

# Using the Semantic Web for the Integration and Publication of Public Procurement Data

José Félix Muñoz<sup>1,2(✉)</sup> and Guillermo Esteban<sup>2</sup>

<sup>1</sup> Aragonese Foundation for Research and Development (ARAID),  
Huesca, Spain

<sup>2</sup> University of Zaragoza, Walqa Technological Park, 22197 Huesca, Spain  
{jfm, gesteban}@unizar.es

**Abstract.** Public procurement is an activity that is common to all administrations, with a major impact on their functioning and that also affects the economy as a whole. This paper presents an experience that shows how the Semantic Web provides appropriate resources to develop data models that can be used both for the management of public contracts and for the publication of information about them. And within that, with a dual objective of improving efficiency by facilitating competitive tendering, and of making easy the monitoring of public contracts by citizens. Firstly, we developed the PPROC ontology in said experience, with the domain of the ontology being the legal institution of public contracting, which includes the procedure for the preparation of contracts. Next, we used the ontology as a basis for the integration and publication of data from various Spanish administrations.

**Keywords:** Public procurement · Semantic web · Ontologies · Open government · Freedom of information

## 1 Introduction

### 1.1 The Right to Free Access to Public Sector Information

One of the third-generation human rights is free access to public sector information, which is now included in the laws of many of the most developed countries [1]. To justify this, it would be sufficient to recognize that this information is created using resources contributed by the general public as a whole through taxes, but that the exercise of this right is usually linked to the following specific purposes:

- The defense of the interests of the affected parties, in the access to administrative files.
- Academic, in the access to historic archives.
- Political, in the transparency conceived as an instrument for the control of the activities of public authorities.
- Economical, in the re-use of public information by the private sector (generally, Open Government Data (OGD) projects form part of this latter purpose).

Nowadays, access to information is regulated by different laws and regulations depending on how it is linked to one purpose or another. This situation is meaningless as the exercise of the right of access has become independent of the purpose and must always be allowed, inasmuch as one of the limitations provided by the law is not applicable. Therefore, it is necessary to draft one unified regulation so that public information spans these four perspectives. This legislation should also encompass the entire life cycle of information and not just access to it [2].

Depending on the purpose of the access, different laws might be applied and the authority might be given to different bodies and institutions. One of the consequences of this fact is that numerous redundancies are generated that, as well as leading to an unnecessary consumption of resources, cause confusion between users and public servants (as can be seen in the case of Italy, in [3]). Another consequence is that very inconsistent strategies and standards will be produced. For example, it is in the access for the re-use of information in which technological tools—such as the Semantic Web—are used to a greater extent, since it is assumed that one of the main objectives is that the re-users can exploit the information by using computerized means. However, as we will see in this paper, the Semantic Web could be a very useful tool for administrations to help the public exercise its right of access to information in all its aspects, and not just from the re-use perspective.

## 1.2 Transparency and Public Procurement

Advertising has always been an essential part of public procurement, in which it must fulfil a dual purpose: on one hand, it is a resource to improve competitive tendering and, on the other, it constitutes an instrument for transparency and for the monitoring of the behavior of the contracting authorities [4]. With the progress of electronic government, the publication of information regarding contracting procedures increasingly began to be performed using electronic means. In 2004, European directives<sup>1</sup> created a specific mechanism called the buyer profile. All public sector entities must have one and publish certain information on it about the contracts that they put out to tender, and therefore, it has become the central information point for companies and the public. However, its use has been severely limited by the major functional and technical differences between the different profiles and the lack of interoperability between them, which makes the integrated processing of the information published on them nearly impossible.

Faced with this problem, the solution that is usually adopted is to make the publication of announcements regarding tender procedures of all administrations mandatory on a single web site. For example, within the scope of the EU there is the Tenders Electronic Daily (TED), which is the online version of the “Supplement to the Official Journal of the EU”, dedicated to European public procurement. In Spain, this is the

---

<sup>1</sup> Today replaced by the directives of the European Parliament and of the Council: 2014/24/EU, of 26 February 2014 on public procurement and repealing Directive 2004/18/EC; and 2014/25/EU, of 26 February 2014, on procurement by entities operating in the water, energy, transport and postal services sectors and repealing Directive 2004/17/EC.

Public Sector Contracting Platform (PSCP). However, only part of the problem is solved using this measure. This is the part related to competitive tendering and, therefore, to economic efficiency since, in order to comply with this objective, it is enough to publish a limited set of announcements. However, transparency requires much more information and, in addition, transparency practices can be very different depending on the policies followed by each authority. Therefore, from the perspective of transparency, the solution involves preparing standards that could be used by administrations to publish all the information that they consider appropriate.

