NATIONAL AND KAPODISTRIAN UNIVERSITY OF ATHENS

SCHOOL OF SCIENCE
DEPARTMENT OF INFORMATICS AND TELECOMMUNICATIONS

POSTGRADUATE STUDIES
“COMPUTER SYSTEMS: SOFTWARE AND HARDWARE”

MASTER THESIS

Modeling and Preserving Greek Government Decisions using Semantic Web Technologies and Permissionless Blockchains

Themistoklis N. Beris

Supervisor: **Manolis Koubarakis,** Professor

ATHENS

FEBRUARY 2018

ΕΘΝΙΚΟ ΚΑΙ ΚΑΠΟΔΙΣΤΡΙΑΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

ΣΧΟΛΗ ΘΕΤΙΚΩΝ ΕΠΙΣΤΗΜΩΝ
ΤΜΗΜΑ ΠΛΗΡΟΦΟΡΙΚΗΣ ΚΑΙ ΤΗΛΕΠΙΚΟΙΝΩΝΙΩΝ

ΜΕΤΑΠΤΥΧΙΑΚΟ ΠΡΟΓΡΑΜΜΑ
“ΥΠΟΛΟΓΙΣΤΙΚΑ ΣΥΣΤΗΜΑΤΑ: ΛΟΓΙΣΜΙΚΟ ΚΑΙ ΥΛΙΚΟ”

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

Μοντελοποίηση και Διατήρηση Ελληνικών Κυβερνητικών Αποφάσεων με χρήση Τεχνολογιών Σημασιολογικού Ιστού και Μπλοκ Αλυσίδων

Θεμιστοκλής Ν. Μπερής

Επιβλέπων: **Μανόλης Κουμπαράκης,** Καθηγητής

ΑΘΗΝΑ

ΦΕΒΡΟΥΑΡΙΟΣ 2018

MASTER THESIS

Modeling and Preserving Greek Government Decisions using Semantic Web Technologies and Permissionless Blockchains

Themistoklis N. Beris

R.N.: M1514

SUPERVISOR: **Manolis Koubarakis,** Professor

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

Μοντελοποίηση και Διατήρηση Ελληνικών Κυβερνητικών Αποφάσεων με χρήση Τεχνολογιών Σημασιολογικού Ιστού και Μπλοκ Αλυσίδων

Θεμιστοκλής Ν. Μπερής

Α.Μ.: Μ1514

ΕΠΙΒΛΕΠΩΝ: Μανόλης Κουμπαράκης, Καθηγητής

ABSTRACT

We present a re-engineering of Diavgeia, the Greek government portal for open and transparent public administration. We study how decisions of Greek government institutions can be modeled using ontologies expressed in OWL and queried using SPARQL. We also discuss how to use the bitcoin blockchain, to enable government decisions to remain immutable. We provide an open source implementation, called DiavgeiaRedefined, that generates and visualizes the decisions inside a web browser, offers a SPARQL endpoint for retrieving and querying these decisions and provides citizens an automated tool for verifying correctness and detecting possible foul play by an adversary. We conclude with experimental results illustrating that our scheme is efficient and feasible.

SUBJECT AREA: Semantic Web, Cryptography

KEYWORDS: linked open data, blockchain, open government, semantic web, bitcoin

ΠΕΡΙΛΗΨΗ

Σε αυτή τη διπλωματική εργασία παρουσιάζουμε έναν ανασχεδιασμό της Διαύγειας, του ελληνικού κυβερνητικού προγράμματος για ανοικτή και διαφανή δημόσια διακυβέρνηση. Μελετάμε τον τρόπο με τον οποίο οι αποφάσεις των δημόσιων φορέων μπορούν να μοντελοποιηθούν χρησιμοποιώντας OWL οντολογίες και εξετάζουμε τον τρόπο με τον οποίο μπορούν να τεθούν SPARQL ερωτήματα πάνω σε αυτές. Με τη χρήση του bitcoin blockchain, αναγκάζουμε τις κυβερνητικές αποφάσεις να παραμείνουν αμετάβλητες. Παρέχουμε μια υλοποίηση ανοικτού λογισμικού, με ονομασία DiavgeiaRedefined, η οποία επιτρέπει τη δημιουργία και την οπτικοποίηση των αποφάσεων σε περιηγητή διαδικτύου, προσφέρει ένα SPARQL τερματικό για τη δημιουργία ερωτημάτων και παρέχει στους πολίτες ένα αυτοματοποιημένο λογισμικό επαλήθευσης ορθότητας των αποφάσεων, ανιχνεύοντας πιθανές ατιμίες ενός κακόβουλου χρήστη. Τέλος, παραθέτουμε πειραματικά αποτέλεσματα, καταλήγοντας ότι οι μηχανισμοί που χρησιμοποιούμε είναι αποτελεσματικοί.

ΘΕΜΑΤΙΚΗ ΠΕΡΙΟΧΗ: Σημασιολογικός Ιστός, Κρυπτογραφία

ΛΕΞΕΙΣ ΚΛΕΙΔΙΑ: ανοικτά διασυνδεδεμένα δεδομένα, blockchain, ανοικτή διακυβέρνηση, σημασιολογικός ιστός, bitcoin

This master thesis is dedicated to my family.

ACKNOWLEDGEMENTS

Firstly, I would like to thank professor Manolis Koubarakis for giving me the opportunity to work on this interesting project and be part of Knowledge Representation, Reasoning and Analytics research group (KR&R). His suggestions about this project were of crucial importance. I would also like to thank the PhD student Dionysis Zindros for the fruitful blockchain-related conversations we had. Finally, I would like to thank the KR&R team for their valuable suggestions regarding the Semantic Web technologies.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

LISTINGS

1. INTRODUCTION

Government decisions which are made by public authorities and institutions, affect significantly the daily lives of ordinary citizens. Therefore, an important dimension of open government is making these government decisions open and easily accessible to the public.

Diavgeia (`https://diavgeia.gov.gr/en`) is a Greek program introduced in 2010, enforcing transparency over the government and public administrations, by requiring that all government institutions have to upload their decisions on the Diavgeia Web portal. The portal is managed by the Ministry of Administrative Reform and E-Governance. Diavgeia is now fully implemented by public authorities. The current rate of uploads in the Diavgeia portal is 16.000 decisions per working day, summing up to a total of 26 million decisions up to now. However, decisions are currently uploaded as PDF files and follow no structuring of their textual content. As a consequence, interested parties (the government, ordinary citizens, non-government bodies, courts, the media, etc.) rely on keyword search over PDF files, in order to find decisions that might effect them in some way or verify that uploaded decisions have been taken according to the law. Also, despite the fact that these decisions are digitally signed, there is no integrity mechanism which ensures the immutability of all decisions over time.

In this work, we aim at revolutionizing the way that decisions of the Diavgeia program are made public, by following the footsteps of other successful efforts in Europe which publish legislative documents as open linked data [8]. By applying Semantic Web techniques, we envision a new state of affairs in which ordinary citizens have advanced search capabilities at their fingertips on the content of public sector decisions. The ability to pose Semantic queries promotes the inclusion of all citizens in the scrutiny of the public sector, leaving considerably less room for governance corruption. In addition, through the use of the disruptive nature of the bitcoin blockchain, we enable decisions to remain immutable, introducing unprecedented levels of transparency to the digital governance of Greece, as we ensure the integrity of the published decisions as open linked data. By implementing automated blockchain tools which verify the tamper-proof property of government decisions, we further promote the inclusion of citizens in the critical observation of the public sector.

Contributions. Towards achieving the aforementioned goals, we provide an open source implementation, named DiavgeiaRedefined[1], which aims to replace the current production implementation of Diavgeia. The DiavgeiaRedefined project consists of the following modules:

1. **Diavgeia ontology**. We follow the latest Semantic Web standards and best practices and develop an OWL ontology, called *Diavgeia ontology*, for modeling the content of decisions uploaded by the Greek public authorities to the Diavgeia website. Using this ontology, decisions can be encoded in RDF and be interlinked with other Greek government data (e.g., legislation in the system Nomothesia [8]), empowering interested parties to pose rich queries over these data sources. The linking of Diavgeia with Nomothesia has the benefit of making sure that the references of public sector decisions refer to valid legislative documents (laws). We also interlink Diavgeia with a dataset encoding the administrative geography of Greece.

[1]The source code of the project can be found on `https://github.com/ThemisB/diavgeiaRedefined`

2. **Web editor and Visualizer**. DiavgeiaRedefined provides web applications to prove that Semantic Web technologies can be used by the Greek government in a user-friendly manner. We develop a Web editor that can be used by public authorities, for authoring their decisions. The result of this procedure is the creation of decisions expressed in RDF, compatible with the Diavgeia ontology. We also develop a Web tool that visualizes the aforementioned decisions.

3. **Blockchain tools**. We employ the disruptive nature of the bitcoin blockchain and we develop the Stamper and the Consistency Verifier tools. The Stamper organizes and aggregates decisions into Merkle trees [9], providing a way to store decisions expressed in RDF on the bitcoin blockchain with very low cost. The Consistency Verifier can be used by interested parties to verify the correctness and detect possible foul play by a participant in the process.

4. **SPARQL Endpoint**. By employing Fuseki server, we empower interested parties to browse, search and pose interesting SPARQL queries to public sector decisions.

5. **Evaluation**. We evaluate DiavgeiaRedefined in two ways: (i) by calculating the blockchain validation time for a month's regular workload, and (ii) by comparing it with the current implementation of Diavgeia in terms of disk space usage.

Organization. The rest of the thesis is structured as follows. Chapter 2 discusses related work in legislative knowledge representation using Semantic Web technologies, presents endeavors that combine linked data with the disruptive blockchain technology and demonstrates some related studies on E-Government, as well as, related European initiatives in the area of digital public administration. Chapter 3 presents Diavgeia. Chapter 4 discusses the Diavgeia ontology, presents the Web editor and Visualizer and some interesting SPARQL queries. Chapter 5 describes the two blockchain tools developed for the preservation and verification of decisions. Chapter 6 presents the evaluation results. Last, Chapter 7 summarizes our contributions and discusses future work.

2. RELATED WORK

This chapter introduces some fundamental notions regarding the Semantic Web and the Bitcoin Blockchain, and discusses the related work which has already been carried out.

2.1 The Semantic Web

The Semantic Web is an extension of the World Wide Web through standards by the World Wide Web Consortium (W3C). The Semantic Web provides a common framework that allows data to be shared and reused across application, enterprise, and community boundaries. The Resource Description Framework (RDF) is a standard model for data interchange on the Web and is a major component in the W3C's Semantic Web activity. RDF is an abstract model with several serialization formats (i.e. file formats), so the particular encoding for resources or triples varies from format to format.

