Linked Open Data for Legislative Domain – Ontology and Experimental Data*

Martin Nečaský, Tomáš Knap, Jakub Klímek, Irena Holubová, and Barbora Vidová-Hladká

Charles University in Prague
Faculty of Mathematics and Physics
necasky@ksi.mff.cuni.cz
http://xrg.cz/

Abstract. In this paper, we present an approach of publishing legislative content as Linked Open Data (LOD). LOD is a set of principles of publishing data on the Web in a machine-readable way so that links between different data sets, possibly published by different publishers, can be created. Therefore, LOD enable not only to publish data but also enrich with other existing data published according to the principles. We present what is the motivation for publishing legislation as LOD and what benefits can be gained. We then introduce a legislative ontology which builds on existing commonly used ontologies. We also show how we converted existing sources of legislation in Czech Republic to LOD.

Keywords: Linked Data, legislation, ontologies, Semantic Web.

1 Introduction

Linked Open Data (LOD) [3] is a set of principles of publishing data on the Web in a machine readable way which enables linking data of various publishers at different places on the Web. Linking means that publishers enrich their data with other data published on the Web. A data consumer can then work not only with the data of a particular publisher but can also work with linked enriching data of other publishers. The links can be created by one of the data publishers of the linked data sets or by any third party publisher.

In this paper, we show how legislation can be published as Linked Open Data. We work with legislation of the Czech Republic, namely acts issued by the National Parliament of Czech Republic, decrees and regulations issued by the National Government of the Czech Republic, and court decisions published by the Supreme and Constitutional Court of Czech Republic. These documents are generally called *sources of law*. They are published by the authorities at various places on the Web in a distributed way in different formats (usually HTML or PDF). At some places only metadata about legislative documents are published. At other places documents themselves are published. Those different places are

* This work was supported by the TAČR grant no. TA02010182.

W. Abramowicz (Ed.): BIS 2013 Workshop, LNBIP 160, pp. 172–183, 2013.

not linked in any way. Therefore, it is hard for citizens to find legislative documents they need and to search for related content. We propose how legislative documents in Czech Republic can be published as LOD and demonstrate advantages of such style of publication. We propose an ontology for legislative documents which is based on existing ontologies already exploited for publishing legislative documents. We enrich them with new specific parts. Then we show how current sources of legislative documents can be transformed to their LOD representation and how the created LOD representation can be exploited by application developers.

There are currently several projects which aim at publishing legislative documents as Linked Data. The project CEN MetaLex [4] provides a standard of how sources of law and references to sources law are to be represented in XML. It also provides a LOD variant of their publication standard. An ontology for publishing legislation as LOD has been provided by the project. MetaLex publication server[1] publishes Dutch legislation according to the standard. A similar project is Legislation.Gov.UK[2] which also publishes legislation as LOD according to MetaLex format. In our work we reuse the rules given by CEN MetaLex and apply them to publishing legislation in Czech Republic. We extend the publication format so that other kinds of relationships not only between acts (as supported by MetaLex) but also other sources of law (e.g., court decisions) can be recorded.

2 Motivation

Because of the complexity of various sources of law, it is very hard for users, i.e. legal professionals as well as common citizens, to search for the required documents. In this section, we demonstrate most common use cases of how professionals as well as citizens work with sources of law.

Sources of law are usually structured to sections which may contain further subsections. Moreover, a source of law may contain references to other sources. A reference may target not only a whole source but also its particular section. Therefore, the structure encoded in sources of law and references between them form a complex network. Moreover, related sources of law are often published by different public authorities. It would be useful to enable to browse this distributed network among different data sources and search for relationships between sources of law and/or their parts. Examples of common use cases are listed in the following list. For the demonstration, we use Figure 1. It shows a part of the network comprising several related acts and court decisions.

1. A user is reading a particular section of an act (e.g., Act Section § `10` of Act `§99/1963`). He would like to see what court decisions have been made in the last decade related to this particular section (i.e., decisions `20 Cdo 1691/2005` and `20 Cdo 1691/2005`).

