



European Public Sector Information Platform

Topic Report No. 2012 / 7

State of Play: Re-use of Public Procurement Data

Authors: Hans Graux and Tom Kronenburg

Published: August 2012

Keywords

PSI, Public Sector Information, Procurement, Public Procurement, Open Government, transparency, accountability, open spending, value for money, competition, internal market

Abstract

Procurements by the public sector represent an enormous part of GDP in the European Union. According to a 2010 study conducted on behalf of the European Commission¹, around 17,30% of the EU GDP, or about 2,16 trillion EUR, is spent on public procurements. Open availability of procurement data represents a significant opportunity for all stakeholders: it facilitates access to procurement markets for aspiring bidding companies, the increased competition provides better value for money for governments in times of austerity, and the transparency on public spending improves democratic legitimization towards European citizens, thus strengthening democracy as a whole. For this reason, the publication of some procurement data has been made mandatory in public procurement legislation, ensuring that at least a baseline of information is available to aspiring re-users.

This paper will show how the re-use of procurement data can serve both to create economic benefits and to evaluate the effectiveness and fairness of public spending, including through specific examples. It will also illustrate the main challenges and gaps that re-users still face, such as the unavailability of data and the difficulties in ensuring its usability.

This report shows that procurement data can have immense value to society. It is therefore imperative that the holders of procurement data are aware of the potential benefits that can be realized, and adopt policies that allow re-users to overcome these barriers. Only by doing so can the full potential of procurement data be unlocked.

¹ Study on the evaluation of the Action Plan for the implementation of the legal framework for electronic procurement, 2010, see http://ec.europa.eu/internal_market/consultations/docs/2010/e-procurement/siemens-study_en.pdf

Table of Contents

Abstract	4
1 Introduction	5
2 Procurement Data	6
3 Publicity and openness in public procurement legislation	7
3.1 <i>Basic rules and principles</i>	7
3.2 <i>Limitations of EU public procurement legislation</i>	8
3.3 <i>Re-use and the PSI Directive as solution</i>	9
4 State of Play: sources of procurement data	10
4.1 <i>At the European level</i>	10
4.2 <i>EU Member States / regional / local</i>	12
5 Re-Users of Public Procurement Data	13
5.1 <i>Commercial Re-use</i>	13
5.2 <i>Civic Hackers and Transparency Activists</i>	15
5.3 <i>Academic Re-use</i>	16
6 Conclusions	17

Abstract

Procurements by the public sector represent an enormous part of GDP in the European Union. According to a 2010 study conducted on behalf of the European Commission², around 17,30% of the EU GDP, or about 2,16 trillion EUR, is spent on public procurements. Open availability of procurement data represents a significant opportunity for all stakeholders: it facilitates access to procurement markets for aspiring bidding companies, the increased competition provides better value for money for governments in times of austerity, and the transparency on public spending improves democratic legitimization towards European citizens, thus strengthening democracy as a whole. For this reason, the publication of some procurement data has been made mandatory in public procurement legislation, ensuring that at least a baseline of information is available to aspiring re-users.

This paper will show how the re-use of procurement data can serve both to create economic benefits and to evaluate the effectiveness and fairness of public spending, including through specific examples. It will also illustrate the main challenges and gaps that re-users still face, such as the unavailability of data and the difficulties in ensuring its usability.

This report shows that procurement data can have immense value to society. It is therefore imperative that the holders of procurement data are aware of the potential benefits that can be realized, and adopt policies that allow re-users to overcome these barriers. Only by doing so can the full potential of procurement data be unlocked.

² Study on the evaluation of the Action Plan for the implementation of the legal framework for electronic procurement, 2010, see http://ec.europa.eu/internal_market/consultations/docs/2010/e-procurement/siemens-study_en.pdf

1 Introduction

Procurements by the public sector represent an enormous part of GDP in the European Union. According to a 2010 study conducted on behalf of the European Commission³, around 17,30% of the EU GDP, or about 2,16 trillion EUR, is spent on public procurements. Due to the significant economic interests involved, public procurement rules and policies need to favour openness, at least to some extent: the more openly and widely procurement opportunities are published, the more bids can be expected from interested tenderers. This increases competition, thus improving value for money.