A collection of RDF statements intrinsically represents a labeled, directed multi-graph. This in theory makes an RDF data model better suited to certain kinds of knowledge representation than are other relational or ontological models. However, in practice, RDF data is often stored in relational database or native representations (also called Triplestores or Quad stores, if context such as the named graph is also stored for each RDF triple).

The Web Ontology Language (OWL) is a family of knowledge representation languages for authoring ontologies. Ontologies are a formal way to describe taxonomies and classification networks, essentially defining the structure of knowledge for various domains: the nouns representing classes of objects and the verbs representing relations between the objects. The OWL languages are characterized by formal semantics. The OWL languages are built upon the RDF.

SPARQL (acronym for SPARQL Protocol and RDF Query Language) is an RDF query language, that is, a semantic query language for databases, able to retrieve and manipulate data stored in RDF format. SPARQL allows for a query to consist of triple patterns, conjunctions, disjunctions, and optional patterns.

Linked data is a method of publishing structured data so that it can be interlinked and become more useful through semantic queries (e.g. using SPARQL). It builds upon standard Web technologies such as RDF, but rather than using them to serve web pages for human readers, it extends them to share information in a way that can be read automatically by computers. Linked data may also be open data, in which case it is usually described as linked open data (LOD).

Tim Berners-Lee, the inventor of the Web and Linked Data initiator, suggested a 5-star deployment scheme for Open Data [3]. The 5 Star Linked Data system is cumulative. Each additional star presumes the data meets the criteria of the previous step(s). This scheme is widely used in order to evaluate the publishment of data on the Web. The ranking of the scheme is the following:

1-star: Data is available on the Web, in whatever format.

2-stars: Available as machine-readable structured data, (i.e., not a scanned image).

3-stars: Available in a non-proprietary format, (i.e, CSV, not Microsoft Excel).

4-stars: Published using open standards from the W3C (RDF and SPARQL).

5-stars: All of the above and links to other Linked Open Data.

2.2 The Bitcoin Blockchain

Bitcoin [20] is the first decentralized digital currency based on a distributed, peer-to-peer consensus network. Transactions propagate through the network in order to be verified and stored in a blockchain. Blockchain is the immutable public distributed ledger which records all bitcoin transactions, forming a chain of blocks. Each block in the blockchain is composed of the previous block in the chain and a payload of transactions. The blockchain technology that bitcoin or other cryptocurrencies offer, has been characterized by a lot of companies as a disruptive technology with a great variety of real world applications[1].

Bitcoin uses a stack-based scripting system for modeling transactions, called Script [4]. Transactions consist of multiple inputs and multiple outputs. Bitcoins are transferred on a transaction input and output, where the input defines where bitcoins are coming from and the output defines the destination. OP_RETURN opcode is a special instruction of Script which allows to save metadata up to 80 bytes on a transaction output [2]. *Miners* are specialized nodes on the network that keep the blockchain consistent, complete, and unalterable. By solving a hard cryptographic problem, miners generate and add new blocks to blockchain. The rest nodes of the network can easily verify and mutually agree that the solution given by the miner is correct and accept the new block.

The consensus algorithm of bitcoin guarantees that, for an attacker to be able to alter an existing block, he must control the majority of the computational resources of the network [15]. As more transactions and blocks are generated, the difficulty of the cryptographic problem increases significantly, which makes the tampering of data written in the blocks very difficult. This security property is often rephrased by saying that the bitcoin blockchain can been seen as an immutable, permissionless data structure. Thus, even if the main goal of bitcoin is to transfer digital currency, there are certifying services which take advantage of the tamper-proof nature of blockchain, by providing users a way to certify existence or ownership of documents (such as *Proof of Existence*[2], *OpenTimestamps*[3] and *Stampery* [10]). In our work, we use the OP_RETURN opcode to embed government data as metadata of a bitcoin transaction, similar to the aforementioned certifying services.

2.3 Legislative Documents as Linked Data

Democratizing access to government and legislative documents has been a primary concern of many governments across the world. Many countries have a government portal where government data is made available free of charge, in some cases as linked data (`https://data.gov.uk/`). The development of information systems archiving the content of legislative documents as linked data has been a common practice towards making legislation easily accessible to public [7]. For example, the MetaLex Document Server [18] hosts all national regulations of the Netherlands, while [14] presents a service for publishing the Finnish legislation as linked open data. Nomothesia (`http://legislation.di.uoa.gr/`) [8] is a research project, which publishes Greek legislation as linked open data

[1] Nasdaq article with title "Blockchain: Disruptive Technology Is Here to Stay": `http://www.nasdaq.com/article/blockchain-disruptive-technology-is-here-to-stay-cm882850`

[2] Proof of Existence: `https://poex.io/about`

[3] OpenTimestamps: `https://opentimestamps.org/`

and it also offers SPARQL endpoint and RESTful API to interested parties for search reasons. All of the above endeavors adopt different vocabularies and ontologies to express the particularities of each country's legislation. Recently, the European Council introduced the European Legislation Identifier (ELI) [11] as a common framework that can be adopted by the national legal publishing systems in order to unify and link national legislation with European legislation. ELI is partly based on the use of Uniform Resource Identifiers (URIs), and partly on a set of structured metadata for referencing European and domestic legislation. All of the afforementioned vocabularies and ontologies, are not a one-size-fit-all model but they have to be extended to capture the particularities of national legislation systems.

2.4 Blockchains in the Semantic Web

The Semantic Web community has just begun to consider applications which take advantage of the distributed, undeletable and immutable nature of the blockchain. Decentralized Semantic Identity [13] examines a semantic approach for W3C WebID, in which the Namecoin blockchain is used to register the user's WebID URI and domain names. In this endeavor, the proposed authentication scheme is outside the control of any single entity.

Recently, A. Sutton and R. Samavi [27] proposed a linked data based method of utilizing blockchain technology to create tamper-proof audit logs. In this work, authors utilize the L2TAP [24] privacy audit log to capture all relevant privacy events and provide SPARQL solutions for major privacy processes such as obligation derivation and compliance checking. The bitcoin blockchain is used to provide non-repudiation to these logs.

M. English et al. [12] discuss what Semantic Web research and development can offer to blockchain research and development and vice versa. Authors demonstrate that blockchain can be applied in practice towards the actualization of a more resilient architecture for the Semantic Web, such as the construction of Secure Resource Identifiers. On the other hand, the Semantic Web can be applied on the blockchain for representing transactions in RDF format, supporting the linking of wallets related by transactions to follow exchange activity around the network. Finally, authors present two use cases of combining blockchain and semantic technologies. The first use case suggests the usage of the blockchain as universal digital ledger for Industry 4.0 transactions [5]. The second one shows a scenario of how a smart contract platform such as Ethereum could support micro and standard accreditation within a higher educational setting.

2.5 Studies on E-Government

Many researchers across the world have suggested different technologies that can be applied on E-Government services. H. Hou [19] describes a use case applied on a city district in southern China, which aims to use blockchain to streamline digital governance services for its one million residents. This project claims to bring more transparency and accessibility of government information, but what the platform looks like and when it can be accessed for actual usage remains unclear[4].

[4] A related article from Coindesk: `https://www.coindesk.com/local-government-china-trials-blockchain-public-services/`

T. Sidoroff and E. Hyvönen [25] present a case study on how semantic search and browsing techniques can be applied to solving the problems of content discovery, aggregation, and linking in e-government portals. This research endeavor indicates that representing content and their linking in terms of semantic web ontologies and logic rules is flexible from the system construction viewpoint, can be used to provide the end-user with useful "semantic" services and can reduce human effort in portal maintenance. Similarly, V. Peristeras and K. Tarabanis [22] use semantic technologies as the enabling infrastructure for creating a set of reusable models for the overall digital governance domain, named Governance Enterprise Architecture (GEA).

2.6 European Initiatives in the Area of Digital Public Administration

SCOOP4C (`https://www.scoop4c.eu/`) is a project funded by the European Union's Horizon 2020 research and innovation programme call "CO-CREATION-05-2015 Coordination and support action (CSA) Co-creation between public administrations: once-only principle". The once-only principle aims at eliminating the administrate burden when citizens are required to provide the same information again and again to public administrations. Instead, public administrations should have the means to re-use information already supplied by citizens in a transparent and secure way. This work investigates how the once-only principle can be effectively realised by governments in 2020, and how further opportunities and innovations can emerge through this realisation throughout all service areas of the public and also the private sector.

TOOP (`http://www.toop.eu/`) is the sister project of SCOOP4C aiming to explore the implementation of the once-only principle for businesses. Through three pilot areas (Cross-border e-Services for Business Mobility, Updating Connected Company Data, Online Ship and Crew Certificates), TOOP wants to enable better exchange of business related data or documents with and between public administrations and reduce administrative burden for both businesses and public administrations. The project is especially ambitious regarding the number of states and areas concerned in these pilots: it aims to connect 38 information systems as data consumers (receiving data) and 32 systems as data providers (sending data to data consumers) in any-to-any transactions. As expected direct impacts, the TOOP pilots foresee time savings and cost reductions for businesses and administrations by reducing the administrative burden. Besides, the generic approach TOOP develops is expected to bring along a reduction in the cost of future e-Government pilots or the setting up of services. Moreover, other national administrations will be able to join the pilots and develop cross-border applications for certain services, thus ensuring demand for Connecting Europe Facility (CEF[5]) Digital Service Infrastructures (DSIs) and building blocks.

ENLARGE (`http://www.enlarge-project.eu/`) is a project led by Istituto per la ricerca sociale (IRS) and funded by the European Union. This project aims to generate and disseminate shared, structured and constructive knowledge on participatory governance (co-design and co-production participatory processes) through an intense process of dialogue and communication among policy makers, civil society actors and practitioners. The specific focus is on sustainable energy and more in general all the solutions aiming at generating more efficiency and effectiveness in public services and public endowments.