[1] `http://doc.metalex.eu/`
[2] `http://www.legislation.gov.uk/`

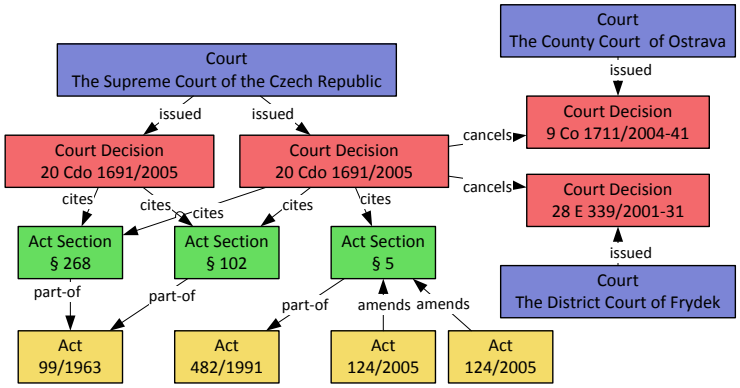


Fig. 1. Sample Links between Sources of Law

2. A user is reading a particular section of an act (e.g., Act Section § `5` of Act `§482/1991`). He would like to find out what amendments correcting the act the chosen section belongs to came to force in 2005 (i.e., amendment of Act `482/1991` defined by Act `124/2005`).
3. A user received a court decision (e.g., `20 Cdo 389/2004`). There are various references to other court decisions and also sections of acts and amendments encoded in the text. He would like to see the reading of each of the referenced decisions and sections.

All these use cases are problematic because sources of law are published as textual documents by various authorities at different places of the Web. Moreover, the sources are not interlinked at all. Their logical structure (sections and their subsections) and links between them are encoded in the text in a way which can only be interpreted by a human. Therefore, the user can only read the sources and has to search for relationships manually. This is very time consuming, cumbersome and the user omits important relationships very easily.

3 Linked Open Data Representation of Czech Legislative Documents

Linked Open Data (LOD) is a set of principles of publishing data on the Web in a machine-readable form which enables to link related data. The links are recorded in a machine readable form and published on the Web as well as part of the data itself. LOD principles are simple:

- Use URIs to denote things.
- Use HTTP URIs so that things can be referred to and looked up by people and machines.
- Provide useful information about the things when its URI is looked up (use standards such as RDF [2] and SPARQL [6]).

– Include links to other related things (using their URIs) when published on the Web.

In our case, things are sources of law and their parts. Useful information about the things are metadata about the sources and their part (e.g., creation date, data from which a source is valid, author/publisher of the source, etc.). Links between the things are relationships between the sources (e.g., a section *is part of* an act, an act *amends* another act, court decision *cites* a section of an act, court decision *cancels* another court decision, etc.). Applying the principles to the legislative domain therefore means assigning HTTP URIs to sources of law and their parts, representing their metadata as RDF, extracting relationships among the sources of law from their original textual expressions and publishing all the metadata together with the links in RDF form so that the sources of law and their parts can be accessed via dereferencing their HTTP URIs or using SPARQL query language.

3.1 LEX Ontology

The first step in our effort to publish sources of law as LOD was designing an ontology. We call it *LEX ontology*. Its core classes and properties are depicted in Figure 2. Before we designed the ontology we analyzed the legislation domain and produced so called domain model. Then we identified what parts of the domain model had already been covered by existing ontologies. We found out that various parts were already covered by ontologies like FRBR[3] or Dublin Core. For these parts we therefore did not created new classes and properties but we reused existing classes and properties from these ontologies. On the other hand, various parts of the domain were not covered by existing ontologies and we had to develop new classes and properties. The resulting LEX ontology is, therefore, a mixture of existing ontologies and our new components.

There also exist various other approaches to describing the legal domain with ontologies. For example, in [5] on ontology for legal concepts was designed. The ontology is called LKIF. It is a detailed ontology of various legal concepts like roles (epistemic roles, functions, person roles, etc.), actions (processes which are performed by some actors in roles), etc. Paper [1] describes the results of a study consisting of two tasks: (i) how the "obligation" Fundamental Legal Concept is differently represented in the FrameNet[4] resource, in terms of Semantic Frames, and (ii) how the concept of "public function" stemmed from the "obligation" Fundamental Legal Concept can be ontologically characterized. However, such ontologies are too detailed for our purposes – our goal is to represent the structure of sources of law and relationships between them. These detailed ontologies aim at representing the semantics encoded in legislation which is a level beyond our current efforts.

