The student's guardian submits to the high school a formal declaration about the legal exercise of guardianship and a photocopy of the student's identity card or a certificate of the municipality in which the student is registered.
11. The head teacher of the high school records the enrolled student's information in the student's Individual Card and the School's Student Register.

“Event”	Invitation issued, Certification of permanent residence submitted, List of students' addresses created and sent, School of students' enrollment designated, List of students' enrollment high schools sent Students' guardians informed, School-leaving certificates of students sent, Required documentation for student enrollment submitted, Student recorded, Enrollment completed
“Control”	Checking students' eligibility for enrolling in the first grade of high school Reviewing submitted registration documents
“Service”	Student enrollment service
“Output”	List of the students' addresses, Recording of student information

4.2.3. Standardize metadata

After extracting the process' critical information, we created a simple list of terms identified in the ministerial decision. Then, we developed a controlled list following the guidelines of [4]. We included unique terms that belonged to the same category, had no overlapping meanings, and were equal in granularity/specificity. These terms were then organized alphabetically into six main categories (Table 5) describing not only the student enrolment process but the whole world of education: 1) Person, 2) Role, 3) Organization, 4) Geographic Locations, 5) Physical Resources, and 6) Activities and Events [11]. This broad categorization was deemed necessary to create education-oriented metadata that could also be used in other administrative processes.

We then further refined our list by creating a synonym ring list. Our choice is justified as the synonym ring list is used for information retrieval, providing access to content represented in texts and other instances of natural, uncontrolled language [4]. This is particularly important in the case of creating a metadata and metaprocess infrastructure, where the user is looking for information such as possible versions of the service and procedure that might be of interest. In compiling our list of synonyms, we included near-synonyms with similar or related meanings rather than limiting the list to only those with true synonymy. This approach allows for a broader range of words that can be used interchangeably in context. Our list of synonym rings was created manually using our expertise as domain experts (Table 6).

Table 5

Controlled list of terms for the student enrollment process (*restricted for space reasons*)

Category	Terms
Person	Citizen, Student, Teacher
Role	Director of Directorate of Secondary Education, Head teacher of high school, Head teacher of primary school, Student guardian
Organization	Directorate of Education, School, ...
Geographic Locations	Neighboring school, School district
Physical Resources	Student Register, Formal declaration, ...
Activities	Apply, Enroll, ...
Events	Enrollment, Start of school year, ...

Table 6

List of synonyms rings (*restricted for space reasons*)

1	Student, pupil, student from foreign countries, handicap student ...
2	Teacher, teaching staff, head teacher ...
3	School, school unit, high school, school district, ...
4	Student guardian, parent, divorced parents...
...	...

In the cases mentioned above, when different terms were used to refer to the same concept, we designated one term as the preferred term and the others as variant terms. The preferred terms were selected based on their frequency in ministerial decisions and their usage by domain experts in their daily work (Table 7).

Table 7

Preferred and variant terms

Preferred term	Variant terms
Teacher (<i>“ekpaideutikos” in greek</i>)	Teacher (<i>“kathigitis” in greek</i>), Teaching staff
School	School unit

In the next step, we defined the specific terms to be included in our controlled list. We relied on domain ontologies (such as ESCO Ontology [12]) and reputable dictionaries (like Wordnet [13]) to provide comprehensive definitions and descriptions for the terms outlined in this paper (Table 8). These resources offer definitions in formats that are understandable both for humans and machines, and their widespread acceptance supports semantic interoperability.

Table 8

Terms definition (*restricted for space reasons*)

Term	Definition	Source
Student	“A learner who is enrolled in an educational institution”	Wordnet
Secondary school head teacher	“Secondary school head teachers are responsible for meeting curriculum standards, which facilitate academic development for the students. [...]”	ESCO Ontology

In the subsequent phase, we enriched our list by introducing hierarchical relationships among the terms to construct a comprehensive taxonomy. To ensure the accessibility of this taxonomy to both humans and machines, we employed the ontology development environment of WebProtégé

(<https://webprotege.stanford.edu/>) for its creation. Within our taxonomy, each term is associated with one or more parent/child (broader/narrower) relationships with other terms, encompassing whole/part, genus/species, and instance relationships (Figure 2).