[5]Read more about the CEF: `https://ec.europa.eu/digital-single-market/en/connecting-europe-facility`

Bloomen (`http://bloomen.io/`) is also an EU-funded research project started in September 2017, aiming to answer the question: "How could blockchain technologies help to compensate creative people for their work?". The goal of Bloomen is to extend the use of blockchain technology to handle different online user transactions, provide an innovative way of content creation, sharing, personalized consumption, monetization and copyrighting. Bloomen has defined three use cases to work on: *Music*, *Media Content* and *WebTV*. The *Music Industry use case* will result in a permissioned blockchain ecosystem, that registers all changes to a database containing the core metadata of musical assets (sound recordings – or master recordings – as well as their underlying musical works – or compositions), their associated rights and their interested parties. It aims at providing improved efficiency and transparency all along the value chain, involving music Rights Management, Claiming of Ownership of Musical Assets and Conflict Resolution. The core idea of the *Media Content use case* is to demonstrate how digital news content could be distributed, tracked and paid for in a decentralized manner, by creating a platform where all content creators (authors, photographers, Radio and video reporters, data visualizers, etc.) can sent pieces of content into the internet, to be rewarded by payments based on a number of factors – such as popularity, number of references to the content, etc. Payments can come from various sources, such as the publisher, or advertisers, or by teaming up with bigger, branded platforms such as the BBC, The Guardian, Le Monde or any other news outlet. Finally, the *WebTV use case* is to demonstrate how WebTV content (on demand and streaming) can be distributed using blockchain technology. At the core of this use case is a paywall for accessing the WebTV content whereby the user can pay with cryptocurrency and consume the content in a multitude of devices (from web streaming to thin clients on his smartphone).

Co-VAL (`http://www.co-val.eu/`) project is a new European research initiative started in November 2017 which aims to discover, analyse, and provide policy recommendations for transformative strategies that integrate the co-creation of value in public administrations. This project will examine cases when citizens, civil servants, private, and third sector organizations voluntarily participate in the development of transformative innovations addressing changing needs and social problems. Co-VAL will push the boundaries of both research and practice by providing (i) a comprehensive and holistic theoretical framework for understanding value co-creation in public services from a service-dominant logic and a service innovation multiagent framework, (ii) measurement and monitoring for transformations in the public sector by using both existing data and new metrics (large-scale survey), (iii) investigation on 4 public-service-related co-creation areas of public sector transformation: digital transformation (including open platforms, big data, and digital service delivery), service design (including service blue-printing), government living labs, and innovative structural relationships (public-private innovation networks and social innovation), and (iv) generation of sustainable impacts in public administration policy and practice by delivering actionable policy recommendations that build on the research findings, by tracking and monitoring how governments' pilot projects and actions are in line with the research-based recommendations , and by facilitating peer to peer knowledge exchange to facilitate implementation.

CoSIE (`https://cosie.turkuamk.fi/`) is also a new European research project started in December 2017 aiming to inspire the beginning of a new era in the co-creation of public services. The CoSIE project states that public service innovations can be achieved by creating collaborative partnerships between service providers and service beneficiaries. During the implementation of CoSIE, the collaborative partners will test and develop the diverse methods of co-creation in the field of public services. The CoSIE partners will

actively search for new ways to use digital tools and open data in order to enable the co-creation of services. With innovative practices the project aims to improve the inclusion of all citizens and to promote their possibilities to act as active members of the society. The co-creation process in this particular project consists of nine individual (but interlinked) pilots. Each pilot has different target groups, service needs and local settings.

SoCaTel (`https://www.tcd.ie/swsp/research/SoCaTel.php`) is a project funded by the European Union's Horizon 2020 research and innovation programme which proposes an approach that will cater to the needs of the growing ageing populations in Europe by improving the accessibility, responsiveness, efficiency, transparency and transferability of social and care services. SoCaTel aims to address gaps in the area of social welfare and social services by introducing a multi-stakeholder platform for the co-creation, and later deployment, of long-term care services. The SoCaTel platform will be designed and conceived through a co-creation process, which will allow and encourage users to interact with Information and Communications Technology (ICT) in a straightforward way. SoCaTel will address the current lack of coordination among primary and community services, hospitals, and other care providers, in order to make accessing and using long-term care (LTC) services easier for older adults and care professionals.

PoliVisu (`https://www.polivisu.eu/`) has also received funding from the Horizon 2020 programme. PoliVisu enhances public involvement and support in urban policy making, by equipping decision makers with the skills and tools - from open (geo) data processing to advanced visualisations - to use big data for collaborative policy experimentation. As a result the city makes better sustainable policy decisions and manages operations more effectively.

Big Policy Canvas (`http://www.bigpolicycanvas.eu/`) is a project funded by the European Union's Horizon 2020 research and innovation programme call "CO-CREATION-06-2017 Policy-development in the age of big data: data-driven policy-making, policy-modelling and policy-implementation". The mission of Big Policy Canvas calls on understanding and exploring both the technological and the socio-economic landscapes that will influence and will be influenced by advanced public sector performance, in order to guarantee not only the sustainability and effective utilisation of the solutions proposed, but also the nurturing of a more data-inspired, politically educated, smart and caring generation within and outside of the public sector boundaries that will work towards improving the overall life experience in Europe.

Finally, the **CITADEL** (`https://www.citadel-h2020.eu/`) project was funded by the EU in October 2016. It is focused on transforming the public sector to make more efficient, inclusive and citizen-centric public services that identify/capture new or unsatisfied needs more quickly and satisfy them more effectively and in an inclusive way, providing also guidelines and features to support new processes. CITADEL is based on Hirschman [17] and Rokkan [23] models, hugely influential across the social sciences, as the schemes to face the challenge of knowing how engaged members in an organization are; how likely they are to remain members; when they might cease to be a member; and why they have not yet become members. To achieve its objectives, the CITADEL ecosystem will combine and promote a set of technologies (e.g., semantics, mobile, analytics, sentiment analysis, open linked data) to both empower public administration to improve its offering and the engagement of citizens and other subjects (i.e., the private sector), as well as foster cooperation among the public administration and the users of its services.

2.7 Public Sector Ontologies

Organization Ontology (ORG): The Organization Ontology (ORG)[6] was created initially by Dave Reynolds of Epimorphics in 2010. It provides the classes, properties and relationships to describe the structure of an organisation and the reporting structures within it. Since 2010 it has been further developed by the W3C and at the time of writing this thesis, the ORG is the only W3C Recommendation regarding the "E-Government" which can be used as a core ontology for organizational structures.

Registered Organization Vocabulary (RegOrg): The Registered Organization Vocabulary (RegOrg)[7], originally known as the Core Business Vocabulary is a profile of ORG for describing organizations that have gained legal entity status through a formal registration process, typically in a national or regional register. The RegOrg includes a minimal number of classes and properties that are designed to capture the typical details recorded by business registers and thereby facilitate information exchange between them, although there is significant variation between business registers in what they record and publish.

Core Public Service Vocabulary (CPSV): The Core Public Service Vocabulary (CPSV)[8] has been created as part of Action 1.1 of the Interoperability solutions for European public administrations (ISA) programme of the European Commission (EC). It is designed to make it easy to exchange basic information about individual public sector services. By using the vocabulary, almost certainly augmented with sector-specific information, organisations publishing data about their services enable (i) the easier discovery of those services with and between countries, (ii) the easier discovery of the legislation and policies that underpin service provision, (iii) the easier recognition of how services provided by a single organization interrelate and are used either by other services or external users and (iv) the easier comparison of similar services provided by different organizations.

2.8 Summary

This chapter concentrated on the fundamental aspects of the Semantic Web and the Bitcoin Blockchain with which readers should be informed in order to understand better the technologies referred in the next chapters. Futhermore, we discussed the related work in legislative knowledge representation using Semantic Web technologies, we presented endeavors that combine linked data with blockchain technology and we demonstrated some related studies on E-Government. Finally, we also discussed recent European initiatives in the area of digital public administration and related public sector ontologies. In the next chapter, we will discuss in great detail the current functionality and implementation of Diavgeia, pointing out its deficiencies.

[6]The Organization Ontology can be found on: `https://www.w3.org/TR/vocab-org/`

[7]The Registered Organization Vocabulary can be found on: `https://www.w3.org/TR/vocab-regorg/`

[8]The Core Public Service Vocabulary can be found on:`https://joinup.ec.europa.eu/rdf_entity/http_e_f_fdata_ceuropa_ceu_fw21_f1e1acc8c_b7bec_b4709_b9430_be26598eb682d`

3. BACKGROUND ON DIAVGEIA

In this chapter, we present the functionality of Diavgeia and its implementation in great detail. We also point out the problems of the current implementation that motivated us to implement DiavgeiaRedefined.

3.1 Diavgeia: The Greek Transparency Program Initiative

Beginning October 1st, 2010, all government institutions are obliged to upload their acts and decisions on the Internet with special attention to issues of national security and sensitive personal data. Each document is digitally signed and assigned a unique Internet Uploading Number (IUN) certifying that the decision has been uploaded at the "Transparency Portal". Following the latest legislative initiative (Law 4210/2013) of the Ministry of Administrative Reform and e-Governance, administrative acts and decisions are not valid unless published online. The main objectives of the Program concern:

1. Safeguarding transparency of government actions.
2. Eliminating corruption by exposing it more easily when it takes place.
3. Observing legality and good administration.
4. Reinforcing citizens' constitutional rights, such as the participation in the Information Society.
5. Enhancing and modernizing existing publication systems of administrative acts and decisions.
6. Making of all administrative acts available in formats that are easy to access, navigate and comprehend, regardless of the citizen's knowledge level of the inner processes of the administration.

Taking the view that the Greek crisis, including its economic manifestations, is also due to the non transparent relationship between the citizens and the state, the transparency program introduced unprecedented levels of transparency within all levels of Greek public administration and established a new "social contract" between the citizen and the state. This initiative has a silent but profound impact on the way officials handle their executive power. The direct accountability brought upon the administration by the radical transparency that the Diavgeia program introduces, leaves considerably less room for corruption, and exposes it much more easily when it takes place since any citizen and every interested party enjoy the widest possible access to questionable acts. Such a collective scrutiny can be extremely effective, since it allows citizens directly involved or concerned with an issue to scrutinize it in depth, rather than leaving public scrutiny to the media, whose choice of issues necessarily may be restricted and oriented towards sensational topics.

Public authorities adopted the Program in three phases: Ministries in October 2010, Extended Public Sector and Independent Authorities in November 2010, Regional and Local Authorities in March 2011. In these 8,5 years, 27.400.000 acts and decisions have been

published on the Transparency Portal from 4.654 public authorities. According to the statistical information provided by the Diavgeia website, the current rate of uploads is 16.000 decisions per working day[1].

