

Euroalert.net: Building a pan-European platform to aggregate public procurement data and deliver commercial services for SMEs powered by open data

José Luis Marín, Ángel Marín and Mai Rodríguez
Gateway Strategic Consultancy Services
Valladolid, Spain

Email: {josmar,anmar,mrodriguez}@gateway-scs.es
Ramón Calmeau EXIS-TI
Valladolid, Spain

Email: ramon.calmeau@exis-ti.com
Jose María Álvarez and José Emilio Labra
WESO RG

Universidad de Oviedo
Oviedo, Spain
Email: {josem.alvarez,jelabra}@weso.es

Abstract—This paper aims to describe a use case of pan-European open data re-use, where Euroalert.net¹ is developing a thriving commercial service fueled by the re-use of public sector information (PSI). It is focused on one of the four major areas of activity developed by Euroalert.net – i.e., the processing of public procurement documents; and describes how a small company can benefit from open data approaches and generate value at a pan-European scale, through innovation. The paper also describes “10ders information Services - 10ders IS”², which is a Euroalert.net R&D project aimed to build a pan-European platform that will aggregate the big data about public procurement information generated by hundreds of thousands of contracting authorities in the EU. One key objective for this new platform is to deliver innovative services based on PSI, specially for SMEs. Finally, the aim of this paper is to highlight some obstacles which arise in the process to commercially re-use public data, not only legal and technical ones but also those caused by the lack of an open data culture among public authorities.

Index Terms—PSI re-use, open data, open government data, linked open data, EU single market, e-procurement

I. INTRODUCTION

EUROALERT.NET (ISSN 1988-3382) is a brand owned by Gateway SCS³, focused in delivering innovative and cost-effective information services. These innovative services are designed to add value to public sector information generated within the framework of public contracts published in EU Member States as well as the information generated by EU Institutions own activity. Euroalert.net monitors, processes, and stores information in four key areas: (1) **News** about EU Institutions activity (2) **EU Legislation**, (3) **EU Funding**

opportunities and (4) **Public Contracts** issued in EU Member States. All the products and services focused on public procurement information, which are mainly designed for EU SMEs, are gathered under the brand “10ders”.

On top of each of these areas, Euroalert.net provides its clients with value added services of different nature which exploit public sector information gathered, indexed, enriched, mashed-up and repackaged with different purposes. This paper is specifically focused on public procurement information which, by itself, is a big challenge to be solved in order to provide new customized services to organizations and companies across Europe.

Government bodies as a whole are the biggest buyers in the European Union, since their purchases together represent over 17% of GDP⁴. However, the lack of a pan-European information service to provide one single access to this huge market of business opportunities estimated at more than 2,155 billion Euros is a great obstacle for companies, especially SMEs.

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) has outlined the following advantages brought by the wider use⁵ 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.

¹<http://euroalert.net/>

²<http://rd.10ders.net>

³<http://gateway-scs.es/en/>

⁴http://ec.europa.eu/bepa/pdf/monti_report_final_10_05_2010_en.pdf

⁵http://ec.europa.eu/internal_market/consultations/docs/2010/e-procurement/green-paper_en.pdf

The European Commission, in its “European Code of Best Practices Facilitating Access by SMEs to Public Procurement Contracts”⁶, identified eight groups of difficulties in this field. Two groups of these difficulties refer to the barriers faced by SMEs in accessing relevant information about tenders. The document states that the “the (big) number of such web portals being used by the government and by regional and local authorities makes it difficult for tenderers to maintain an overview”. *10ders IS* deals with this problem and intends to improve a situation affecting more than 20 million non-financial companies established in the EU.

In the past two years an increasing amount of governments in Europe and all over the world have created data catalogues for Open Government Data (OGD) to ease the access to public information. The Digital Agenda for Europe⁷, Europe’s strategy for a flourishing digital economy by 2020, is also supporting Open Data strategies. Furthermore, one of Digital Agenda’s Big Ideas for the 1st Digital Agenda Assembly entitled “Beyond raw data: public sector information, done well”⁸, is especially focused on this topic.

The availability of public data enables⁹ greater transparency; delivers more efficient public services; and encourages greater public and commercial use and re-use of government information. Initiatives such as Linking Open Data (LOD)¹⁰ and the semantic web provide additional background to exploit this new context of publishing public data.

Euroalert.net is aware of this context and interacts with the open data community as well as with governments (at local, regional, national and EU level) who, in the end, are the holders of the data used by *10ders IS*, with the aim to promote the release of new datasets in machine-readable formats.