But the drive behind openness of procurement data is not purely motivated by economics. Since the budgets being spent originate from tax payers, there is also a strong motivation to ensure that money is spent fairly on the basis of equal competition. Therefore, openness is not only required as a way of optimising the participation in tendering opportunities, but also to show tendering outcomes to the public: who won a public contract, and under which terms? On this point too, a certain degree of openness is required.

As we shall see below, the EU level procurement legislation contains certain measures to ensure the public availability of this data, separate from PSI rules, and does not contain any inherent restrictions on the usage of the data. As a result, one might reasonably expect that several interesting re-use cases have emerged over the past years.

In the sections below, we will briefly explore the concept and scope of procurement data, the extent to which such data is legally required to be opened up under public procurement rules, examples of procurement data re-use, and finally some challenges and gaps.

³ Study on the evaluation of the Action Plan for the implementation of the legal framework for electronic procurement, 2010, see http://ec.europa.eu/internal_market/consultations/docs/2010/e-procurement/siemens-study_en.pdf

2 Procurement Data

While there is no formal definition of procurement data, as a working definition it can broadly be defined as any data relating to the organisation, participation or outcome of public tenders, including information on:

- The object of the tender: what is being tendered for (goods, works, or services), and what are the specific requirements (e.g. required functionality, quality of the goods-works-services, timing for delivery, etc);
- Identification information on the participants, such as the contracting authority (i.e. the public administration making the purchase), the economic operator (i.e. the tenderer), and any intermediaries (e.g. central purchasing bodies making purchases on behalf of other public administrations, or the operators of procurement services or portal sites);
- Information on the permitted competition within the tender, determining which tenderers are eligible to participate and under which conditions, e.g. through an open procedure, restricted procedure, competitive dialogue or negotiated procedure;
- Financial/budget related information, notably the monies transferred, reserved or budgeted for the procurement, and the sum of the winning bid after the contract is awarded;
- Information on the outcome of the tender, notably the identity of the winning tenderer.

Procurement data has a natural link to spending data, as much of government spending is the result of procurements. However, procurement data has a number of specific peculiarities. One of the most prominent differences is the existence of a European harmonized legal framework on public procurements, which directly impacts openness and availability of procurement data. In the sections below, we will firstly examine how and to what extent the European legal framework for public procurement encourages openness.

3 Publicity and openness in public procurement legislation

3.1 Basic rules and principles

As noted in the introduction, the European Union has adopted specific rules to harmonize public procurement legislation in the Member States. Currently, the primary legal instruments are Directives 2004/17 and 2004/18 EC. The former Directive (2004/17) applies only to the utilities sectors (water, energy, transport and postal services), whereas the latter (2004/18) contains the general rules for other types of procurements. These Directives were to be implemented before January 2006, although transposition efforts continued for some time afterwards in most Member States.

The Directives contain certain rules that aim ensure public awareness of procurements, as a way of encouraging participation and thus competition, ultimately improving the value for money obtained in procurements. The primary instruments of openness in the legal framework include:

- The basic principle that procurement opportunities must be advertised through contract notices. The content and format of these notices has been standardized at the European level⁴. Their usage is mandatory above EU thresholds, as will be discussed below.
- The publication of contract award notices after a procurement has been concluded. As with contract notices, their content and format has been standardized.
- The semantic standardization work that has been done to ensure that procurements can be published and understood homogeneously across the EU, through the Common Procurement Vocabulary.⁵
- The development of a technical infrastructure that allows contracting authorities to send standardized XML messages in order for their procurements to be published

⁴ See http://simap.europa.eu/buyer/forms-standard/index_en.htm

⁵ See http://simap.europa.eu/codes-and-nomenclatures/codes-cpv/codes-cpv_en.htm

at the European level⁶.

