

INNOVATIVE SERVICES TO EASE THE ACCESS TO THE PUBLIC PROCUREMENT NOTICES USING LINKING OPEN DATA AND ADVANCED METHODS BASED ON SEMANTICS

Jose María Álvarez and José
Emilio Labra
WESO-Universidad de Oviedo
Facultad de Ciencias, C/Calvo Sotelo,
S/N 33007
Oviedo, Spain
josem.alvarez@weso.es,
jelabra@weso.es

Ramón Calmeau
Exis TI
C/ Fray Luis de León, 11 - 2ª, 28012
Madrid, Spain
ramon.calmeau@exis-ti.com

Ángel Marín and José Luis
Marín
Paseo Zorrilla 89 - 1A
47007
Valladolid, Spain
anmar@gateway-scs.es,
josmar@gateway-scs.es

Abstract – The aim of this paper is to describe a public procurement information platform which provides a unified pan-european system that exploits the aggregation of tender notices using linking open data and semantic web technologies. This platform requires a step-based method to deal with the requirements of the public procurement sector and the open government data initiative: 1) modeling the unstructured information included in public procurement notices (contracting authorities, organizations, contracts awarded, etc.), 2) enriching that information with the existing product classification systems and the linked data vocabularies; 3) publishing relevant information extracted out of the notices following the linking open data approach; 4) implementing enhanced services based on advanced algorithms and techniques to exploit the information in a semantic way. This system is supposed to be especially relevant for SMEs that want to tender in the European Union (EU), easing their access to the information of the notices and fostering their participation in cross-border public procurement processes across Europe. Finally an example of use is provided to evaluate and compare the goodness and the improvement of the proposed platform regarding to the existing ones.

1. – Introduction

In the European e-Procurement context there is an increasing commitment to boost the use of electronic communications and transaction processing by government institutions and other public sector organizations. The European Commission (EC) outlines the following advantages in the wider use¹ of e-Procurement: increased accessibility and transparency, benefits for individual procedures, benefits in terms of more efficient procurement administration and potential for integration of EU procurement markets. However several interlinked challenges to fulfill a successful transition to e-Procurement are missing: overcoming inertia and fears on the part of contracting authorities and suppliers, lack of standards in e-Procurement processes, no means to facilitate mutual recognition of national electronic solutions, onerous technical requirements, particularly for bidder authentication and managing multi-speed transition to e-Procurement.

¹ http://ec.europa.eu/internal_market/consultations/docs/2010/e-procurement/green-paper_en.pdf

The first action to ease the interconnectivity and interoperability within Europe's emerging e-Procurement landscape was the creation of TED² ('Tenders Electronic Daily'). It is the on line version of the 'Supplement to the Official Journal of the European Union', dedicated to European public procurement (1500 new procurement notices every day) but an unified information system pan-European dealing with: 1) dispersion of the information; 2) duplication of the same notice in more than one source; 3) different publishing formats; 4) problems regarding to a multilingual environment and 5) aggregation of low-value procurement opportunities, is required. Other set of actions of the EU in the eGovernment context are the conceptual/terminological maps of particular domains available in RAMON³, the Eurostat's metadata server: in the Health field, the "European Schedule of Occupational Diseases" or "International Classification of Diseases"; in the Education field, thesauri as "European Education Thesaurus"; European Glossary on Education; in the Employment field, the "International Standard Classification of Occupations"; in the European Parliament activities the "Eurovoc Thesaurus" and in the E-procurement field the "Common Procurement Vocabulary", hereafter CPV, among others. The structure and features of these systems are very heterogeneous, although some common aspects can be found in all of them: 1) hierarchical relationships between terms or concepts, 2) multilingual character of the information. These knowledge organization systems (KOS) enable users to annotate information providing an agile mechanism for performing tasks such as exploration, searching, automatic classification or reasoning.

Although the EC tries to encourage the creation of strategies to improve access to public procurement markets by SMEs. A question about the further steps is not yet answered: "*What further steps might be taken to improve the access of all interested parties, particularly SMEs, to e-Procurement systems?*"¹. In that sense, some governments have created data catalogues under the principles⁴ (complete, primary, timely, accessible, machine processable, non-discriminatory, non-proprietary, and license-free) of Open Government Data (OGD) initiative to make it easy for the public the access to public information. This public data enables⁵ greater transparency; delivers more efficient public services; and encourages greater public and commercial use and re-use of government information. Initiatives like Linking Open Data (LOD) and the semantic web provides the needed background to answer the aforementioned question exploiting this new context of publishing public data

Obviously in the e-Procurement field, the public information published by governmental contracting authorities are a suitable candidate to apply the LOD approach and the semantic web technologies providing a framework to develop advanced algorithms and deal new enhanced services oriented to SMEs. This growing commitment to the reuse of public sector information (PSI) and initiatives like semantic web, LOD and the use of KOS provide building blocks for an innovative unified pan-European information system that: encourages standardization of key processes and systems, gives economic operators the tools to overcome technical interoperability and can be considered a new step to encourage the participation of the SMEs in both the e-Procurement and the traditional tendering processes.

