Semantic Digital Libraries in Public Administration: A Knowledge Graph Approach to Certificate Request Management

Valentina Albano[1], Giovanni Carau[2], Donatella Firmani[3,§], Elio Gullo[4], Claudia Ilardi[5] and Luigi Laura[6,*]

[1]Dipartimento della Funzione Pubblica
[2]Università Telematica Internazionale Uninettuno
[3]Sapienza Università di Roma
[4]Dipartimento della Funzione Pubblica
[5]Formez
[6]Università Telematica Internazionale Uninettuno

Abstract

Digital libraries in public administration often struggle with complex, paper-based bureaucratic procedures for certificate requests. This study presents a semantic digital library approach that leverages knowledge graphs to simplify and automate these processes. The research, conducted as part of an exploratory project for the Department of Public Function [1], demonstrates how semantic technologies can transform traditional document management in Public Administration (PA) [2].

The semantic foundation of our digital library is built upon a custom ontology [3] that defines a comprehensive taxonomy of certificates, required documents, and their relationships within the public administration domain. Our research shows that this semantic digital library approach is not only easily adoptable but also naturally extensible, offering a promising direction for modernizing public service delivery and reducing paper-based workflows.

The implementation follows the Resource Description Framework (RDF) model, organizing the digital library's resources through subject-predicate-object triples. This semantic structure enables rich relationships between digital resources and facilitates intuitive navigation through certificate-related data and documents based on their semantic connections.

Looking ahead, this semantic digital library framework has the potential to evolve into a Linked Open Data (LOD) ecosystem [4], enabling seamless information exchange between different public administration entities. This evolution would create an interconnected network of digital libraries, further streamlining certificate request procedures and document validation processes.

Keywords

Semantic Digital Libraries, Knowledge Graphs, Public Administration Services, Certificate Management Systems

1. Introduction

Public Administrations around the world manage vast digital libraries of citizen documents, certificates, and administrative records that are fundamental for daily civic life. However, these digital repositories often evolve as decentralized silos, creating a fragmented and chaotic information landscape that citizens must navigate to access essential services. From birth certificates to business permits, from tax documents to educational credentials, citizens frequently find themselves lost in a maze of disconnected

IRCDL 2025: 21st Conference on Information and Research Science Connecting to Digital and Library Science, February 20-21 2025, Udine, Italy

*Corresponding author.

§The work of Donatella Firmani has been partially supported by HORIZON Research and Innovation Action 101135576 INTEND "Intent-based data operation in the computing continuum".

v.albano@governo.it (V. Albano); giovanni.carau@gmail.com (G. Carau); donatella.firmani@uniroma1.it (D. Firmani); e.gullo@governo.it (E. Gullo); cilardi@formez.it (C. Ilardi); luigi.laura@uninettunouniversity.net (L. Laura)
0000-0003-0358-3208 (D. Firmani); 0000-0001-6880-8477 (L. Laura)

systems, redundant requests, and unclear procedures. This fragmentation not only frustrates citizens, but also increases administrative overhead and reduces the efficiency of public services.

The situation is particularly challenging in Italy, where the complex interplay between national, regional and municipal administrations creates additional layers of complexity in document management and service delivery. In this work, we propose a semantic approach based on knowledge graphs to bring order to this chaos, offering a unified and intuitive way to organize and access public administration's digital libraries. By implementing a structured knowledge representation of certificates and their requirements, we aim to transform the current scattered landscape into an interconnected, easily navigable system that better serves both citizens and public administrators. In the digital era where we live, the transformation of public administration represents a top priority. In Italy, this transformation is guided and regulated by the Digital Administration Code (CAD) [5], established by the legislative decree 82/2005 [6].

This process has required a profound reform in the culture of public administration, transforming it from a self-referential entity that might seem closed and hostile to citizens, into a model focused on service provision, open to citizen participatory instances, with a clear focus on efficiency, effectiveness, transparency, and legality goals [7].

The procedure of citizens requesting certificates from public administrations often appears as a complex and bureaucratic process and therefore represents one of the processes within Public Administrations, a candidate for digital transformation.

In the context of this experimental project, the digital transformation process for the citizen's certificate request process was explored in detail, focusing on the principles, challenges, and opportunities linked to its digitalization.

In the initial analysis phase, the main actors involved in the management of certificate requests were identified as follows:

1. The Citizen: the end user who, after authenticating themselves with their digital identity on the reference Public Administration portal, wishes to submit a certificate request by completing a specific request form.
2. The Reference Public Administration: represents the administration responsible for providing the certificate requested by the citizen. The Reference Public Administration exposes its defined catalog of certificates, dynamically managing the information associated with each certificate request, including required attachment documents.
3. Cooperating Public Administrations: represent the administrations that interact with the Reference Public Administration and provide data and certificates to the Reference Public Administration itself, information necessary for processing the citizen's certificate request.

The goal was to improve the transparency and efficiency of the entire certificate request processing among the various involved actors: the citizen submitting the request, the Public Administration involved in processing the certificate request for subsequent issuance, and the cooperative Public Administrations responsible for providing mandatory data, documents, and certificates for the certificate issuance.

A graph-based knowledge model accessible to all Public Administrations was proposed, which enables the identification and determination of information. Thus, the future scenario of certificate requests could be further simplified, allowing a Public Administration involved in processing the certificate to retrieve, through an open platform (Linked Open Data) [8], the necessary certificates and information held by other public administrations, avoiding the citizen from the laborious process of acquiring and retrieving the required certificates, resulting in increased efficiency in terms of time and streamlining the bureaucratic process.

The main objective of this transformation has been to overcome this sluggishness, making the administration more efficient, transparent, and service-oriented, both towards citizens, businesses, and other public institutions.

In the context of certificate requests, a key aspect is the Semantic Web [9], which indicates the transition from a Web of Contents understandable only by humans to an environment of data interpreted directly by machines, allowing the automation of complex tasks for users.

The main technologies involved include

1. RDF (Resource Description Framework) [10]: A structure using identifiers (URI) to identify resources and connect data in subject-predicate-object triples.
2. Linked Data [4]: A standard for representing and accessing data on the web, aiming to make data interconnected and accessible using URIs to identify resources, allowing the publication of these URIs via HTTP to access resources, and ultimately using links between resources to discover additional information from other sources.

2. Related Works

In recent years, there has been significant interest in applying semantic technologies to digital libraries and public administration. For instance, Haslhofer et al. discuss the role of knowledge graphs in libraries and digital humanities, highlighting their potential to enhance data integration and retrieval [11]. Similarly, Ebeid and Pierce introduce MedGraph, an experimental semantic information retrieval method using knowledge graph embedding for biomedical citations indexed in PubMed [12]. Xu et al. present PubMed Knowledge Graph 2.0, which connects papers, patents, and clinical trials in biomedical science, demonstrating the utility of knowledge graphs in integrating diverse data sources [13].

