Towards a Decentralized, Trusted, Intelligent and Linked Public Sector: A Report from the Greek Trenches

Themis Beris
Dept. of Informatics and Telecommunications, National and Kapodistrian University of Athens
Athens, Greece
tberis@di.uoa.gr

Iosif Angelidis
Dept. of Informatics and Telecommunications, National and Kapodistrian University of Athens
Athens, Greece
iosang@di.uoa.gr

Ilias Chalkidis
Athens University of Economics and Business
Athens, Greece
ihalk@di.uoa.gr

Charalampos Nikolaou
Dept. of Computer Science, University of Oxford
Oxford, United Kingdom
babis.nikolaou@cs.ox.ac.uk

Christos Papaloukas
Dept. of Informatics and Telecommunications, National and Kapodistrian University of Athens
Athens, Greece
christospap@di.uoa.gr

Panagiotis Soursos
Dept. of Informatics and Telecommunications, National and Kapodistrian University of Athens
Athens, Greece
psoursos@di.uoa.gr

Manolis Koubarakis
Dept. of Informatics and Telecommunications, National and Kapodistrian University of Athens
Athens, Greece
koubarak@di.uoa.gr

ABSTRACT

This paper is a progress report on our recent work on two applications that use Linked Data and Distributed Ledger technologies and aim to transform the Greek public sector into a decentralized, trusted, intelligent and linked organization. The first application is a re-engineering of Diavgeia, the Greek government portal for open and transparent public administration. The second application is Nomothesia, a new portal that we have built, which makes Greek legislation available on the Web as linked data to enable its effective use by citizens, legal professionals and software developers who would like to build new applications that utilize Greek legislation. The presented applications have been implemented without funding from any source and are available for free to any part of the Greek public sector that may want to use them. An important goal of this paper is to present the lessons learned from this effort.

CCS CONCEPTS

• Information systems → Distributed storage; **Semantic web description languages**; **• Security and privacy**; **• Computing methodologies → Knowledge representation and reasoning**; **• Applied computing → E-government**; **• Software and its engineering → Peer-to-peer architectures**;

KEYWORDS

linked open data; distributed ledgers; blockchain; open government; semantic web; bitcoin; tamper-proof; public sector services

WWW '19 Companion, May 13–17, 2019, San Francisco, CA, USA

ACM ISBN 978-1-4503-6675-5/19/05.
https://doi.org/10.1145/3308560.3317077

ACM Reference Format:
Themis Beris, Iosif Angelidis, Ilias Chalkidis, Charalampos Nikolaou, Christos Papaloukas, Panagiotis Soursos, and Manolis Koubarakis. 2019. Towards a Decentralized, Trusted, Intelligent and Linked Public Sector: A Report from the Greek Trenches . In *Companion Proceedings of the 2019 World Wide Web Conference (WWW '19 Companion), May 13–17, 2019, San Francisco, CA, USA.* ACM, New York, NY, USA, 10 pages. https://doi.org/10.1145/3308560.3317077

1 INTRODUCTION

Decisions made by governments and public authorities affect significantly the daily lives of ordinary citizens. Therefore, an important dimension of open and trusted government is making these decisions *open and easily accessible* to the public, so that citizens can trust the public organizations that make these decisions, and can easily check whether taxpayer's money is spent wisely.

We believe that Semantic Web, Linked Data and Distributed Ledger Technologies can help realize this vision of open and trusted government. However, the vision cannot be realized by these technologies alone: we also need other technologies (especially other Artificial Intelligence technologies) as we will discuss here.

Our experience in using ontologies, linked data and distributed ledger technologies in the public sector has been gained through project DiavgeiaRedefined[1], an internal project of our team in the National and Kapodistrian University of Athens.[2]

Diavgeia (https://diavgeia.gov.gr/en) [3] 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. Diavgeia is now fully implemented by public authorities in Greece. The current rate of uploads in the portal is up to 42,000 decisions during busy working days, summing up to a total of 32.5 million up

[1] http://pyravlos-vm5.di.uoa.gr/diavgeia/
[2] http://kr.di.uoa.gr/
[3] διαύγεια is the word for transparency in Greek.

to now. These decisions come from more than 4,200 public sector organizations in Greece with over 91,000 accounts on the portal.

Decisions are currently uploaded in Diavgeia as PDF files and follow no structuring of their textual content. As a consequence, interested parties (the government, public authorities, ordinary citizens, non-government bodies, courts, the media, etc.) rely on keyword search over PDF files, in order to find decisions that might affect 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.

Interaction with Diavgeia can take a lot of the precious time of public servants as our administrative colleagues at the National and Kapodistrian University of Athens have told us repeatedly. Combined with the bureaucratic apparatus of Greece, time spent on Diavgeia is, not only slowing down administrative processes, but also making the working life of public sector employees miserable. As an example, our colleagues estimate that getting authorization to travel to an international conference today at our university takes at least 15 days and involves at least 5 very experienced employees for at least 3 person hours.