2. – Related Work

The state of the art of this proposal is summarized according to three points of views: 1) LOD and OGD approaches to model, enrich and publish data included in public procurement

² <http://ted.europa.eu/>

³ <http://ec.europa.eu/eurostat/ramon>

⁴ http://resource.org/8_principles.html

⁵ <http://www.w3.org/TR/gov-data/>

notices. 2) Thesauri, taxonomies and product scheme classifications to classify the information about public procurement notices. 3) Semantic methods to exploit the new structured information providing enhanced services.

In the scope of LOD and OGD there are projects trying to exploit the information of public procurement notices like LOTED⁶ (“Linked Open Tenders Electronic Daily”) where they use the RSS feeds of TED. UK government⁷ is doing a great effort to promote its information sources using the LOD approach. They have published datasets from different sectors: transport, defense, NUTS geographical information⁸, etc. Most of the public administrations in the different countries are also betting for OGD approach to make public their information: Spain (Aporta project⁹), New Zealand¹⁰, Australia¹¹, USA¹², etc. Regarding the use of LOD and organizations there is a new ontology for modeling the information about organizations¹³ and recently it has been released “The Open Database Of The Corporate World”¹⁴.

In the field of the semantic web technologies and for modeling the domain knowledge there are several options: RDF (*W3C Recommendation 10 February 2004*), RDF(S) (*W3C Recommendation 10 February 2004*), OWL 2 (*W3C Recommendation 27 October 2009*) or SKOS (*W3C Recommendation 18 August 2009*) among others. They provide a common format and data model for sharing and linking knowledge organization systems via the web. This information can be retrieved using SPARQL (*W3C Recommendation 15 January 2008*), a query language and a protocol to retrieve the information of datasets published via an endpoint. Currently, there is a working group defining a vocabulary and a set of instructions that ease the discovery and usage of linked datasets (voID¹⁵), the new specification of SPARQL (1.1¹⁶) enables a method for discovering and vocabulary for describing SPARQL services made available via an endpoint and ELDA¹⁷ an implementation of *linked-data-api*¹⁸ that provides a configurable way to access RDF data using simple RESTful URLs that are translated into queries to a SPARQL endpoint. Other approach to consume “semantic data” consists on querying OWL models with SPARQL [18] and an extension for georeasoning with SPARQL has been developed in [19].

Product Scheme Classifications (also known as PSCs) are taxonomies, thesauri or controlled vocabularies that have been built to solve specific problems of interoperability and communication in e-commerce [5, 16]. The aim of a PSC is to be used as a standard de facto by different agents for information interchange in marketplaces [1, 8]. Any PSC, as well as other classification systems can be interpreted as: 1) domain-ontologies [4] or 2) conceptual schemes [10] comprised of conceptual resources. Finally, Good Relations¹⁹ is ontology for the e-commerce developed by Martin Hepp et. al.