Below we will describe an experience that includes the integration and publication of data about public contracts and the procurement procedures of various Spanish Administrations following this strategy. We carried out the experience within the framework of the ContSem project<sup>2</sup>—led by the company iASoft (Oesia)—and we based it on semantic technology (Sect. 2). Their main objective has been publish the information as linked RDF, which holds the five stars level from the rating system for open data established by the W3C.<sup>3</sup> To do this we developed the PPROC ontology,<sup>4</sup> the principal characteristics of which are described in Sect. 3. Using this ontology, we integrated (Sect. 4) and semantically labeled and published (Sect. 5) the data of two local administrations and of the entire state administration of Spain. Finally, as a conclusion, we present some observations drawn from the experience.

## 2 Data Models in Electronic Administration

### 2.1 Management and Interoperability

Over the short history of electronic administration, new purposes or objectives for the processing of data by administrations have been added, which has led to changes in the methodology used in the design of data models.

We can assume that the first reason why public administrations used computers was management. If we analyze the data models prepared for management applications, we can see that their designs meant that their structure was closely related to the temporal succession in which the information was being generated or received by the management bodies. This was due to the way that the computer analysts obtain information from the administrative managers, as these latter tend to request that the different fields be incorporated sequentially, in the order that they are required to record certain information. This leads to an organization of data that we could call “procedural”, as it is closely connected to the procedure and, therefore, to the way of working of each administrative service. Its reflection in the applications consists of a succession of screens where the entire sequence of a specific case (of a public contract, for example) can be followed from start to finish.

---

<sup>2</sup> Financed by the Spanish Ministry of Industry, Commerce and Tourism through the project “Optimization of public procurement using semantic technologies”, TSI-020606-2012-4.

<sup>3</sup> <http://www.w3.org/DesignIssues/LinkedData.html>.

<sup>4</sup> <http://contsem.unizar.es/def/sector-publico/pproc>.

The objective of interoperability was later requested from the information systems of public administrations, which led to the development of data models whose objective was to achieve interoperability, by addressing communication between systems. Public procurement is a good example of this, as the announcements that are published on these web sites that, as we have just seen, both the EU and national governments have created to centrally publish information about public tenders are among the first exchanges of information performed electronically using structured messages. There are various initiatives whose purpose is to create standards for electronic procurement, including, within the scope of the EU, OpenPEPPOL<sup>5</sup> and CEN BII.<sup>6</sup> In both cases, XML formats that make it possible to structure the messages exchanged by the various agents involved in electronic procurement are defined. However, the administrations chose to create “de facto” standards, such as the one established for the TED eSenders<sup>7</sup> or CODICE,<sup>8</sup> defined for the Spanish PSCP. In the design of these standards, computer analysts use documents (messages) that are exchanged between organizations as their reference, and these determine the structure of the information. Consequently, we can call these XML standards “document-oriented”.

## 2.2 Re-Use

Recently, laws highlight the re-use of information, both by the different bodies of each administration and by the administrations as a whole, as one of the principles that should govern electronic administration. From a technical perspective, this objective relates to the application of the principles of unique data and shared data. However, while interoperability solely addresses communication between systems and ignores how information is organized inside each of them, re-use requires consistency between the parts of a system and between the various systems themselves, so that information can be shared between multiple agents and for various purposes.

One of the purposes will be to facilitate the exercise of the right of free access to public sector information. From the first moment, this right must be considered as a basic requirement in the design of systems and, therefore, public access will have to be considered as one of the possible uses of the information in the design of data models. Consequently, it will not be enough to propose common data structure for public organizations; instead it will be necessary to define models which includes all the information that citizens may require. Going further, the final objective must be that the information used for management and the information for public access be the same, which does not mean it is necessary to publish all information, but does mean that no conversion process be required for publication.

Numerous projects are being developed within the context of the Semantic Web, with the objective of creating models that will make it possible to represent legal concepts. It is calculated that in 2011 over 60 ontologies focused on legal knowledge

---

<sup>5</sup> <http://www.peppol.eu>.

<sup>6</sup> <http://www.cenbii.eu>.

<sup>7</sup> [http://simap.europa.eu/ojs\\_esenders/sending\\_xml\\_notices/index\\_en.htm](http://simap.europa.eu/ojs_esenders/sending_xml_notices/index_en.htm).

<sup>8</sup> <https://contrataciondelestado.es/wps/portal/codice>.

had been completed. These are very varied ontologies, both from a granularity point of view and regarding their degree of formalization and the methods used for their development [5]. Regarding the sources of knowledge used, the main legal concepts prepared by the philosophy of law have been used as a basis for representations of legal knowledge. Therefore, some place legal texts at the heart of representation, and others are focused on the activity of legal experts, while there are also those focused on the modeling of discourse, which is the basis of legal argument.