- The development of TED (Tenders Electronic Daily), a European portal site on which all of the contract notices are published⁷.

Thus, a significant framework already exists that imposes the use of standardized forms, along with a set of procedures to publish procurements and their outcomes at the European level. This has been accompanied by technical standardization and implementation efforts, ensuring that they can be used in practice.

These efforts have not been without effect: as indicated in the 2010 Evaluation study, in 2009 90,2% of notices sent to the OJEU were sent in a structured electronic format, with several countries nearing 100%.

Thus, certain basic procurement information is made widely and freely available, both through national portals and through TED. However, this does not imply that the data can be downloaded and re-used without restrictions. Licensing restrictions can apply, typically involving a cost to the re-users, as will be further discussed below.

3.2 Limitations of EU public procurement legislation

Even disregarding possible licensing costs, there are several other limitations to the availability and possibility to re-use procurement data.

The first relevant restriction is that EU public procurement legislation does not apply to all procurements. Applicability of European norms is determined by financial thresholds, which are defined at the European level through frequently issued regulations. The thresholds are defined in terms of contract value: only if the threshold is exceeded by the procurement (i.e. if the value of the contract matches or exceeds the EU threshold) will European rules apply.

The height of the threshold varies depending on the public authority, i.e. whether it is a central or sub-central authority, as well as on the type/nature of the procurement (e.g. works have a significantly higher threshold than services contracts). Presently⁸, the transpositions of the EU Directives will never apply to contracts with a value of less than 130.000 EUR. Thus, a significant portion of lower value contracts are not governed by the

⁶ See http://simap.europa.eu/ojs_esenders/sending_xml_notices/index_en.htm

⁷ See <http://www.ted.europa.eu/TED/main/HomePage.do>

⁸ See http://ec.europa.eu/internal_market/publicprocurement/rules/current/index_en.htm

European rules. They may still be subject to national publication requirements, but they will at any rate not be required to follow European laws and processes. In practical terms, TED was said to have published notices in 2009 representing approximately 3,60% of EU27 GDP⁹; given that procurements as a whole account for approximately 17,30% of the GDP, this implies that about 85% of public procurement budgets is not announced via TED.

A second restriction is that the publication requirements have their own thresholds, which are set in the Directives and can vary quite broadly. The use of prior information notices – a form of abbreviated contract notices used in advance of the actual procurement announcement – is only required when the value of the contract exceeds 750.000 EUR, depending again on the nature of the procurement. Exceptions are defined as well, leading to a rather unpredictable publications regime.

Finally, the Directives do not address the issue of re-use themselves. While the requirement to publish notices implies at a minimum the right for the public to access notices in order to obtain information on procurements and possibly to participate in them, it is unclear what restrictions could be imposed on further re-use. In the sections below, we will take a look at whether/how the PSI Directive solves this issue.

3.3 Re-use and the PSI Directive as solution

The PSI Directive does not refer to procurement data or to procurement in general, other than by noting that the definitions of a ‘public sector body’ and ‘body governed by public law’ are taken from prior public procurement directives¹⁰. As such, the PSI Directive will apply to procurement data whenever its general requirements are met, without further specific rules or exceptions.

As a result, most of the types of procurement data mentioned above will fall within the scope of the PSI Directive. Offers submitted by tenderers however would likely be excluded from the scope of the PSI Directive, as they will typically contain commercially confidential information, which is excluded from the scope of the PSI Directive under Article 2 (c). Access to information on losing bids will thus often be unavailable, and information on the winning bid will be available only to the extent permitted through public procurement regulations (such as e.g. the publication of contract award notices as referenced above).