⁶ <http://loted.eu:8081/LOTED1Rep/>

⁷ <http://data.gov.uk>

⁸ <http://nuts.psi.enacting.org/>

⁹ <http://www.aporta.es>

¹⁰ <http://cat.open.org.nz/>

¹¹ <http://data.australia.gov.au/>

¹² <http://www.aporta.es/>

¹³ <http://www.epimorphics.com/public/vocabulary/org.html>

¹⁴ <http://opencorporates.com/>

¹⁵ <http://vocab.deri.ie/void/guide>

¹⁶ <http://www.w3.org/TR/sparql11-service-description/>

¹⁷ <http://code.google.com/p/elda>

¹⁸ <http://code.google.com/p/linked-data-api/>

¹⁹ <http://www.heppnetz.de/projects/goodrelations/>

The use of semantic methods to exploit the data from the semantic web like Spreading Activation (SA) techniques and Rule Based Systems (RBSs) is widely used. The main application of SA techniques is focus on Document and Information Retrieval [3]. These techniques have been also used in semantic search based on hybrid approaches [2, 9] and user query expansion combining metadata and user information to improve web data annotations. RBSs have been used a long time to decision support, diagnosis, etc. in different fields. In the semantic web area and due to the apparition of OWL 2-RL, SPARQL Rules! and RIF (*Rule Interchange Format-W3C Recommendation 22 June 2010*), these systems are growing in their use to deal with the web of data but a clear approach to mix datasets and RBSs is missing. They can also be applied to SA techniques to handle the activation and propagation of the concepts. Finally, they are several algorithms (like *MapReduce* [12, 14, 17]) and approaches available to process a huge amount of data in searching and reasoning systems which are based on incremental indices [13], sync/async parallel search [15] and semantic search [2, 11].

3. – Main Contributions

The proposed work aims to apply the LOD approach and semantic web technologies to improve and ease the access to public procurement notices addressing the principles of OGD for the e-Procurement sector. Following, the main contributions are highlighted:

- Transforming vocabularies and PSCs developed by governments such as CPV²⁰, CPC²¹, Eurovoc²² (now available in SKOS), etc. to RDF, RDF(S), SKOS or OWL.
- Modeling the information inside the public procurement notices as web information resources and enriching them with the aforementioned controlled vocabularies, geographical information and the information now available in the linked data cloud.
- Publishing the information via an SPARQL endpoint providing a linked data node based on standards.
- Providing enhanced services (search and sort, matchmaking, geo reasoning, statistics, etc.) that exploit this semantic information through advanced algorithms based on Spreading Activation (SA) techniques, rule based systems (RBS) and a mixing of them.
- Exploiting the enhanced services and the information of the public procurement notices using standards, easing the access to the organizations (support to multilingual and multicultural issues), lowering prices and building new business models, especially interesting for SMEs.

The system with the aforementioned features is able to answer questions like the next one (this motivation example is developed in Section 5.1):

Which public procurement notices are relevant to Dutch companies (only SMEs) that want to tender for contracts announced by local authorities with a total value lower than 170K € to procure “Road bridge construction work” and a two year duration in the Dutch-speaking region of Flanders (Belgium)?

²⁰ Common Procurement Vocabulary-
http://europa.eu/legislation_summaries/internal_market/businesses/public_procurement/122008_en.htm

²¹ International Standard Industrial Classification of All Economic Activities-
<http://unstats.un.org/unsd/cr/registry/isic-4.asp>

²² EuroVoc is a multilingual, multidisciplinary thesaurus covering the activities of the EU-
<http://eurovoc.europa.eu/>

Finally, this platform partially addresses the challenges and objectives of the EU in its Work Programme for ICT (2011/2012)²³: *Objective FI.ICT-2011.1.7-Technology foundation: Future Internet Core Platform and Challenge 4: Technologies for Digital Content and Languages: Objective ICT-2011.4.1-SME initiative on Digital Content and Languages, Objective ICT-2011.4.4-Intelligent Information Management, Reactive algorithms, etc.*

4. – Public Procurement Platform-Functional Architecture Overview

The functional architecture, see Figure 1, illustrates the processes for retrieving the information about public procurement notices, *RDFizing* this information and linking together. Afterwards it is enriched with existing vocabularies from LOD initiative. Finally, the information is published via the SPARQL endpoint and can be retrieved through enhanced services based on semantic methods or directly querying the service of the endpoint with the SPARQL language.