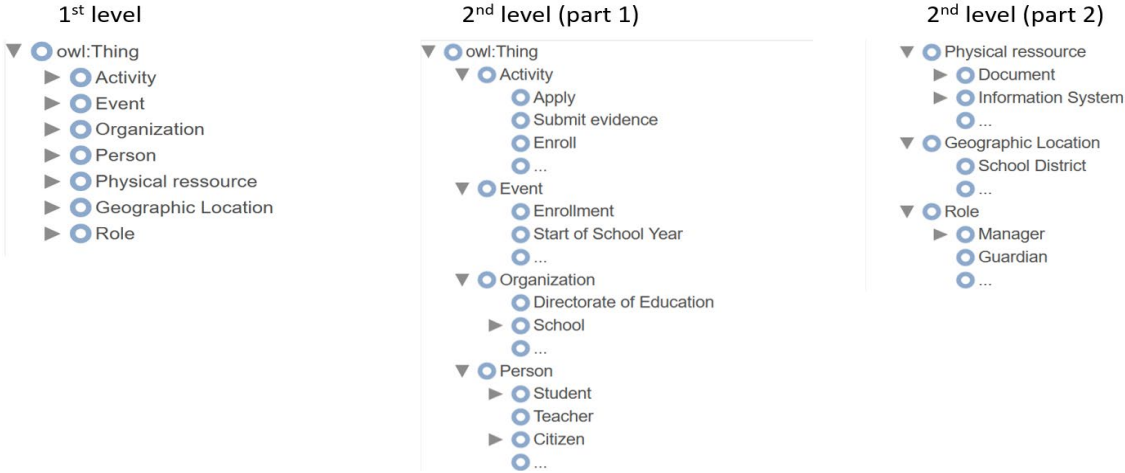


Figure 2: Taxonomy of student enrollment-related terms.

4.2.4. Standardize metaprocesses

After annotating the legal text and creating a taxonomy of student enrollment-related terms, we proceeded with the hierarchical documentation of the process. This was necessary to accurately document the student enrollment process within the intricate education system and resolve issues of confusion and ambiguity in process terminology. For the documentation of the enrollment process, we used the hierarchical process structure scheme from the APQC Process Classification Framework (PCF) - Cross-Industry - Excel Version 7.4 [7]. According to this, processes are structured into five levels. The Level 1-Category represents the highest level of processes. The Level-2 Process Group denotes a group of processes. The Level-3 Process represents a single process. The Level - 4 Activity designates the key steps to execute a process. Finally, the Level - 5 Task describes elements of work that go into executing an activity.

Applying the APQC scheme to our case, we have implemented the process hierarchical structure. At Level 1, we have established the overarching category of "Student Enrollment". At Level 2, we have delineated the process group pertaining to secondary education, denoted as "Student Enrollment in Secondary Education School." Moving down to Level 3, we have specified a particular version of the aforementioned process group, labeled as "Student Enrollment in High School." Upon discovering variations in the enrollment process within the Greek educational system, we introduced an additional Level 4, encompassing the process version "Student Enrollment in the First Grade of High School." In Level 5, we added the main activities of the process that make up its internal structure. To move towards standardizing the structure of the school registration process, we looked into "MITOS" for processes related to citizen registration, not only in schools, but in state registers in general. We found four basic steps that are repeated in almost all procedures: 1. "Submission of an application for enrollment", 2. "Checking the application", 3. "Decision on the application", and 4. "Notification of decision". Finally, in Level 6, we captured the process' tasks (Figure 3).

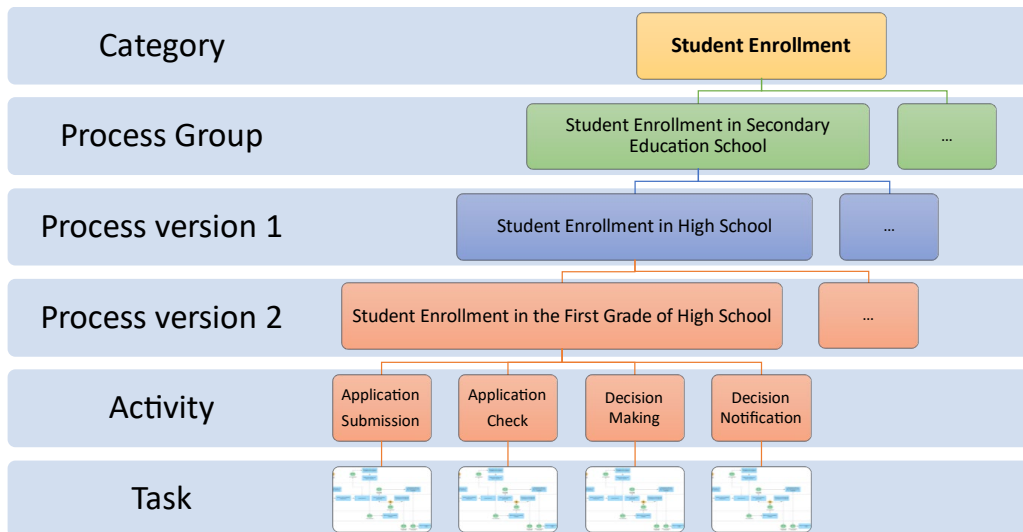


Figure 3: Hierarchical decomposition of Student Enrollment process.