3.2 Greek public sector decisions and relevant laws

Public sector decisions cover a broad spectrum of activities in Greece. The Greek government has enacted 34 different decision types that may be uploaded on Diavgeia. The decision type is chosen by the government institutions according to the context of the decision. Despite the many different decision types, we observed that the majority of them follow the same pattern. A decision starts by refering to a number of different Greek laws on which is based and then gives the main text of the decision. The following figure illustrates an example of an *Appointment* decision type that adheres to the aforementioned pattern.

Appointment decision type

Appointment of R.F. as Full Professor
In accordance with:

1. The provisions of Law 3549/2007, article 25, paragraph 1.
2. The provisions of Presidential Decree 2011/54.
3. The provisions of Law 4386/2016, article 70, paragraph 4.

We decide:

1. The appointment of R.F. as Full Professor at the X department, at the Y University, on the subject of “Semantic Web”.

Figure 3.1: Example of an *Appointment* decision type

Despite the fact that this pattern can be used to define a common format for the different types of decisions, for the time being, public sector authorities upload their decisions as PDF files which follow no structuring of their textual content. Futhermore, citizens have no guarantee that the legislative references of a decision exist and are valid (such as *Laws* and *Presidential Decree* of the appointment example). By using the 5-star rating model for data [3], Greek public sector decisions are marked as 1-star.

In this work, we improve the current way of publishing Greek public sector decisions on the Web, by expressing them as 5-star open linked data. We take advantage of the aforementioned pattern and develop the Diavgeia ontology based on it. Technically, we view decisions as a collection of legal RDF documents with this standard structure.

We also employ the Nomothesia in order to ensure that the references to Greek legislation exist. Nomothesia has so far published 5 primary types of Greek legislation (*Constitution*, *Presidential Decrees*, *Laws*, *Acts of Ministerial Cabinet*, and *Ministerial Decisions*), as well as, 2 secondary ones (*Legislative Acts* and *Regulatory Provisions*). Nomothesia structures all legal documents, by using persistent URIs according to the template proposed

[1] Diavgeia provides up to date statistical results on `https://diavgeia.gov.gr/stats`

by ELI `http://www.legislation.di.uoa.gr/eli/{type}/{year}/{id}`. For instance, for the first provision of the appointment example, a linking of Diavgeia with the Nomothesia URI `http://legislation.di.uoa.gr/eli/law/2007/3549/article/25/paragraph/1` can be made. By integrating Nomothesia into Diavgeia, we also give citizens the ability to simply click to the legislative references of public sector decisions and see instantly the relevant passage of Greek legislation.

3.3 Metadata of Decisions

In addition to the uploading of the PDF file, public sector authorities also have to fill metadata information which describe the decision. The metadata used vary according to the type of decision. For instance, the metadata of the *ExpenditureApproval* decision type, holds important information about government's expenditure (such as the sender and receiver VAT registration numbers, the expense amount, etc). Diavgeia offers an OpenDataAPI (`https://diavgeia.gov.gr/api/help`) that can be used as an endpoint to query over the metadata. Despite the fact that OpenDataAPI is a step towards promoting transparency, inconsistency between decision text and metadata information is possible[2]. In our work, we embed metadata information into the RDF document, eliminating the possibility of inconsistency.

3.4 Identifiers and Modifications of Decisions

Each decision is assigned a unique Internet Uploading Number (IUN), certifying that the decision has been uploaded on Diavgeia. IUN is of significant importance, since citizens and other public authorities can use decisions, by solely referring to their unique number. In addition to IUN, each decision is also assigned a unique version token. Government institutions can upload a new version of a decision by claiming a new version token, but maintaining the same IUN. Diavgeia functions in an **append-only** manner, as it maintains the original decision with all its subsequent modifications. This is exactly why Diavgeia is amenable to blockchain technologies as we discuss in Chapter 5.

3.5 Summary

In this chapter, we presented Diavgeia as the Greek Transparency Program initiative and examined its main objectives and its ambitious goal to eliminate the corruption in the public sector. Then, we discussed the common pattern that the majority of the decisions follow, providing an example of an appointment. We considered the problem of the inconsistent metadata and we presented the identifier tokens that a decision gets upon its submission on the Diavgeia portal. In the next chapter, we will start describing the way that DiavgeiaRedefined adopts Semantic Web Technologies to tackle some of the aforementioned issues.

[2] An article about inconsistent metadata on Diavgeia website: `https://eellak.ellak.gr/2016/07/06/veltionontas-tin-piotita-dedomenon-stin-diavgia/`

4. MODELING DECISIONS USING SEMANTIC WEB TECHNOLOGIES

In this chapter, we develop an OWL ontology for modeling decisions of Diavgeia. We call our ontology *Diavgeia ontology* and we discuss its current version that adopts the ELI framework and the Nomothesia ontology. We present the Web editor and Visualizer components of DiavgeiaRedefined that generate and visualize the decisions expressed in RDF, respectively. Futhermore, we describe the linking of decisions with other datasets and we pose interesting SPARQL queries which take advantage of this interlinking.

4.1 The Diavgeia ontology

The ontology of Diavgeia is based on the pattern followed by public sector decisions, as discussed in Chapter 3.1. It imports and uses properties that are defined in the ELI ontology and the Nomothesia ontology. The core[1] of Diavgeia ontology is shown on Figure 4.1. The 34 different decision types can be viewed as legal documents (class `LegalResource` of the ELI ontology). A decision (`LegalResource`) `changes` itself, by generating a new `version` and maintaining its unique `iun`. A `LegalResource` is composed of multiple `Considerations` and `Conclusions`. The `Consideration` class models the passages that are used to prove the validity of the decision (e.g. the three provisions of the appointment example), while `Conclusion` models the passages that are used as conclusions of the decision (that is the final passage of the appointment example). Both `Consideration` and `Conclusion` classes, use the `cites` property of the ELI ontology, to make a reference either to Greek Legislation (Nomothesia) or to another decision of Diavgeia. The `has_text` property is used to describe the text body of either `Consideration` or `Conclusion`.

The Diavgeia ontology offers 121 properties to cover all the particularities of different decision types. In addition to `Consideration` and `Conclusion`, the ontology provides classes which describe important public sector activities. For instance the class `Expenses` links an expense of a public authority to an individual or business. For the time being, this crucial information is expressed as metadata of the PDF decision, underlying the possibility of metadata inconsistency as described in Chapter 3.2. By merging metadata and decision text in a single RDF file, this possibility is eliminated.

In order to identify legal resources, we also need appropriate URIs. Persistent URIs is a strongly recommended best practice [1], according to ELI. It is very important to have reliable means to identify the public sector decisions. Based on what is stated in Chapter 3.3, we can structure the persistent URIs of decisions according to the template `http://www.diavgeia.gov.gr/eli/{iun}/{version}`. Modifications of a decision result to the generation of a new URI which has the same `iun` and a new `version` number. Thus, the version of an enacted decision can be seen as the decision which has the most recent `date_publication` for a specific `iun`.

[1] The full Diavgeia ontology is available on: `https://github.com/ThemisB/diavgeiaRedefined/blob/master/rdf/diavgeia.owl`

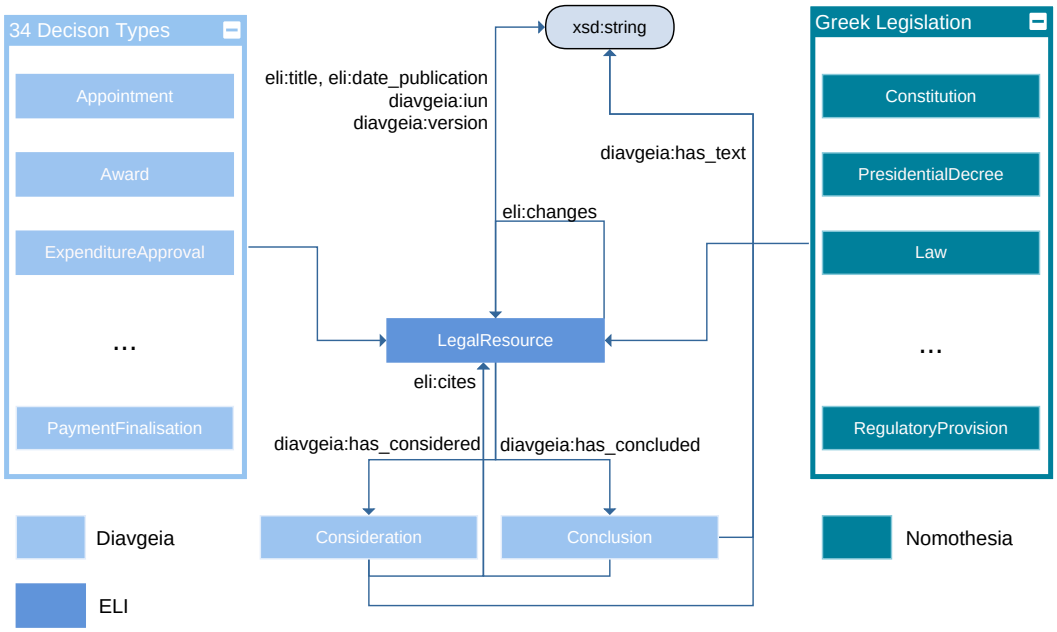


Figure 4.1: The core of Diavgeia ontology

4.2 Web Editor

DiavgeiaRedefined offers two main Web components in order to transparently adopt Semantic Web technologies to the production implementation of Diavgeia. The first one is a web editor for decisions, used exclusively by public sector authorities. The Web Editor is a well-structured HTML form that government institutions can use in order to write their decisions. The HTML elements of the form are associated with the properties and classes of Diavgeia ontology. By submitting the form, the decision is stored both as a compressed Notation3 file in the filesystem of Diavgeia and in Jena Apache's triple store.

Figure 4.2 illustrates the Web Editor tool, as used by the government institution which is responsible for the authoring of the aforementioned appointment decision example. Firstly, the government institution has to fill in some general information, such as its Department, phone number, etc. The main body of the decision starts after the "Author your decision" section. Users have to choose one out of the 34 different decision types. According to their choice, the rest of the HTML form loads dynamically capturing the particularities of the decision type.

Figure 4.3 presents the authoring of an award decision type in the Web Editor component. A public authority has to upload an award decision in order to publish the individual/company which has been chosen to carry out a public work (such as a road construction). This type of decision contains financial information regarding the public expense (the VAT number of the assignees and the estimated budget of the work).