Ontologies also differ regarding their objectives. In some, called core ontologies, the objective is the creation of models of general legal concepts [6]. In other cases, the objective is not to describe general concepts but to design a specific social mechanism. This is the case with our project, in which the intention is to represent the social mechanism used to connect the contracting process of public sector entities. For the development of these models we consider that an approach based on the “theory of the institution” is appropriate. According to this approach, the central focus of the model would be public procurement, considered as a legal institution whose purpose is the attainment of a “product”: a public contract [7]. This approach helps to demarcate the field of knowledge to be represented, as the institutions can also be seen as systems with a well-defined interface with an environment [8].

It seems logical that a closely studied formalization of a certain legal institution—an “institution-oriented” model—could be useful for its representation within the information system, both for management and publication requirements. Although these do not totally coincide, there are no differences in terms of the organization and the high-level entities; instead they appear, above all, because the management requires certain additional, internal information. This information has not been included in the ontology, as it is not its objective to describe all of the entities and properties that are required in management. These data are not of interest to the public and it is sufficient to carry out an extension of the ontology to incorporate the data that are necessary in the back-office.

### 3 The PPROC Ontology

#### 3.1 Structure and Sources of Knowledge

The perspective focused on the “institution” firstly determines the semantic relationships of the model. Some systems of legal concepts are organized vertically from the most general concepts to the most specific ones. In this case, the relationships are about belonging. Other systems, known as operational families, gather together the elements related to a specific item [9]. An institution-based model belongs to this second type and its semantic relationships are organized according to the role that each concept plays within the institution that is represented. In order to identify and define these relationships, the science of the law can be used, which is devoted to studying and organizing the legal elements that comprise institutions and the relationships between them.

Secondly, considering an institution as the scope of the model will help us to establish the nature of the entities that will be able to form part of the model, as all the

realities involved in the public contract formation process will have to be included, and these comprise theoretical, social and physical realities [10]. This varied nature of the entities also means that the sources of knowledge will be different. We will firstly need to use legal texts, as public procurement forms part of public law, and within this, of administrative law. In this respect, it is necessary to consider that the durability of the representation contained in the model could be severely reduced through excessive connection to legal texts as, in general, administrative laws vary much more over time than those of private law—especially those of civil law, where many of the core figures of the legal system are defined with a greater degree of generalization and abstraction. In this sense, by focusing the model on the “legal institution”, it will be possible to identify the elements of the institution that have become established over the years, remaining constant in successive regulations. Likewise, this strategy is also consistent with the objective of the ontology—the publication of information thereby facilitating understanding by the public—as, in order to achieve the objective, not only is it unnecessary to strictly adhere to details by reflecting the literal contents of laws, it could even prove to be counterproductive.

According to Ferraris, social realities are always related to documents, so much so that being a document is sufficient to be a social fact [10]. Given that the publication of public information is always performed through documents, each of the acts that comprise contracting procedures will be a social reality. Therefore, the second source of knowledge to be used will be the documents through which administrations publish information about these procedures and contracts. Laws establish part of their content but above all, it is determined by the practices of the various administrations and, within the scope of e-procurement, by the standardization needs. Finally, the third source of knowledge to be used is that regarding physical realities, and the most important for the model are those that appear within the object of the contracts. It is also necessary to include other physical facts that we could call secondary (such as the physical location of an office, for example).

### 3.2 Use of Semantic Technology in Public Procurement

The growth in procurement by electronic means has led to what is called e-procurement, in which both XML standards and semantic technology are widely used [11]. Among the developments in the European context, there are some that consist of the preparation of ontologies for the representation of knowledge about public procurement.

One of these experiences is LOTED2 [12], which bases most of its content on the two directives (2004/17/EC and 2004/18/EC) that at the time of development regulated public contracts in Europe. The result is a thorough study of legal documents. This means that the legal content of European procurement is heavily present and rigorously represented in LOTED2. Another initiative focused on public procurement is MOLDEAS (Methods On Linked Data for E-procurement Applying Semantics) [13], an ontology focused on the representation of information contained in the announcements about public tenders. The objective of this ontology was to provide a pan-European standard about public procurement data, enriching it with the classifications of already-existing products and publishing it by following established open data guidelines.

A third initiative in the EU is the Public Contracts Ontology (PCO), implemented within the framework of the LOD2 Project [14]. PCO models the main aspects of public contracts, although not in great depth. The ontology considers “only the information that is publicly available in existing systems on the Web [...], mainly produced during the tendering phase”. Hence, the result is a lightweight ontology that re-uses widely accepted ontologies and vocabularies such as VCard, Payments Ontology, schema.org, Call for Anything vocabulary and GoodRelations.

