

# NextProcurement: Challenges in Public Procurement in Spain

María Navas-Loro<sup>1</sup>

<sup>1</sup>*Ontology Engineering Group, Universidad Politécnica de Madrid, Spain*

## Abstract

Public procurement accounts for 14% of the annual budget of the various governments of the European Union. Despite its importance, there are currently several challenges in data processing in this domain. This paper identifies the main challenges detected in the first phase of the European NextProcurement project, which aims to create a platform to help harmonize and enrich public procurement data in the European Union, focusing on the specific case of Spain.

## Keywords

Public Procurement, NextProcurement, Challenges, Textual documents, Natural Language Processing, Spanish Public Procurement

## 1. Introduction

Public authorities in the European Union spend around 14% of the annual Gross Domestic Product (about 2 trillion euros) on the purchase of services, utilities and supplies.<sup>1</sup> Free access to this data allows facilitating accountability and transparency in Europe. Many governments, therefore, provide this data on their own national open data portals (this is for instance the case of the Spanish portal PLACE/PLASCP<sup>2</sup>), and different platforms have been developed to improve both efficiency and transparency in European public procurement by exploiting this information<sup>3</sup> [1].

There are several steps in the process from the publication of a tender to its execution and payment. These steps, in turn, involve various administrative documents, such as the bid itself and the technical criteria to be met, as well as the evaluation criteria for the different proposals. The automatic processing of this information would greatly simplify and improve the transparency of the public administration.

However, the processing of these documents entails certain challenges for today's data processing and natural language processing, as it requires some domain knowledge. This paper breaks down some of the main challenges, extracted from document analysis and from domain

---

*Joint Proceedings of ISWC2022 Workshops: the International Workshop on Artificial Intelligence Technologies for Legal Documents (AI4LEGAL) and the International Workshop on Knowledge Graph Summarization (KGSum), October, 2022*


✉ [mnavas@fi.upm.es](mailto:mnavas@fi.upm.es) (M. Navas-Loro)

🌐 <https://marianavas.linkeddata.es> (M. Navas-Loro)

🆔 0000-0003-1011-5023 (M. Navas-Loro)



© 2022 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

 CEUR Workshop Proceedings (CEUR-WS.org)

<sup>1</sup>[https://ec.europa.eu/growth/single-market/public-procurement\\_en](https://ec.europa.eu/growth/single-market/public-procurement_en)

<sup>2</sup><https://contrataciondelestado.es>

<sup>3</sup><https://opentender.eu/es/about/about-opentender>

experts' experience in Spanish public procurement in the context of the NextProcurement project<sup>4</sup>, aimed to develop an open harmonized and enriched public procurement data platform. We will consider different factors for each of these challenges, such as the need for experts and/or annotators to undertake them, related work in the public administration domain if available, or the feasibility of solving the challenges, either in terms of the resources required or the cost required versus the benefits of undertaking the tasks to address the challenges.

The rest of the paper is organized as follows. Section 2 introduces the public procurement tendering process in Spain, explaining the different stages in it. Section 3 explains the documents involved in the tendering process. Section 4 and 5 present the main challenges identified, organizing them according to the stage of the process in which they appear from among those explained in the previous section. Finally, Section 6 concludes the most interesting tasks to tackle from the analysis performed.

## 2. Public Procurement Process in Spain

Figure 1 shows the different stages through which a public tender passes in Spain. Each of these phases is briefly explained below:

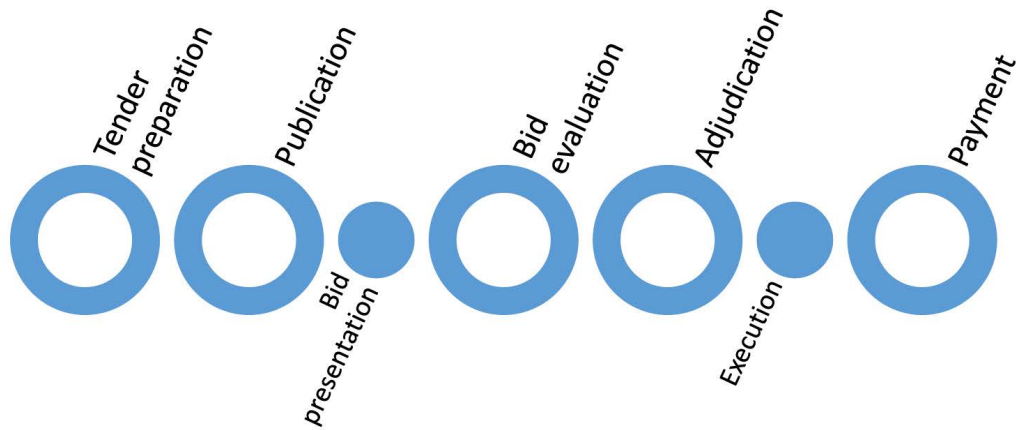
1. Tender preparation: The public administration detects a necessity (whether for services or material resources) and prepares a tender, which involves the drafting of documentation. This documentation includes a series of criteria that the bids must meet to be selected, both from a technical and an economic point of view. This documentation must as well, as far as possible, detail how the evaluation of the bid will be carried out. Possible improvements at this stage include suggesting wording to the public employee based on similar existing tenders, and also recommendations of keywords and classification within a taxonomy.
2. Tender publication: once the tender is ready, it is published on the various public platforms available. To facilitate access to potential bidders, it is important to include metadata correctly and, as much as possible, to publish it as semantic information.
3. Bid presentation: During a limited period of time, potential suppliers can send their offers to the public administration. The number of potential suppliers of different sizes, as well as their volume, will depend, among other factors, on the classification and accessibility of the tender in the previous stages. An accurately classified tender will reach target bidders, while a wrongly classified one will be difficult to find.
4. Bid evaluation: Once the deadline for submission of bids has expired, the public administration in charge evaluates the different bids based on the previously specified criteria. The ability to detect patterns of behaviour or even illegalities such as *collusion*<sup>5</sup> among bidders would facilitate this process.
5. Adjudication: The previous stage will lead to the award to the bidder with the best evaluation, who will have to adhere to the specified conditions.

---

<sup>4</sup><http://nextprocurement-project.com/>

<sup>5</sup>According to Cambridge Dictionary, "agreement between people to act together secretly or illegally in order to deceive or cheat someone".

6. Execution: The awarded bidder will supply the materials or services within the stipulated execution period.
7. Payment: Following completion of the service, the public administration proceeds to pay for the service.



**Figure 1:** Timeline of a public tender in Spain. Big circles represent steps performed by the public administration, while the small ones are done by potential/awarded suppliers.

We could additionally consider a latter phase, consisting of the consultation of data from different past tenders by the administrations for internal use.

### 3. Documentation

Documentation of a tender usually contains the main document (that can be in the form of a document *per se* or in the form of metadata included in a platform) that details the service to be provided, usually including a title, a description and information such as its CPV classification. CPV codes (Common Procurement Vocabulary codes)<sup>6</sup> help classify public procurement processes in the European Union across different languages. Each European public procurement process must be classified with at least one CPV, among the thousands of possible codes (more than 9000) in the CPV taxonomy.

There can be additional documents attached to the tender. They are not compulsory, and if present they are usually referred to with urls that go to scanned documents that must be read using OCR techniques. The most important ones are named “Pliegos” in Spanish, and are detailed below.

<sup>6</sup><https://simap.ted.europa.eu/web/simap/cpv>

### **3.1. Pliegos**

#### **3.1.1. General Administrative Clause Specifications (Pliego Cláusulas Administrativas Generales, PCAG)**

PCAG includes those rules and conditions that apply to all the contracts from a certain Public Administration, in a general manner. Regarding how the clauses will be later evaluated, we can distinguish two different types:

- Non-automatically evaluable criteria: Quantification depends on a value judgment. Mathematical formulas are not applicable to them and they are therefore considered subjective criteria, although the public administration must make an effort to specify the aspects to be valued in each one of them.
- Automatically evaluable criteria: Mathematical formulas are applied, therefore being objective. It includes the economic offer (a formula is always applied), and those technical criteria will also be evaluated through the use of formulas (e.g., the memory of a computer).