So, does this means that Diavgeia needs to be abolished? We don't think so! Transparency is an essential element of public life and we need to fight for it as citizens and professionals. It is very important, especially in Greece, which has made a bad name for itself after the 2008 financial crisis, which resulted in the financial bailout of the country and the resulting unprecedented austerity imposed on the Greek people by the troika of lenders. We believe that disruptive technologies from Artificial Intelligence and Distributed Ledgers can actually help alleviate the inefficiencies, both technical and social, that are currently present in Diavgeia and lead to a state where Diavgeia fulfills its original purpose: to enable Greek taxpayer's money to be spent transparently, effectively and within the boundaries of the law.

With our work in DiavgeiaRedefined, we aim at revolutionizing the way that decisions of the Diavgeia program are made public, by following the footsteps of another successful research effort of our team, which we also discuss briefly in this paper. This effort has resulted in the development of the platform Nomothesia [4] that makes Greek legislation available on the Web as open linked data[5] [1, 6]. By applying Artificial Intelligence techniques, we envision a new state of affairs in which public sector employees and ordinary citizens have user-friendly speech- and natural language-based interfaces to government portals, such as Diavgeia, that can help them interact with public sector organizations in the most efficient and effective way. In addition, through the use of techniques from Distributed Ledgers, we enable public sector organizations to decentralize their decision making while at the same time guaranteeing unprecedented levels of trust and transparency. For example, in DiavgeiaRedefined, we use the bitcoin blockchain to enable Greek public sector decisions to remain immutable, introducing a great level of transparency to the Diavgeia program and ensuring the integrity of the published decisions as open linked data.

[4]Νομοθεσία is the word for legislation in Greek.
[5]http://legislation.di.uoa.gr/

Technically, we view public sector decisions as a collection of legal documents which are organized according to an ontology and are encoded in RDF. As a result, decisions can be interlinked among themselves and with other open data in complex ways. For instance, a decision might refer to an earlier version of itself or it might refer to other legislative documents (such as laws or other public sector decisions) that are related to it. In addition, decisions might refer to other open linked data sources, e.g., the administrative divisions of Greece. Linking to other linked open data is crucial, because a software developer is then able to pose queries such as "Find all funds provided by the Greek government to municipalities with a population over 100,000 people" which targets data from the Diavgeia portal but also data from the Greek administrative geography dataset.[6] Our aim is to re-engineer the existing Diavgeia portal so that, not only it enables software developers to develop applications easily using data from Diavgeia, but it also gives ordinary citizens a way to verify public sector decisions are taken according to the Greek law, and that prudent administration and governance practices are followed. By adopting the tamper-proof nature of bitcoin in Diavgeia, we build tools that are able to commit Diavgeia's daily decisions in a single blockchain transaction and are also able to automatically validate the decisions' integrity.

The goal of this paper is to discuss the current state of play of our research efforts DiavgeiaRedefined and Nomothesia. We give more information about DiavgeiaRedefined since it fits perfectly with the theme of this workshop. We also discuss Nomothesia, since the two projects have contributed and continue to contribute to each other. In previous published work by our team, DiavgeiaRedefined has been presented in [2] and Nomothesia in [1, 6].

The rest of the paper is organized as follows. Section 2 discusses related works that combine linked data with blockchain technology. Section 3 presents the current Diavgeia portal and discusses its weaknesses. Section 4 discusses the DiavgeiaRedefined ontology, presents its web editor and visualizer modules and some interesting SPARQL queries. Section 5 describes the two blockchain tools developed for the preservation and verification of decisions. Section 6 presents our evaluation results that show that DiavgeiaRedefined can offer more functionality and efficiency than the current Diavgeia portal. Section 7 discusses the current state of play of DiaveiaRedefined's sister project Nomothesia. Section 8 discusses what other Artificial Intelligence technologies are required to realize the vision of intelligent, trusted, decentralized, linked public administrations and governments. Lastly, Section 9 summarizes the lessons we learned from DiavgeiaRedefined and Nomothesia and our efforts to make them available for free to the Greek public sector.

2 RELATED WORK

The Semantic Web and Linked Data research community has contributed a lot in the last few years in the domain of developing intelligent solutions for the public sector especially in Europe. Currently, the main such activity in Europe is ISA[2] which studies the interoperability of information systems deployed in the public sector in Europe.[7].

[6]http://linkedopendata.gr/dataset/greek-administrative-geography
[7]https://ec.europa.eu/isa2/home_en

The Semantic Web and Linked Data community has also recently begun to consider applications which leverage the distributed, immutable nature of blockchains. In this context, one workshop examining the interaction of these two areas took place in the Web Conference of 2018 in Lyon[8] and it is continued in 2019 by the present workshop.

The article [9] was the first to discuss what Semantic Web research and development can offer to Blockchain research and development and vice versa. The paper [10] examines a semantic approach for decentralized semantic identities (in the context of the W3C WebID effort), 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. The paper [22] proposed a linked-data-based method of utilizing blockchain technology to create tamper-proof audit logs that provide proof-of-log manipulation and non-repudiation.