Similarly, access to procurement data can also be limited or redacted on the basis of data protection concerns under Article 1.5 of the PSI Directive. While most procurements will

⁹

See

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

¹⁰ Specifically Directives 92/50/EEC, 93/36/EEC and 93/37/EEC, and 98/4/EC; as indicated in recital 10 of the PSI Directive.

intrinsically contain only a limited amount of personal data (since procurements are normally issued in a A2B-B2A relationship), it is none the less possible to obtain some potentially sensitive information on individual citizens through procurement data. An interesting example of such a use case will be discussed in section 5.2 below on civic re-use of procurement data. In these cases, data protection laws will have to be respected both by the public sector bodies providing the procurement data and by the aspiring re-users.

Generally, the re-use of procurement data could be expected to be facilitated due to the public availability of key information under the procurement rules as described above. Essentially, public procurement is a sphere of public policy where there is a legal obligation as well as a policy incentive to make information widely available: openness facilitates competition and therefore increases value for money and democratic legitimacy.

None the less, this does not imply that all potentially interesting procurement data is available for free re-use. The publication of key procurement data through procurement websites (at the regional, national or European level) operated by public sector bodies as such is not a form of re-use, as it is a part of the initial purpose within the public task for which the documents were originally produced. However, any further use of such procurement data (or other procurement data that would not be routinely published on such websites) is a form of re-use, which can be made subject to prior licensing under the conditions of the PSI Directive, as we shall see below. Thus, restrictions can still apply.

4 State of Play: sources of procurement data

Given the legal requirement to advertise certain procurement data, at least above a certain threshold, it is not surprising that there is a relatively large number of sources of procurement data. Obviously, it would not be feasible to provide an exhaustive overview in this section, as conceptually any public sector body launching a procurement will be a holder of procurement data. However, there are certain portal sites and entities whose core business is the dissemination of procurement data. This section will provide a summary and non-exhaustive overview of these sources. As noted above, these sources should not be considered re-users, since the making available of procurement data is a part of the initial purpose within the public task for which the documents were originally produced.

4.1 At the European level

The primary source of procurement data at the European level is the European portal site Tenders Electronic Daily (TED)¹¹. TED publishes the notices that public sector bodies are legally required to submit to the European Union under the requirements of the public

¹¹ See <http://www.ted.europa.eu/>

procurement directives. It acts as a specialized electronic supplement to the Official Journal of the European Union, in which procurement opportunities are advertised.

The TED site is freely searchable without prior registration, but registered users have the opportunity to create custom profiles, receive periodic notices of procurements which are of interest to them, and search the TED archives for past procurements. TED is purely an information dissemination and aggregation site, and does not directly allow participation in procurements.

The procurement data made available through TED is relatively extensive: on the day of writing (31 July 2012), TED saw the publication of 710 active contract notices, 5 design contests, 79 prior information notices, and 3 qualification systems with call for competition. However, as noted above this data cannot be freely reused in its entirety. While free searches are allowed via TED, the European Union claims copyright protection on its website, and notes that “[a]nyone wishing to download the contents of TED, e.g. to offer their own alert service, is invited instead to sign a licence agreement, giving access to an FTP server or delivery by e-mail. [...] The delivery of files (in XML format via, for example, an FTP server), namely for exploitation and commercial re-dissemination, is subject to the conclusion of a priced license agreement issued on behalf of the European Union by the Publications Office, to which applications must be addressed in writing [...]”¹².

Thus, re-use of procurement data on a significant scale will imply a cost. While obtaining data through site scraping may be technically possible¹³, the legality of this is unpredictable given the copyright claims made by the European Union. However, alternative routes exist: the recent LOTED proofs-of-concept (Linked-Open Tenders Electronic Daily)¹⁴ relies on TED’s RSS feeds to collect procurement data from recent advertisements, which is thereafter linked with data from other sources to enrich the original information. LOTED will be further discussed in section 5.3 below.

Several European projects have been working on public procurement interoperability, and specifically on eProcurement. The most well known of these is arguably the large scale pilot project PEPPOL¹⁵. However, PEPPOL (like other procurement projects) does not act as a source of procurement data in its own right; rather, it aims to develop tools and components that can be re-used by existing or new procurement systems at the national, regional or local level, in order to facilitate cross border participation in public procurements and to facilitate the cross border exchange of procurement data, e.g. through common implementations of standards and interfaces¹⁶.