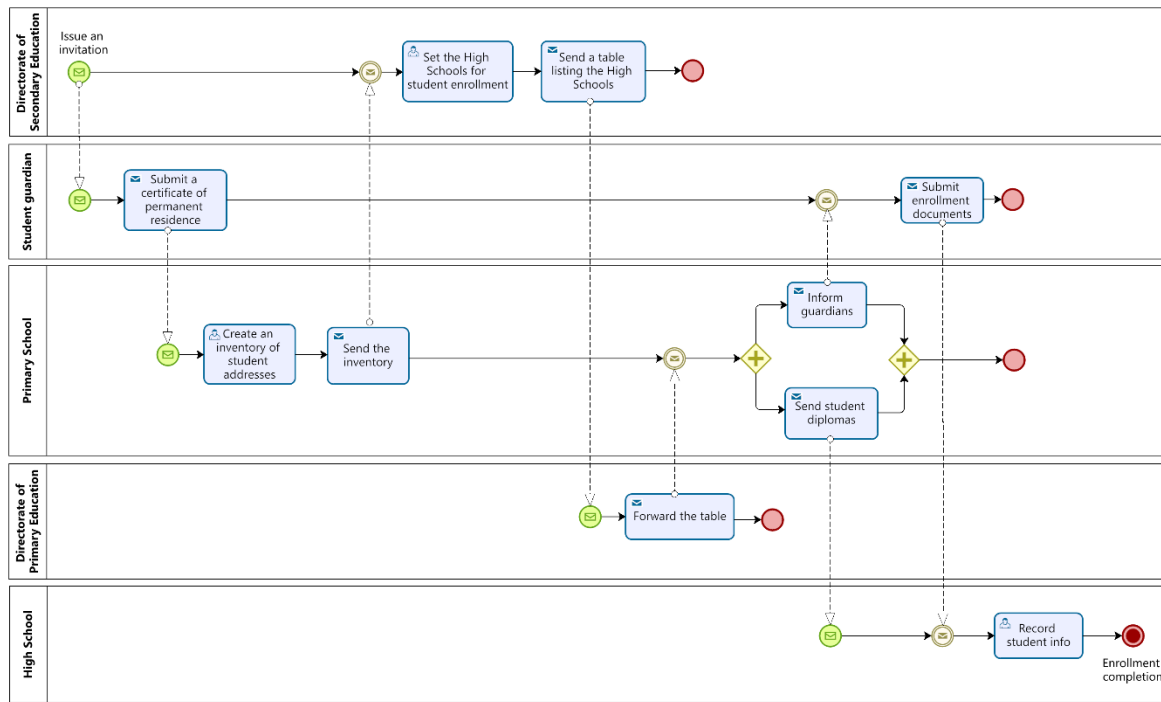
4.2.5. Create process model

Using the information obtained from the ministerial decision annotation, the metadata, and the metaprocesses, we modeled the process of student enrolment in the first grade of high school in the BPMN language (Figure 4).

4.2.6. Ontologize process model

In this phase, we moved on to ontologizing the model. To semantically enrich our model, we primarily utilized ontologies and core vocabularies designed for Public Administration needs. If specific public sector ontologies were not available, we utilized domain-neutral ontologies. For instance, public organizations involved in the process (e.g., the Secondary Education Directorate) are described using the Core Public Organization Vocabulary developed by the European Commission [14]. When structuring information about individuals, such as students and their guardians, we employed the Core Person Vocabulary of the European Commission [15]. For occupations like teachers and school head teachers, we used the ESCO ontology, which was developed to describe "European Skills, Competences, Qualifications and Occupations"[12].