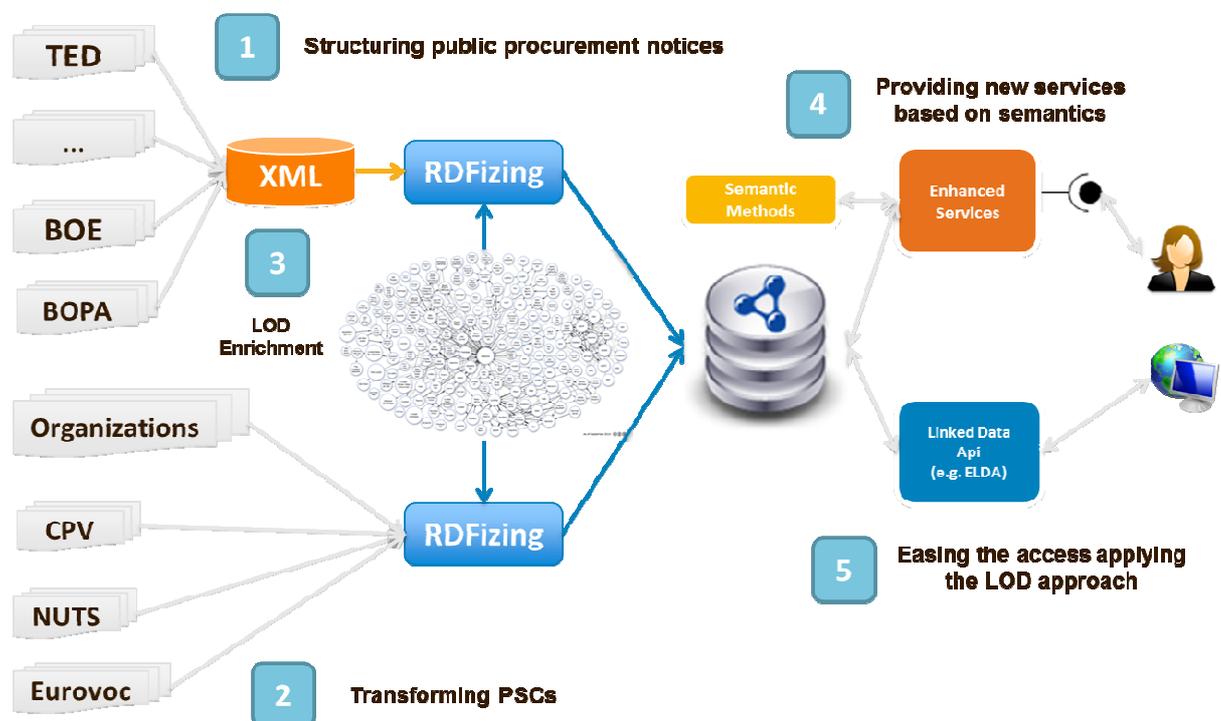


Figure 1 Functional architecture of the public procurement notices platform.

Following, we describe the building blocks and basis processes of the architecture:

- Public Procurement Notices. Source of documents published by official European organisms about public contracts. They are extrated out from all online available sources in European (e.g. TED- European public procurement journal), national (e.g. BOE-Official State Gazzete of Spain) and regional scopes (e.g. BOPA- Autonomous Community's Official Newspaper, Principality of Asturias-Spain).
- Public Procurement Notices Database. Local database containing the retrieved information from different sources. The format of this information is available in XML as intermediate language.

²³ ftp://ftp.cordis.europa.eu/pub/fp7/ict/docs/ict-wp-2011-12_en.pdf

- PSCs. Set of classifications used to classify the public procurement notices. Currently, they are available in different formats: MSEXcel, PDF, etc. They are coded using RDF/OWL.
- RDFizing. It is the process to transform the data available in the databases to generate a RDF view. This process link together all the information, enrich with the existing vocabularies and store it in a triple store.
- RDF Store. It is a triple store such as Virtuoso²⁴ or Sesame²⁵ in which the RDF data is stored and published via the SPARQL endpoint.
- Semantic Methods. It is the application of the libraries such as ONTOSPREAD²⁶ and RIFle²⁷ to exploit the information and provide enhanced services to the clients.
- Linking Open Data. It is the set of existing linked data vocabularies to be used in the enrichment of the information.
- Linked Data Api (e.g. ELDA). It is an API to publish information about datasets, URI schemes specification, etc.

The combination of these building blocks seeks for creating a new innovative way to exploit the information included in public procurement notices in the context of the semantic web and LOD initiatives reusing the existing technologies, vocabularies, etc. Following, the key points of this approach are summarized: 1) existing PSCs, information about public procurement notices and organizations are published as linked data and 2) the application of SA techniques, RBSs and a mixing of them to provide enhanced services adding value to original information.

5. – Providing Innovative Services for Public Procurement Notices

In the e-Procurement information domain, one of the targeted services to be improved is the “search of public procurement notices according to a profile”. In the context of searching, matchmaking refers to the procedure of retrieving a relevant list of results that matches with the intentions of an organization that wants to tender in a specific activity sector. Other interesting service on e-Procurement is the extraction of statistics to generate reports about the history of some place, organization or contracting authority. They can be exploited through temporal series, weighted aggregation operators [20, 21] or statistical inference, specifically predictive inference. In the next subsection we focus on the enhanced service for searching, see motivating example in Section 3, and a step-based improvement process is illustrated.