PCO and MOLDEAS describe the main concepts of public procurement without examining details very much and, as a consequence, some specific relations, roles or behaviors are not strictly represented (e.g. the contracting body or distinguish between objective and subjective award criteria). On the other hand, LOTED2 represents almost every aspect of public procurement with the result that his the model is closely related to the text of the 2004 directives. None of the three ontologies studied had transparency as its primary goal and the main objective we chased with the use of the ontology was to improve the transparency of public contracting processes. Consequently, although taking PCO as our basis, we decided to develop the PPROC ontology, the main objective of which is to facilitate access for all parties interested in information regarding public contracts, which means that not only contracting powers and tenderer companies, but also the general public as a whole has been considered in its design.

### 3.3 Description of the Ontology

The PPROC ontology is composed of 78 classes and 129 properties that are divided into four blocks, each one of which includes the classes that are directly related to the following points of the contract: (1) the object, which is the supply that the contract covers; (2) the parts, which are the agents that participate in the procurement process and, when appropriate, in the contract—the contracting authority, tender, awarded tender, etc.; (3) the procedure, composed of the steps taken until the end of the contract; and (4) the fulfilment, which includes actions that must be taken after the contract formalization.

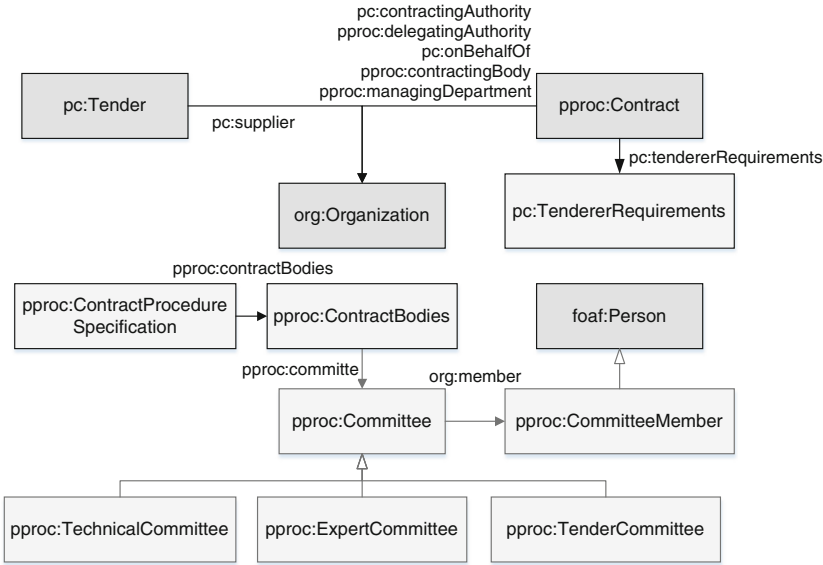
The class `pproc:Contract` is the main class all contracts begin with (see Fig. 1). It contains the basic information about the contract and serves as an entry point to link to other classes. In order to define the object of the contract, PPROC can use two different (non- exclusive) approaches. The first way to define the object is to use `pproc:object` or `pc:mainObject`, properties that are especially appropriate for using the Common Procurement Vocabulary (CPV).<sup>9</sup> It is exclusively used in EU procurement and consists of a main vocabulary for defining the object of a contract without entering into great detail, and a supplementary vocabulary for adding further qualitative information. However, the object in public contracts is not different from the object in other business situations, and therefore vocabularies developed for business can be re-used. Consequently, we have established a second way, and to further define

---

<sup>9</sup> [http://simap.europa.eu/codes-and-nomenclatures/codes-cpv/codes-cpv\\_en.htm](http://simap.europa.eu/codes-and-nomenclatures/codes-cpv/codes-cpv_en.htm).







**Fig. 2.** Parties of the contract

Next, to describe proposals made by the suppliers, the class `pc:Tender` is re-used. PCO uses two properties to link tenders to their related contract: `pc:tender` and `pc:awardedTender`. We created subclasses to further define a `pc:Tender` (`pproc:AwardedTender`, `pproc:AcceptedTender`, `pproc:ExcludedTender` and `pproc:FormalizedTender`). Finally, tenderers are also defined through the Organization Ontology and linked using the `pc:supplier` property from a `pc:Tender`.

Another block of information is referred to the procedure. This includes the kind of procedure and all the information about it that could be useful to any party and the public, such as the tender requirements, briefing meetings or information about remedies. However, the information about the procedure is very important for the control of contracting and, therefore, the ontology also includes classes to describe other points, such as the people that participate in the procedure or possible resources and their result. It is also necessary to know if the type of procedure used is the one related to the contract, and the ontology has specific properties to do this, such as `pproc:assumptionProtectingProcedureType`.