To delineate the process activities we used the ISO 18629 PSL, a standardized language for specifying and exchanging process information [16]. To represent the events, we employed the SEM ontology [17], which defines the primary components of an event, such as time and place. It is particularly beneficial for characterizing elements like geographic locations (e.g., school district) or periods (e.g., school year). For the controls conducted during the enrollment process (e.g., verifying supporting documents), we used the Core Criterion and Core Evidence Vocabulary [18] formulated by the European Commission to standardize data about criteria and evidence. Lastly, to describe the public services facilitated by this process, we utilized the Core Public Service Vocabulary [19] developed by the European Commission, which outlines public services and their ecosystem, encompassing inputs, outputs, and the administering organization.



Powered by
 Modeler

Figure 4: The process of student enrollment in the first grade of high school in BPMN.

4.2.7. Publish information

Based on the above-mentioned ontologies the final step produces a machine-readable documentation of the selected administrative process. The selected language for the implementation is RDF, however any other language could also be used. Figure 5 shows an excerpt of the produced code that represents the enrollment process (`ex:EnrollmentService`) and some of the Tasks of the process (Tasks 2 – 4 as described in Table 4). A human-readable view can also be produced based on the produced RDF using dedicated templates. This will allow the presentation of critical information of the process in a structured way (e.g., at MITOS).

Each task includes information such as the description (`dct:description`), involved actors (separating them to actors – `ex:actor` - and recipients – `ex:recipient`), required/produced (`ex:require`, `ex:produce`) evidences, the period of time (`ex:time`) that a task should be performed according to the law and information about the order of the task (`ex:next`, `ex:order`). The prefix “`ex`” is used as an example. In some cases, the code could be further enhanced with more concepts from the ontologies mentioned above, however “`ex`” concepts are used to improve the code readability. For example, the concept “`headTeacherPrimarySchool`” is defined at ESCO as “`esco: c9d89e09-d57f-4980-9164-880e60711e97`”.

Based on the produced code a number of queries can be executed to retrieve valuable information. The queries are expressed in SPARQL that is a query language for RDF. For example, the left query at Figure 6 returns the tasks (including their order, description and time) that involve the student guardian as an actor and the relevant evidences of these tasks. The right query returns that tasks (including the order, description and time) that involve the head teacher of a primary school either as actor or receiver. The output of the queries is presented at Figure 8 and 9 respectively.

```

ex:task2 a ex:Task;
  dct:description "Submit a certification of permanent residence";
  ex:actor ex:StudentGuardian;
  ex:recipient ex:headTeacherPrimarySchool;
  ex:require ex:declPermanentResidence,ex:proofPermanentResidence;
  ex:time greg:December;
  ex:next ex:task3;
  ex:order "2"^^xsd:integer.
ex:task3 a ex:Task;
  dct:description "Create a list of the student addresses.";
  ex:actor ex:headTeacherPrimarySchool;
  ex:produce ex:listOfStudenAddresses;
  ex:next ex:task4;
  ex:order "3"^^xsd:integer.
ex:task4 a ex:Task;
  dct:description "Send the list";
  ex:actor ex:headTeacherPrimarySchool;
  ex:recipient ex:DirectorateSecondaryEducation;
  ex:require ex:listOfStudenAddresses;
  ex:next ex:task5;
  ex:order "4"^^xsd:integer.
ex:declPermanentResidence a cv:Evidence ;
  dct:title "Formal declaration of permanent residence".
ex:proofPermanentResidence a cv:Evidence ;
  dct:title "Proof of permanent residence".
ex:listOfStudenAddresses a cv:Evidence ;
  dct:title "List of student addresses".
ex:EnrollmentService a cpsv:PublicService;
  ex:involveTask ex:task2, ex:task3, ex:task4;
  dct:title "Student enrollment process".

```

Figure 5: Excerpt of the RDF representation of the student enrollment process

```

PREFIX ex: <http://www.exampe.org/>
PREFIX dct: <http://purl.org/dc/terms/>
SELECT ?order, ?taskDescription,
       ?EvidenceTitle, ?time
WHERE{
  ex:EnrollmentService ex:involveTask ?task.
  ?task ex:actor ex:StudentGuardian.
  ?task dct:description ?taskDescription.
  ?task ex:require ?req.
  ?req dct:title ?EvidenceTitle.
  ?task ex:order ?order
  OPTIONAL{?task ex:time ?t.
           ?t rdfs:label ?time}
}ORDER BY ?order
PREFIX ex: <http://www.exampe.org/>
PREFIX dct: <http://purl.org/dc/terms/>
SELECT ?order, ?actorTask, ?recipientTask, ?time
WHERE{
  ex:EnrollmentService ex:involveTask ?task.
  OPTIONAL{?task ex:actor ex:headTeacherPrimarySchool.
           ?task dct:description ?actorTask.
           ?task ex:order ?order}
  OPTIONAL{?task ex:recipient ex:headTeacherPrimarySchool.
           ?task dct:description ?recipientTask.
           ?task ex:order ?order}
  OPTIONAL{?task ex:time ?t.
           ?t rdfs:label ?time}.
}ORDER BY ?order