5.1. – Example of Use

The first example shows the process to enhance the search service from a simple query according to the profile of an organization and the CPV codes of the public procurement notices. In the second example, the query is extended to get results more relevant for this organization in the target sector. Finally, a system with features of georeasoning and fuzzy logic on public procurement notices is presented to get a more accurate query according to the information available about the public procurement notices, the previous history (statistics of the participation in public procurement processes) and the preferences of the organization. The URI prefixes of this example come from the “Prefix.cc²⁸”, a service to look up prefixes

²⁴ <http://virtuoso.openlinksw.com/>

²⁵ <http://www.openrdf.org/>

²⁶ API Java implementing Spreading Activation techniques. Available at: <http://code.google.com/p/ontospread>

²⁷ API Java implementing the RIF dialects: Core, PRD and DTB. Available at: <http://rifle.sf.net>

²⁸ <http://prefix.cc/>

for RDF developers. Also, *muo*²⁹ and *ppn* (namespace for public procurement notices) prefixes are also added to do a more human-readable example.

Let be *E* an organization that wants to tender in a public procurement process, the representation using N3, this information is provided by the process of RDFizing according to the information extracted out of the public procurement notices and other sources, is the next one:

```

<http://mydutchcompany.com/> a v:VCard ;
  v:fn "Dutch Company Inc." ;
  v:org [ v:organisation-name "Dutch Company Inc." ;
         v:organisation-unit "Corporate Division" ] ;
  v:adr [ rdf:type v:Work ;
         v:country-name "Netherlands" ;
         v:locality "Amsterdam" ;
         v:postal-code "1016 XJ" ;
         v:street-address "Lijnbaansgracht 215" ] ;
  v:geo [ v:latitude "52.36764" ;
         v:longitude "4.87934" ] ;
  v:tel [ rdf:type v:Fax, v:Work ;
         rdf:value " +31 (10) 400 48 00" ] ;
  v:email <mailto:company@mydutchcompany> ;
  v:logo <http://mydutchcompany.com/logo.png> .

<http://purl.org/weso/units/euro> a muo:UnitOfMeasurement;
  muo:measuresQuality <http://purl.org/weso/physicalQuality/Money>.
  muo:altSymbol "€" ;
  muo:prefSymbol "€" .

<http://purl.org/weso/ppn/noticeValue> a muo:QualityValue;
  muo:numericalValue "170.000";
  muo:inTime "2011-01-12" ;
  muo:measuredIn <http://purl.org/weso/units/euro>.

<http://purl.org/weso/units/year> a muo:UnitOfMeasurement;
  muo:measuresQuality <http://purl.org/weso/physicalQuality/Time>.
  muo:altSymbol "year" ;
  muo:prefSymbol "year" .

<http://purl.org/weso/ppn/noticeDuration> a muo:QualityValue;
  muo:numericalValue "2";
  muo:inTime "2011-01-12" ;
  muo:measuredIn <http://purl.org/weso/units/year>.

<http://purl.org/weso/organizations#dutchOrganization> a
org:FormalOrganization;
  org:purpose cpv:45221111-3 ;
  org:purpose cpv:45221113-7 ;
  org:purpose <http://purl.org/weso/ppn/noticeValue> ;
  org:purpose <http://purl.org/weso/ppn/noticeDuration> ;
  org:purpose <http://sws.geonames.org/50.85_43.49/ > ;
  skosxl:prefLabel "Dutch organization" ;
  org:classification <http://purl.org/organizations#SME>;
  org:hasSite <http://mydutchcompany.com/> ;
  org:siteAddress <http://mydutchcompany.com/> ;
  ... .
    
```

Table 1 Information about an organization in N3.

²⁹ MUO-Measurement Units Ontology by Fundación CTIC. Available at: <http://idi.fundacionctic.org/muo/>

Let be N a set of public procurement notices with the next features (all of them are *Active*):

ID (dct:identifier)	Product Description (CPV code)	Total Value (€)	Located in (lat/lon) NUTS-id (level)	Duration (years)	Classification
PPN-1	cpv:45221111-3 Road bridge construction work	160,000	lat "50.85" lon "43.49" NUTS-NL326 (3) 13860 Groot-Amsterdam	2	SME
PPN-2	cpv:45221113-7 Footbridge construction work	180,000	NUTS-BE2 (3) 290 VLAAMS GEWEST	3	SME
PPN-3	cpv:45221110-6 Bridge construction work	100,000	lat "50.85" lon "43.49"	2	SME
PPN-4	cpv:45221111-3	80,000	NUTS-BE2	2	SME
PPN-5	cpv:45221100-3 Construction work for bridges	190,000	NUTS-B3 (3) 300 RÉGION WALLONNE	4	SME
PPN-6	cpv:452211137	300,000	lat "50.85" lon "43.49"	3	SME
PPN-7	cpv:45221000-2 Construction work for bridges and tunnels, shafts and subways	200,000	NUTS-1025 (2) NL NEDERLAND	3	SME
PPN-8	cpv:45221113-7	150,000	lat "50.85" lon "43.49"	2	SME
PPN-9	cpv:45221111-3	100,000	NUTS-BE2	2	Large Company
PPN-10	cpv:45221114-4 Construction work for iron bridges	200,000	NUTS-275 (2) BE BELGIQUE-BELGIË	2	SME