In the context of digital libraries, Ferilli and Redavid propose an ontology and knowledge graph infrastructure for knowledge representation, emphasizing the importance of structured semantic frameworks in enhancing information retrieval [14]. Kruk et al. introduce MarcOnt, an integration ontology for bibliographic description formats, facilitating interoperability between different metadata standards [15]. Soergel explores the intersection of digital libraries and knowledge organization, discussing how semantic technologies can improve information access and management [16].

Limani et al. discuss the development of a Scholarly Artifacts Knowledge Graph, outlining use cases for digital libraries and demonstrating how knowledge graphs can support advanced scholarly communication services [17]. Ferro and Crestani provide an overview of digital libraries, focusing on quality information provision and the role of semantic technologies in achieving this goal [18].

These studies collectively underscore the transformative potential of semantic technologies and knowledge graphs in enhancing the functionality and interoperability of digital libraries and public administration systems.

3. A semantic digital library for certificate request management in public administration

In this section we present our results; we first briefly recall the methodologies and frameworks used, then we explain the process of the creation (and manipulation) of the Knowledge Graph; finally we provide some examples.

3.1. Methodologies and Frameworks

The aim of this phase was to analyze a knowledge graph model to simplify and automate the request process. The activities included:

1. An analysis of the process aimed at designing an advanced, configurable, and maintainable Knowledge Graph, making it accessible to all, reducing the costs and time of certificate request and issuance.
2. The evolution of the graph towards cooperative participation among Public Administrations via Linked Open Data was also envisioned.

During the experimental study, several technologies and tools were evaluated, focusing on:

1. Ontology and OWL Language [3]: Tools used to model entities and relationships within the knowledge graph.
2. RDF Framework [10]: Employed to represent structured information within the graph through resource identifiers or URIs, processed in subject-predicate-object triples.
3. Frameworks and libraries for managing RDF-based graphs: Evaluation and comparison among Apache Jena [19], RDFLib [20], and dotNetRdf [21].
4. Frameworks for rendering Knowledge Graphs: Various options were examined, both for network visualization only (Gephi [22], Graphviz [23], Cytoscape [24], Pajek [25], SocNetV [26]) and for graph rendering and analysis (python-igraph [27], NetworkX [28], visNetwork [29], JuliaGraphs [30], and Pyvis [31]).
5. Linked Data [32] and Digital Transformation: Analysis of the application of Linked Data principles to ensure data interoperability among different Public Administrations.

The selection of technologies was based on criteria such as flexibility, efficiency, and adherence to semantic web standards (RDF, RDFS, OWL, etc). The criteria for choosing frameworks and technologies were applied in the following subprocesses of the overall process:

1. Definition and Publication of Ontology: Using an ontology that defines specific concepts, entities, and relationships for the context under examination.
2. Adherence to the RDF Model: Easy implementation of a knowledge graph based on RDF to handle triples (subject, predicate, object).
3. File Configuration Ease: Allowing PA operators to compile files with rules and configurations to build a subgraph for a certificate request.
4. Easy Graph Loading and Maintenance: Enabling easy construction, extension, and maintenance of the graph.
5. Customization of Partitioning Criteria and Data Visualization: Providing capabilities to customize visualization based on partitioning criteria and application of metrics.
6. Interactive Graph Visualization: Offering interactive visualization with specific criteria and filters.
7. Open and Collaborative Evolution: Facilitating easy evolution following the Linked Open Data approach.

Based on these criteria, the following frameworks were selected:

1. WebVOWL [33]: For ontology creation.
2. OnOntology [34]: For ontology publication.
3. Excel Template: For manually preparing the dataset.
4. RdfLib (Python) [20]: For loading the Knowledge Graph.
5. NetworkX [28]: For data navigation and application of partitioning criteria.
6. Pyvis [31]: For graph rendering and user interaction.

3.2. Building the Knowledge Graph

The adopted approach divides the construction of the knowledge graph into specialized subprocesses, utilizing dedicated frameworks for specific functions, enhancing efficiency compared to a centralized approach such as Apache Jena [19].
The ontology definition begins with the identification of relevant resources and entities, along with their relationships and attributes. With the help of the WebVOWL framework [33], the primary OWL classes associated with the context entities were identified, such as:

- **Entity (Ente)**: can be specialized into Public Administration (PA) and Private Entity (entity name = Ente Privato), representing entities issuing certificates.
- **Certificate Request (entity name = Richiesta di Certificato)**: a central element linked to a specific PA.

- **Certificate (entity name = Certificato)**: an entity representing the result of managing certificate requests, which can also be specialized as "Certificate for Non-EU Citizens" (entity name= “Certificato per ExtraComunitario”) and "Certificate for Minors" (entity name= “Certificato per Minorenne”) to handle additional information associated with specific categories of subjects.
- **Requesting Group (entity name = Gruppo Richiedenti)**: a group of people involved in the certificate request.
- **Requester (entity name = Richiedente)**: information associated with individuals involved in the certificate request.
- **Company (entity name = Azienda)**: in cases where the requester owns a company, and the certificate requests concern the company.
- **Property (entity name = Immobile)**: if the requester owns property, and the certificate requests concern property ownership.
- **Address (entity name = Indirizzo)**: specified as Domicile (entity name = Domicilio), Residence (entity name = Residenza), Company Address (entity name = Indirizzo Azienda), Entity (entity name = Ente), and Property(entity name = Immobile), depending on the type of entity involved.

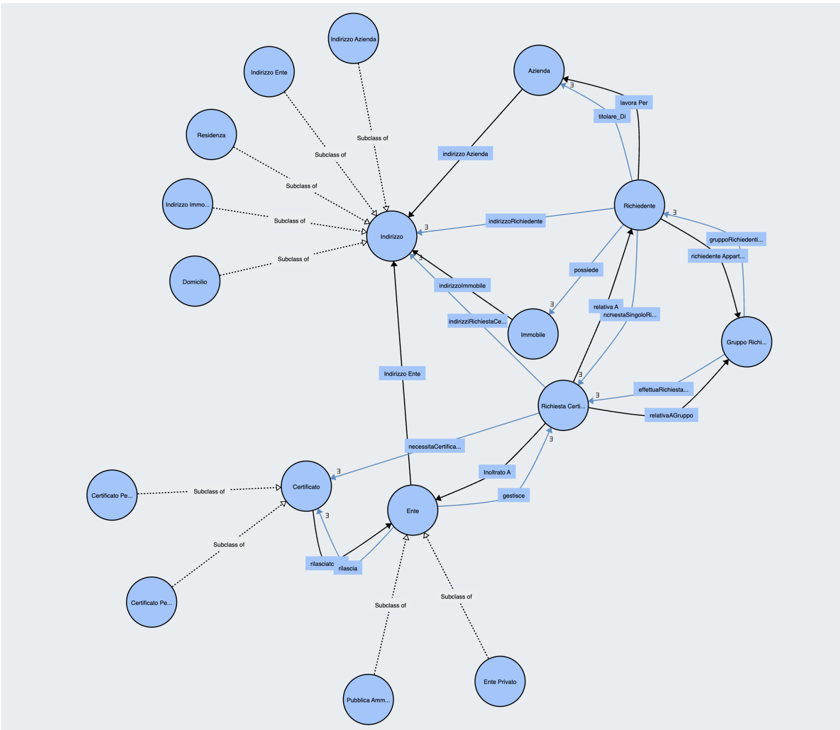


Figure 1: WebVOWL: ontology definition for certificate request management.