Let us now describe the most important classes and properties of LEX ontology in a more detail. As we have already noted, legislation is defined by written

[3] Functional Requirements for Bibliographic Records (FRBR) `http://www.loc.gov/cds/downloads/FRBR.PDF`

[4] `https://framenet.icsi.berkeley.edu/fndrupal/`

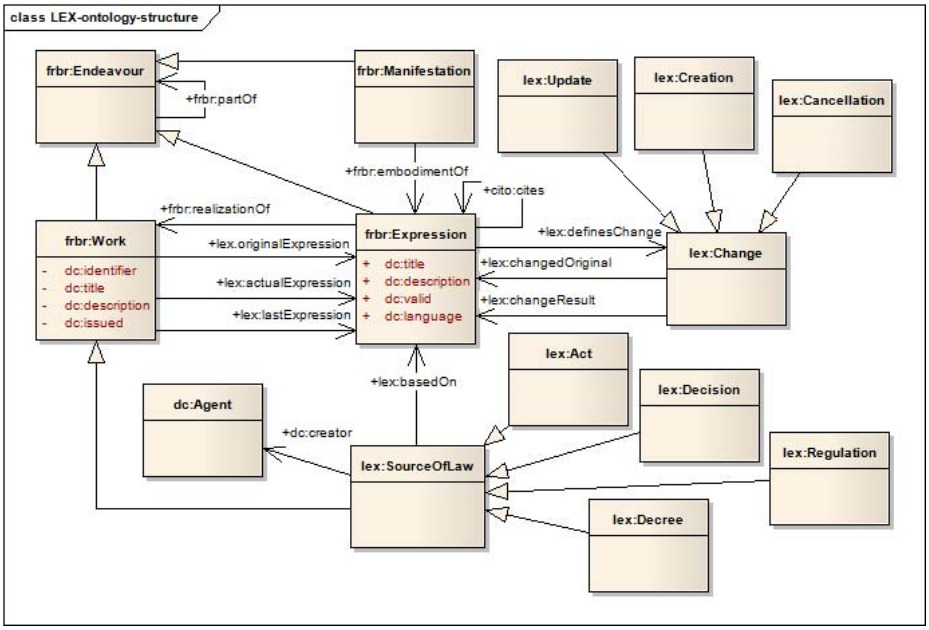


Fig. 2. LEX Ontology

documents called *sources of law*. There are different kinds of sources of law for which we introduce the following classes in our ontology: `lex:SourceOfLaw` for sources of law of all kinds (superclass of all other classes), `lex:Act` for acts, `lex:Decree` for decrees, `lex:Regulation` for regulations, and `lex:Decision` for court decisions.

Sources of law of most kinds (except of court decisions) exist in different versions. Some versions are outdated, at most one version is currently valid, and some versions are enacted but have not come to force yet. A particular version may have several physical embodiments, i.e. may be embodied in different physical documents (of different formats, e.g. HTML, PDF, XML, RDF, etc.). From this viewpoint, it is reasonable to represent a source of law as an abstract notion of intellectual creation which is independent of particular versions of the source. Moreover, each version of the source as well as its each physical embodiment should have representation on its own. This logic is built into the LEX ontology. However, we do not introduce own ontological constructs but reuse the Functional Requirements for Bibliographic Records (FRBR) ontology. We reuse the following three FRBR classes: `frbr:Work` for abstract notions of an intellectual creation which are sources of law, `frbr:Expression` for particular versions of sources of law, `frbr:Manifestation` for particular documents which are physical embodiments of particular versions of sources of law.

The usage of FRBR allows to distinguish a source of law itself, its particular versions and their physical embodiments. From the linked data point of view, it is therefore possible to link and query the source of law as an abstract entity

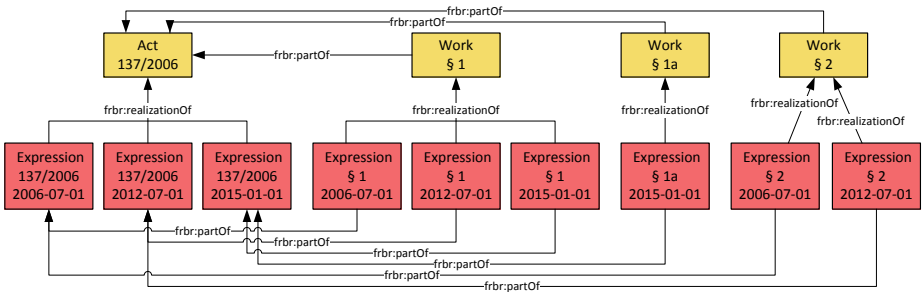


Fig. 3. Sample LEX Representation of an Act

which is independent of particular versions of the source. It is also possible to link and query its particular versions.