The objective of building a pan-European platform which aggregates the huge amount of tender notices issued by contracting authorities in every EU country presents a set of overwhelming challenges, which Euroalert.net is currently in the path to solve within the context of its project *10der IS*. These challenges can be summed up in some main issues such as: (1) Getting the data, (2) No unified data structure, (3) Storing big data, (4) Multiple languages, (5) Exploitation in services of different nature.

In conclusion, this project represents a big investment in R&D which intends to develop an efficient architecture on top of which Euroalert.net will build more viable commercial services powered by public sector information.

II. 10DERS INFORMATION SERVICES, THE PLATFORM AND THE R&D PROJECT

The development of “Tenders Electronic Daily” - TED¹¹, can be considered as the biggest EU-wide effort made to ease

interoperability within Europe’s e-Procurement framework. TED is the online version of the S Series of the Official Journal of the European Union – OJEU, which is a supplement to the Journal dedicated to European public procurement. TED publishes over 1,500 new over EU-threshold value procurement¹² notices every day. According to Eurostat estimates, in 2009 TED covered 3.60% of EU27 GDP¹³. This means that below threshold tenders – i.e., low value procurement, represented over 13% EU27 GDP, a huge amount of opportunities targeted by *10ders IS* which are out of the scope of TED.

In order to accomplish this overwhelming challenge, *10ders IS* is developing innovative approaches for the five main difficulties previously mentioned:

- 1) **Getting the data.** Tender notices are published in a wide range of public sources at European, national, regional and local levels. Furthermore, lower-value tenders following restricted publication procedures are posted on the websites of each institution, agency or public authority. *10ders IS* is working on an efficient architecture for tracking and monitoring thousands (estimate) sources of data. The system will be capable of identifying new opportunities posted daily in all those sources and will also be able to add new sources as they are recognized. This step represents a technological problem, but also a negotiation issue with the main data holders that will later be described in the section “Approach to getting the data”.
- 2) **No unified data structure.** Although the data concerning tenders are public - though not always reusable, there is no unified format for the notices, not even a limited number of possibilities for expressing the information. Downloading and processing documents from each source in order to be aggregated involves a number of complex problems to understand the structured or unstructured information in the notices and to extract data that can be mixed for all these sources. Furthermore, the available data and their format can vary over time given that most public sources do not provide a guarantee of stability and future compatibility of the formats used to publish them.
- 3) **Storing big data.** The amount of information to be processed by the system makes it necessary to solve the issue about how to store and efficiently exploit it. In this regard, it can be pointed that, only the OJEU itself, ads more than 20,000 documents every day. The storage strategy faces the additional challenge to be scalable in order to include new sources of information at a large scale.
- 4) **Multiple languages.** As an added challenge *10ders IS* deals with the fact that tender documents may be published in one or more of the 23 official languages of the European Union. In addition, each country sometimes use other co-official languages in their territory. All algorithms and systems are able to process, aggregate and understand information which is written in a large number

⁶ http://ec.europa.eu/internal_market/publicprocurement/docs/sme_code_of_best_practices_en.pdf

⁷ http://ec.europa.eu/information_society/digital-agenda/index_en.htm

⁸ http://ec.europa.eu/information_society/events/cf/dae1009/item-display.cfm?id=5_257

⁹ <http://www.w3.org/TR/gov-data/>

¹⁰ <http://www.w3.org/wiki/SweoIG/TaskForces/CommunityProject-s/LinkingOpenData>

¹¹ <http://ted.europa.eu/>

¹² http://simap.europa.eu/supplier/opportunities-in-europe/index_en.htm

¹³ <http://epp.eurostat.ec.europa.eu/tgm/table.do?tab=table&init=1&plugin=1&language=en&pcode=tsier09>

of languages and also deal with cultural specificities such as currency, tax system or the way postal addresses are encoded.

- 5) **Exploitation in different services.** A unified information system with structured data from public procurement announcements offers multiple possibilities of exploitation, for different products, with a wide range of operational requirements. Some of these services will even be designed in the future, so original documents will have to be re-processed without penalizing the performance of existing services.