In navigating the graph, three elements were used in accordance with the RDF triple system:

1. **Subject**: the resource from which the arc originates (represented by owl:Class).
2. **Predicate**: the property labeled on the arc (represented by owl:ObjectProperty).
3. **Object**: the resource/entity (represented by owl:Class) or the literal value pointed to by the arc (rdf:Literal).

Observing the ontology shown in Fig.1, we highlight two possible scenarios regarding navigation between the "Public Administration" (entity name = Pubblica Amministrazione) and "Certificate Request" (entity name = Richiesta Certificato) entities obtained through triple management (subject, predicate, object) in an OWL/RDF perspective:

1. The "Public Administration" entity (entity name = Pubblica Amministrazione) represents the subject from which the arc originates; the predicate labeled "manages" (predicate name = gestisce) identifies the arc pointing to the "Certificate Request" entity (entity name = Richiesta Certificato) as the object.
2. The "Certificate Request" (entity name = Richiesta Certificato) entity represents the subject from which the arc originates; the predicate labeled "forwarded To" (predicate name = inoltrato A) identifies the arc pointing to the "Public Administration" entity (entity name = Pubblica Amministrazione) as the object.

In the case of considering the first option, a second level of classification will be defined, which would describe the triple as "a Public Administration manages one or more Certificate Requests."

Hence, by navigating the graph in this direction, we would first have an initial selection of the PA (first level of classification or filter), and subsequently, we could identify all the Certificate Requests related to that PA and select one (second level of classification/filter). At this point, the navigation, once a specific Certificate Request is selected, would proceed to identify only the entities necessary for processing the Certificate Request.

Within the context of the involved entities, the "Requester" (entity name = Richiedente) represents the individual for whom a certificate is requested. Using RDF triples, starting from the "Certificate Request" (entity name = Richiesta Certificato) entity as the subject, two scenarios can be delineated leading from the central entity to the related "Requester" (entity name = Richiedente) entity:

- Certificate for an individual: Here, the connection between the entities is defined by a predicate called "related To" (predicate name = relativa A), represented by a direct link from the "Certificate Request" (entity name = Richiesta Certificato) to the "Requester" (entity name = Richiedente).
- Certificate for a group of individuals (e.g., adoption certificate, divorce, cessation of cohabitation). In this case, the path between the entities is mediated by "Requesting_Group" (entity name = Gruppo_Richiedenti), which provides details on the relationship and the number of requesters involved in the request. This involves two sequential steps:
  - Link between "Certificate Request" (entity name = Richiesta Certificato) and "Requesting Group" (entity name = Gruppo_Richiedenti) using the predicate "relatedToGroup" (predicate name = relativaAGruppo).
  - Link between "Requesting Group" (entity name = GruppoRichiedenti) and "Requesters" (entity name = Richiedente) through the predicate "groupIncludesRequesters" (predicate name = Gruppo_Richiedenti).

The combination of these two navigations will be subsequently managed by a dynamic rules engine embedded in a Rule Engine. Within the "Requester" (entity name = Richiedente) entity, all relevant attributes are included. During dataset loading, the Rule Engine will select the appropriate attributes solely based on the specific Certificate Request (entity name = Richiesta Certificato), thereby reducing the set of attributes to only those relevant to that specific "Requester" (entity name = Richiedente) entity.

After converting the ontology into serialized Turtle (TTL) format [35], the publication will be carried out using the OnToology website, following a specific procedure that allows the serialized OWL file, representing the ontology, to be published on a specific URI [36] as depicted in Fig. 2.

During the Data Set Preparation phase, Excel sheets are prepared with essential rules to feed the knowledge graph, in anticipation of the subsequent dataset loading through the Rule Engine integrated in RDFLib. These Excel sheets, compiled by Public Administration operators, have two distinct types and different objectives:

1. First Excel Sheet: Catalog of Certificate Requests for each Public Administration, as shown in Fig. 3.
2. Second Excel Sheet: Configuration of attributes of significant entities in each Certificate Request, with the definition of specific navigation rules, as depicted in Fig. 4.

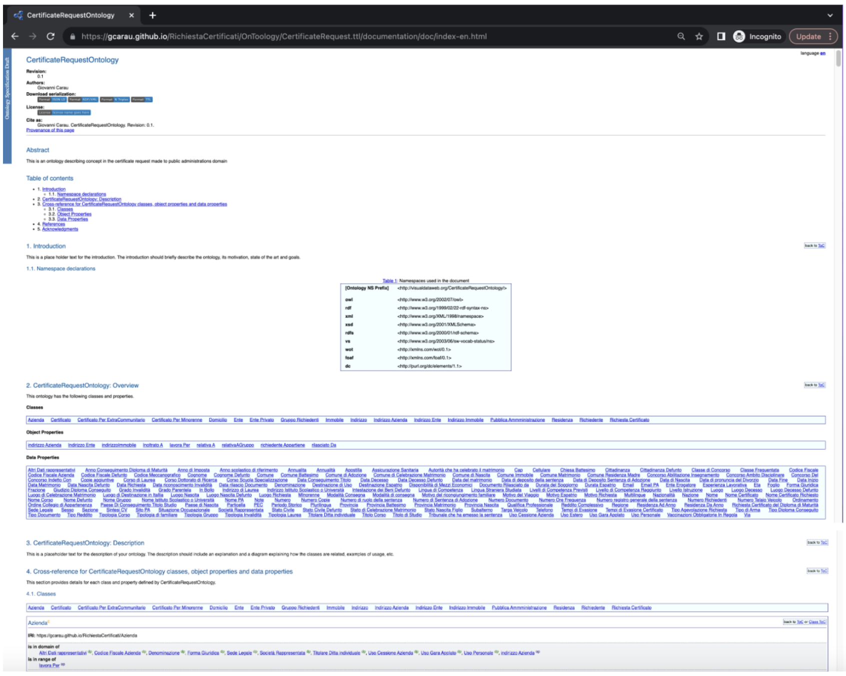


Figure 2: Certificate Request Ontology: publication on a specific URI.

For each certificate request, in the second Excel sheet, relevant entities and attributes are identified, forming a dataset of 106 certificates and 2560 associated entity and attribute elements.
The last step involves converting the two Excel sheets into CSV files, which will constitute the dataset for loading the knowledge graph after any data transformation and cleansing processes.