We also reuse two FRBR properties: `frbr:realizationOf` to link a version (member of `frbr:Expression`) to its source of law (member of `frbr:Work`), and `frbr:embodimentOf` to link a document (member of `frbr:Manifestation`) to a version of a source of law it is embodiment of (member of `frbr:Expression`).

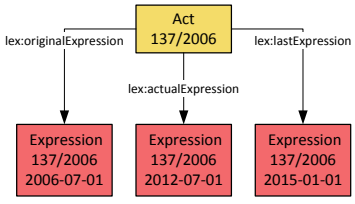
Figure 3 shows a sample representation of a *Public Contracts Act* which is a part of the law of Czech Republic. The boxes in the upper part (yellow ones) are resources which are instances of `frbr:Work`. The left-most box is a resource representing the act itself. Therefore, it is not an instance of `frbr:Work` directly but an instance of its subclass `lex:Act` which is intended for acts. The other boxes at the upper levels represent parts of the act. These are represented simply as instances of `frbr:Work` linked to the resource representing the whole act. The fact that a given resource represents a part of an act cannot be recognized from its type. It is recognized from the fact that it is an instance of `frbr:Work` and that it is a part of a resource which is an instance of `lex:Act`. Next, there are various instances of `frbr:Expression` at the lower layer of the figure depicted (red boxes). Each such resource is linked to its corresponding instance of `frbr:Work` (or its subclass `lex:Act`, respectively) via the property `frbr:realizationOf` and models a particular version of that instance. In other words, the instances of `frbr:Expression` linked to the resource representing the act represent different versions of the act. Concretely, there is depicted a version of the act valid from `2006-07-01`, another version valid from `2012-07-01` and a version which will be valid from `2015-01-01`. The same is for the parts of the act. For part §1, we have three versions. In other words, the part has been present in all versions of the act. On the other hand, the part §2 has only expressions valid from `2006-07-01` and `2012-07-01`. In other words, this part will be canceled in the version of the act valid from `2015-01-01`. Part §1a was added to the version of the act valid from `2012-07-01`. It was not present in the first version valid from `2006-07-01` as can be seen from the figure.

The example demonstrates the advantages of our proposed representation of sources of law and their parts. Anyone can link his or her data to as an

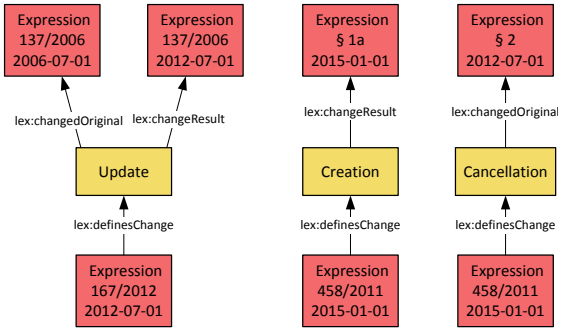
abstract concept representing a source of law or any of its parts independently of a particular version. For example, it is possible to link §2 of *Public Contracts Act* independently of a particular version. But, it is also possible to link a particular version if necessary, e.g., §2 of *Public Contracts Act* valid from `2012-07-01`. This reflects real situations. In some situations we only know that, e.g., an act or its part is referenced. In that case we need to link an abstract concept (i.e. an instance of `frbr:Work`). In other situations we know exactly which version of the source of law is referenced. In that case we need to link a particular version (i.e. an instance of `frbr:Expression`).

For a given source of law we need to know its currently valid version, original version (i.e. the first version), and the last enacted version (which have not necessarily needed to come to force yet). For this, there is no corresponding property provided by FRBR. Therefore, we introduced three new properties in the LEX ontology: `lex:originalExpression` to link the original (first) version to the respective source of law, `lex:actualExpression` to link the currently valid version to the respective source of law, and `lex:lastExpression` to link the last enacted version to the respective source of law.