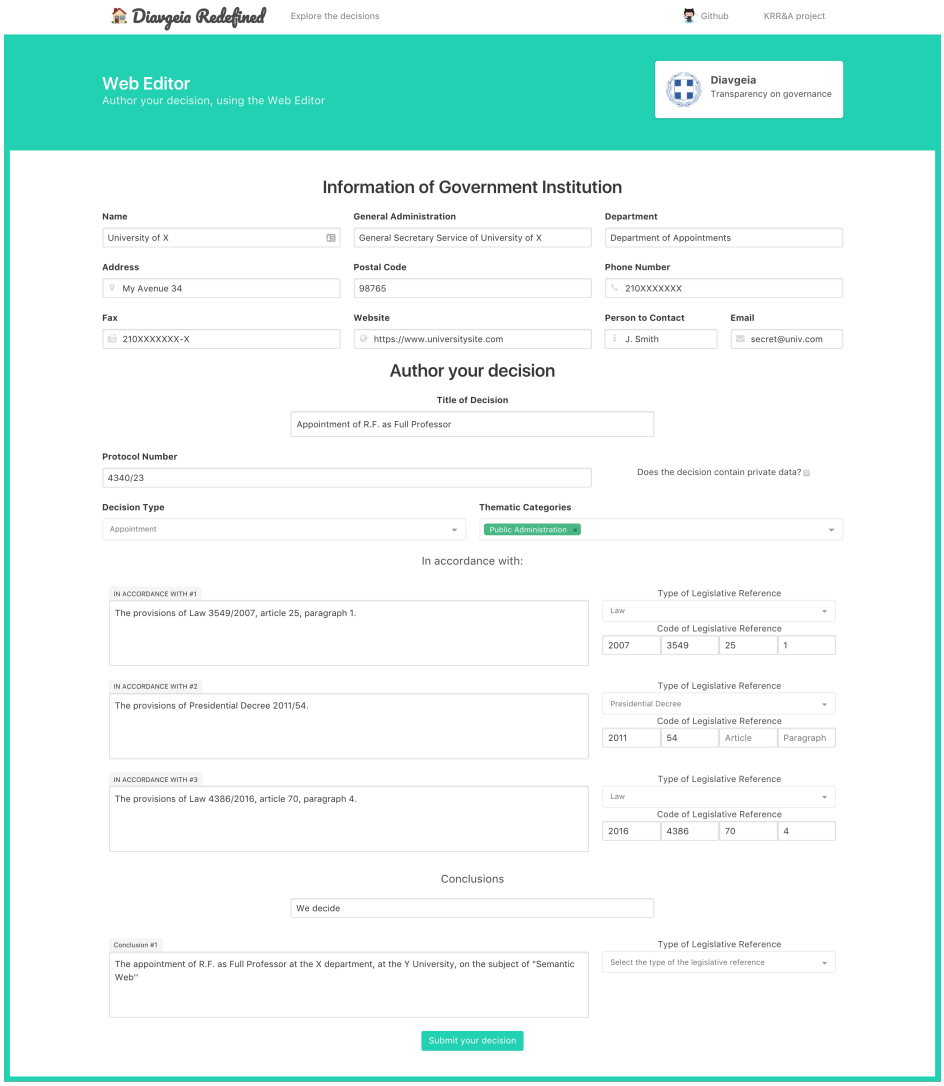


Figure 4.2: Authoring the appointment example using the Web Editor

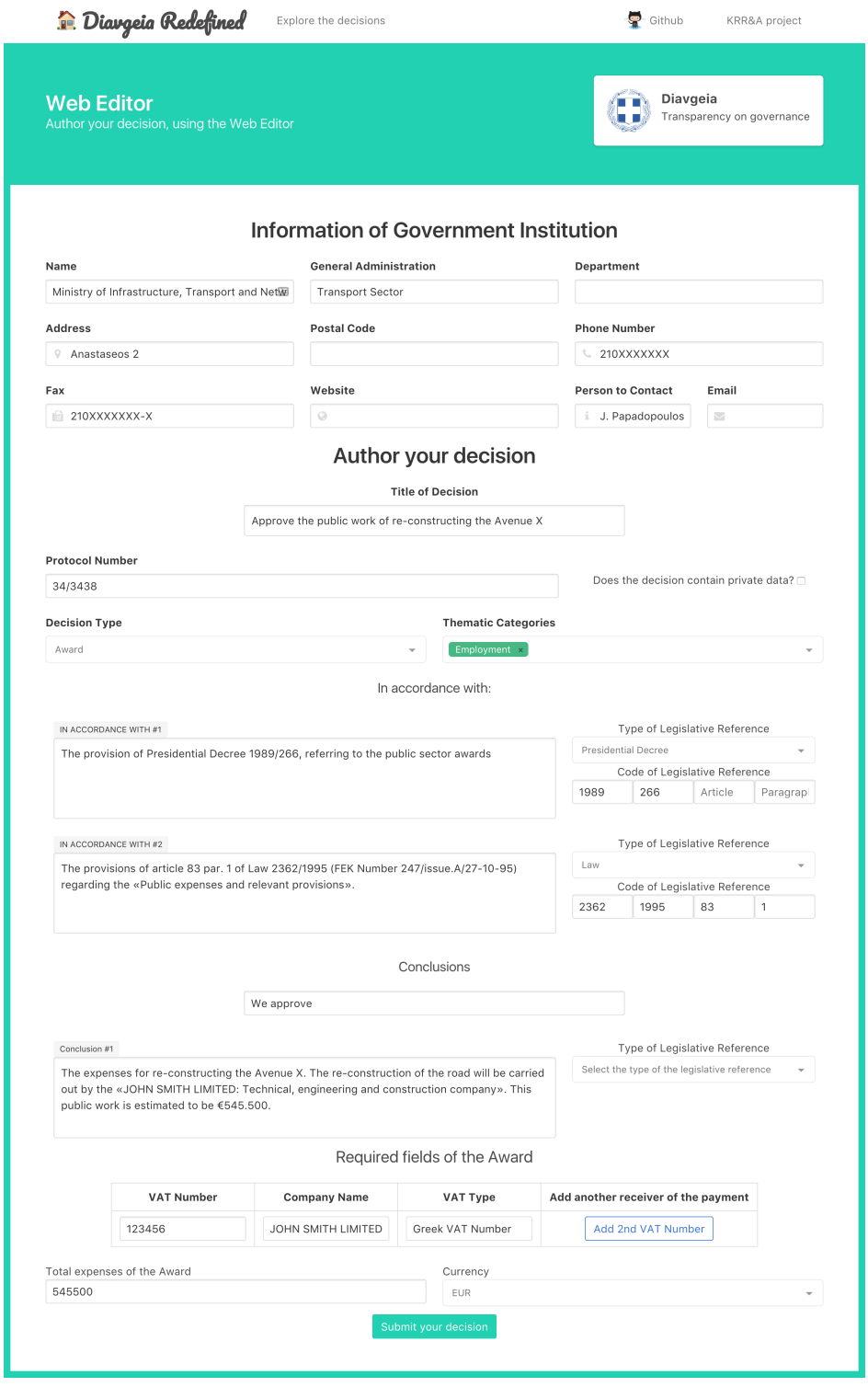


Figure 4.3: Authoring an award example using the Web Editor

4.3 Visualizer

The Visualizer is another component of DiavgeiaRedefined which can be used both by public authorities and citizens. Its purpose is to provide a visualization of the decisions expressed in RDF inside a Web browser. Users provide the persistent URI of the decision they want to visualize and the decision is displayed in the browser in a user-friendly manner. The figure 4.4. demonstrates the visualization of the appointment decision which

was generated by the Web editor in the previous section, while figure 4.5 demonstrates the visualization of the award example.