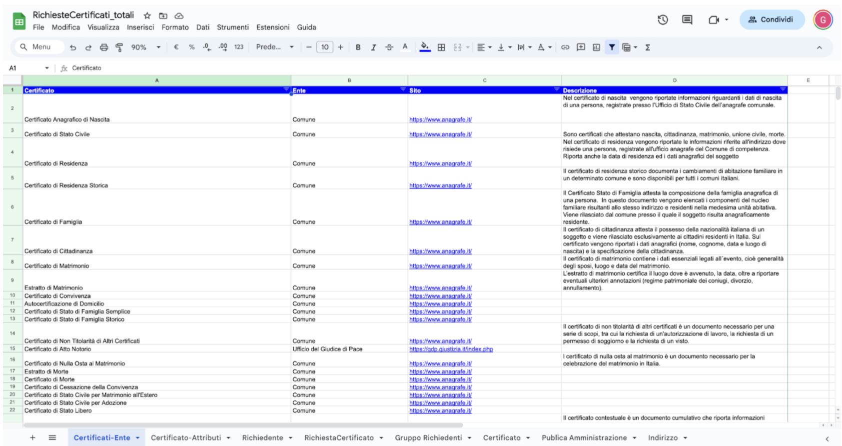


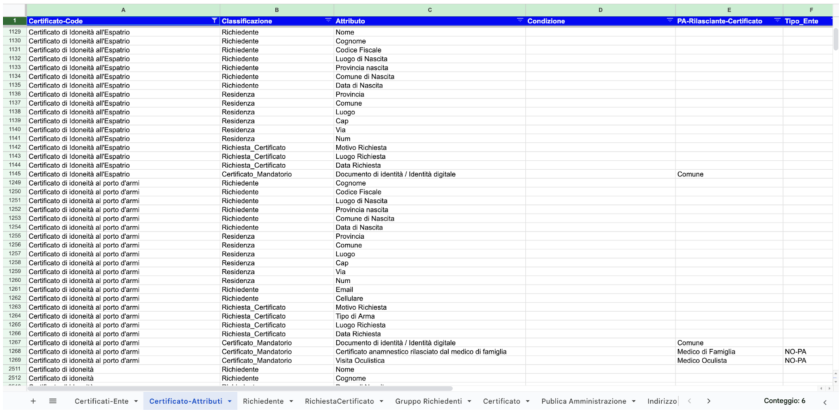
Figure 3: Excel sheet containing the catalog of certificates aggregated by Public Administration.

In the design solution, the choice of an RDF library is crucial to meet technical and business requirements. The RDF framework, based on the triple paradigm (subject, predicate, object), promotes the expansion of the knowledge graph towards Linked Open Data (LOD), encouraging cooperation among Public Administrations and facilitating the publication of structured data on the web.

After an in-depth analysis, three RDF libraries were evaluated:

1. **Apache Jena** [19], in Java.
2. **DotNetRDF** [21], in DotNet (C#).
3. **RDFLib** [20], in Python.

Based on a comparative study conducted by the International Journal on Emerging Technologies



	Certificato-Code	Classificazione	Attributo	Condizione	PA-Rilasciante-Certificato	Tipo_Ente
1129	Certificato di Idoneità all'Espatrio	Richiedente	Nome			
1130	Certificato di Idoneità all'Espatrio	Richiedente	Cognome			
1131	Certificato di Idoneità all'Espatrio	Richiedente	Codice Fiscale			
1132	Certificato di Idoneità all'Espatrio	Richiedente	Luogo di Nascita			
1133	Certificato di Idoneità all'Espatrio	Richiedente	Provincia nascita			
1134	Certificato di Idoneità all'Espatrio	Richiedente	Comune di Nascita			
1135	Certificato di Idoneità all'Espatrio	Richiedente	Data di Nascita			
1136	Certificato di Idoneità all'Espatrio	Residenza	Provincia			
1137	Certificato di Idoneità all'Espatrio	Residenza	Comune			
1138	Certificato di Idoneità all'Espatrio	Residenza	Luogo			
1139	Certificato di Idoneità all'Espatrio	Residenza	Cap			
1140	Certificato di Idoneità all'Espatrio	Residenza	Via			
1141	Certificato di Idoneità all'Espatrio	Residenza	Num			
1142	Certificato di Idoneità all'Espatrio	Richiesta_Certificato	Motivo Richiesta			
1143	Certificato di Idoneità all'Espatrio	Richiesta_Certificato	Luogo Richiesta			
1144	Certificato di Idoneità all'Espatrio	Richiesta_Certificato	Data Richiesta			
1145	Certificato di Idoneità all'Espatrio	Certificato_Mandatorio	Documento di identità / Identità digitale		Comune	
1249	Certificato di idoneità al porto d'armi	Richiedente	Cognome			
1250	Certificato di idoneità al porto d'armi	Richiedente	Codice Fiscale			
1251	Certificato di idoneità al porto d'armi	Richiedente	Luogo di Nascita			
1252	Certificato di idoneità al porto d'armi	Richiedente	Provincia nascita			
1253	Certificato di idoneità al porto d'armi	Richiedente	Comune di Nascita			
1254	Certificato di idoneità al porto d'armi	Richiedente	Data di Nascita			
1255	Certificato di idoneità al porto d'armi	Residenza	Provincia			
1256	Certificato di idoneità al porto d'armi	Residenza	Comune			
1257	Certificato di idoneità al porto d'armi	Residenza	Luogo			
1258	Certificato di idoneità al porto d'armi	Residenza	Cap			
1259	Certificato di idoneità al porto d'armi	Residenza	Via			
1260	Certificato di idoneità al porto d'armi	Residenza	Num			
1261	Certificato di idoneità al porto d'armi	Richiedente	Email			
1262	Certificato di idoneità al porto d'armi	Richiedente	Cellulare			
1263	Certificato di idoneità al porto d'armi	Richiesta_Certificato	Motivo Richiesta			
1264	Certificato di idoneità al porto d'armi	Richiesta_Certificato	Tipo di Arma			
1265	Certificato di idoneità al porto d'armi	Richiesta_Certificato	Luogo Richiesta			
1266	Certificato di idoneità al porto d'armi	Richiesta_Certificato	Data Richiesta			
1267	Certificato di idoneità al porto d'armi	Certificato_Mandatorio	Documento di identità / Identità digitale		Comune	
1268	Certificato di idoneità al porto d'armi	Certificato_Mandatorio	Certificato anamnestico rilasciato dal medico di famiglia		Medico di Famiglia	NO-PA
1269	Certificato di idoneità al porto d'armi	Certificato_Mandatorio	Visita Oculistica		Medico Oculista	NO-PA
2511	Certificato di idoneità	Richiedente	Nome			
2512	Certificato di idoneità	Richiedente	Cognome			

Certificati-Ente · Certificato-Attributi · Richiedente · RichiestaCertificato · Gruppo Richiedenti · Certificato · Publica Amministrazione · Indirizzo · Conteggio: 6

Figure 4: Excel sheet containing the configuration of entities, attributes, and navigation for a specific Certificate Request.