In [18], Janowicz et al. discuss the value proposition of distributed ledger technologies and crypto-currencies for academic publishing. An informal model of how the Semantic Web journalâĂŹs peer-review workflow could benefit from distributed ledger technologies is provided, while the challenges of implementing a system of this kind are documented. While the problem addressed here is oriented towards academics, the authors argue that the proposed methodology can be applied to any similar science-related challenge.

Finally, Hoffman et al. tackle the challenge of implementing a permissioned blockchain-based system[17] aimed at eliminating losses that tax authorities are currently struggling with. A key innovation of this work is the proposal of a VAT Invoice 2.0 modelled as a Linked Data document. The main idea is that, having an ontology-based tax document and the right permissions, one can investigate the entire commercial chain for any taxable item.

3 BACKGROUND ON DIAVGEIA

In this section, we present the current Diavgeia portal in detail and point out the problems of the current implementation.

3.1 Greek public sector decisions

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[9] 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.

[8]https://sites.google.com/view/lddl-3/home

[9]The following is a fun example. A recent decision, listing the proposals that will be funded under a particular research and innovation call, starts with references to 33 (!) Greek laws. The good news was that 176 proposals were funded; one of them was our new project Choronomothesia which we discuss in Section 7.

Example of an *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".

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 use the aforementioned pattern as a basis to develop the Diavgeia ontology. 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 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.

3.2 Metadata of Decisions

In addition to the uploading of the PDF file, public sector authorities also have to fill metadata information describing 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 OpenDataAPI being a step towards promoting transparency, inconsistency between decision text and metadata information is possible[10]. In our work, we embed metadata information into the RDF document, enforcing consistency.

3.3 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 important, since citizens and other public authorities can use decisions, by solely referring to their unique number. In addition to IUN, each decision is 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, so it is amenable to blockchain technologies as we discuss in Section 5.

4 MODELING DECISIONS USING SEMANTIC WEB TECHNOLOGIES

In this section, we present 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 Section 3.1. It imports and uses properties that are defined in the European Legislation Identifier (ELI) ontology and the Nomothesia ontology. The core[11] of Diavgeia ontology is shown on Figure 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` consists of multiple `Considerations` and `Conclusions`. The `Consideration` class models the passages used to prove the validity of the decision (e.g., the three provisions of the appointment example), while `Conclusion` models the passages 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 describes 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 Section 3.2. By merging metadata and decision text in a single RDF file, this possibility is eliminated.

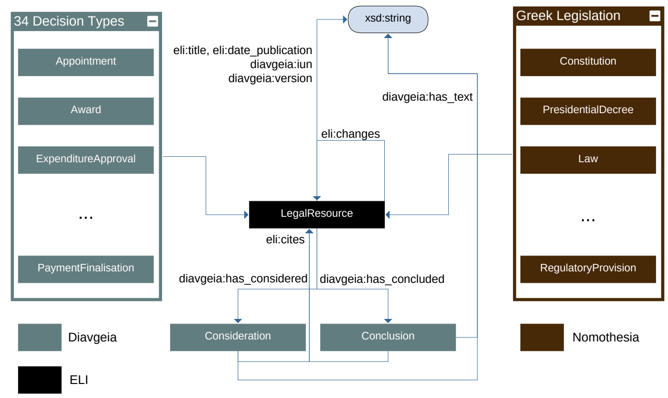


Figure 1: The core of Diavgeia ontology

In order to identify legal resources, we also need appropriate URIs. Persistent URIs is a strongly recommended best practice according to ELI. It is very important to have reliable means to identify the public sector decisions. Based on what is stated in Section 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`.

4.2 Web editor and Visualizer

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 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 as a compressed N3 file in the filesystem of Diavgeia and in Jena Apache's triple store.

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

4.3 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 Section 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}.

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

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

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

4.4 Querying the Resulting RDF Data

By employing the Fuseki Server, we enable the formulation of complex queries over decisions of Diavgeia. This provides interested parties a mechanism to query the decisions of public sector organizations in a much easier way than in the current portal. The following query is such an example:

```
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
```

Retrieve the decisions taken in 2017 involving the 5 highest total amounts.

Our administrative colleagues from our university have told us that the same task in the current Diavgeia portal would take *months* to complete.

5 PRESERVING DIAVGEIA DECISIONS USING DISTRIBUTED LEDGERS

In this section, 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 N3 files in its filesystem and in the triple store.
(2) The administrator of Diavgeia has to decide on the length of time intervals t for the stamping procedure, 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 N3 decisions which have not been stamped yet. Stamper organizes and aggregates these decisions into a Merkle Tree [7], 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[12], maintaining only a chain, a pool, and a *hierarchical deterministic (HD)* wallet [14] based on BIP44 [19].

[12] 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).

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[13], 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.