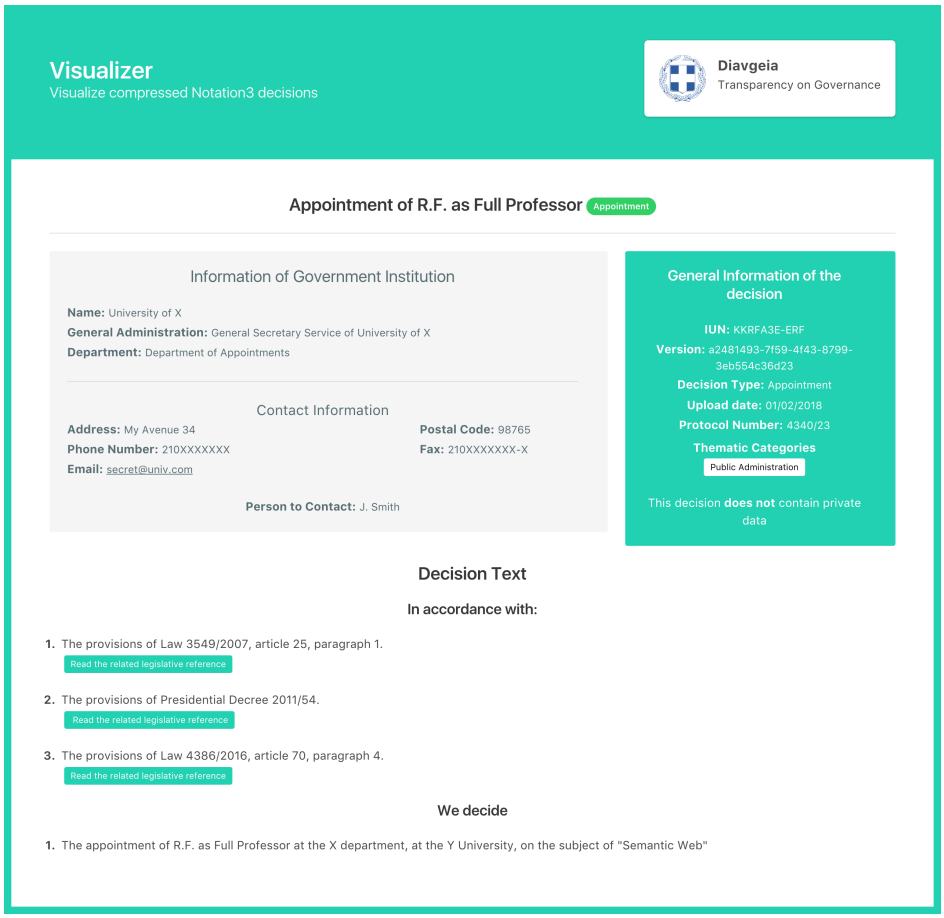


Figure 4.4: The visualization of the appointment example

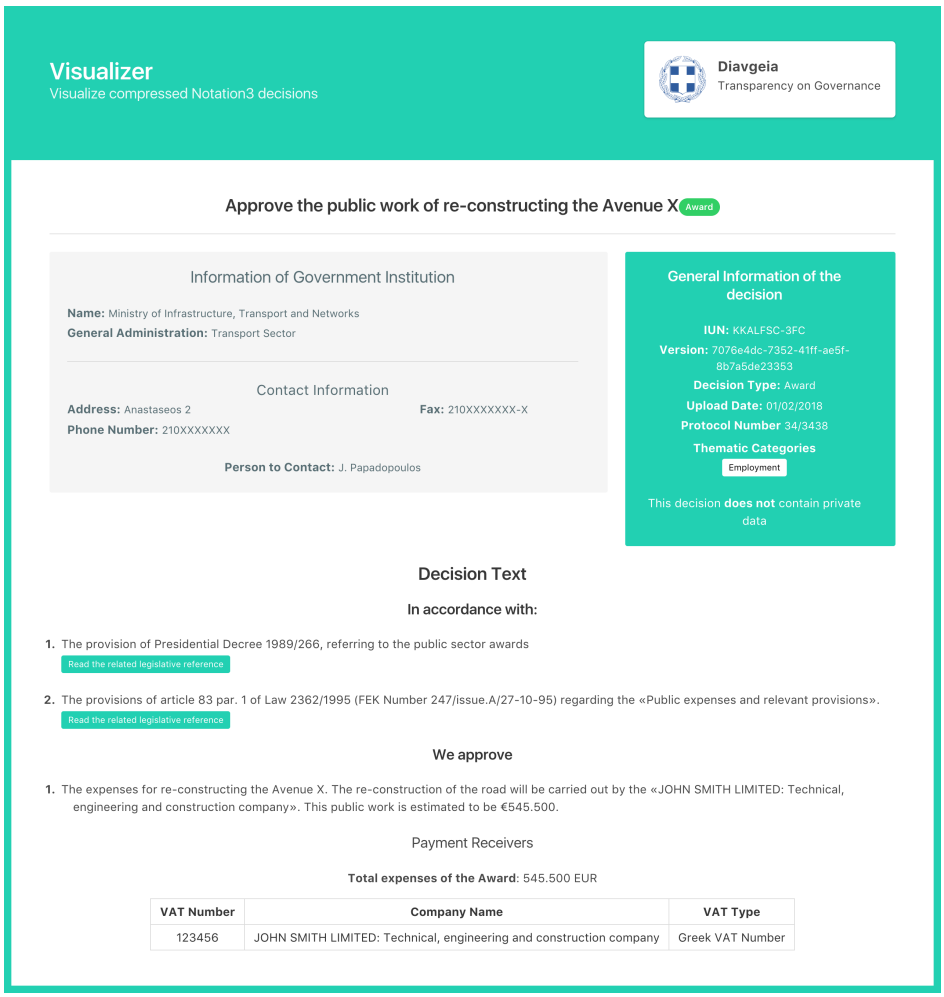


Figure 4.5: The visualization of an award

4.4 Linking decisions with other public sector data

The linking of decisions with other public sector data, can be done by public authorities, using the Web editor component of DiavgeiaRedefined. Firstly, the `Consideration` or `Conclusion` classes of a decision, may make reference to the Greek Legislation of Nomothesia, as mentioned in Chapter 4.1. Linking Diavgeia with Nomothesia is easy, since the latter provides persisent URIs according to template `http://www.legislation.di.uoa.gr/eli/{type}/{year}/{id}`.

Public authorities have also to link `SpatialPlanningDecisions` decision type, with the dataset of administrative geography of Greece[2]. Linking decisions with it is also easy and it is achieved through the construction of constant mappings.

4.5 Querying the Resulting RDF Data using SPARQL

By employing the Fuseki Server, we enable the formulation of complex queries over decisions of Diavgeia. Semantic queries promote the inclusion of all citizens or other in-

[2]The Greek Administrative Geography dataset and the ontology are available on: `http://linkedopendata.gr/dataset/greek-administrative-geography`

terested parties in the scrutiny of the public sector, leaving considerably less room for governance corruption. We present some interesting queries one can pose. The gag prefix refers to the Greek Administrative Geography dataset.

```
SELECT ?decision WHERE {
  ?decision diavgeia:has_expense ?expense;
            eli:date_publication ?date.
  ?expense diavgeia:expense_amount ?amount.
  FILTER (?date >= "2017-01-01"^^xsd:date &&
  ?date <= "2017-12-31"^^xsd:date)
} ORDER BY DESC(?amount) LIMIT 5
```

Listing 4.1: Retrieve the decisions with the 5 highest government expenses of 2017.

This query of Listing 4.1 empowers citizens or other interested parties to find the decisions with the 5 highest government expenses of 2017. The queries which consider government expenses are of crucial importance, since they enable the tracking of economic transactions and promote the transparency in the financial sector. This query can be limited to certain thematic categories with the use of the `diavgeia:thematic_category` property. Thus, interested parties (e.g. a researcher) can search for the highest expenses based on their interest (e.g. the thematic category of “Science”).

```
SELECT DISTINCT ?decision WHERE {
  ?signatory foaf:name "I.Stournaras".
  ?nomothesiaLegislation eli:passed_by ?signatory.
  ?reference eli:cites ?nomothesiaLegislation.
  {?decision diavgeia:has_concluded ?reference} UNION
  {?decision diavgeia:has_considered ?reference}
}
```

Listing 4.2: Find decisions which make a reference to greek legislative documents that have been signed by the ex-Greek Minister of Finance I. Stournaras.

The query illustrated in Listing 4.2 enables interested parties to search decisions which refer to greek legislative documents that have been signed by a particular politician. By employing Nomothesia [8], we are able to obtain the legislative documents of a specific signatory. Those types of queries are also important, because interested parties can see if a politician signs legislative documents which are not widely used by the majority of the public authorities. Legislative documents which are only used by an extremely small subset of the public authorities might be an indication of corruption between a politician and a public authority.

```
SELECT ?organizationId ?sponsoredVatNumber
(COUNT(*) as ?timesPreferred) SUM(?expenseAmount) WHERE {
  ?decision a diavgeia:Award;
            diavgeia:has_expense ?awardExpense;
            diavgeia:organization_id ?organizationId.

  ?awardExpense diavgeia:expense_amount ?expenseAmount;
                diavgeia:expense_currency "EUR";
                diavgeia:has_sponsored ?sponsored.

  ?sponsored diavgeia:afm ?sponsoredVatNumber.
  FILTER (?expenseAmount > 100000)
} GROUP BY ?organizationId ?sponsoredVatNumber
  ORDER BY DESC(?timesPreferred)
```

Listing 4.3: Find the public administrations which prefer to fund certain individuals for expensive public works (e.g. over 100.000 Euros) and calculate the sum of expenses that the public administrations have so far approved for those individuals.

The query of Listing 4.3 retrieves the public administrations which fund certain individuals for expensive public works, it calculates the sum of the related expenses and orders the results based on the times that the public administration funded the certain individual. This is another query which tries to track the corruption in the financial sector. The Greek public authorities often create a call for contests regarding public works. If a public authority assigns multiple works to a certain individual, that may also be an indication for a non-meritocratic selection process.

```
SELECT ?decision WHERE{
  ?municipality gag:has_official_name "MUNICIPALITY␣OF␣CHANIA".

  ?decision diavgeia:has_municipality ?municipality;
            diavgeia:thematic_category "Energy";
            eli:date_publication ?date;
            diavgeia:spatial_planning_decision_type
              "ConcessionOfPublicLandProperty".

  FILTER(?date >= "2018-01-01"^^xsd:date)
}
```

Listing 4.4: Find the decisions published after 2018 which refer to the municipality of Chania and concede public land property to solve energy related issues.

The query illustrated in Listing 4.4 presents the exploratory capabilities that Linked Data offer. Consider an individual which lives in the municipality of Chania, located in the island of Crete. The individual is interested in the renewable energy, trying to find land which will enable him to build his own solar panels. By posing the aforementioned query, the individual is able to retrieve the decisions which are published after 2018 and concede public land property for energy related issues. The interlinking of Diavgeia with the Greek Administrative dataset enables interested parties to pose such queries with significant geographical importance.

```
SELECT (COUNT(DISTINCT ?fekIun) as ?numberFekDecisions)
       (COUNT(DISTINCT ?iun) as ?numberTotalDecisions)
       ( xsd:float(?numberFekDecisions)/
       xsd:float(?numberTotalDecisions) AS
       ?averageOfPrintedDecision )
       WHERE {
         ?fekDecisions diavgeia:fek_issue ?fekIssue;
                       diavgeia:iun ?fekIun.
         ?decisions diavgeia:iun ?iun.
       }
```

Listing 4.5: Calculate the average amount of the decisions which are uploaded on Diavgeia and also printed by the National Printing House.

The Greek government can also pose SPARQL queries similar to the last one to measure a wide range of statistical information or to supervise the public authorities. The query presented in Listing 4.5 empowers the Greek government to calculate the average amount of the decisions which are uploaded on Diavgeia and also printed by the National Printing House of Greece. Another query that may be also posed by the government is the measurement of the work which has been carried out by the different Greek public authorities to load balance the work among them.

4.6 Summary

In this chapter, we presented in great detail the *Diavgeia ontology* used for the modeling of the decisions of Diavgeia. We discussed its current version which integrates Nomothesia, ELI and the Greek Administrative dataset. By implementing the Web editor and Visualizer components of DiavgeiaRedefined, we are able to generate and visualize the decisions expressed in RDF, respectively. The importance of adopting Semantic Web technologies in public sector is evident from the SPARQL queries we pose; those queries empower interested parties to track the financial transactions of the public sector, discover possible cases of government corruption, easily explore certain decisions of interest (e.g. the solar systems example) or even extract statistical from these decisions for better public administration management. The use of Semantic queries will drastically enhance the inclusion of citizens in the inspection of the public sector, as they will be able to pose complex queries of their interest over the government data. In the next chapter, we will describe the use of the bitcoin blockchain on Diavgeia in order to create a cheap integrity mechanism for the decisions which is easily verifiable.

5. PRESERVING DECISIONS USING BITCOIN BLOCKCHAIN

In this chapter, we describe the use of the bitcoin blockchain on Diavgeia. We present in detail the two blockchain tools that DiavgeiaRedefined offers, called Stamper and Consistency Verifier.

5.1 Stamper

Stamper is the tool which should be used by the administrators of Diavgeia in order to store public sector decisions on the bitcoin blockchain. The stamping procedure is described as follows:

1. Government institutions upload their decisions on Diavgeia. The backend of Diavgeia stores decisions as compressed Notation3 files in its filesystem and in the triple store.