[37], the choice was narrowed down between Apache Jena and RDFLib, excluding DotNetRDF due to performance issues in loading large-scale graphs.

Although Apache Jena offers advantages, particularly in Java, it is not ideal for big data contexts in the data cleaning and data transformation phases. In this sense, the Python platform has proven to be more suitable.

Within this framework, the Python platform was preferred over Java. The RDFLib, NetworkX, and Pyvis frameworks were chosen, each specialized in a specific domain, conforming to an architecture based on the Model-View-Controller (MVC) design pattern [38].

The Model-View-Controller (MVC) pattern implies that the Controller modifies the Model based on user input, and these modifications are reflected in the View.

For each component of the Model-View-Controller pattern, specific roles were assigned to the involved frameworks in the defined architecture:

1. **Controller**: The Controller consists of three distinct Python frameworks, each performing a specific role in the architecture:
   a) **RDFLib**: Loads the graph through the use of the Rule Engine based on the structured dataset.
   b) **NetworkX**: Implements partitioning criteria for graph visualization navigation.
   c) **Pyvis**: Generates an interactive HTML file representing the knowledge graph.
2. **Model**: Each phase (graph loading, traversal and partitioning, user interaction) has unique content and graph representation for the framework involved in that phase.
3. **View**: In the final stage of the process, Pyvis generates an interactive HTML page that serves as the visual interface for the user.

In summary, as shown in Fig. 5, all three frameworks act, in specific process phases, as elements of the controller, performing tasks such as loading new records or receiving user interactions through the view. The activities of each controller element influence both the Model and the View in the MVC architecture.

In response to Python RDFLib's lower performance compared to Apache Jena, the graph traversal, criteria management, and measurements tasks were transferred to NetworkX, known for its excellent performance.

Figure 6 emphasizes the subdivision of the entire process into sub-processes, with reference to the involved frameworks. The last swimlanes highlight the application part, represented by the overall controller, formed by the combination of controllers executed by each framework, namely RDFLib, NetworkX, and Pyvis, which are executed sequentially.

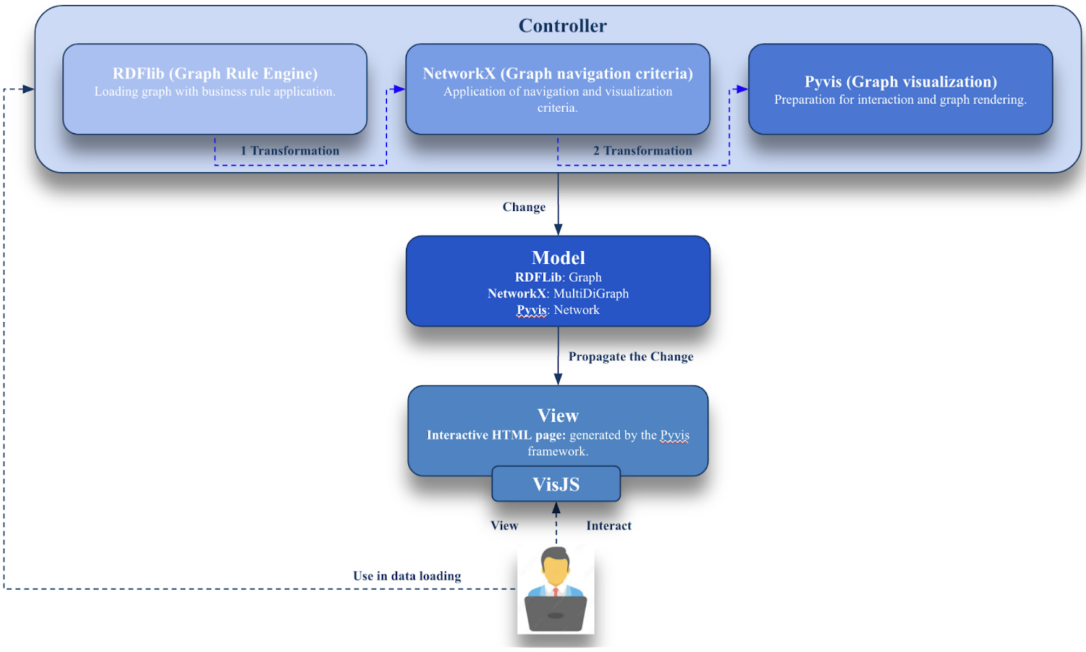


Figure 5: Solution Architecture: Python Frameworks used in the MVC architecture of the solution.

The output of each controller constitutes the input for the next. This architectural solution ensures a better specialization and performance of the controller in fulfilling a specific task (loading, partitioning, preparing for rendering).

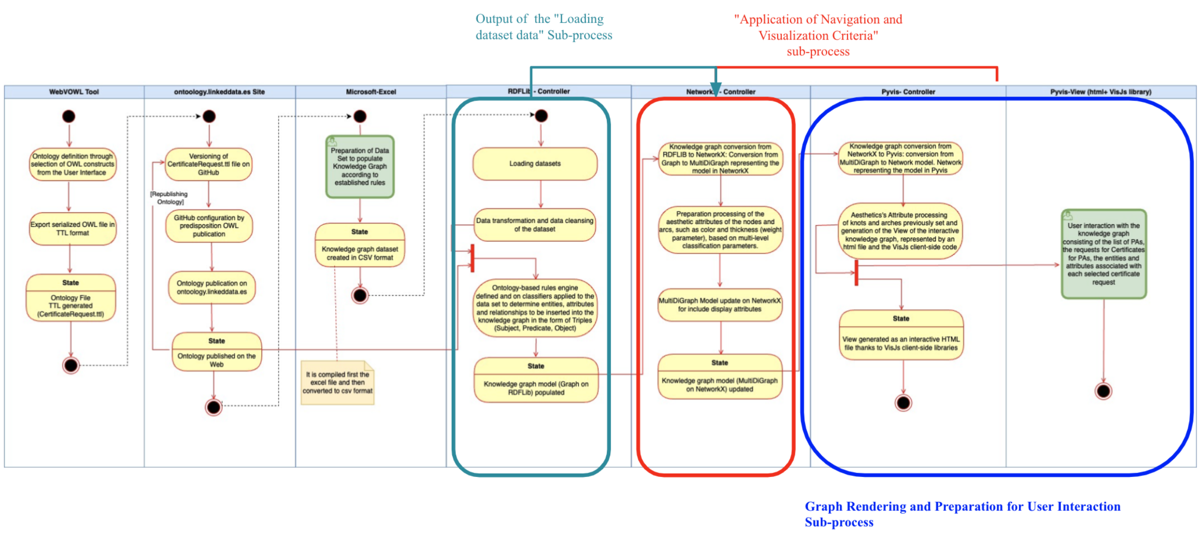


Figure 6: Activity diagram: decomposition of the overall process into subprocesses with reference to the involved frameworks.