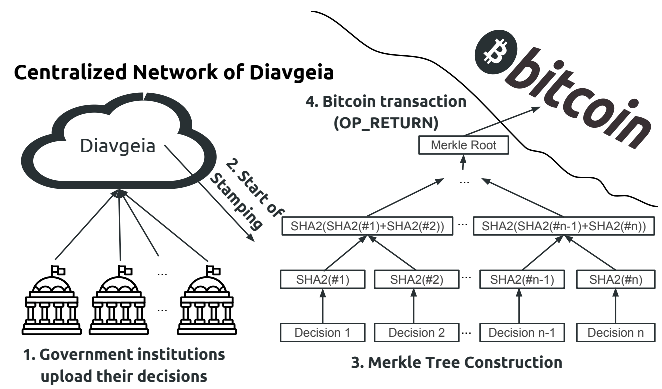


Figure 2: The Stamping procedure

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 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 Section 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.

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

As mentioned in Section 5.1, Stamper uses the open source bitcoin library (bcoin) 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 foul play, these services might be accused of having modified the Merkle root in the first place.

5.3 Consistency Verifier

Consistency Verifier is the tool used by the interested parties to verify the immutability of decisions over time. Algorithm 1 formalizes the steps it takes to verify the integrity of decisions.

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;
        end
    end
end
return true;
```

Algorithm 1: Consistency Verification procedure

The first step is to download the compressed N3 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.

6 EXPERIMENTAL EVALUATION

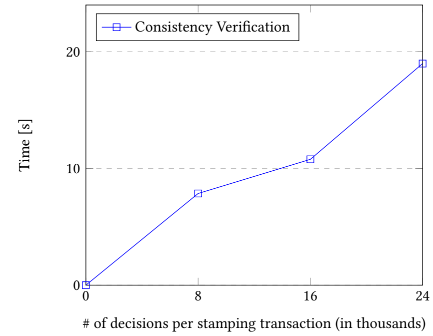
Here, we present a scalability evaluation of the Consistency Verifier tool, discussing the disk space reduction gzip compression of N3 files offers. Section 6.1 describes the synthetic dataset used in the Consistency Verifier experiment.

6.1 Dataset

To simulate the consistency verification process, we generated synthetic gzip N3 decisions[14], 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 also 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 gzip N3 decisions, respectively.

[14] 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 3: Evaluation of the Consistency Verifier

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 elapsed verification time is measured by Javascript's *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 to gather bitcoin transactions, by making requests to *chain.so*. To account for variability in the testing environment, the average of five independent runs is being reported.

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

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.

However, we acknowledge bitcoin's limitations in terms of cost, speed, and scalability [21]. Therefore, we would like to apply Stamper and Consistency Verifier to other blockchain technologies, like Ethereum [4].

6.4 Disk Space Reduction

Diavgeia currently hosts over 31 million decisions, leading to disk space limitations. The average size of a PDF-decision is about 2.5MB, summing up to a total of many TBs. We have created a sample, consisting of equivalent PDF and compressed gzip N3 files, for each different decision type of Diavgeia ontology[15]. For the sample, we find compressed N3 decisions about 86 times smaller, when compared to their PDF equivalent. Hence, decisions encoded in RDF allow for sophisticated SPARQL querying and also save space. In future work we also plan to use ideas from HDT [11] to lower storage requirements even further.

7 THE LINKED-DATA-ENABLED PORTAL NOMOTHESIA

Let us now present the Nomothesia platform which is the sister project of DiavgeiaRedefined.

The development of information systems archiving the content and metadata of legal documents is a common practice towards making legislation easily accessible to the public [5]. To name a few examples, the MetaLex document server [16] offers Dutch national regulations published by the official portal of the Dutch government, while the United Kingdom publishes legislation on its official portal. In the same spirit, [13] presents a service that offers Finnish legislation as linked data. Last, the Publications Office of the EU has developed a central content and metadata repository, called CELLAR, for storing official publications and bibliographic resources produced by the institutions of the EU [12].

Following in the footsteps of other successful efforts in Europe, we aim at modernizing the way Greek legislation is made public. We envision a new state of affairs in which ordinary citizens have advanced search capabilities at their fingertips on the content of legislation. We also envision that legislation is published in a way so that developers can consume it, and so that it can be also combined with other open data to increase its value for interested people. Currently, there is no other effort in Greece that takes this perspective on legislation and related decisions made by government institutions and administration alike.

Unfortunately, Greek legislation documents, published at the web site of the National Printing House of Greece[16], can be found mostly in PDF format, TXT formatted documents existing only for 2000 onwards. To make matters worse, while there are clearly defined instructions on how to write such a legal document, they are rarely followed, rendering any potential attempt to structure the legal text in a more manageable form particularly hard.

To address this, our group has developed a web platform called Nomothesia [6] (http://legislation.di.uoa.gr/), which consists of many components, each specifically dedicated to a proper task in order to create a pipeline of modern technologies. The idea is to automate the process of converting greek legal documents in a useful RDF knowledge graph any interested parties (and programmers) can use to extract useful information, or provide services with.