¹² See <http://www.ted.europa.eu/TED/misc/legalNotice.do>

¹³ See "Downloading the EU", Michael Friis, <http://friism.com/downloading-the-eu>

¹⁴ See <http://loted.eu> or a full discussion paper at <http://ceur-ws.org/Vol-631/paper6.pdf>

¹⁵ See <http://peppol.eu/>

¹⁶ See http://www.peppol.eu/about_peppol/peppol-standards for a list of relevant supported

4.2 EU Member States / regional / local

The public procurement directives require advertisements for procurement opportunities to be published, which is not only done at the European level, but also at the national, regional or local level. This is often more intuitive for local tenderers than European level publications, as they'll be confronted with an infrastructure that is likely more familiar to them, and which may in addition allow them to immediately participate in procurements (i.e. by submitting specific bids via a portal, rather than just obtaining relevant information).

The precise number of eProcurement portals making procurement data available in the EU is hard to quantify precisely, since any administration is free to publish its procurement notices via any means that it considers to be appropriate (notwithstanding its legal obligation to use at a minimum the European mechanisms if the EU thresholds are met). A commune launching a large procurement might e.g. publish the specifications on its own website, a regional procurement site, a national site, and TED. Overlaps are thus certainly possible.

None the less, certain estimates on the number of national, regional and local sources are available. Based on a study presented to the European Parliament in June 2012¹⁷, 329 major e-procurement portals were identified across the EU, with at least one portal existing in each Member State. All of these at least provided notices (i.e. access to procurement data in the form of notices of opportunities), and 150 offered e-submission functionalities (i.e. allowed tenderers to submit offers). The less recent 2010 Impact Assessment identified and listed 129 procurement portals¹⁸, covering all Member States with the exception of Greece.

However, the impact of this significant amount of potentially available procurement data should not be overestimated. Like the TED portal, the national, regional and local portals also tend to make their information available for case-by-case searching, but not for free and large scale information retrieval, making the websites themselves less useful as tools for re-use. Furthermore, it should be noted that information notices to be submitted to the European level have been harmonized, but that this is not the case for information published on national, regional or local websites. This implies that data scraping from public

standards.

¹⁷ Study on e-Procurement Measurement and Benchmarking, The e-Procurement Landscape in the EU, Dr. Gabriella Cattaneo, IDC European Government Consulting; given at the European Conference on e-Procurement, Brussels, June 26 2012; See <http://www.europarl.europa.eu/document/activities/cont/201207/20120710ATT48610/20120710ATT48610EN.pdf>

¹⁸ See http://ec.europa.eu/internal_market/consultations/docs/2010/e-procurement/siemens_annexes_en.pdf, Annex B (page 6 and following) for a full listing.

website is not only legally ambiguous, but also time consuming and complicated, as there is no certainty that information is provided in a homogeneous way. This remains one of the main challenges for the re-use of procurement data.

5 Re-Users of Public Procurement Data

Given the wide availability of procurement data sources, one might reasonably expect that various successful re-use projects would be available. In this section, we will examine some of the most notable cases of commercial, civic and academic re-use of procurement data.

5.1 Commercial Re-use

One of the main distinguishing features of procurement data is that it has an immediate commercial value. For companies that wish to participate in public procurements, keeping track of all potential procurement opportunities themselves is time consuming and expensive, and typically not a part of their core business. While portals such as TED can alleviate this problem to some extent, it was already highlighted above that TED does not include all possible relevant procurements, as it is only exhaustive for procurements exceeding the EU thresholds, which only accounts for about 15% of all public procurement budgets. To get a reasonably complete overview, one would have to monitor all of the hundreds of procurement data sources highlighted above. Thus, there is a commercial opportunity for companies to sell access to more comprehensive database of tendering opportunities.