In Fig. 7, the graph produced in HTML format is represented, including all the client-side logic (JavaScript) through the VisJS library, enabling user interaction.

Pyvis offers an interactive visualization that allows the use of specific criteria and filters for the selective display of nodes, edges, and attributes of the knowledge graph.

Within the Knowledge Graph, we have the main node (colored blue), representing the container of Public Administrations (level 0), with adjacent colored nodes representing instances of different Public

Administrations (level 1) being addressed.

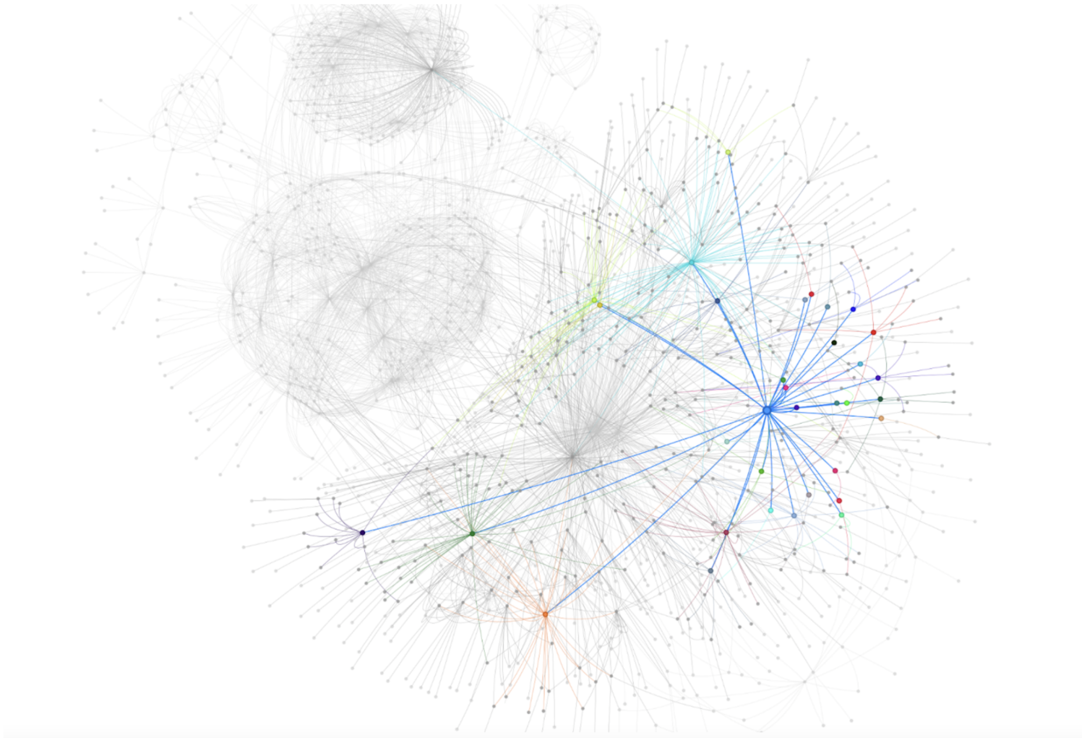


Figure 7: Knowledge Graph: visualization of Public Administrations as colored nodes (1st level classification).

3.3. An example

Let's consider the Use Case of the "Certificato di iscrizione alla Cassa integrazione guadagni" issued by INPS [39] as PA.
Starting from the excel configuration template, shown in Fig. 8, the operation of the Public Administration operator was simulated, configuring the significant entities involved in the "Certificato di iscrizione alla Cassa integrazione guadagni".

A	B	C	D	E	F
	Classificazione	Attributo	Condizione	PA-Rilasciante-Certificato	Tipo_Ente
Certificato di iscrizione alla Cassa integrazione guadagni	Richiedente	Nome			
Certificato di iscrizione alla Cassa integrazione guadagni	Richiedente	Cognome			
Certificato di iscrizione alla Cassa integrazione guadagni	Richiedente	Codice Fiscale			
Certificato di iscrizione alla Cassa integrazione guadagni	Richiedente	Luogo di Nascita			
Certificato di iscrizione alla Cassa integrazione guadagni	Richiedente	Provincia nascita			
Certificato di iscrizione alla Cassa integrazione guadagni	Richiedente	Comune di Nascita			
Certificato di iscrizione alla Cassa integrazione guadagni	Richiedente	Data di Nascita			
Certificato di iscrizione alla Cassa integrazione guadagni	Richiedente	Email			
Certificato di iscrizione alla Cassa integrazione guadagni	Richiedente	Cellulare			
Certificato di iscrizione alla Cassa integrazione guadagni	Residenza	Provincia			
Certificato di iscrizione alla Cassa integrazione guadagni	Residenza	Comune			
Certificato di iscrizione alla Cassa integrazione guadagni	Residenza	Luogo			
Certificato di iscrizione alla Cassa integrazione guadagni	Residenza	Cap			
Certificato di iscrizione alla Cassa integrazione guadagni	Residenza	Via			
Certificato di iscrizione alla Cassa integrazione guadagni	Residenza	Num			
Certificato di iscrizione alla Cassa integrazione guadagni	Azienda	Ragione Sociale Azienda	Per_Impiegato		
Certificato di iscrizione alla Cassa integrazione guadagni	Azienda	Data Assunzione	Per_Impiegato		
Certificato di iscrizione alla Cassa integrazione guadagni	Azienda	Qualifica Professionale	Per_Impiegato		
Certificato di iscrizione alla Cassa integrazione guadagni	Azienda	Livello Inquadramento	Per_Impiegato		
Certificato di iscrizione alla Cassa integrazione guadagni	Azienda	Orario Lavoro	Per_Impiegato		
Certificato di iscrizione alla Cassa integrazione guadagni	Azienda	Retribuzione	Per_Impiegato		
Certificato di iscrizione alla Cassa integrazione guadagni	Azienda	Tipo CIG	Per_Impiegato		
Certificato di iscrizione alla Cassa integrazione guadagni	Azienda	Periodo Fruizione	Per_Impiegato		
Certificato di iscrizione alla Cassa integrazione guadagni	Azienda	Numero Ore CIG Autorizzate	Per_Impiegato		
Certificato di iscrizione alla Cassa integrazione guadagni	Azienda	Importo CIG Giornaliera	Per_Impiegato		
Certificato di iscrizione alla Cassa integrazione guadagni	Richiesta_Certificato	Luogo Richiesta			
Certificato di iscrizione alla Cassa integrazione guadagni	Richiesta_Certificato	Data Richiesta			
Certificato di iscrizione alla Cassa integrazione guadagni	Richiesta_Certificato	Motivo Richiesta			
Certificato di iscrizione alla Cassa integrazione guadagni	Certificato_Mandatorio	Documento di identità / Identità digitale		Comune	
Certificato di iscrizione alla Cassa integrazione guadagni	Certificato_Mandatorio	Codice Fiscale		Agenzia delle Entrate	
Certificato di iscrizione alla Cassa integrazione guadagni	Certificato_Mandatorio	Contratto di Lavoro		Datore di Lavoro	NO-PA
Certificato di iscrizione alla Cassa integrazione guadagni	Certificato_Mandatorio	Domanda di CIG presentata dal Datore di Lavoro		INPS	

Figure 8: Significant Entities and Attributes for the Certificate Request of registration in the Income Support Fund (Richiesta di Certificato di iscrizione alla Cassa integrazione).