#### **3.1.2. Particular Administrative Clause Specifications (Pliego Cláusulas Administrativas Particulares, PCAP)**

Rules specific to the tender in hand, award criteria and the legal conditions of the contract. The elements of the contract include:

- The object, budget and cost, execution time, capacity and solvency required to bid, etc.
- The awarding of the contract. The procedure, how and where the bids must be submitted, how the contracting company will be selected and how the contract will be formalized.
- The execution of the contract, including labour, social and economic obligations the contractor company has to comply with.
- The prerogatives of the Administration, Jurisdiction and Remedies. Here are defined the privileges and jurisdiction that will be applied in the event of disagreement between the parties and how the appeals should be handled.
- The criteria of valuation of the offers, among other annexes.

#### **3.1.3. Technical Prescription Clause Specifications (Pliego Cláusulas Prescripciones Técnicas, PPT)**

This *pliego* includes the minimum technical conditions required in the project, supply, service, work, etc. These specifications will detail the characteristics of the supply, work or service required by the Administration, as well as those aspects of improvement that will be evaluated in the bids. Anything that is not specified in the specifications can be evaluated, except in the case that it appears in the criteria 'other improvements', where those things that have not been specified in each section can be evaluated.

## 4. Challenges in Tender Preparation

### 4.1. Tender Drafting

The wording of the tenders may present inaccuracies that are difficult to manage. Many imprecise words, such as "solvency", are frequently used without being clear about what they imply. Additionally, tenders are expressed in many different ways depending on the writer, even if they ask for the same products or services, and the same work of drafting is done multiple times, incurring great inefficiency.

Thus, it would be desirable to be able to detect the worst written parts of existing tenders and create a taxonomy of the worst errors. For this purpose, words that tend to be imprecise, such as "solvency", could be detected. Similarly, a style recommender could also be generated on the basis of correct tenders. If, for example, a public administration wants to publish a tender looking for drugs of a certain type, it would be desirable to locate similar texts in order to use them as a reference, based on text similarity but also other similar aspects (such as budget). Finally, also a metric could be designed to measure the clarity and readability.

The main problem in dealing with these issues is the need for annotations and expert knowledge to detect and classify main errors. Natural Language Processing could help in these tasks but would require annotated corpora to correctly identify domain particularities of the domain.

### 4.2. CPV Code Assignment

As previously mentioned, Common Procurement Vocabulary codes (CPVs) help classify public procurement processes in the European Union. Thanks to CPVs, decision-makers can easily explore contracting processes across Europe, and potential suppliers from different countries may use them to detect procurement processes of interest, independently of their size or country of origin.

Each public procurement process must be classified with at least one CPV. However, manual CPV classification presents three main challenges [2]. First, there are thousands of possible codes (more than 9000), some of them with similar purposes, making it difficult for those assigning or curating them to decide which codes better suit a specific process. This problem is worsened with short descriptions, which are abundant among tenders. Second, different administrations follow different classification guidelines. Third, since CPVs are organized in a hierarchy, and thus annotated at different levels of granularity according to the annotator's or department's criteria, some codes are often overgeneralized. More information about these problems and proposed solutions can be found in literature [2], since it has been previously targeted both in academia [3, 4, 2, 5] and industry [6].

The main problems reported regarding CPV classification are human error and less specific CPV being commonly used. A system that receives the description of a tender and recommends the best fitting CPV codes could help with these issues. For this kind of classification task there is no need for human annotation, but just classified tenders that can be found in different public repositories (in the case of Spain, Hacienda<sup>7</sup>, PLACE, and also the European TED<sup>8</sup>), it is an

---

<sup>7</sup><https://www.hacienda.gob.es/es-ES/GobiernoAbierto/Datos%20Abiertos/Paginas/LicitacionesContratante.aspx>