Figure 3.1 (a) shows *Public Contracts Act* as an instance of `lex:Act` and links to its original, actual and last expression. The links specifies that the first (original) version was the expression valid from `2006-07-01`. The actually valid expression is the one valid from `2012-07-01`. It will be later replaced by the last enacted expression in `2015-01-01`.



(a) Sample representation of original, actual and last expression of a work according to LEX ontology



(b) Sample representation of changes in an act according to LEX ontology

A source of law can change another source of law. For example, an act can be an amendment of one or more another acts. The same is true for other kinds of sources of law. A regulation or decree can change another regulation or decree, respectively, and a court decision can cancel another court decision (in case of court decisions, the situation is more specific – only cancelation is possible). For representing changes LEX ontology proposes a new class `lex:Change`. A change (instance of `lex:Change`) is defined by some source of law

in some version. Therefore, the instance is referred from the respective expression (instance of `frbr:Expression`) which represents that version (via property `lex:definesChange`). Each change says that a given version of a given source of law (i.e. instance of `frbr:Expression`) is changed and a new version of the same source of law (i.e. another instance of `frbr:Expression` linked to the same instance of `frbr:Work` as the previous version) is created. Therefore, the resource representing the change is linked to the expression which is changed via property `lex:changedOriginal` and to the expression which is the result of the change via property `lex:changeResult`.

We distinguish three kinds of changes. LEX ontology proposes three respective subclasses of `lex:Change`: `lex:Creation` to express that an expression of a new source of law or new part of a source of law has been created by the defining expression; only `lex:changeResult` is present to link to the created expression, `lex:Cancellation` to express that an expression of an existing source of law or existing part of a source of law has been canceled by the defining expression; only `lex:changedOriginal` is present to link to the canceled expression, and `lex:Update` to express that an expression of an existing source of law or existing part of a source of law has been amended (updated) by the defining expression; both `lex:changedOriginal` and `lex:changeResult` are present to link the changed expression and the resulting expression.

Figure 3.1 (b) shows sample changes. The left-most update is an update. It is defined by an expression of act `167/2012` valid from `2012-07-01`. The update changes an expression of the act `137/2006` valid from `2006-07-01` to a new expression of the same act valid from `2012-07-01`. The second change is creation. It is defined by act `458/2011` valid from `2015-01-01`. The change creates a new expression which represents a new part §`1a` of the act `137/2006` which has not existed previously. The last change is cancelation. It is again defined by act `458/2011` valid from `2015-01-01`. The change cancels an expression which represents an existing part §2 of the act `137/2006`.

3.2 Experimental Linked Data

On the base of LEX we transformed available sources of law available in Czech Republic to Linked Data representation according to the ontology. There are several web sites where data about sources of law can be accessed in Czech Republic. First, it is the web site of The National Parliament. Here, meta-data about each act, decree and regulation can be obtained in a form of HTML pages. It is also possible to get information that a given source of law changes another source of law (i.e. updates, creates or changes). However, this is only at the level of whole sources of law. The parliament does not provide information about parts of the sources of law. Another source of information is the National Governmental Portal. Here, it is possible to access original expressions of all acts, regulations and decrees. Also, consolidated versions of some of these sources of law can be accessed here (approx. 20 % of all of published sources of law). However, the national portal only provides texts, there are no information about changes published.

We have already processed both sources of data and translated them to LOD. The third place where sources of law can be obtained are web sites of courts in Czech Republic, namely the Supreme Court of the Czech Republic and The Constitutional Court of the Czech Republic. Both publish their decisions on their web sites. Other courts in the Czech Republic do not publish their decisions on their web sites. However, we have not processed this sources yet.