There are several undertakings who offer access to procurement opportunities as a commercial service. Euroalert¹⁹ is one of the more prominent examples, as a portal operated by the Spanish Gateway SCS3, selling various reporting and alert services²⁰ in relation to European procurements at the EU and national/local/regional level. Information is obtained from some of the primary procurement data sources mentioned above, and repackaged into formats that are arguably more accessible and user friendly to tenderers, including particularly SMEs. Information services are offered as analysis reports, e-mail alerts and newsletters, and more recently mobile apps.

Interestingly, Euroalert has made the conscious strategic choice not to use site scraping techniques, and only aggregates data from sources that provide machine readable datasets. Frequently, this requires the conclusion of licensing agreements, which Euroalert noted is “not always possible [...] 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

¹⁹ See <http://euroalert.net/en/>

²⁰ For an overview, see <http://euroalert.net/en/buy.aspx>

identified as the most important barrier to re-use public sector information, once the technical issues have been solved.”²¹ Despite this difficulty, Euroalert profiles itself as a “thriving commercial service fuelled by the re-use of public sector information (PSI)”.

It is worth noting that Euroalert also actively supports Open Data, by making available part of the data sets that it uses via its own website²², under a Creative Commons Attribution-ShareAlike 3.0 Unported license²³, in an OpenDocument format. These data sets include lists of contracting authorities from Germany, Belgium, Spain, France, Ireland and the UK, and have been created by Euroalert on the basis of TED announcements.

Other commercial services are of course also available in various European countries, although Euroalert arguably takes the Open Data philosophy the furthest. Several examples of other commercial services can be found in the 2010 assessment study, which noted that “in a number of countries with strong public/private sector collaboration, fee based websites were also available. This was noted to be the case in all Scandinavian Member States (Denmark, Finland, Sweden), the Baltic countries (Estonia, Latvia, Lithuania), Germany, Poland, Portugal and the UK, as well as those with more decentralised procurement markets, such as Austria and France. In all likelihood, the number of countries with commercial procurement aggregators is bigger in reality, but the collected country reports only refer to commercial platforms if these are considered a key part of national public procurement strategies.”²⁴ Thus, commercial aggregation is a relatively common service, with examples including the various MerCell platforms²⁵, the French E-Marchéspublics.com²⁶, or the Romanian eLicitatie²⁷. Clearly, procurement data is a form of PSI with evident commercial value.

²¹ “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”; see <http://share-psi.eu/papers/Euroalert.pdf>, p.3-4

²² See http://euroalert.net/en/help_tenders_opendata_contributions_authorities.aspx

²³ See <http://creativecommons.org/licenses/by-sa/3.0/> for the full license text.

²⁴ See http://ec.europa.eu/internal_market/consultations/docs/2010/e-procurement/siemens-study_en.pdf, p. 163

²⁵ See www.mercell.dk, www.mercell.ee, www.mercell.lv, www.mercell.lt, www.mercell.no and www.mercell.de

²⁶ See <http://www.e-marchespublics.com/>

²⁷ See <http://www.e-licitatie.ro>

5.2 Civic Hackers and Transparency Activists

It was already noted in the introduction above that openness of procurement data not only serves a clear economic purpose – namely optimizing competition in public procurements, thus ensuring optimal value for tax money – but it also caters to a more fundamental societal need, namely ensuring transparency of public expenditures. If improperly managed, public procurements can lead to corruption, with contracts being granted to suboptimal bidders. This harms public budgets and hurts public faith in their governments.

Openness of procurement data can support accountability in public spending, by ensuring that crucial information on procurements is accessible to the general public. In this way, the people can assess whether their tax money has been spent responsibly, and can search for any anomalies in public spending. To some extent, this need is already met by contract award notices, as published at the European level. However, this approach is somewhat flawed: contract award notices are only required for a relatively small segment of public procurements, and will not provide information on smaller contracts. Furthermore, the notices will not necessarily contain all relevant information: a single person operating in multiple companies could win a large amount of contracts spread across his/her various companies without being detected. Aggregate data is thus largely hidden.