```

Figure 6: SPARQL queries based on the RDF code.

order	taskDescription	EvidenceTitle	time
2	"Submit a certification of permanent residence"	"Formal declaration of permanent residence"	"December"@en
2	"Submit a certification of permanent residence"	"Proof of permanent residence"	"December"@en
11	"Submit documentation to complete enrollment"	"Formal declaration about the legal exercise of guardianship"	
11	"Submit documentation to complete enrollment"	"Certificate of the municipality in which the student is registered"	
11	"Submit documentation to complete enrollment"	"Photocopy of the student's identity card"	

Figure 7: Output of SPARQL left

order	actorTask	recipientTask	time
2		"Submit a certification of permanent residence"	"December"@en
3	"Create a list of the student addresses."		
4	"Send the list"		
8	"Inform the students' guardians."		
9	"Send the school-leaving certificates of students."		

Figure 8: Output of SPARQL right.

5. Conclusions and discussion

The paper discusses the challenges that have arisen following the introduction of the National Registry of Administrative Procedures "MITOS". Although the registry was designed to standardize administrative processes, improve transparency, and provide accurate information on public administration operations, difficulties have arisen in accurately documenting processes, leading to structural variations and inconsistent content. To tackle these challenges, the paper proposes an approach based on establishing a common reference infrastructure of standardized metadata and metaprocesses.

As a case study, we examine the administrative process of enrolling students in their first year of high school. Our approach appears to address the challenges related to both the structure and the content of the current process documentation. Specifically, utilizing metamodels like TOGAF it enables public servants to comprehensively extract crucial information from legal texts, including actors and steps, with no omissions. Furthermore, by employing language models like ChatGPT, it improves the completeness of the extracted information. Moreover, the establishment of standardized metadata ensures accurate process documentation and minimizes inconsistencies. Similarly, outlining the hierarchical structure of the process and standardizing its steps creates a precise structure for capturing, naming, and situating processes within the complex web of administrative processes in the field of education. Finally, the suggested approach emphasizes the practical and domain-specific application of generic core vocabularies such as CPSV, thereby facilitating the implementation of public processes.

In terms of practical usefulness, our proposed approach can enhance the quality of information published in "MITOS" and reduce the resources needed for manually studying and annotating legal texts. As only a limited number of civil servants are mandated to document and publish administrative processes, training them using models such as TOGAF and ChatGPT is quite feasible. Furthermore, the gradual addition of metadata and metaprocesses to the common reference infrastructure will provide them with standard elements for documenting processes. This will help in publishing reliable information based on common metadata and metaprocesses, contributing to a unified picture of public administration and making it easier for citizens to find the information they need. Additionally, publishing information in a machine-readable format can lay the groundwork for creating applications that aid public administration in its work, such as identifying and streamlining steps and processes, and providing tools for citizens to extract information quickly and reliably, for example, by submitting queries to the SPARQL endpoint.

The challenge is to implement the proposed method on a large scale within the domain of public administration. Since there are a large number of public agencies, specific organizational actions should be taken to facilitate communication, negotiation, and the establishment of commonly accepted metadata and metaprocesses among the involved parties. It's important to note that we cannot accurately predict the challenges and threats of applying our approach on a large scale because our proposal has not yet been implemented in a real environment and on a large sample, which is a limitation of our study. As future work, we plan to extend our approach to a broader array of administrative processes in education. This will facilitate the development of more complex controlled vocabularies, such as a thesaurus, which will serve as the foundation for creating an ontology-based business process metamodel for the domain of education management. Our objective is for this model to encompass the entire spectrum of administrative processes in education, enabling the creation of low-level business process models and supporting automated analysis, decision-making, and digitalization. Lastly, we will explore the possibility of applying the developed ontology-based model to other areas of public sector.

References

- [1] B. A. P. Nguyen and H. Scholta, A Method for the Collaborative and Semi-automated Generation of Conceptual Models from Legal Regulations in Public Organizations, in: M. R.