Further to all these achievements, some of the publicly available contributions made by the platform to the community will be:

- The transformation of vocabularies developed by governments such as CPV¹⁴, Eurovoc¹⁵ (now available in SKOS), etc. to RDF, RDF(S), SKOS or OWL. Euroalert.net will publish through its infrastructure the information via an SPARQL endpoint providing a linked data node based on standards.
- A model for the information inside public procurement notices as web information resources, enriched with controlled vocabularies, geographical information and the information now available in the linked data cloud. Euroalert.net will release its model (code name opXML) for representing tender information, so it can be adopted by any interested stakeholder in the public procurement ecosystem, and potentially standardised by any interested body.

III. INNOVATIVE SERVICES AND MAIN CLIENTS

The ultimate objective of *10ders IS* effort is to exploit the information contained in public procurement notices via enhanced services, by easing the access to the organizations, lowering prices of services, and building new business models especially focused on SMEs.

Euroalert.net is currently offering both free and subscription services¹⁶, packaged in many different innovative products which are already in the market. Some of these products, specially intended to be cost-effective for SMEs, are:

- **Tenders alerting systems**¹⁷. Euroalert.net helps companies to keep track of business opportunities in EU public sector which are relevant for their products or services. Companies can customize a profile with dozens of criteria which then helps the system algorithms to identify relevant opportunities.
- **Tenders Product Reports**¹⁸. Euroalert.net also helps its clients to get into new markets, by offering a simple and clear report listing major public buyers for their products and services. The report also includes data regarding the most successful contract winners who can be their competitors, or maybe appropriate partners for alliances or

trade agreements. This is an excellent strategic planning and market analysis tool for Internationalization Strategic Plans.

- **Listings of companies winning contracts.** Euroalert.net offers to its clients the possibility to obtain lists of operators that have been awarded public tenders. This winning companies can be extracted by customized criteria – either winning in a particular country, established in a certain country, related to a certain group of products or services, etc.
- **Outsourcing of alerting systems and customized newsletters.** Euroalert.net offers the institutions and organizations it works with, the possibility to outsource information systems on public procurement, as all its services can be purchased in private-branded configurations.
- **Free Widgets or One Web Full Content Widgets.** Euroalert.net provides customized widgets which allow its subscribers to automatically publish in their own website any content distributed by Euroalert.net. Contents can be filtered according to the client's criteria and fully adapted to the layout of their website. Thus, the website can be fully updated with no maintenance nor dedication of additional resources.
- **Customized data mining for market intelligence.** Euroalert.net provides raw statistics on public procurement, tailored to the interests of each client. It helps in finding and extracting knowledge and trends from the hundreds of thousands of business opportunities with the public sector which are published every year.
- **Web Services.** Contents are also provided so they can be automatically stored in clients' databases. This allows them to combine the information provided by Euroalert.net with other sources of information and enrich the chain of information re-use.

Some of these products have already been in the market for a few years and are benefiting from the increasing amount of sources of information aggregated by *10ders IS*, which makes them much more valuable for clients. This has allowed Euroalert.net to achieve a growing clients portfolio¹⁹ across Europe, ranging from Regional Governments to Chambers of Commerce, Business and Trade Associations, as well as Agencies for International Trade and Investment Promotion, apart from private held companies.

Some outstanding examples of the organizations that are currently trusting Euroalert.net services in this field are the London Chamber of Commerce (UK), the Birmingham Chamber of Commerce (UK), Enterprise Ireland (IE), Poland Trade and Investment (PL), Promo Madrid (ES), Yorkshire Regional Development Agency (UK) or Junta de Castilla y León (ES).

IV. APPROACH ON GETTING DATA

Although for re-users it would certainly be good news if data holders released datasets according to the eight Open Governments Data (OGD) principles²⁰ (complete, primary, timely, accessible, machine processable, non-discriminatory,

¹⁴http://europa.eu/legislation_summaries/national_market/businesses/public_procurement/122008_en.htm /inter-

¹⁵<http://eurovoc.europa.eu/>

¹⁶<http://euroalert.net/en/buy.aspx>

¹⁷<http://euroalert.net/en/tenders-alerts>

¹⁸<http://euroalert.net/en/Tenders-Reports>

¹⁹<http://euroalert.net/en/clients-portfolio.aspx>

²⁰<http://www.opengovdata.org/home/8principles>

non-proprietary, and license-free), reality is still far and away from this point. Most of Open Data Portals initiatives are still posting public sector information in not machine-readable formats (even in PDF and HTML) and unfortunately not many datasets can be marked high according to the five-star ranking system²¹ proposed by Sir Tim Berners-Lee for OGD. This is with no doubt a big obstacle for commercial re-users.