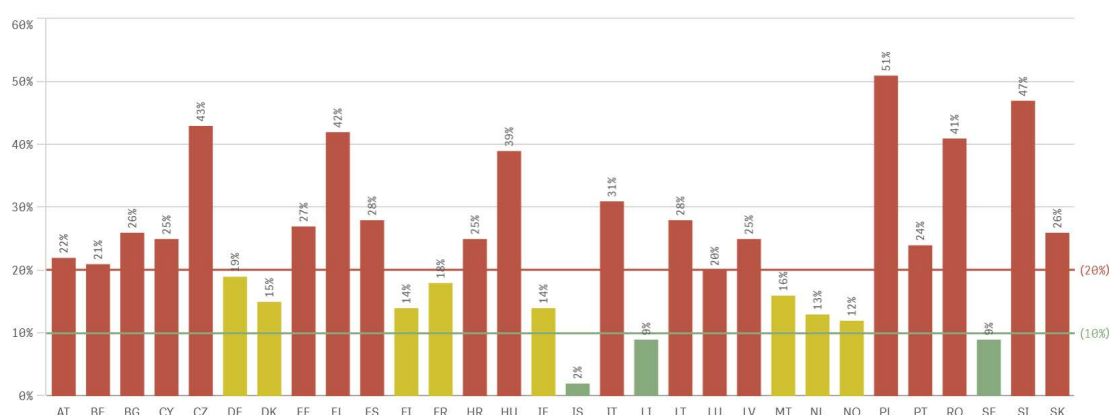
<sup>8</sup><https://simap.ted.europa.eu/>

affordable task from the point of view of the required specialized manpower, since no domain annotators are needed. One of the problems we may encounter is the under-representation of some of the CPV codes, but this can be solved by explicitly searching for these codes to arrive at a balanced dataset, or even with data augmentation techniques.

## 5. Challenges in Bid Evaluation

### 5.1. Pattern Detection

Several of the indicators used at the European level to evaluate the performance of the different states in terms of public procurement are related to the number of bidders that apply for a tender. Figures 2 and 3 show different statistics per country that illustrate some of the problems in current European public procurement.

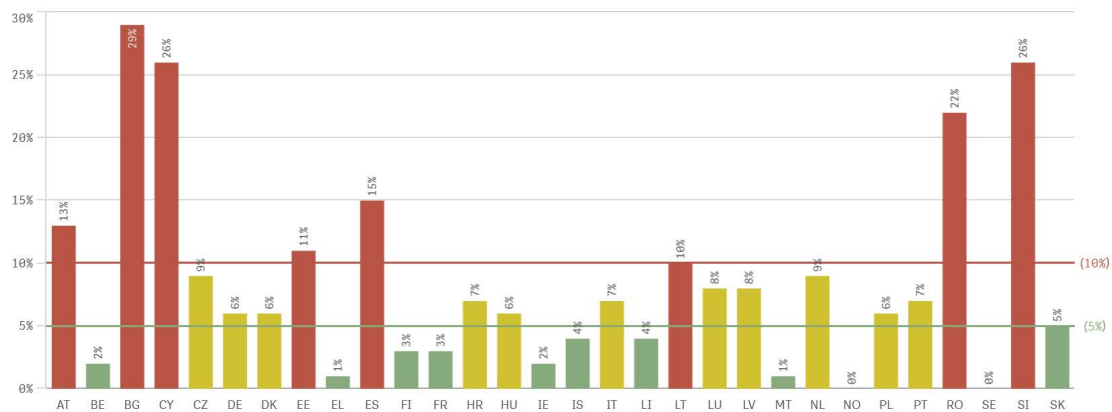


**Figure 2:** Proportion of contracts awarded where there was just a single bidder (excluding framework agreements). Figure extracted from [https://single-market-scoreboard.ec.europa.eu/policy\\_areas/public-procurement\\_en](https://single-market-scoreboard.ec.europa.eu/policy_areas/public-procurement_en)

The more companies that apply for a tender, the higher the quality of the final adjudication. Additionally, with this low participation, it is difficult to avoid problems such as collusion or awarding of contracts to the same companies without justified reasons. The analysis of competition would gradually improve the publication of tenders.

Possible solutions to these issues include building clustering taking into account the amount and type of clauses (e.g. social) to understand low bidding. Additionally, doing some profiling of bidders (e.g. defining typologies among bidders and different levels of participation) would allow an analysis of the relation between the type and amount of bidders and tenders (object, clauses, etc). This could also lead to a network of bidders/tenders that would facilitate collusion detection.

Regarding requirements and feasibility, basic Machine Learning techniques could be tested against available data; no annotation is required, just expert knowledge for the design of the typology.