Table 2 List of Public Procurement Notices.

The process to provide an enhanced service of searching notices is the next one: 1) the system builds directly a query, see Table 3, according to the profile (intentions that are turned to filters in the query) of an organization (CPV Codes, location, amount and duration) to be executed via the SPARQL endpoint, and 2) the query returns a result set with the public procurement notices matching the purposes of the client (organization, etc.)

```
SELECT ?id ?description ?amount ?duration WHERE{
  ?notice rdf:type ppn:PublicProcurementNotice .
  ?notice dct:identifier ?id .
```

```
?notice dct:description ?description .
?notice hasStatus ppn:Active .
?notice org:classification <http://purl.org/organizations#SME> .
?notice wgs84_pos:lat ?lat.
?notice wgs84_pos:lon ?long .
?notice ppn:hasAmount ?amount.
?amount muo:measuredIn <http://purl.org/weso/units/euro> .
?notice ppn:hasDuration ?duration.
?duration muo:measuredIn <http://purl.org/weso/units/year> .
FILTER (
((?notice hasCPVcode cpv:45221111-3) OR (?notice hasCPVcode cpv:45221113-
7))
and (?lat == "50.85") and (?long == "43.49")
and (?amount <= 170,000^xsd:double) and (?duration <= 2) )}
```

Table 3 Basic SPARQL query.

The results of this request, see Table 4, are similar to a search process based on a classical relational database query (e.g. using SQL) and it does not suppose neither advance in the state of the art nor an enhanced service for searching. It only matches the profile of the company with the characteristics of the public procurement notices, if there is an exact match then the public procurement notice is selected in the result set.

ID (dct:identifier)	Product Description (CPV code)	Total Value (€)	Duration (years)
PPN-1	cpv:45221111-3 Road bridge construction work	160,000	2
PPN-8	cpv:45221113-7 Footbridge construction work	150,000	2

Table 4 Result set of the first SPARQL query.

- In the second example, the system builds a SPARQL query but previously the Spreading Activation techniques are performed on CPV with the purpose codes as input parameters and a value for each relation in the CPV (*skos:broader*, *skos:narrower*, etc.) to get a list of most representative concepts according to the input ones, see Table 5. This new query uses this rank list of concepts (CPV codes) to build an extended SPARQL query that finally is executed via the endpoint.

CPV Code	Product Description	Activation Value (normalized)
cpv:45221110-6	Bridge construction work	0,95
cpv:45221113-7	Footbridge construction work	0,65
cpv:45221111-3	Road bridge construction work	0,65
...
cpv:45221114-4	Construction work for iron bridges	0,25

Table 5 Ranking of concepts after Spreading Activation.

Currently, the SPARQL language does not allow to specify that a triple match has more relevance to sort the results than other triple match as existing syntactic search engines do like

Apache Lucene³⁰ (boosting terms). This situation implies that a query for each CPV code must be performed, see Table 6. Finally, a merge process must gather the results and sort them according to the relevance of the specific CPV code. That is why a query is built for each CPV code.

```
SELECT ?id ?description ?amount ?duration WHERE{
  ...
  ?notice ppn:hasCPVcode cpv:45221110-6 .
  FILTER ( (?lat == "50.85") and (?long == "43.49")
    and (?amount <= 170,000^xsd:double) and (?duration <= 2) ) }
```

Table 6 One of the enhanced SPARQL queries

The results of this request, see Table 7, are different from the first ones (a new public procurement notice is added) because the search has been spread using new CPV codes not presented in the input parameters but representative to the preferences of the organization.

ID (dct:identifier)	Product Description (CPV code)	Total Value (€)	Duration (years)
PPN-3	cpv:45221110-6 Bridge construction work	100,000	2

Table 7 Result set of the second SPARQL query.