The use of scrapping techniques to extract data from PDFs or HTML documents and the post-cleaning of these datasets, is probably the most wide spread approach among public sector information re-users to get raw datasets. Furthermore, many commercial services are even still based in manual copy-paste approaches from original sources of information.

This kind of data-acquisition models involve many limitations in order to build accurate, advanced and cost-effective commercial services, as datasets which come from the extraction process are full of errors or very difficult to keep up-to-date. Certainly, this is not a good scenario to build a business on top of it.

Euroalert.net has decided to follow a pure open data approach to build its tenders services and is not aggregating any source which does not provide machine readable datasets in origin. This strategy results in less data at first, but better data, which in the end represent more valuable data and better services for Euroalert.net client, both sustainable and scalable at a pan-European level.

The steps followed by Euroalert.net while aggregating a new source of public procurement information to *10ders IS* can be described as: (1) Formal request to data holder to release the datasets, (2) Sign license agreement (either free or fee-based), (3) Implementation of technical process, (4) Upgrade of existing services with new data.

Once a non-exclusive license is granted by the data holder, Euroalert.net launches a technical process in order to understand the structure of the dataset and how it can be integrated with *10ders IS* platform.

Unfortunately, it is not always possible to sign a license agreement as some data holders are not still aware about open data policies or are simply not willing to release data for commercial re-use. In many cases this has been identified as the most important barrier to re-use public sector information, once the technical issues have been solved.

V. CONCLUSIONS AND FUTURE WORK

Long before the publication of the 2003 EU Directive on the re-use of public sector information²², when Euroalert.net began providing business information services for public authorities and companies, its management board was already aware of the potential that the increasing amount of information available in reusable formats could represent. It was already clear for Euroalert.net that this was a great opportunity for growth and a possibility to develop more advanced services, generating value and creating new jobs, an idea which is

²¹<http://epp.eurostat.ec.europa.eu/tgm/table.do?tab=table&init=1&plugin=1&language=en&pcode=tsier090>

²²<http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:32003L0098:ES:NOT>

line with the objectives set by the Europe 2020 Strategy put forward by the Commission²³.

Thanks to the strategy applied by the company over the past years, which has been described in the previous sections of this paper, Euroalert.net has been able to launch since 2009 a full range of new value added services. Applying this processes, the quality of the information has been considerably improved and contains fewer errors. The amount of the information which is processed has grown and now comes from different public authorities in different countries. This increases the possibility to add value to information by combining data coming from multiple sources.

In this point, Euroalert.net intends to follow the path described in this paper in order to reach the full potential of the platform, which lies upon reducing the current degree of dispersion of information, standardizing data models and formats, enriching public procurement notices as well as the implicit support to multilingual and multicultural issues. This process is already helping to raise awareness among European SMEs about business opportunities available for them and brought by tendering information, allowing them to take full advantage of all opportunities. All this is already contributing to the objectives set by the European Commission in the Digital Agenda for Europe as well as those laid down in the Small Business Act for Europe²⁴.

However, and despite the important steps made over the past two years, a scenario where all governments work together with re-users to get the maximum value of the information they are holding, is still a distant future.

Being aware of it, and trying to contribute to create the appropriate environment for the commercial re-use of public sector information, Euroalert.net has been active in disseminating open data culture. In so doing, the company launched its project *ShowMeTheTenders.net*²⁵ and presented it at the PSI Meeting 2010 held in Madrid²⁶, and is also releasing the knowledge acquired in the process of opening up new sources of public procurement data.

Euroalert.net hopes that all these actions made in its domain of interest and the joint efforts contributed to the Open Data community, will help to rise awareness among public authorities so that the opening of data sources and the procedures to request raw datasets are put in place and clarified.

Acknowledgements

This paper is part of “10ders Information Services project” co-funded by the Spanish Ministry of Industry, Commerce and Tourism (TSI-020100-2010-919) and the European Regional Development Fund (EFDR). The project is led by Gateway Strategic Consultancy Services which holds Euroalert.net brand, and developed in cooperation with the company “Exis TI”²⁷ and “WESO Research Group”²⁸ from the University of Oviedo.

²³http://ec.europa.eu/europe2020/index_en.htm

²⁴ <http://ec.europa.eu/enterprise/policies/sme/small-business-act/>

²⁵<http://showmethetenders.net/>

²⁶<http://www.aporta.es/web/guest/psimeeting2010eng>

²⁷<http://www.exis-ti.com/>

²⁸<http://purl.org/weso>