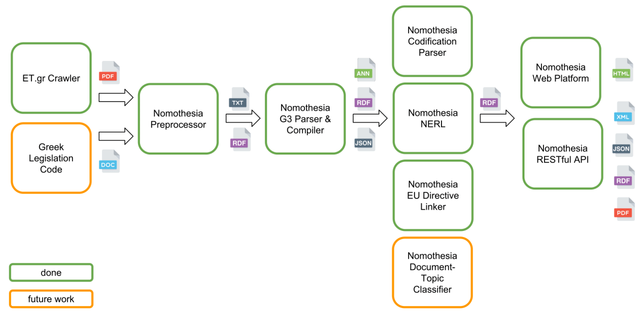


Figure 4: Nomothesia's Pipeline

The first phase of the pipeline, document processing, is achieved by crawling the web site of the National Printing House of Greece, gathering the documents, pre-processing and parsing them in order to produce structured RDF data; expanding upon ELI's ontology, each document is split into increasingly specific parts (books, parts, sections, chapters, articles, paragraphs, indents, lineas, citations), therefore allowing us to generate a complete structure of the original document. During this phase, metadata are also extracted from each document. Finally, two important aspects related to laws are addressed: its timeline and its codification.

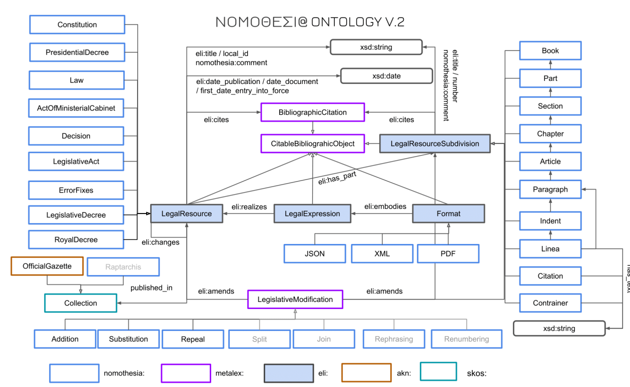


Figure 5: ELI's (expanded) ontology

There are 3 main dates involving a legal document: the date of publication, the date it was signed and the day it is set in force. While in many cases all 3 dates coincide with each other, there are some where they differ. When dealing with the temporal dimension of a law, it is very important to capture all 3 dates to do proper reasoning and inferencing. This is even more evident when tackling the codification of a law.

The codification basically represents all modifications a law has been subjected to, with major types being additions, substitutions and repeals. By having the 3 main dates of a law, it is possible to create a timeline where a law can be shown in the form it had during

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

[16] http://www.et.gr/

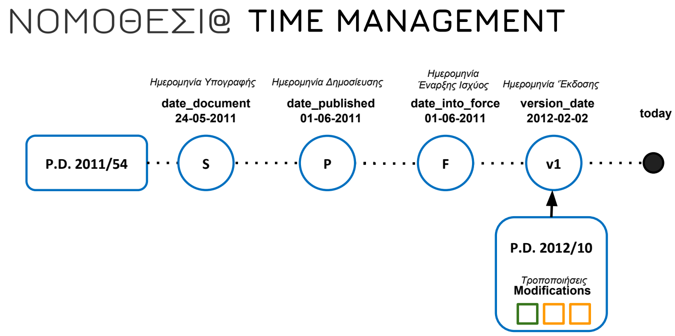


Figure 6: Nomothesia's timeline

a specific point in time. We take advantage of this knowledge in our web platform in order to showcase the legal text of a law in the exact form it is meant to have with all modifications applied to it up to the point in time the user wishes. Also, depending on the kind of modification in question, a different color is being used.

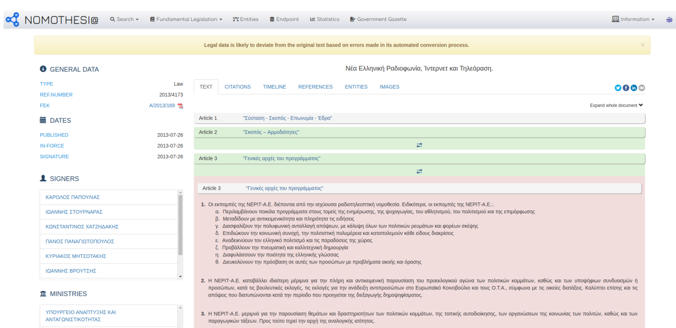


Figure 7: Nomothesia platform: visualizing law/2013/4173

The next phase addresses the information extraction and linking, both of which we thoroughly document in our work of [1]. Employing deep learning techniques, we predict for each token in some text its potential of being a reference to an entity. That reference is realized within a part of the text, but before it is of any use, it must be post-processed due to many abbreviations, cases and accents the greek language contains. When a normalized form is determined for each entity, we use the tool Silk [23] in order to interlink the current data with the third-party datasets of DBpedia Greek politicians, Greek administrative geography and European legislation. The final knowledge graph combines all that and, as a result, provides great querying capabilities.

Currently, Nomothesia contains about 5 million RDF triples which are interlinked in our knowledge graph. About 12,000 documents have been successfully codified, while our knowledge graph contains approximately 195,000 references to entities. We provide all legal documents of issues A and D of the National Printing House for the years 1990-2018, the penal and civil codes of Greece and all european directives and treaties extracted from EUR-Lex[17] in Greek. We provide persistent URIs for all rFesources of the knowledge graph, so developers can utilize RESTful API calls (we offer HTML, PDF, JSON, RDF, XML formats), as well as a SPARQL endpoint to allow direct querying on our knowledge graph, while we make the individual datasets we interlinked publicly available[18].

Currently, in collaboration with a company of landscape engineers in a Greek project called Choronomothesia, we aim to extract geospatial information as well, including maps, tables of coordinates etc. This is even harder than information extraction from text, because most maps or tables of coordinates are within images, occasionally handwritten and frequently maps are split among pages. This necessitates the usage of image analysis techniques to tackle the problem. However, extreme caution is required, because everything else is being extracted from text, while this information will be from images and, in order to end up with a meaningful knowledge graph, the two must be integrated. To the best of our knowledge, all the challenges faced by Choronomothesia are open problems in the relevant literature so far.

8 OTHER ARTIFICIAL INTELLIGENCE TECHNOLOGIES FOR INTELLIGENT PUBLIC ADMINISTRATIONS AND GOVERNMENTS

The previous sections discussed how linked data and distributed ledger technologies can help public administrations and governments become decentralized, trusted and linked with the aim of serving citizens in a more effective way. We now turn to discuss other Artificial Intelligence technologies that can also help towards the vision of making public administrations more *intelligent* than they are today. In our opinion, the following three intertwined technologies can offer very much in this direction:

- *Extracting information from legacy public sector documents.* Even if the Greek public sector decided to adopt DiavgeiaRedefined today, the decisions submitted as PDFs to the current portal in the last eight years are an untapped source of information. What can we do to extract useful information from these PDFs? We have already tackled this problem in Nomothesia platform using deep learning techniques [1] as we discussed in Section 7 already. Unfortunately, it is far from clear how to do this for Diavgeia. While it is challenging to address the intricacies of each issue of legal documents of the National Printing House in Nomothesia, as well as extracting information from PDF format, at least the documents' structure follows patterns that can make it easier to at least classify which documents would contain what information and in which potential places. In the documents of Diavgeia, this is not the case at all, as the documents are unstructured, each can contain classes of information others do not and, as a result, pose the challenge of finding a way to extract useful information from them reliably.
- *Knowledge graphs.* General knowledge graphs like DBpedia, Yago and Wikidata should be interlinked with public administration and government knowledge graphs. This will allow e.g., for sophisticated queries like "Which European countries invest in Computing and Artificial Intelligence less than what a United States technological university, based

[17] https://eur-lex.europa.eu/homepage.html

[18] http://legislation.di.uoa.gr/data

in Boston, is investing in a new college covering the same research area?". To answer this query, one has to combine information, e.g., from Wikidata, that shows that MIT will be investing 1 billion dollars to a new Computing and Artificial Intelligence college[19] and compare this amount with the total investment in the same research area by various European countries. The former information can probably be found in a general knowledge graph such as Wikidata, while the latter can be found in various government knowledge graphs.

- *Multilingual question answering.* Ordinary citizens and public sector employees do not speak SPARQL, although we wish they did; instead, they converse in natural language. Hence, current work in question answering over knowledge graphs and linked data [8] is fundamental in making public administration and government information more easily available to citizens. Work on question answering has evolved a lot in the last 10 years [8], but there are at least three kinds of questions that the currently available question answering systems do not deal with effectively: temporal questions (see TEQUILA for a nice start in this area [24]), questions involving quantities [24] and multilingual questions (see AMUSE [15] for a successful effort in this area). As an example, being able to deal with quantities is needed if one wants the question answering engine to be able to answer questions like the one mentioned in the previous paragraph, which involves finding information about a quantity (investment) and computing the total of such quantities.
- *Spoken dialogue systems.* When a citizen interacts with a public sector employee today, the interaction is through spoken dialogue. For example, if you request information about your submitted tax return in your local government office, the relevant public sector employee might ask for your tax identification number. Subsequently, if the answer to your enquiry was not the one expected, you might rephrase your question and ask again. Enter spoken dialogue systems! Spoken dialogue systems is a technology which can have immense applications in the public sector. Spoken dialogue systems, e.g., chatbots, can be there all the time, even late at night, when a citizen wants to request information about a transaction with the government e.g., information about one's tax return. To the best of our knowledge, spoken dialogue systems are not currently used in any part of the Greek public sector although their use, in the form of chatbots, is widespread in the private sector (e.g., in banks or telecom organizations etc.). However, chatbots are already in use in the public sector of other countries. For example, EMMA is a chatbot utilized by the U.S. Citizens and Immigration Services of the Dept. of Homeland Security.[20]
  Developing spoken dialogue systems is an important area in Artificial Intelligence research today, and advances to the state of the art in this topic are presented in many high quality conferences and journals, most prominently at SIGDIAL, the annual conference of the joint ACL/ISCA Special Interest Group on Discourse and Dialogue. The paper [20] which appeared in last year's SIGDIAL discusses spoken dialogue systems for navigating information expressed as linked data, and its ideas are certainly applicable to spoken dialogue systems for the public sector in Greece and elsewhere.

Our current efforts in both DiavgeiaRedefined and Nomothesia are concentrating on the above four research areas.

9 CONCLUSIONS