- Johannessen, C. Csáki, L. Danneels, S. Hofmann, T. Lampoltshammer, P. Parycek, G. Schwabe, E. Tambouris, and J. Ubacht (Eds.), *Electronic Participation*, Springer Nature Switzerland, 2024, pp. 194–208. doi: 10.1007/978-3-031-70804-6_13.
- [2] K. Hinkelmann, B. Thönssen, and D. Wolff, Ontologies for E-government, in: R. Poli, M. Healy, and A. Kameas (Eds.), *Theory and Applications of Ontology: Computer Applications*, Springer, Dordrecht, 2010, pp. 429–462. doi: 10.1007/978-90-481-8847-5_19.
 - [3] M. Mauch, S. Bachinger, P. Bornheimer, S. Breidenbach, D. Ehrhardt, L. Feddoul, H. Legner, F. Löffler, Fr. Löffler, M. Raupach, S. Schindler, J. Schröder, B. König-Ries, From Legal Texts to Digitized Services for Public Administration. In: *Language Models: Legal Parrots or more? Proceedings of the 27th International Legal Informatics Symposium IRIS 2024*, Salzburg, Österreich.
 - [4] P. Harpring, *Introduction to Controlled Vocabularies: Terminology for Art, Architecture, and Other Cultural Works*. Getty Publications, 2010.
 - [5] CEDS, Common Education Data Standards (CEDS). URL: <https://ceds.ed.gov/Default.aspx>
 - [6] Z. Mohseni, I. Masiello, R.M. Martins, A technical infrastructure for primary education data that contributes to data standardization. *Educ Inf Technol* (2024). <https://doi.org/10.1007/s10639-024-12683-2>.
 - [7] APQC, APQC Process Classification Framework (PCF) - Cross-Industry - Excel Version 7.4, 2024. URL: <https://www.apqc.org/resource-library/resource-listing/apqc-process-classification-framework-pcf-cross-industry-excel-11>
 - [8] K. Hinkelmann, A. Gerber, D. Karagiannis, B. Thoenssen, A. van der Merwe, and R. Woitsch, A new paradigm for the continuous alignment of business and IT: Combining enterprise architecture modelling and enterprise ontology, *Computers in Industry*, vol. 79, pp. 77–86, Jun. 2016, doi: 10.1016/j.compind.2015.07.009.
 - [9] A. Deguchi *et al.*, “What Is Society 5.0?,” in *Society 5.0: A People-centric Super-smart Society*, Singapore: Springer, 2020, pp. 1–23. doi: 10.1007/978-981-15-2989-4_1.
 - [10] TOGAF, The TOGAF Standard, Version 9.2 - Content Metamodel. URL: <https://pubs.opengroup.org/architecture/togaf9-doc/arch/chap30.html>
 - [11] D. C. Hay, *Data Model Patterns: Conventions of Thought*. Dorset House Pub., 1996.
 - [12] European Commission, The ESCO ontology. URL: <https://ec.europa.eu/esco/lod/static/model.html>
 - [13] Princeton University, WordNet: A Lexical Database for English. URL: <https://wordnet.princeton.edu/>
 - [14] European Commission, Core Public Organisation Vocabulary. URL: <https://joinup.ec.europa.eu/collection/semantic-interoperability-community-semic/solution/e-government-core-vocabularies/core-public-organisation-vocabulary>
 - [15] European Commission, Core Person Vocabulary. URL: <https://joinup.ec.europa.eu/collection/semantic-interoperability-community-semic/solution/core-person-vocabulary/release/100>
 - [16] L.C. Pouchard, A.F. Cutting-Decelle, J.J. Michel, M. Grüninger, ISO 18629 PSL: A standardised language for specifying and exchanging process information. *IFAC Proceedings Volumes* 38.1 (2005): 37-45., doi: 10.3182/20050703-6-CZ-1902.01525.
 - [17] W. R. van Hage, V. Malaisé, R. Segers, L. Hollink, and G. Schreiber, “Design and use of the Simple Event Model (SEM),” *Journal of Web Semantics*, vol. 9, no. 2, pp. 128–136, Jul. 2011, doi: 10.1016/j.websem.2011.03.003.
 - [18] European Commission, Core Criterion and Core Evidence Vocabulary. URL: <https://joinup.ec.europa.eu/collection/semantic-interoperability-community-semic/solution/e-government-core-vocabularies/core-criterion-and-core-evidence-vocabulary>
 - [19] European Commission, Core Public Service Vocabulary. URL: <https://joinup.ec.europa.eu/collection/semantic-interoperability-community-semic/solution/core-public-service-vocabulary>