Also, the term of the contract does not end with the formalization, which is the time when the contracting procedure is considered as finished. Contracts are often modified at a later time through specific procedures, which often change points such as the price or the term for completion. These modifications can be used to breach the principles of the contracting and, therefore, a fourth block is dedicated to this phase, which we call fulfilment. This contains classes that make it possible to represent the conditions and limits that possible modifications to the contract are subject to. Finally, clauses are also

included that make it possible to represent what the final result of the contract is, if one or more modifications are made.

Regarding its future use by local administrations, PPROC is recommended as the ontology to be used by smart cities offering their public contract data according to the proposed technical norm from the Spanish Association for Standardization and Certification (AENOR), UNE 178301 on Open Data for Smart Cities.

## 4 The Integration of Data Regarding Procurement

Procurement is a horizontal task, as there are many services of a public administration that require contracts to be prepared to implement its functions. The way in which this task is distributed varies greatly in administrations composed of various departments and services. In the cases that we studied we observed that there is normally a central department specializing in procurement, but only contracts that exceed a determined figure are handled through this. For example, in Zaragoza City, the Council is responsible for major contracts (those worth over 20,000 Euros) while each service deals with the minor contracts that it promotes. It is also very common for there to be bodies that are linked to the administration but that act autonomously, in the form of foundations, state companies, etc. These entities manage their own contracts, regardless of their figures. The result is that there is little control over procurement, not only by the public but also by the managers and politicians responsible for this.

In the last two years, Zaragoza City Council has developed a project for the implementation of a workflow tool for the management of procurement files, which replaces the management applications used in recent years. This platform is unique for all services that handle contracts and for all contract types, meaning that only autonomous bodies are omitted from integration in this first phase. The decision taken by the parties responsible for this process was to use the PPROC ontology as a basis for the data structure of the new tool. This was also done by the Provincial Government of Huesca for the development of its contracting platform. In both cases it was considered that the use of the ontology as a starting point for the design of the data structure would enable the data organization to be independent of any temporal or secondary aspect, and instead would be related to the entities that comprise the “institution” and the relationships between them.

The data integration starts discovering the relevant public procurement data aimed to transparency and, as explained in Sect. 3, capturing its model as an ontology. Afterwards comes the matching of the data from different sources, mapping the classes and properties of the ontology with the entities of the relational databases that were being used for the management of contracts and for the publication of announcements in the buyer profile. This step is the most expensive, as requires knowledge of both the original database and ontology schemas. The W3C has a language recommendation that express customized mappings from relational databases to RDF called R2RML<sup>12</sup> which we used in both our use cases. Finally, the data is transformed using the

---

<sup>12</sup> <http://www.w3.org/TR/r2rml/>.

generated map and an existing R2RML implementation that automatically connects to the given relational database and retrieves the information as RDF.

The object of another integration task was the announcements published on the PSCP. All contracts from the state public sector already had to be necessarily published on this online platform, and the Single Market Guarantee Act 20/2013 of 9th December broadened its scope to include contracts from all Spanish public administrations, that is, autonomous and local administrations are now also included. As we already mentioned, the PSCP uses its own XML dialect, named CODICE, to enable administrations to publish data on their website. We therefore prepared a mapping between CODICE and the PPROC ontology, the purpose of which is to serve as a basis for the translation of XML messages into RDF triples.<sup>13</sup>

Additionally, we carry out monitoring of the classes used by each entity, with the document “PPROC usage”, which describes the classes and properties of the ontology that are used by each authority.<sup>14</sup> According to this monitoring, of the three cases studied 58 classes or properties (28 %) of the ontology are used. One of the reason is that contracting authorities still publish an important part of the information in non-structured formats (mainly PDF documents). Moreover, there are other information that they do not publish nowadays. But, the final goal of transparency regulations is to improve transparency through the publication of all the information regarding public contracts. And, from the open data perspective, the aim is to publish all this information using linked RDF. In consequence the PPROC domain covers all the data that can be published by a contracting authority. In next years, we hope that the progressive implementation of electronic processing in public procurement increases the ratio of structured data which could be directly published using the ontology.

Also, as a result of this monitoring, we have been able to confirm that the information available in the different administrations varies greatly, given that of these 58 elements only 11 are used by the three administrations, 32 are used by two and 15 by just one. Obviously, this represents a major obstacle when integrating data for the realization of queries and for their analysis. We therefore consider that this monitoring could help to improve this situation, as its purpose is that it will be used as a source of guidance for user administrations regarding entities to be published on their profiles—thereby creating a kind of core vocabulary based on the practices of administrations—and also used as a source of information for the public regarding information points and the content of each of them. And, in short, to encourage re-use by creating a common information archive regarding public contracting.