Let us now summarize what we learned from developing DiavgeiaRedefined and Nomothesia. Since we have already discussed technical improvements that we plan in these two systems in our previous papers [1, 2], this section focuses on the *social* aspects of our work, especially the lessons learned regarding the adoption of DiavgeiaRedefined and Nomothesia by the Greek public sector.

The first lesson that we learned is that, once you develop a system which disruptively effects in the public sector of your country and publish a couple of papers about it, you are not done. In fact, this is when the hard work of persuading the public sector to deploy an operational system based on your research and development starts. This is especially hard in a country like Greece where, in our opinion, the bureaucratic apparatus of the public sector typically fails to easily embrace innovative ideas coming from outsiders, including Greek universities. Our DiavgeiaRedefined and Nomothesia experiences are a case in point, as we explain above.

The first implementation of Nomothesia was completed by our group in September 2014. Since then, four more undergraduate theses and one M.Sc. thesis resulted in the creation of the current version of the portal. In April 2016, Nomothesia was presented and received an award in the 1st IT4GOV contest, which was organized by the Greek Ministry of Administrative Reform and Electronic Governance. Among the contest jury members were three ministers and people from the industry, who judged Nomothesia positively. Following this contest, we had been in contact with officials from the House of Parliament and the Ministry of Administrative Reform and Electronic Governance, who have expressed interest in deploying Nomothesia internally in the ministry. We have also presented Nomothesia to the Expert Group on European Legislation Identifier in Brussels on October 16, 2017. In all presentations, we have received very encouraging comments that motivated us to continue working on this topic. In March 2018, the Ministry of Administrative Reform (now without the "Electronic Governance" in its name) participated with us in a joint grant proposal in which DiavgeiaRedefined and Nomothesia would be the starting point of our joint work. The grant proposal was ranked very positively by the reviewers, but it was not selected for funding by the the funding agency due to the small budget of the call. In December 2018, we made a presentation to the Athens Bar Association about our platform. The association has an operational system called Isocrates [21] which is used by many lawyers in Greece today. In this meeting we explored various avenues for collaboration. In December 2018, we also made a detailed presentation to the Ministry of Administrative Reform (now under a new minister). We again stressed that we are happy to make the Nomothesia platform available to them for free, and that we think that both our research group and the ministry

[19] http://news.mit.edu/2018/mit-reshapes-itself-stephen-schwarzman-college-of-computing-1015
[20] https://www.uscis.gov/emma
[21] http://www.dsanet.gr/

would benefit from our joint collaboration. In general, through our interaction with the Greek public sector, we have realized that the Greek public sector is now ready to introduce Semantic Web and Linked Data technologies in the whole chain of law production which starts at the relevant ministry and ends with the approval of the law from the Hellenic Parliament, but there is no such readiness (or approved government policy) regarding the re-engineering of the transparency portal of Diavgeia.

The second lesson that we learned is that ordinary Greek citizens, with some knowledge of what Semantic Web, Linked Data and Distributed Ledger technologies can do for the public sector today, immediately understood the value of our proposals and were very supportive. Although we have received mostly positive feedback from people we did not personally know, we also received some constructive criticism. Some, for example, suggested that if the DiavgeiaRedefined ideas are adopted in the operational system of the Greek public sector, the only people to benefit will be Greek journalists working in the national yellow press who, together with a SPARQL expert, would then be able to pose queries to the portal and retrieve results much more easily. These results would then be interpreted in any way which fits the headlines that they want to have in their newspapers. This made us realize that the introduction of disruptive technologies, like Artificial Intelligence and Distributed Ledgers, in the public sector needs to be approached using a *participatory design approach* so that its impact becomes apparent to all stakeholders right from the start.

Finally, the third lesson that we learned is that it is important to teach Semantic Web and Linked Data technologies to graduate students in our universities. The desire to see Semantic Web and Linked Data technologies taught as part of the standard curriculum in Computer Science departments was also expressed by the participants of the recent panel “Enterprise Knowledge Graphs” which was organized by Natasha Noy at ISWC 2018. [22]