An interesting civic use case of procurement data re-use that addresses this issue can be found in the Slovakian ZNasichDani.sk service. As has already been previously discussed on the EPSIPlatform website²⁸, Znasichdani.sk (which translates roughly to “From our Taxes”) was created in March 2011 by the Fair-Play Alliance, a Slovakian NGO, as a tool to detect potential corruption in public procurement contracts. It combines data from two Slovakian databases that are publicly available: a database of public procurement contracts (the Bulletin of Public Tenders), which indicates which entities have won specific bids, and a company register (the Business Register of the Slovak Republic), indicating which individuals have controlling roles in specific entities.

The service thus allows users to enter the name of specific individuals, and obtain information on any procurements that have been won by companies in which that person holds some form of official mandate. In this manner, ZNasichDani.sk allows users to find indications of possible anomalies in public purchasing behaviour, irrespective of the legal entity that was used to participate in a procurement.

The application was critically acclaimed, receiving an award in the Open Data Challenge from EU Commissioner Neelie Kroes, and a Trailblazer award from the EPSIPlatform. However, it also received significant pushback. At the request of one of the people that could be targeted by Znasichdani.sk, the court of Bratislava ordered the removal of

²⁸ See notably <http://epsiplatform.eu/content/slovak-privacy-commissioner-not-impressed-open-data> and <http://epsiplatform.eu/content/znasichdanisk-making-procurement-data-matter-more>

information about her and her company from the Znasichdani.sk site in June 2011.

While it is just a single case, the example of Znasichdani.sk shows how the re-use of procurement data can offer societal value as well as commercial value. At the same time however, it also indicates how the combination of open data to obtain new knowledge can create sensitive results. From a privacy protection perspective, it is clear that the re-use of procurement data can raise significant new challenges.

5.3 Academic Re-use

Finally, re-use of procurement data could also be interesting for academic purposes, specifically to measure the implementation of policy choices (i.e. where/how money is being spent), how this is effected in practice (i.e. where money ultimately ends up), and how new applications can be built on existing data (i.e. what knowledge can be extracted from existing data and what can be done with this?). An interesting example is the aforementioned LOTED (Linked-Open Tenders Electronic Daily)²⁹ proof of concept.

LOTEd aimed to explore ways of dynamically linking TED procurement data, obtained via TED's RSS-feed, to other data sets in order to extract new knowledge from existing information and to build new applications. LOTED has developed its own ontology³⁰, allowing it to reformat data extracted from TED into a harmonized format which is suitable for linking with other datasets. For the purposes of this proof of concept, tender information was automatically linked to two other datasets: geonames³¹ (indicating where the contracting authority is geographically located on the basis of the city name and country code indicated in the notices) and DBPedia³² (containing relevant information on that city as extracted from Wikipedia).

Thus, the proof of concept mainly links tenders to additional geographic information on the location of the public sector body, and allows it to be presented in a clear and intuitive format (including e.g. maps of the region). This could potentially be useful to quickly assess the feasibility and appeal of a procurement to a specific tenderer. E.g. a Polish vendor might be interested in delivering products to a public sector body in the northeast of Germany, but less so in the southwest of Germany, due to the simple logistics and costs of delivery.

The authors of LOTED argue that the value of the proof of concept is not in its current functionality, but rather in its ability to demonstrate that TED data can be reformatted to a

²⁹ See <http://loted.eu> or a full discussion paper at <http://ceur-ws.org/Vol-631/paper6.pdf>

³⁰ See <http://loted.eu/ontology>

³¹ See <http://www.geonames.org/>

³² See <http://dbpedia.org>

more useful container and thereafter linked to relevant datasets. The proof of concept provides several highly interesting examples of how new knowledge can be obtained after this exercise is completed; e.g. through LOTED, users can identify specifically which procurements are offered in a given region, including which types of procurements and budgets have been allocated over a given period of time. As with the example of Znasichdani.sk, another use case is identifying potential indications of corruption, or at least of irrational political imbalances: based on DBPedia data, LOTED can detect which political party is governing cities or regions in which larger or smaller than average budgets are allocated, thus identifying potentially imbalanced spending. Such profiling can also be simply interesting to analyse policy biases, by indicating in which sectors (e.g. education, road maintenance, environment, ...) specific political parties actually spend their procurement budgets, irrespective of what their official party agenda says on the topic. None of these questions could be answered through TED, geonames or DBPedia on its own.