Finally, the third improvement consists in the application of georeasoning (like the *findNearby* service of GeoNames³¹) and fuzzy logic techniques (e.g. aggregation operators) to evaluate the history (statistics) of the company in its previous participations in public procurement processes to establish an extended margin on the total value and duration of the contract announced in the public procurement notices. These two processes generates the specific values to build an extended semantic query in SPARQL, see Table 8, keeping the use of the Spreading Activation to previously create a ranking of most representative concepts. The formal mathematical models of these functions are out of the scope of this paper and they will be reported to the appropriate logical/mathematical venue.

```
SELECT ?id ?description ?amount ?duration WHERE{
  ...
  ?notice nuts:containedBy ?place .
  ?notice hasCPVcode cpv:45221113-7 .
  FILTER ( ( (?place nuts:containedBy nuts:NUTS-NL326 ) or
    (?place nuts:containedBy nuts:NUTS-1025) or
    (?place nuts:containedBy nuts:NUTS-B3) or
    (?place nuts:containedBy nuts:NUTS-BE2) or
    ) and (?duration <= 3) and (?amount <= 200,000^xsd:double)) }
```

Table 8 One of the enhanced SPARQL queries with georeasoning and fuzzy logic

The final results of this request, see Table 9, (taking into account the aforementioned situation with SPARQL queries) are the next ones (results from first and second queries are skipped).

ID (dct:identifier)	Description (CPV code)	Total Value (€)	Duration (years)
PPN-2	cpv:45221113-7 Footbridge	180,000	3

³⁰ <http://lucene.apache.org>

³¹ <http://www.geonames.org/export/web-services.html>

	construction work		
PPN-4	cpv:45221111-3 Construction work for bridges	80,000	2
PPN-7	cpv:45221000-2 Construction work for bridges and tunnels, shafts and subways	200,000	3

Table 9 Result set of the third SPARQL query.

The public procurement notices PPN-5, PPN-6, PPN-9 and PPN-10 are never retrieved due to: 1) and 2) the duration and the total value of the public contract are out of the permitted limit, 3) the public contract requires a large company and 4) the geographical scope of PPN-10 is Belgium and it cannot be ensured that this public procurement notice matches with the geographical area desired by the company.

6. – Validation and experimentation

Currently we are finishing the process for publishing the PSCs and the information extracted from public procurement notices as linked data. We are tuning up the SA techniques for working with rules and generating SPARQL queries to be executed in a SPARQL endpoint. Finally we are designing the experiment to validate the goodness and the improvement of the system regarding to existing systems. In that sense, the experiment apart from the selected service to be tested depends on two main variables: 1) the amount of information used and 2) the number of tests that should be carried out. From the first variable point of view 10^6 public procurement notices (provided by Gateway SCS-Euroalert.net³²) and over 400,000 organizations are available. On the second one, we are studying how many tests would be appropriate to provide a correct evaluation but the information about how many queries are requested per day in the existing public systems can be a right trail.

On the other hand, taking into account that the service of searching or matchmaking is the most relevant in this kind of system we are preparing a test suite with the aforementioned information (search queries and expected results) to compare the precision and recall of existing public systems (free text and advanced key fields search of TED) to the proposed one (LOD+SA+RBS+SPARQL). The expected result of this evaluation will validate our approach for improving the access and retrieval of the information about public procurement notices using the LOD approach.

7. – Conclusions and Future Work

According to the EC, e-Procurement may, by its nature, be more compatible or facilitate the use of procurement budgets in support of EU 2020 objectives. Some countries have adopted the recommendations of the EU introducing strategies to encourage SMEs participation in the procurement processes but the intended participation of the SMEs is still far from the EU expectations. The proposed platform tries to ease the access to the public procurement notices using LOD and semantic methods. The proposed approach partially fulfills the EU expectations and is supposed to provide a new way to exploit the information published inside public procurement notices applying advanced algorithms on LOD. Following the advantages

³² <http://euroalert.net/>

of this approach are presented: decreasing of the information's dispersion, unification of the data models and formats, implicit support to multilingual and multicultural issues, enrichment of the public procurement notices, alignment with the Digital Agenda for Europe, raising awareness on public procurement opportunities among SMEs and deployment of innovative and enhanced services on public procurement notices based on standards.

Regarding the future work, the results of this study are intended to be exploited by a commercial service like EuroIert.net [6, 7] and we are also interested in report the results to The Internal Market and Services Directorate General (DG MARKT) of the EC, The Information Society and Media Directorate General (DG INFSO) of the EC, the LOD and OGD initiatives among others.