The associated entities include: Request_Certificate (entity name = Richiesta_Certificato), Requester (entity name = Richiedente), Mandatory_Certificate (entity name = Certificato_Mandatorio), Residence (entity name = Residenza), and Company (entity name = Azienda), with their respective attributes relevant to the specific Certificate Request (entity name = Richiesta di Certificato).

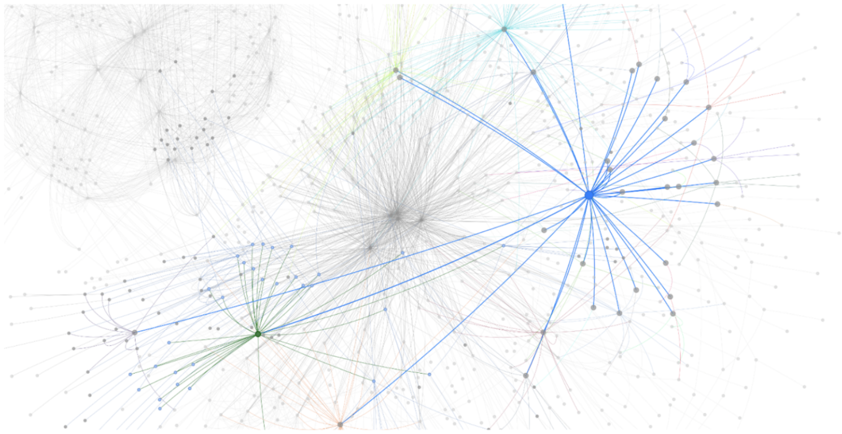


Figure 9: Knowledge Graph: Selection of INPS among the Censused Public Administrations.

Once the graph is populated with the subgraph instance containing the significant information for the "Certificato di iscrizione alla Cassa integrazione guadagni," it is visualized using the Pyvis framework in an interactive HTML page.

Next, we select INPS as an eligible Public Administration and observe the set of possible certificate requests highlighted with sky-blue dots (Fig. 9).

Finally, let's select the "Certificate Request of registration in the Income Support Fund" (certificate name = Richiesta di Certificato di iscrizione alla Cassa integrazione) from the catalog of certificate requests associated with INPS (Fig. 10).

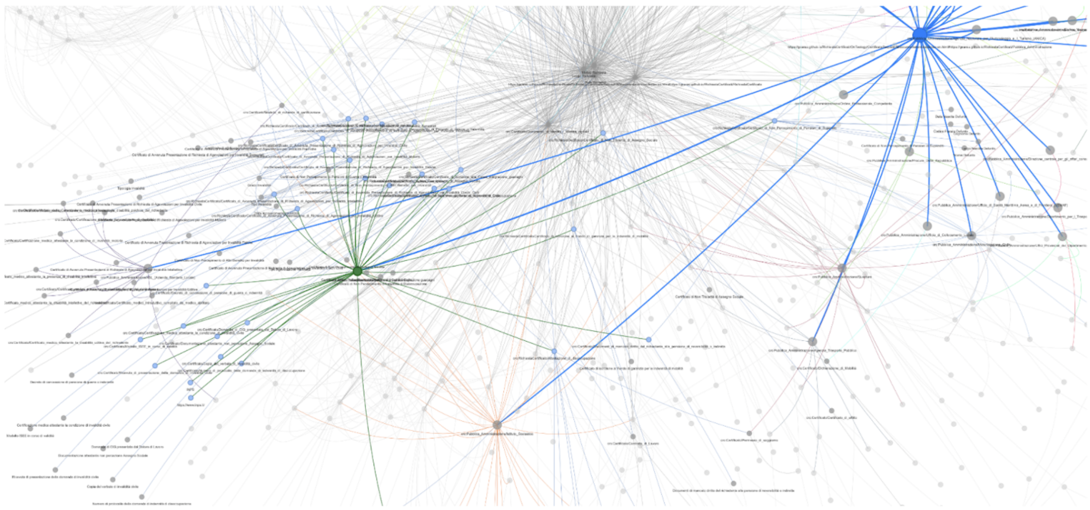


Figure 10: Knowledge Graph: Selection of the Certificate Request for registration in the Income Support Fund (certificate name = Richiesta di Certificato di iscrizione alla Cassa integrazione) from the catalog of associated certificate requests with INPS.

The ultimate goal of this knowledge graph approach is to automate and streamline the complex bureaucratic processes involved in certificate requests within public administration. By leveraging the power of semantic technologies and linked data, this system aims to reduce the administrative burden on citizens and government employees alike. Through intelligent automation and data integration, the knowledge graph enables a more efficient, transparent, and user-friendly experience for those navigating the often-complex landscape of public services.

4. Conclusion

This study has explored a novel semantic digital library approach for certificate request management in public administration, leveraging knowledge graphs to streamline and automate complex bureaucratic processes. By constructing a custom ontology that defines a comprehensive taxonomy of certificates, required documents, and their relationships, we have demonstrated the potential of semantic technologies to transform traditional document management in the public sector.

The knowledge graph approach presented here offers several key advantages. First, it is easily adoptable, as it can be built upon existing digital library infrastructures and document repositories. Second, it is naturally extensible, allowing for the seamless addition of new certificate types, requirements, and relationships as administrative processes evolve. Third, the semantic structure of the knowledge graph enables rich, intuitive navigation and discovery of certificate-related information, reducing the burden on citizens to understand complex bureaucratic procedures.

Looking ahead, we envision this semantic digital library framework evolving into a fully realized Linked Open Data ecosystem. By publishing structured data about certificates and their requirements using RDF and other semantic web standards, public administrations can enable seamless information exchange and interoperability. This would allow the automatic retrieval and validation of required documents across organizational boundaries, further streamlining the certificate request process for citizens.