2. The administrator of Diavgeia has to decide that the stamping procedure should take place at predefined time intervals t, ensuring the integrity of decisions. Thus, the backend of Diavgeia starts a new stamping procedure every t time units.

3. At the start of the stamping procedure, we find all the compressed Notation3 decisions which have not been stamped yet. Stamper organizes and aggregates these decisions into a Merkle Tree [9], using the hash function SHA-256. The root of the Merkle tree represents the fingerprint of the decisions which will be included in the forthcoming bitcoin transaction. By applying the SHA-256 hash function on the Merkle tree construction, the resulting root has a constant size of 32 bytes.

4. The next step is to create a Bitcoin transaction and broadcast it to the rest of the network. DiavgeiaRedefined uses the bcoin library (`http://bcoin.io/`), offering Diavgeia an spv node[1], maintaining only a chain, a pool, and a *hierarchical deterministic (HD)* wallet [16] based on BIP44 [21].

   A stamping transaction in our model consists of two outputs and one input. The first output contains the OP_RETURN opcode followed by the Merkle root in the *scriptPubKey* output (*scriptPubKey* = OP_RETURN + Root). This output guarantees the immutability of decisions. The second output is a pay-to-pubkey-hash[2], having as *pubKey* the next derived public address of the HD wallet. The input *scriptSig* consists of Diavgeia's signature and the current *publicKey* derived from HD wallet (*scriptSig* = *signature* + *publicKey*). The size of a stamping transaction is 267 bytes. In order to have certain guarantees that our transaction will be written into the next block and confirmed nearly immediately, mining fees can cost up to 120,150 satoshi (0.00125 bitcoin), which at the time of writing roughly amounts to $16.84.

After the end of each stamping transaction, Diavgeia publishes to its website the transaction identifier (Txid) and the order of decisions, as used for the Merkle tree construction. It also publishes once, the Master Public Key of its HD wallet. By publishing Diavgeia's

[1] A method for verifying if particular transactions are included in a block without downloading the entire block (`https://bitcoin.org/en/developer-guide#simplified-payment-verification-spv`).

[2] `https://en.bitcoin.it/wiki/Script#Standard_Transaction_to_Bitcoin_address_.28pay-to-pubkey-hash.29`

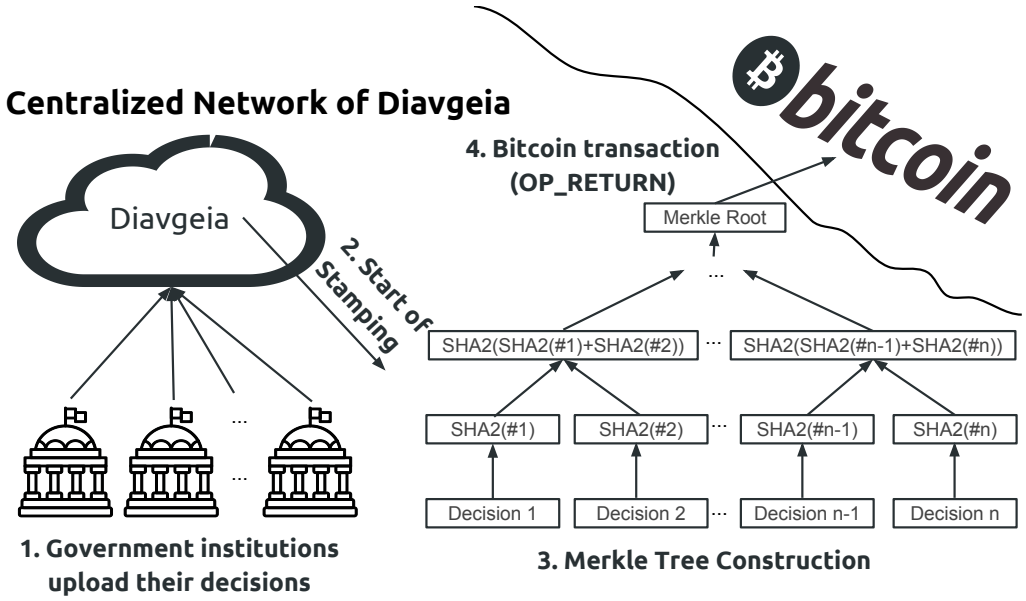


Figure 5.1: The Stamping procedure

master public key, interested parties are able to track the sequence of public keys and stamping transactions of Diavgeia. These publications are necessary to be made for the proper functionality of the Consistency Verifier (see Chapter 5.3).

5.2 Guarantees of Stamper

The Stamper tool provides high levels of immutability guarantees, especially when t value is configured to be small. Generally, the threat of a decision's modification or deletion appears on the time gap between two consecutive stampings. Small t values imply more stamping invocations and as a result Diavgeia creates more stamping transactions, but this comes at a higher cost. We consider a t value ranging from 3 hours to 1 day, to be an affordable solution for the government, since the daily cost of the usage of the blockchain will range from 0.00125 to 0.005 bitcoin ($16.84 - $134.72). The threshold for a decision's modification is also small, since an adversary (the administrators of Diavgeia, the government or other public authorities) are able to modify the decision in the next 3 hours to 1 day after its publishment.

As mentioned in Chapter 5.1, Stamper uses the open source bitcoin library (bcoin) in order to create the stamping transactions and relay them to the network. DiavgeiaRedefined does not use existing blockchain timestamping services (such as Stampery or OpenTimeStamps) because, in case of a foul play by an adversary, these third-party services might be accused of having modified the Merkle root in the first place.

5.3 Consistency Verifier

Consistency Verifier is the tool which can be used by the interested parties in order to verify that decisions have remained immutable over time. Algorithm 1 formalizes the steps

Consistency Verifier takes to verify the integrity of decisions.

Algorithm 1: Consistency Verification procedure

Data: Decisions included in stamping transaction *i*: d_i, Master Public Key: mpk
Result: Boolean result of verification.

```
foreach usedPublicAddress of mpk do
    transaction ← getTransactionBySigScript(usedPublicAddress);
    if transaction has OP_RETURN in the scriptPubKey output then
        merkleTree ← constructMerkleTree(d_i);
        if merkleTree→merkleRoot != transaction→merkleRoot then
            return false;
return true;
```

The first step is to download the compressed Notation3 decisions which have been included in stamping transactions. Afterwards, the verifier downloads in ascending time order all bitcoin transactions (using the *chain.so* bitcoin block reader, available at `https://chain.so/`), related to the used public addresses derived from Diavgeia's master public key. In case of a stamping transaction, the verifier constructs the Merkle tree using the decisions of the first step. If the computed Merkle root is equal to the Merkle root found on the stamping transaction, decisions have remained unmodified.

5.4 Summary

In this chapter, we described the use of the disruptive nature of the bitcoin blockchain on Diavgeia. By implementing Stamper, we presented an inexpensive procedure that may be applied by the administrators of Diavgeia in order to provide immutability to the decisions over time. We discussed the security guarantees of this tool, pointing out the importance of the time gap between two consecutive stamping procedures. By implementing the Consistency Verifier tool we enable interested parties to verify the integrity of the decisions which have been stored on blockchain in an automated way. In the next chapter, we will evaluate the Consistency Verifier tool and we will present the disk space reduction which is achieved by expressing decisions in RDF.

6. EXPERIMENTAL EVALUATION

This chapter presents a scalability evaluation of the Consistency Verifier tool and discusses the disk space reduction that compressed Notation3 files offer. Chapter 6.1. describes the synthetic dataset used in the Consistency Verifier experiment. In Chapter 6.2, we illustrate the details of the test environment and in Chapter 6.3. we discuss the results of the experiment. The disk space reduction is presented in Chapter 6.4.

6.1 Dataset

To simulate the consistency verification process, we generated synthetic compressed Notation3 decisions[1], according to the *Diavgeia ontology*. Firstly, synthetic decisions have 7-17 `Consideration` and `Conclusion` class entities, each one of them has 150-350 random bytes as text part. Moreover, we have included several common-used properties, such as protocol number and thematic categories of a decision, as well as, information related to the departments of government institutions which upload them (phone number, address, etc).

We examine the time it takes an interested party to verify the consistency of Diavgeia in a month's common workload. We consider the scenario in which Diavgeia stores decisions on bitcoin blockchain, once a day. According to the Webpage of Diavgeia (`https://diavgeia.gov.gr/en`), the current rate of uploads is 16000 decisions per working day and assuming a month has 22 working days, we make 22 bitcoin stamping transactions. To examine the scalability of the verifier, we provide 3 different datasets, containing 8000, 16000 and 24000 decisions per day, summing up to 176000, 352000, 528000 compressed N3 decisions, respectively.

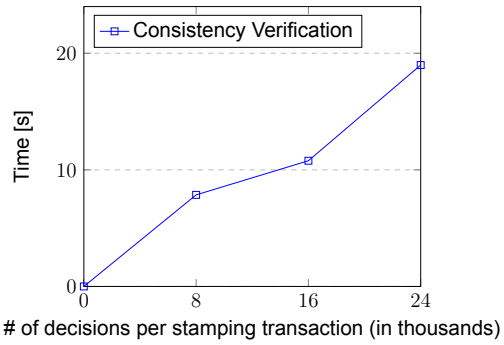
6.2 Test Environment

The verification experiment was run on a MacBook Pro with a 2.9 GHz Intel Core i5 processor and 8GB of memory, since this process may be executed by interested parties with a standard modern computer. The Javascript methods used to measure the elapsed verification time is *console.time* - *console.timeEnd*. The execution time measures the time needed to create the 22 Merkle trees and compare the computed roots with the roots extracted from the stamping transactions. The recorded time does not take into account any network time; the time needed to download synthetic decisions from our Web server or the time needed to gather bitcoin transactions, by making requests to *chain.so*. To account for variability in the testing environment, each reported elapsed time is the average of five independent executions.

6.3 Experimental Results

The experiment consisted of retrieving all synthetic decisions from our Web server and bitcoin transactions from *chain.so* and then compare the corresponding Merkle roots for

[1]Datasets are available in: `https://doi.org/10.6084/m9.figshare.5729292.v1`



#of decisions per stamping tx	Verification Time (s)
8000	7.853
16000	10.780
24000	18.986

Figure 6.1: Evaluation of the Consistency Verifier

validity, as presented in Algorithm 1. We use the 3 different datasets described in Chapter 6.1. The elapsed verification time is plotted in Figure 6.1.