**Figure 3:** Proportion of procurement procedures that were negotiated with a company without any call for bids. Figure extracted from [https://single-market-scoreboard.ec.europa.eu/policy\\_areas/public-procurement\\_en](https://single-market-scoreboard.ec.europa.eu/policy_areas/public-procurement_en)

## 5.2. Clauses and Evaluation

As previously reported, when drafting a proposal it is common for many of the clauses to repeat or resemble those of previous tenders. This similarity could also be used to automate the evaluation of similar clauses. Developing a clause topology (e.g., social, environmental, or innovation related) that organizes and helps generate a repository of clauses would significantly streamline the entire process of drafting and evaluating tenders, facilitating also the creation of evaluation models.

Additionally, in some cases, we have tenders that could include clauses that were not included. Sometimes, besides these clauses we find “special execution conditions”, but it is not clear how they affect bidding and execution; are there fewer bidders? How are these special conditions later evaluated?

As mentioned, a possible solution would be to identify and generate a repository of clauses related to objects/topics/CPVs... this way they could be easily reused and statistics could be retrieved. The same applies to the validation criteria related to clauses and special conditions. With this information, we may be able to recommend clauses given the object of a new clause because similar tenders included them. This work would lead to the creation of a Knowledge Graph that relate all this information with information extracted for other problems (e.g., bidders).

Nevertheless, although the obvious usefulness of these solutions, the main issue when tackling these tasks is the need for expert knowledge.

## 5.3. Questions over tenders

In order to know the impact of the tenders, to avoid collusion and similar problems, and to improve transparency, public administrations want to be able to make queries to tender platforms, such as “How many pills for X disease were bought last year?”, or “For how many

*tenders was company Y awarded?”.*

Different Natural Language Processing tools could be used in tenders to derive triples, as well as Named Entity Recognition and Disambiguation. Once all the information is stored as triples, a Question Answering system could be built to query the knowledge graph. Additionally, in order to facilitate querying the system, SPARQL queries could be expressed as natural language.

This work would depend on solutions to the problems previously presented, such as processing tenders, clauses, and bidders. Once these are tackled, different Question Answering systems could be tested, such as ElasticSearch<sup>9</sup>.

## 6. Conclusions

The list of challenges presented is not exhaustive, but summarizes the main problems detected in the first phase of the NextProcurement project.

As previously mentioned, not all of them are equally feasible, and some of them require expert knowledge that is difficult to obtain. In addition, annotation tasks and the creation of typologies are often cumbersome and time-consuming for people not familiar with these tasks, which significantly complicates their achievement. Therefore, the first steps to be taken in the context of the NextProcurement project will focus on the CPV code assignment (already started [2]) and the identification of patterns, since the data required for them is already available. In the meantime, work will be done on the possible expert annotation of clauses in tenders in order to try to address the other challenges identified.

## Acknowledgments

This work has been supported by NextProcurement European Action (grant agreement INEA/CE-F/ICT/A2020/2373713-Action 2020-ES-IA-0255).

## References

- [1] A. Soyulu, O. Corcho, B. Elvesæter, C. Badenes-Olmedo, F. Yedro-Martínez, et al., Data quality barriers for transparency in public procurement, *Information* 13 (2022). URL: <https://www.mdpi.com/2078-2489/13/2/99>. doi:10.3390/info13020099.
- [2] M. Navas-Loro, D. Garijo, O. Corcho, Multi-label text classification for public procurement in spanish, *Procesamiento del Lenguaje Natural* 69 (2022) 73–82. URL: <http://journal.sepln.org/sepln/ojs/ojs/index.php/pln/article/view/6429>.
- [3] O. Ahmia, Assisted strategic monitoring on call for tender databases using natural language processing, text mining and deep learning, Ph.D. thesis, Université de Bretagne Sud, 2020.
- [4] S. Kayte, P. Schneider-Kamp, A mixed neural network and support vector machine model for tender creation in the european union ted database, in: *Proceedings of the 11th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management*, INSTICC, SciTePress, 2019, pp. 139–145. doi:10.5220/0008362701390145.

---

<sup>9</sup><http://www.elastic.co/products/elasticsearch>



- [5] A. Suta, Multilabel text classification of public procurements using deep learning intent detection, Master's thesis, KTH, Mathematical Statistics, 2019.
- [6] Deloitte, Study on up-take of emerging technologies in public procurement, Technical Report, Deloitte, 2020.