While challenges remain, particularly in terms of data governance, privacy, and the organizational changes required to fully leverage semantic technologies, the benefits are clear. By embracing knowledge graphs and Linked Open Data, public administrations can not only improve the efficiency and transparency of their operations, but also deliver higher quality, more responsive services to the public. As governments around the world seek to digitally transform and modernize, semantic approaches like the one described in this study offer a promising path forward.

References

[1] Ministro per la pubblica amministrazione., 2023. URL: https://www.funzionepubblica.gov.it/dipartimento-della-funzione-pubblica.

[2] Dipartimento della funzione pubblica., 2023. URL: https://www.funzionepubblica.gov.it/.

[3] Web ontology language (owl)., 2023. URL: https://www.w3.org/OWL/.

[4] Linked open data., 2023. URL: https://www.w3.org/egov/wiki/Linked_Open_Data.

[5] Codice amministrazione digitale., 2023. URL: https://www.agid.gov.it/it/agenzia/strategia-quadro-normativo/codice-amministrazione-digitale.

[6] Decreto legislativo 7 marzo 2005, n. 82., 2023. URL: https://www.bosettiegatti.eu/info/norme/statali/2005_0082.htm.

[7] Legge 7 agosto 1990, n. 241., 2023. URL: https://www.bosettiegatti.eu/info/norme/statali/1990_0241.htm.

[8] T. B. Lee, The semantic web. w3c., 2023. URL: https://www.w3.org/2000/Talks/0906-xmlweb-tbl/text.htm.

[9] T. B. Lee, Semantic web, 2023. URL: https://www.w3.org/2000/Talks/0906-xmlweb-tbl/text.htm.

[10] Resource description framework (rdf)., 2023. URL: https://www.w3.org/RDF/.

[11] B. Haslhofer, A. Isaac, R. Simon, Knowledge graphs in the libraries and digital humanities domain, Encyclopedia of Big Data Technologies (2018) 1–8. URL: https://link.springer.com/referenceworkentry/10.1007/978-3-319-63962-8_291-1. doi:10.1007/978-3-319-63962-8_291-1.

[12] I. A. Ebeid, E. Pierce, Medgraph: An experimental semantic information retrieval method using knowledge graph embedding for the biomedical citations indexed in pubmed, arXiv preprint arXiv:2112.06348 (2021). URL: https://arxiv.org/abs/2112.06348. doi:10.48550/arXiv.2112.06348.

[13] J. Xu, C. Yu, J. Xu, Y. Ding, V. I. Torvik, J. Kang, M. Sung, M. Song, Pubmed knowledge graph 2.0: Connecting papers, patents, and clinical trials in biomedical science, arXiv preprint arXiv:2410.07969 (2024). URL: https://arxiv.org/abs/2410.07969. doi:10.48550/arXiv.2410.07969.
[14] S. Ferilli, D. Redavid, An ontology and knowledge graph infrastructure for digital library knowledge representation, in: Proceedings of the 1st International Workshop on Open Data and Ontologies for Cultural Heritage (ODOCH-2019), volume 2375, 2020, pp. 49–60. URL: https://www.researchgate.net/publication/338744232_An_Ontology_and_Knowledge_Graph_Infrastructure_for_Digital_Library_Knowledge_Representation.
[15] S. R. Kruk, M. Synak, K. Zimmermann, Marcont–integration ontology for bibliographic description formats, in: Proceedings of the International Conference on Dublin Core and Metadata Applications, 2005, pp. 231–234. URL: https://dcpapers.dublincore.org/pubs/article/download/829/825.
[16] D. Soergel, Digital libraries and knowledge organization, in: Semantic Digital Libraries, Springer, 2009, pp. 9–39. URL: https://link.springer.com/chapter/10.1007/978-3-540-85434-0_2. doi:10.1007/978-3-540-85434-0_2.
[17] I. Limani, L. Lahti, H. Toivonen, Scholarly artifacts knowledge graph: Use cases for digital libraries, in: Metadata and Semantic Research: 15th International Conference, MTSR 2021, Virtual Event, November 29–December 3, 2021, Revised Selected Papers, Springer, 2021, pp. 3–14. URL: https://researchr.org/publication/LimaniLT21-0. doi:10.1007/978-3-030-98879-3_1.
[18] N. Ferro, F. Crestani, Digital Libraries: Providing Quality Information, volume 9469 of *Lecture Notes in Computer Science*, Springer, 2015. URL: https://books.google.com/books/about/Digital_Libraries_Providing_Quality_Info.html?id=3BcpCwAAQBAJ. doi:10.1007/978-3-319-27974-9.
[19] Apache jena - home., 2023. URL: https://jena.apache.org/.
[20] Getting started with rdflib., 2023. URL: https://rdflib.readthedocs.io/en/stable/gettingstarted.html.
[21] dotnetrdf, 2023. URL: https://dotnetrdf.org/.
[22] Gephi - the open graph viz platform., 2023. URL: https://gephi.org/.
[23] Graphviz., 2023. URL: https://graphviz.org/.
[24] What is cytoscape?, 2023. URL: https://cytoscape.org/what_is_cytoscape.html#section-image1.
[25] Pajek., 2023. URL: http://vlado.fmf.uni-lj.si/pub/networks/pajek/doc/pajekman.pdf.
[26] Socnetv (social network visualizer)., 2023. URL: https://socnetv.org/.
[27] igraph 0.10.8 documentation., 2023. URL: https://python.igraph.org/en/0.10.8/.
[28] Tutorial — networkx 3.2 documentation., 2023. URL: https://networkx.org/documentation/stable/tutorial.html.
[29] Introduction visnetwork., 2023. URL: https://datastorm-open.github.io/visNetwork/.
[30] Graphs.jl., 2023. URL: https://juliagraphs.org/Graphs.jl/stable/.
[31] Documentation — pyvis 0.1.3.1 documentation., 2023. URL: https://pyvis.readthedocs.io/en/latest/documentation.html.
[32] Linkeddata., 2023. URL: https://www.w3.org/wiki/LinkedData.
[33] Webvowl (web-based visualization of ontologies)., 2023. URL: https://service.tib.eu/webvowl/.
[34] Using ontoology: Tutorial., 2023. URL: https://ontoology.linkeddata.es/tutorial.
[35] Rdf 1.2 turtle., 2023. URL: https://www.w3.org/TR/rdf12-turtle/.
[36] Universal resource identifiers in www., 2023. URL: https://www.w3.org/Addressing/URL/uri-spec.html.
[37] M. A. Bamboat, (pdf) performance of rdf library of java, c and python on large rdf models [scopus index y category journal]., 2021. URL: https://www.researchgate.net/publication/349483361_Performance_of_RDF_Library_of_Java_C_and_Python_on_Large_RDF_Models_Scopus_index_Y_Category_Journal.
[38] Model-view-controller., 2023. URL: https://en.wikipedia.org/wiki/ModelâĂŞviewâĂŞcontroller.
[39] Portale inps - home., 2023. URL: https://www.inps.it/.