We close this paper with some advice to the Semantic Web and Linked Data community of which we are proud to be members. While the rest of the world seems determined to equate Artificial Intelligence with Deep Learning, we, as researchers working on Semantic Web and Linked Data, should stay alert and make a strong case for our technical contributions in the last 20 years that have resulted in robust technologies that are now in operation in many countries (especially in Europe) and in many applications. We should try to collaborate with researchers form other disciplines e.g., researchers from Distributed Ledger technologies and try to apply our technologies to areas where they can improve the lives of our fellow citizens. Our team will try to do this in the next three years in the context of the large European project AI4EU which involves 79 partners from academia, research and industry in Europe and which has a focus on *human-centered* Artificial Intelligence.[23]

10 ACKNOWLEDGEMENTS

The present work was co-funded by the European Union (ERDF) and Greek national funds through the Operational Program “Competitiveness, Entrepreneurship and Innovation”, under the call “RESEARCH - CREATE - INNOVATE” (GeoREgC, project code: T1EDK 01000). We thank our colleagues Giorgos Apostolopoulos, Realdo Gatzi and Stavros Yannakis for their contributions to the most recent version of Nomothesia. Manolis Koubarakis would also like to thank Martin Kaltenböck, Pierre Maret and Stavros Vassos for interesting discussions regarding the Artificial Intelligence techniques discussed in Section 8 and their use in the public sector. Last, but not least, we would like to thank our colleagues Lydia Themeli, Olivia Katrakazi and Mirto Tsapatsari, who have spent an enormous number of hours fighting the Greek bureaucracy.

[22] http://iswc2018.semanticweb.org/panel-enterprise-scale-knowledge-graphs/
[23] http://ai4eu.org/

REFERENCES

[1] I. Angelidis, I. Chalkidis, and M. Koubarakis. 2018. Named Entity Recognition, Linking and Generation for Greek Legislation. In *JURIX*.
[2] T. Beris and M. Koubarakis. 2018. Modeling and Preserving Greek Government Decisions Using Semantic Web Technologies and Permissionless Blockchains. In *ESWC*.
[3] T. Berners-Lee. 2006. 5 star deployment scheme. https://www.w3.org/DesignIssues/LinkedData.html.
[4] Vitalik Buterin et al. 2014. A next-generation smart contract and decentralized application platform. *white paper* (2014).
[5] P. Casanovas, M. Palmirani, S. Peroni, T. M. van Engers, and F. Vitali. 2016. Semantic Web for the Legal Domain: The next step. *Semantic Web* 3 (2016).
[6] I. Chalkidis, Ch. Nikolaou, P. Soursos, and M. Koubarakis. 2017. Modeling and Querying Greek Legislation Using Semantic Web Technologies. In *ESWC*.
[7] J. Cucurull and J. Puiggali. 2016. Distributed Immutabilization of Secure Logs. In *STM*.
[8] D. Diefenbach, V. López, K. D. Singh, and P. Maret. 2017. Core techniques of question answering systems over knowledge bases: a survey. *Knowl. Inf. Syst.* (2017).
[9] M. English, S. Auer, and J. Domingue. 2016. Block Chain Technologies & The Semantic Web: A Framework for Symbiotic Development. In *CSCUBS*.
[10] J. G. Faísca and J. Q. Rogado. 2016. Decentralized Semantic Identity. In *SEMANTICS*.
[11] J. D. FernÃąndez, M. A. MartÃŋnez-Prieto, C. GutiÃľrrez, A. Polleres, and M. Arias. 2013. Binary RDF Representation for Publication and Exchange (HDT). *Journal of Web Semantics* (2013).
[12] E. Francesconi, M. Wilhelm Küster, P. Gratz, and S. Thelen. 2015. The Ontology-Based Approach of the Publications Office of the EU for Document Accessibility and Open Data Services. In *EGOVIS*.
[13] M. Frosterus, J. Tuominen, M. Wahlroos, and E. Hyvönen. [n. d.]. The Finnish Law as a Linked Data Service. In *ESWC 2013 Satellite Events*.
[14] G. Gutoski and D. Stebila. 2014. Hierarchical deterministic Bitcoin wallets that tolerate key leakage. *IACR* (2014).
[15] S. Hakimov, S. Jebbara, and P. Cimiano. 2017. AMUSE: Multilingual Semantic Parsing for Question Answering over Linked Data. In *ISWC*.
[16] R. Hoekstra. 2011. The MetaLex Document Server - Legal Documents as Versioned Linked Data. In *ISWC*.
[17] M. R. Hoffman. 2018. Can Blockchains and Linked Data Advance Taxation. In *WWW*.
[18] K. Janowicz, B. Regalia, P. Hitzler, G. Mai, S. Delbecque, M. Fröhlich, P. Martinent, and T. Lazarus. 2018. On the prospects of blockchain and distributed ledger technologies for open science and academic publishing. *Semantic Web* (2018).
[19] M. Palatinus and P. Rusnak. 2018. Multi-Account Hierarchy for Deterministic Wallets. https://github.com/bitcoin/bips/blob/master/bip-0044.mediawiki.
[20] A. Papangelis, P. Papadakos, Y. Stylianou, and Y. Tzitzikas. 2018. Spoken Dialogue for Information Navigation. In *SIGdial*.
[21] Manu Sporny. 2017. LD-DL'17 Workshop Keynote Talk By Many Sporny: Building Better Blockchains Via Linked Data. In *WWW (Companion Volume)*.
[22] A. Sutton and R. Samavi. 2017. Blockchain Enabled Privacy Audit Logs. In *ISWC*.
[23] J. Volz, Ch. Bizer, M. Gaedke, and G. Kobilarov. 2009. Silk - A Link Discovery Framework for the Web of Data *(CEUR)*.
[24] G. Weikum. 2018. Beyond Knowledge Graphs: New Frontiers of Machine Knowledge. The Web Conference Workshops.