This experiment validates the linear time growth of the Consistency Verifier. The integrity check for a month's regular workload, consisting of 16000 decisions per day, takes about 11 seconds. Even for the extreme case of 24000 decisions per day, the verifier takes approximately 19 seconds to perform the integrity check. These results validate the scalability of our blockchain solution and demonstrate that interested parties can efficiently perform integrity checks over the data of Diavgeia.

6.4 Disk Space Reduction

Diavgeia currently hosts over 26 million decisions, leading to disk space limitations. The average size of a PDF-decision is about 2.5MB, summing up to a total of 65TB. We have created a sample, consisting of equivalent PDF and compressed Notation3 files, for each different decision type of Diavgeia ontology[2]. For the aforementioned sample, we notice that compressed Notation3 decisions are about 86 times smaller, compared to the equivalent PDF files. Hence, encoding decisions in RDF not only allows for sophisticated SPARQL querying, but it also saves space.

6.5 Summary

In this chapter, we presented a scalability evaluation of the Consistency Verifier tool concluding that our scheme is feasible in practice. We open sourced the dataset used on this evaluation and described the testing environment of the experiment in order to easily reproduce the results. By expressing decisions in RDF, Diavgeia can reduce its disk space needs up to 86 times, drastically limiting its operating costs. In the next chapter, we will provide a summary of the work which has been done on DiavgeiaRedefined and we will discuss its possible future work extensions.

[2]The sample is available on: `https://github.com/ThemisB/diavgeiaRedefined/tree/master/rdf/samples`

7. CONCLUSIONS AND FUTURE WORK

In this paper we presented how Greek public sector decisions can be published as linked open data using Semantic Web technologies. We also discussed how to use the bitcoin blockchain to guarantee decision immutability over time. The Diavgeia ontology we employed, is based on the latest European standards and captures the particularities of Greek public sector decisions. We highlighted the importance of intelinking Diavgeia with other publicly available open data, by posing interesting SPARQL queries. We implemented the Web editor and Visualizer components in order to transparently adopt Semantic Web technologies in a user-friendly manner. We also introduced two blockchain tools; Stamper is responsible for storing government data on the bitcoin blockchain and Consistency Verifier provides citizens an automated way to verify the integrity of decisions. Finally, we evaluated the Consistency Verifier and measured the disk space reduction which compressed Notation3 decisions offer over the current PDF-format.

As an initial step of our future work, we would like to proceed with an evaluation of DiavgeiaRedefined, based on user experience of government institututions and citizens and study possible improvements. As a part of functional improvements, we would like to implement a question answering system which will translate natural language queries to SPARQL queries. This system will offer ordinary citizens a way to examine the legality and good administration of Diavgeia, without posing SPARQL queries by themselves.

We acknowledge bitcoin's limitations in terms of cost, speed, and scalability [26]. We utilized bitcoin since it provides a well-established storage mechanism for integrity proof and to demonstrate the feasibility of our solution applied to Linked Data. We would like to apply Stamper and Consistency Verifier to other blockchain technologies, such as Ethereum [6]. In our future work, we will further develop Consistency Verifier. Firstly, we will offer a slower, but safer option of downloading the blockchain, in order to replace the requests made on *chain.so* explorer. We will also implement an inclusion mechanism, to verify that a given decision has remained unchanged. Moreover, the verifier does not take into account the data available through the SPARQL endpoint, meaning that a modification to this data will go unnoticed. We will extend the verifier with the option to perform a full verification procedure which will ensure that data offered through SPARQL endpoint is the same with the compressed Notation3 decisions and therefore same with the stamping transactions of bitcoin.

ABBREVIATIONS - ACRONYMS

W3C	World Wide Web Consortium
RDF	Resource Description Framework
SPARQL	SPARQL Protocol and RDF Query Language
OWL	Web Ontology Language
LOD	Linked Open Data
CSV	Comma-Separated Values
ELI	European Legislation Identifier
L2TAP	Linked Data Log to Transparency, Accountability and Privacy
CSA	Coordination and Support Action
SCOOP4C	Stakeholder Community Once-Only Principle for Citizens
TOOP	The Once - Only Principle Project
CEF	Connecting Europe Facility
DSI	Digital Service Infrastructure
ENLARGE	ENergies for Local Administrations to Renovate Governance in Europe
COSIE	Co-creation of Service Innovation in Europe
ICT	Information and Communication Technology
PDF	Portable Document Format
URI	Uniform Resource Identifier
ISA	Interoperability Solutions for European Public Administrations
EC	European Commision

BIBLIOGRAPHY

[1] Phil Archer, Stijn Goedertier, and Nikolaos Loutas. Study on persistent URIs, with identification of best practices and recommendations on the topic for the MSs and the EC. https://joinup.ec.europa.eu/sites/default/files/document/2013-02/D7.1.3-StudyonpersistentURIs.pdf, 2012.

[2] Massimo Bartoletti and Livio Pompianu. An Analysis of Bitcoin OP_RETURN Metadata. In *Financial Cryptography and Data Security - FC 2017 International Workshops*, volume 10323 of *Lecture Notes in Computer Science*, pages 218–230. Springer, 2017.

[3] Tim Berners-Lee. 5 star deployment scheme. https://www.w3.org/DesignIssues/LinkedData.html.

[4] Bitcoin Wiki. Script Manual. https://en.bitcoin.it/wiki/Script.

[5] Paul Brody and Veena Pureswaran. Device democracy: Saving the future of the internet of things. *IBM, September*, 2014.

[6] Vitalik Buterin. A Next-Generation Smart Contract and Decentralized Application Platform. https://github.com/ethereum/wiki/wiki/White-Paper.

[7] Pompeu Casanovas, Monica Palmirani, Silvio Peroni, Tom M. van Engers, and Fabio Vitali. Semantic Web for the Legal Domain: The next step. *Semantic Web*, (3):213–227, 2016.

[8] Ilias Chalkidis, Charalampos Nikolaou, Panagiotis Soursos, and Manolis Koubarakis. Modeling and Querying Greek Legislation Using Semantic Web Technologies. In *The Semantic Web - 14th International Conference, ESWC 2017*, volume 10249 of *Lecture Notes in Computer Science*, pages 591–606, 2017.

[9] Jordi Cucurull and Jordi Puiggali. Distributed Immutabilization of Secure Logs. In *Security and Trust Management - 12th International Workshop*, volume 9871 of *Lecture Notes in Computer Science*, pages 122–137. Springer, 2016.

[10] Adán Sánchez de Pedro Crespo and Luis Iván Cuende García. Stampery Blockchain Timestamping Architecture (BTA) - Version 6. *CoRR*, 2017.

[11] ELI Task Force. ELI implementation methodology. http://data.europa.eu/doi/10.2830/813167.

[12] Matthew English, Sören Auer, and John Domingue. Block Chain Technologies & The Semantic Web: A Framework for Symbiotic Development. In *Computer Science Conference for University of Bonn Students, J. Lehmann, H. Thakkar, L. Halilaj, and R. Asmat, Eds*, pages 47–61, 2016.

[13] José G. Faísca and José Q. Rogado. Decentralized Semantic Identity. In *Proceedings of the 12th International Conference on Semantic Systems*, pages 177–180. ACM, 2016.

[14] Matias Frosterus, Jouni Tuominen, Mika Wahlroos, and Eero Hyvönen. The Finnish Law as a Linked Data Service. In *The Semantic Web: ESWC 2013 Satellite Events - ESWC 2013 Satellite Events*, volume 7955 of *Lecture Notes in Computer Science*, pages 289–290. Springer, 2013.

[15] Juan A. Garay, Aggelos Kiayias, and Nikos Leonardos. The Bitcoin Backbone Protocol: Analysis and Applications. In Elisabeth Oswald and Marc Fischlin, editors, *Advances in Cryptology - EUROCRYPT 2015*, volume 9057 of *Lecture Notes in Computer Science*, pages 281–310. Springer, 2015.

[16] Gus Gutoski and Douglas Stebila. Hierarchical deterministic Bitcoin wallets that tolerate key leakage. *IACR Cryptology ePrint Archive*, page 998, 2014.

[17] Albert O Hirschman. *Exit, voice, and loyalty: Responses to decline in firms, organizations, and states*, volume 25. Harvard university press, 1970.

[18] Rinke Hoekstra. The MetaLex Document Server - Legal Documents as Versioned Linked Data. In *The Semantic Web - ISWC 2011 - 10th International Semantic Web Conference*, volume 7032 of *Lecture Notes in Computer Science*, pages 128–143. Springer, 2011.

[19] Heng Hou. The Application of Blockchain Technology in E-Government in China. In *26th International Conference on Computer Communication and Networks*, pages 1–4. IEEE, 2017.

[20] Satoshi Nakamoto. Bitcoin: A Peer-to-Peer Electronic Cash System, 2008.

[21] Marek Palatinus and Pavol Rusnak. Multi-Account Hierarchy for Deterministic Wallets. `https://github.com/bitcoin/bips/blob/master/bip-0044.mediawiki`.

[22] Vassilios Peristeras and Konstantinos A. Tarabanis. Reengineering public administration through semantic technologies and a reference domain ontology. In *Semantic Web Meets eGovernment, Papers from the 2006 AAAI Spring Symposium, Technical Report SS-06-06, Stanford, California, USA, March 27-29, 2006*, pages 99–110. AAAI, 2006.

[23] Stein Rokkan. Entries, voices, exits: Towards a possible generalization of the hirschman model. *Information (International Social Science Council)*, 13(1):39–53, 1974.

[24] Reza Samavi and Mariano P Consens. L2tap+ scip: An audit-based privacy framework leveraging linked data. In *Collaborative Computing: Networking, Applications and Worksharing (CollaborateCom), 2012 8th International Conference on*, pages 719–726. IEEE, 2012.

[25] Teemu Sidoroff and Eero Hyvönen. Semantic e-goverment portals-a case study. In *Proceedings of the ISWC-2005 Workshop Semantic Web Case Studies and Best Practices for eBusiness SWCASE05*, volume 7, 2005.

[26] Manu Sporny. LD-DL'17 Workshop Keynote Talk: Building Better Blockchains Via Linked Data. In *Proceedings of the 26th International Conference on World Wide Web*, page 1429. ACM, 2017.

[27] Andrew Sutton and Reza Samavi. Blockchain Enabled Privacy Audit Logs. In *The Semantic Web - ISWC 2017 - 16th International Semantic Web Conference*, volume 10587 of *Lecture Notes in Computer Science*, pages 645–660. Springer, 2017.