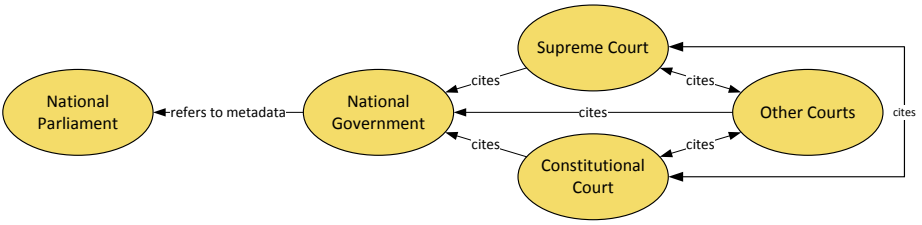


Fig. 4. Datasets converted to LOD according to LEX ontology

The goal of our work was to transform these available sources to Linked Data. For each data source mentioned above we created its RDF equivalent. In the following subsections, we describe translation of each data source in a more detail. Data sources and links between them are visualized in Figure 4. Currently, we have fully translated sources of law from the National Parliament of Czech Republic and the National Government of the Czech Republic. The resulting numbers of resources and RDF statements are depicted in Table 1 summarizes the amounts of resources and statements obtained from each data source. The National Parliament publishes metadata for each act, regulation, decree and decision of the Constitutional Court since 1945. It also contains information about updates of acts. The National Government publishes content of acts, regulations, decrees and decisions, some of them in current consolidated versions. It does not publish structured information about updates and metadata. As shown in Figure 4, there are links from the governmental data set to the parliamentary data set. Each instance of `lex:Act`, `lex:Regulation`, `lex:Decree` and `frbr:Expression` in the governmental data set is linked to its equivalent in the parliamentary data set where further meta data and information about changes in acts (amendments) are published.

Let us note that LOD representation of court decisions published at Supreme and Constitutional Court web sites is not listed in the table. The reason is that we still work on the translation and currently we have only several (approx. 500) experimental decisions translated. This is not a complete set of court decisions.

Parliament of the Czech Republic. From this data source we obtained by a proprietary web scraper meta data about each act, decree and regulation issued in our country since 1945. We represented each obtained source of law as an RDF resource. In other words, we had to create a HTTP URI for each such resource and create RDF triples recording metadata for each source. For constructing URIs we chose the following pattern:

Table 1. Number of resources and statements in the created LOD representation of sources of law obtained from various public authorities. The numbers at each row are presented in the following order: # `frbr:Work`, # `lex:Act`, # `lex:Regulation`, # `lex:Decree`, # `lex:Decision`, # `frbr:Expression`, `lex:Change`, # RDF statements.

Authority data set	# 1	2	3	# 4	# 5	# 6	7	8
National Parliament	0	6372	8238	3149	1069	33238	20640	710950
National Government	552690	6372	8238	3149	0	617445	0	4789725

`http://linked.opendata.cz/resource/legislation/cz/`
`{sourceOfLawKind}/{validFromYear}/{number}`

where `sourceOfLawKind` is `act`, `decree`, or `regulation`. `ValidFromYear` is the year in which the source of law became or will become valid. And, `number` is the number of the source of law issued by the parliament. All this information is present at the parliament web site and we were able to scrape it using our scrapers. Moreover, for each source of law we were able to scrape the list of other sources of law changed by this source. In the list there is the number of the changed source of law and the kind of change (i.e. creation, cancelation, or update).

As the result, we created an RDF data set where each act, decree and regulation issued in Czech Republic since 1945 is represented as an instance of the respective subclass of `lex:SourceOfLaw` with RDF statements specifying the number of the source of law, its title and validity date. From the list of changes for each source of law we were able also to reconstruct RDF resources representing all expressions of the sources. For constructing URIs of the expressions, we chose the following pattern:

`URIofSourceOfLaw/version/cz/{validityDate}`

where `URIofSourceOfLaw` is the HTTP URI of the source of law the expression realizes. `ValidityDate` is the date from which the expression is valid.

Figure 5 shows a part of the obtained RDF representation of the act `137/2006`. The representation is serialized in Turtle format.

National Portal of Czech Government. We built a proprietary scraper which downloads sources of law published by the portal (acts, regulations, decrees). We parsed this representation so that parts and sections of each source of law were obtained. However, the obtained textual representations were only original expressions of the sources of law. The other expressions (i.e. versions) of the sources could not be obtained in most cases because the Czech government does not produce consolidated texts of all acts (actual consolidated version of only 20 % of acts is published by the government). Moreover, no historical consolidated versions are published.

```
<http://linked.opendata.cz/resource/legislation/cz/act/2006/137-2006>
    rdf:type lex:Act ; dcterms:title "Zakon o verejnych zakazkach"@cs ;
    dcterms:identifier "137/2006" ;
    dcterms:issued "2006-04-19" ;
    lex:actualExpression <http://linked.opendata.cz/resource/
                            legislation/cz/act/2006/137-2006/
                                version/cz/2012-07-01> .

<http://linked.opendata.cz/resource/legislation/cz/act/2006/
        137-2006/version/cz/2012-07-01> rdf:type frbr:Expression ;
    dcterms:title "Zakon o verejnych zakazkach"@cs ;
    dcterms:issued "2012-07-01" ;
    frbr:realizationOf <http://linked.opendata.cz/resource/
                            legislation/cz/act/2006/137-2006> .
```