Thus, the proof of concept shows the ability of open data sets to be interlinked to gain new relevant insights, and even to improve public policy making and democratic legitimacy. The LOTED team also identified specific challenges, notably the incompleteness of data within specific data sets, and the complexity of establishing links between data sets due to the lack of a pre-existing framework for data discovery, data cleaning etc. Having to start largely from scratch, the process proved to be resource intensive. Further work would be needed to develop such a framework and to more easily conduct analysis on the links established between datasets.

6 Conclusions

The report has shown through various examples that a fairly large amount of procurement data is currently available through sources at the European, national, regional and local level, and that interesting re-use cases already exist. Commercial re-usage has already existed for some time, and shows the possibility of extracting economic value from open procurement data. Arguably however, the civic use cases are even more innovative and have a potentially greater value to society: they show how open data and links between data sets can be used to evaluate the effectiveness and fairness of public spending. They allow citizens to determine whether their taxes are spent wisely and efficiently, and where anomalies lie. A greater availability and maturity of such applications would certainly increase the transparency and democratic legitimacy of public spending.

Some clear challenges and gaps have been identified as well, however. Firstly, data is not universally available. Procurement data is subject to various publicity requirements, and certainly at the EU level, there is a very substantial dark number of procurements which is never published through TED. This means that re-users need to trawl a large number of additional sources, and are even then unlikely to get a comprehensive picture of procurement policies. Data is in other words still very fragmented.

Aside from availability, a second issue is the usability of data. Several of the use cases above noted that the available data was only semi-structured, and that formats differed from source to source. This makes it complex and resource intensive to identify, extract, reformat and link data sets. Usability is not only hampered by this technical/semantic challenge, but also by legal barriers: while site scraping is usually technically feasible, its legality is not always clear. Larger scale information extraction and re-use on the other hand is often made subject to licensing and payment restrictions. To the extent that such processes are strictly necessary, they should at least ensure that data is available as easily and cheaply as possible, in accordance with the philosophy and principles of the PSI Directive.

It is clear that procurement data can have immense value. It is therefore imperative that the holders of procurement data are aware of the potential benefits that can be realized, and adopt policies that allow re-users to overcome these barriers. Only by doing so can the full potential of procurement data be unlocked.

About the Authors

Hans Graux is a bar lawyer and founding partner at the Brussels based law firm time.lex (www.timelex.eu), which specializes in ICT law and ICT policy challenges. In addition, he is an affiliated researcher at the Interdisciplinary Centre for Law and ICT (www.icri.be) at the K.U.Leuven. He also acts as the independent legal advisor to the Vlaamse Toezichtcommissie (Flemish Supervisory Committee - <http://www.vlaamsetoezichtscommissie.be/>), which supervises personal data exchanges within Flemish public sector bodies.

Tom Kronenburg is a consultant with Zenc B.V. based in the Netherlands. He specialises in information as a solution to societal problems. Tom is one of the curators of the EPSI Platform website and travels throughout the European Union to connect PSI holders and re-users, citizens and governments.

Copyright information

© 2012 European PSI Platform. This document and all material therein have been compiled with great care. However, the author, editor and/or publisher and/or any party within the European PSI Platform or its predecessor projects the ePSIplus Network project or ePSINet consortium cannot be held liable in any way for the consequences of using the content of this document and/or any material referenced therein. This report has been published under the auspices of the European Public Sector Information Platform.



The report may be reproduced providing acknowledgement is made to the European Public Sector Information (PSI) Platform. The European Public Sector Information (PSI) Platform is funded under the European Commission eContentplus programme.