Acknowledgements

This work is part of “10ders Information Services project”³³ partially funded by the Spanish Ministry of Industry, Commerce and Tourism (TSI-020100-2010-919) and the European Regional Development Fund (EFDR), led by “Gateway Strategic Consultancy Services” and developed in cooperation with “Exis TI”³⁴ and “WESO Research Group”³⁵.

References

- [1] G. Alor-Hernández, JM. Gómez Berbís, and A. Rodríguez González et al. “HYDRA: A Middleware-Oriented Integrated Architecture for e-Procurement in Supply Chains”. T. Computational Collective Intelligence, 1:1_20, 2010.
- [2] D. Berrueta, JE. Labra, and L. Polo. “Searching over Public Administration Legal Documents Using Ontologies”. In JCKBSE, pages 167_175, 2006.
- [3] PR. Cohen and R. Kjeldsen. “Information Retrieval by Constrained Spreading Activation in Semantic Networks”. Inf. Process. Manage., 23(4):255_268, 1987.
- [4] M. Hepp. “Possible Ontologies”. IEEE-Internet Computing, (1):90_96, 2007.
- [5] J. Leukel and V. Schmitz et al. “Exchange Of Catalog Data In B2B Relationships - Analysis And Improvement”.
- [6] JL. Marín. “Web 2.0. Una descripción muy sencilla de los cambios que estamos viviendo”. Netbiblio, 2010.
- [7] JL. Marín and JE. Labra. “Doing Business by selling free service”. In P. Ordoñez et al., editor, Web 2.0: The Business Model, part 6, pages 89_102. Springer, 2009.
- [8] B. Omelayenko and D. Fensel. “An Analysis of B2B Catalogue Integration Problems”. In Proc. of the International Conference ICEIS-2001, Setúbal, Portugal, 2001.
- [9] C. Rocha and D. Schwabe et. al. “A Hybrid Approach for Searching in the Semantic Web”. In WWW, pages 374_383, 2004.
- [10] JM. Álvarez, E. Rubiera and L. Polo. “Promoting Government Controlled Vocabularies for the Semantic Web: the EUROVOC Thesaurus and the CPV Product Classification System”. SIEDL 2008.
- [11] R. Baeza-Yates, M. Ciaramita, P. Mika. and H. Zaragoza. "Towards Semantic Search" Natural Language and Information Systems (NLDB), Volume 5039, London, UK, p.4-11 (2008) .
- [12] H. Karlo, S. Suri and Vassilvitskii. "S. A Model of Computation for MapReduce". Symposium on Discrete Algorithms (SODA) (2010)
- [13] R. Blanco, E. Bortnikov, F. Junqueira, R. Lempel, L. Telloli and H. Zaragoza "Caching Search Engine Results over Incremental Indices". SIGIR'2010 (2010)
- [14] T. Condie, N. Conway, P. Alvaro, JM. Hellerstein, K. Elmeleegy and R. Sears. "MapReduce Online". The 7th USENIX Symposium on Networked Systems Design and Implementation (NSDI '10) (2010)

³³ <http://rd.10ders.net>

³⁴ <http://www.exis-ti.com/>

³⁵ <http://purl.org/weso>

- [15] M. Marin "Sync/Async Parallel Search for the Efficient Design and Construction of Web Search Engines". Parallel Computing, Elsevier (2010).
- [16] S. Méndez, J. de Andrés Suárez, JE. Labra Gayo and P. Ordóñez. "A Semantic Based Collaborative System for the Interoperability of XBRL Accounting Information". In: WSKS 1 (2008), p. 593-599.
- [17] J. Dean and S. Ghemawat. "MapReduce: Simplified Data Processing on Large Clusters". OSDI'04., San Francisco, CA, December, 2004.
- [18] E. Sirin, B. Parsia, B. Cuenca, Kalyanpur, A. & Katz, Y. Pellet: A practical OWL-DL reasoner. Web Semantics Science Services and Agents on the World Wide Web 5, 51-53 (2007).
- [19] X. Lopez. "GeoSPARQL - A geographic query language for RDF data. A proposal for an OGC Draft Candidate Standard. Slides at: http://www.ogcnetwork.net/system/files/Spatial_SPARQL_Lopez.pdf
- [20] V. Torra and Y. Narukawa, Modeling Decisions. Information Fusion and Aggregation Operators. Berlin, Germany: Springer, 2007.
- [21] M. Grabisch, J.-L. Marichal, R. Mesiar, and E. Pap, Aggregation Functions. New York: Elsevier, 2009.