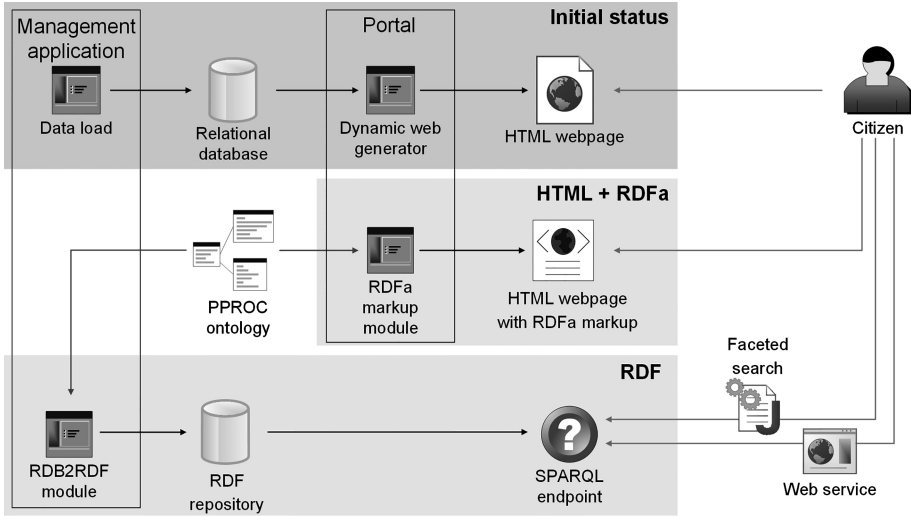
## 5 The Publication of Data Regarding Procurement

Once the PPROC was prepared, the following task was the publication in RDF of the buyer profiles of Zaragoza City Council and of the Provincial Government of Huesca. When it comes to the release of structured data already published on the website (as

---

<sup>13</sup> <http://dx.doi.org/10.6084/m9.figshare.1327549>.

<sup>14</sup> <http://contsem.unizar.es/def/sector-publico/usage.html>.



**Fig. 3.** Options for the publication of the contracting party profile

buyer profiles), a decision has to be made between two non-exclusive alternatives (see Fig. 3). The first is to publish an endpoint that enables users to perform queries directly to the data graph, using SPARQL query language. The second is to semantically label the content of the web pages, using the RDFa markup,<sup>15</sup> which provides a specification to express structured data directly in any markup language (e.g. HTML). Both administrations chose the first option. We installed the SPARQL endpoints<sup>16</sup> into the servers of the administrations, using a Virtuoso Universal Server.<sup>17</sup>

As we have already mentioned, the contracts of the PSCP have also been published, for which we developed software that generates RDF triples based on the CODICE format using the R2RML specification. This information has been made publicly available in a SPARQL endpoint<sup>18</sup> that, at the time of the drafting of this paper, contains approximately 12 million RDF triples, including information about 199,611 contracts.

One of the benefits of the integration task is that multiple contracting authorities can publish its own set of contracts through a SPARQL endpoint. As all these data follows PPROC model, it is possible to build agents that retrieve and aggregate this information. As an example, we have developed a web service that can recover and add information from various SPARQL endpoints, which are indicated in a file that contains the URL of each of them. The web service is implemented using REST architecture and has two different calls, which are made using JSON messages. The first call

<sup>15</sup> <http://www.w3.org/TR/rdfa-syntax/>.

<sup>16</sup> <http://www.zaragoza.es/ciudad/risp/sparql.html> and <http://www.dphuesca.es/sparql>.

<sup>17</sup> <http://semanticweb.org/wiki/Virtuoso>.

<sup>18</sup> <http://pproc.unizar.es:8890/sparql/>.

(facetQuery) uses the set of facet-value pairs chosen by the user as input, returning a list of contracts (identifiers and some basic information) that match these facet-value pairs. The second call (contractQuery) uses the contract identifier provided by the first call as input, and returns all the information about the contract in a JSON-LD format. Based on this service we have published a faceted search web that is also kept active in a public URL, available to whoever wished to use it.<sup>19</sup>