Fig. 5. RDF representation of a sample act (serialized in Turtle)

```
<http://linked.opendata.cz/resource/legislation/cz/act/2006/137-2006>
    dcterms:description [CONTENT-OF-ORIGINAL-VERSION] .

<http://linked.opendata.cz/resource/legislation/cz/act/2006/137-2006/
        section/1>  a frbr:Work ; dcterms:identifier "1" ;
    frbr:partOf <http://linked.opendata.cz/resource/legislation/cz/
        act/2006/137-2006> .

<http://linked.opendata.cz/resource/legislation/cz/act/2006/137-2006/
        section/2> a frbr:Work ; dcterms:identifier "2" ;
    frbr:partOf <http://linked.opendata.cz/resource/legislation/cz/
        act/2006/137-2006> .
```

Fig. 6. RDF representation of a sample act (serialized in Turtle)

Therefore, the result is that we gathered original expressions of all acts, regulations and decrees and obtained the RDF representation of all their parts and sections according to the LEX ontology. This could not be gathered from the parliament. On the other hand we do not have information about changes. These are present only at the web site of the parliament. The resulting RDF representation is linked to the representation obtained from the parliament based on the Linked Data principles.

Figure 6 shows a part of the obtained RDF representation of the act 137/2006. The representation is serialized in Turtle format. It shows how the act is structured to its sections (only the first two main sections are depicted, the act contains 1481 sections in total). The figure also demonstrates how links to the previous data set are expressed – a given source of law has the same HTTP URI in both data sets. Therefore, the link is realized implicitly. This is possible because we store both data sets in the same database and publish them in the same HTTP domain. If we would publish them in two different domains,

resources representing the same source of law in different data sets would have differnt HTTP URIs. These URIs could be linked on each other using the OWL predicate `owl:sameAs`.

4 Conclusions

We presented an experiment with publishing legislative documents as Linked Open Data. We demonstrated the benefits LOD principles bring to users of legislation (common citizens, lawyers, etc.). We proposed an ontology for publishing structured legislative documents and relationships between them. We also converted two existing data sources in Czech Republic to the proposed representation. In the future work we will concentrate on publishing court decisions in the Czech Republic and linking them to the already published LOD data sets.

References

1. Agnoloni, T., Fernández Barrera, M., Sagri, M.T., Tiscorni, D., Venturi, G.: When a FrameNet-Style Knowledge Description Meets an Ontological Characterization of Fundamental Legal Concepts. In: Casanovas, P., Pagallo, U., Sartor, G., Ajani, G. (eds.) AICOL-II/JURIX 2009. LNCS, vol. 6237, pp. 93–112. Springer, Heidelberg (2010), http://dx.doi.org/10.1007/978-3-642-16524-5_7
2. Beckett, D.: RDF/XML Syntax Specification (Revised). W3C (February 2004), http://www.w3.org/TR/rdf-syntax-grammar/
3. Bizer, C., Heath, T., Berners-Lee, T.: Linked Data – The Story So Far. International Journal on Semantic Web and Information Systems 5(3), 1–22 (2009)
4. Hoekstra, R.: The metalex document server - legal documents as versioned linked data. In: Aroyo, L., Welty, C., Alani, H., Taylor, J., Bernstein, A., Kagal, L., Noy, N., Blomqvist, E. (eds.) ISWC 2011, Part II. LNCS, vol. 7032, pp. 128–143. Springer, Heidelberg (2011)
5. Rubino, R., Rotolo, A., Sartor, G.: An owl ontology of fundamental legal concepts. In: Proceedings of the 2006 Conference on Legal Knowledge and Information Systems: JURIX 2006: The Nineteenth Annual Conference, pp. 101–110. IOS Press, Amsterdam (2006)
6. Harris, A.S.S.: SPARQL 1.1 Query Language. W3C (March 2013), http://www.w3.org/TR/sparql11-query/