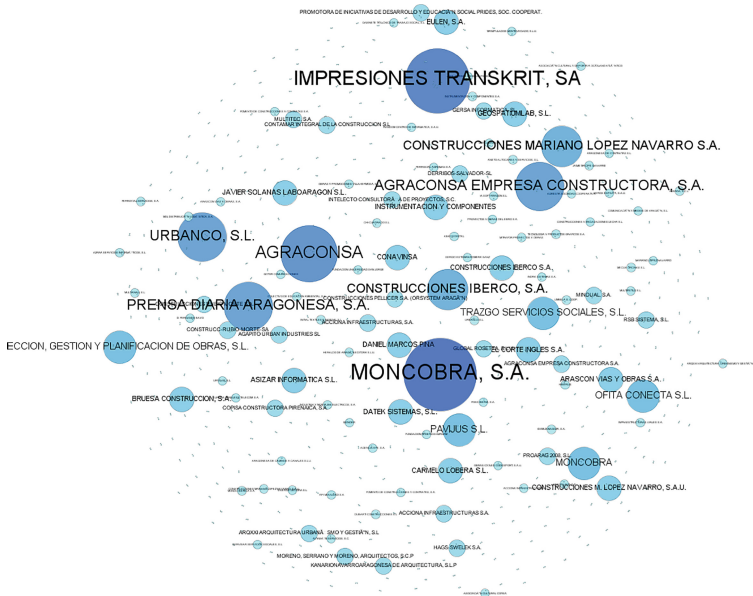
The set of facets was defined by the experts in procurement and some of them are representative of the difficulty entailed in obtaining valid results when managing information from different sources and where different policies have been followed for data entry and management. For example, the first facet is the object of the contract, that is, the provision (product or service) that is contracted. As we have said, there is a specific code (the CPV), but often this field is not completed and other times it is completed with a very low disaggregate level. To do this, we also implemented a search for words describing the object. Although in the ontology there is a specific field for the textual description of the object, this search is performed not only in this field but also in the title and in the description of the contract, as very often these fields are the only places where the object is specified.

The second facet refers to the economic aspects, where it is necessary to consider that there are several sums associated with just one contract. One is the estimated value, which is included in the announcement prior to the tendering (provided that this is published, as it is optional). In all cases there is a tender budget, which is the sum for which the tender is published, a sum for which the contract is awarded and a final sum that may be the same or different from this depending on whether there have been changes made or not. The final amount is also higher in contracts that accept possible extensions. The third facet is the status of the procedure, which is determined by searching to ascertain if certain steps have been performed, such as the approval of the case, the publication of the tender announcement and the awarding or formalization of the contract. The fourth and fifth facets refer to the parties of the contract, as these are the contracting authority and the provider. Finally, the sixth facet comprises the dates of some steps and milestones and the seventh is the type of procedure.

Based on these tools, our group has begun to analyze some aspects of the public procurement in Spain. One of these aspects refers to the companies that are contracted by administrations. For example, Fig. 4 shows the main Zaragoza City Council suppliers and the number of contracts that were awarded from 2008 to 2014. In this regard, one of the next planned actions is to integrate this information with that from databases that contain information about the companies, in order to inter-relate the companies that belong to the same groups. In this first steps, we do not use semantic web tools such as reasoning, agents or logic, but, as a basis for future works, we are now identifying some questions about public contracts which can be studied using these tools. Also, our group collaborates with other groups and re-users interested in the exploitation of the linked open data about public contracts.

---

<sup>19</sup> <http://contsem.unizar.es/docs/facetedsearch/>.



**Fig. 4.** Zaragoza City Council suppliers according to the number of contracts awarded (2008–2014)

## 6 Conclusion

The Semantic Web provides the opportunity to agree upon models that will represent the various realities and “institutions”. In this way, the data structures used for various functions within the same system and between different information systems can be progressively harmonized. In addition, it is possible that the distinction between the information that administrations use to perform their work and the information that it prepares (currently) to make it available to the public will disappear, or at least lessen significantly. Both circumstances are conducive to the progressive implementation in public administrations of the principles of unique data and shared data.

In the experience that we have described, we have seen how the work initially carried out to facilitate public access to information can have a direct influence on a major improvement to the information system as a whole and, therefore, to the business intelligence within the organization itself. This represents one more piece of evidence that the design and development of information systems must be done in such a way that one sole set of data is used for different purposes, both current and future.

These purposes include that of facilitating citizens’ access to information. From a legal perspective, this access, which is a fundamental right, should be regulated as a single unit and, therefore, regardless of the use that is to be made of the information. In order for the legislation that regulates administrations to be consistent with all these objectives, it is necessary to have new laws based on a unique and comprehensive concept of public information. Thus, based on this concept, these new laws will

regulate all aspects related to the obtaining, management and exploitation of public information. If this is done, not only could the right to free access to public information be greatly strengthened, but so could the efficiency and quality of the information systems of administrations as a whole.

Finally, we should state that from the point of view of interaction between information technology and law, that the application of the Semantic Web to public information could lead to numerous scenarios—as has been shown in this article—in which the representations of legal knowledge will be truly useful in practical terms. This fact would help to reduce the lack of trust that in some (or many) cases legal experts have regarding the applications of artificial intelligence to their field of knowledge.

**Acknowledgements.** The authors would like to thank Miguel Angel Bernal and Francisco Serón (co-authors of the PPROC ontology), and Oscar Corcho, Carlos Bobed and Carlos Becana (contributors). Also to those responsible for procurement and the buyer's profile and the technicians of the participating administrations in the project: the Zaragoza City Council (Maria Jesús Fernandez, Ana Budría, Laura Fernando, Víctor Morlán and Rubén Notivol), the Provincial Government of Huesca (Cristina de la Hera, Montserrat Rodríguez and Javier Casado) and the Regional Government of Aragón (Ricardo Cantabrana, José María Subero and Eva Sanz).

## References

1. Ackerman, J.M., Sandoval-Ballesteros, I.E.: The global explosion of freedom of information laws. *Adm. Law Rev.* **58**(1), 85–130 (2006)
2. Muñoz Soro, J.F., Bermejo Latre, J.L.: The redefinition of the objective scope of transparency and the right of access to public sector information. In: Valero Torrijos, J., Fernandez Salmeron, M. (eds.) *Legal Regime of Transparency in the Public Sector*, pp. 189–239. Aranzadi, Cizur Menor (2014)
3. Palmirani, M., Martoni, M., Girardi, D.: Open government data beyond transparency. In: Andrea, K., Francesconi, E. (eds.) *EGOVIS 2014. LNCS*, vol. 8650, pp. 275–291. Springer, Heidelberg (2014)
4. Miroslav, M., Miloš, M., Velimir, Š., Božo, D., Đorđe, L.: Semantic technologies on the mission: preventing corruption in public procurement. *Comput. Ind.* **65**(5), 878–890 (2014)
5. Casellas, N.: *Legal Ontology Engineering*. Springer, Dordrecht (2011)
6. 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)
7. MacCormick, N.: Norms, Institutions, and Institutional Facts, *Law and Philosophy* **17**, 301–345 (1998)
8. Boer, A.: *Legal Theory, Sources of Law and the Semantic Web*. IOS Press, Amsterdam (2009)
9. Fernandez-Barrera, M., Sartor, G.: The legal theory perspective: doctrinal, conceptual systems vs. computational ontologies. In: Sartor, G., Casanovas, P., Biasiotti M.A., Fernandez-Barrera M. (eds.) *Approaches to Legal Ontologies*, pp. 15–47. Springer, Dordrecht (2011)

10. Ferraris, M.: Social ontology and documentality. In: Sartor, G., Casanovas, P., Biasiotti, M. A., Fernandez-Barrera, M. (eds.) *Approaches to Legal Ontologies*, pp. 83–97. Springer, Dordrecht (2011)
11. Alvarez-Rodríguez, J.M., Labra, J.E., Ordoñez de Pablos, P.: New Trends on e-Procurement Applying Semantic Technologies. Current Status and Future Challenges. *Comput. Ind.* **65**, 800–820 (2014)
12. Distinto, I., d’Aquin, M., Motta, E.: LOTED2: An Ontology of European Public Procurement Notices (2012). <http://www.semantic-web-journal.net/content/loted2-ontology-european-public-procurement-notice-1>
13. Alvarez-Rodríguez, J.M., Labra, J.E., Cifuentes, F., Alor-Hernandez, G., Sánchez, C., Luna, J.A.G.: Towards a pan-European e-procurement platform to aggregate, publish and search public procurement notices powered by linked open data: the MOLDEAS approach. *Int. J. Software Eng. Knowl. Eng.* **22**(03), 365–383 (2012)
14. Nečaský, M., Klímek, J., Mynarz, J., Knap, T., Svátek, V., Stárka, J.: Linked data support for filing public contracts. *Comput. Ind.* **65**(5), 862–877 (2014)
15. Hepp, M.: GoodRelations: an ontology for describing products and services offers on the web. In: Gangemi, A., Euzenat, J. (eds.) *EKAU 2008. LNCS (LNAI)*, vol. 5268, pp. 329–346. Springer, Heidelberg (2008)
16. Miles, A., Matthews, B., Wilson, M., Brickley, D.: SKOS Core: simple knowledge organisation for the web. In: *International Conference on Dublin Core and Metadata Applications*, pp. 3–10 (2005)



Electronic Government and the Information Systems  
Perspective

4th International Conference, EGOVIS 2015, Valencia,  
Spain, September 1-3, 2015, Proceedings

Kő, A.; Francesconi, E. (Eds.)

2015, XII, 391 p. 98 illus., Softcover

ISBN: 978-3-319-22388-9