EU Contract Hub: Towards a More Accessible Public Procurement

Virginia Ramón-Ferrer[1,*,†], Álvaro Fontecha[1,†], Carlos Badenes-Olmedo[1] and Oscar Corcho[1]

[1]Universidad Politécnica de Madrid, Campus de Montegancedo, Boadilla del Monte, Comunidad de Madrid, 28660, Spain

Abstract

Public procurement is one of the pillars of the proper functioning of a country, given the direct impact it has on its economy. The European Union places great importance on the transparency and accessibility of data related to public procurement, giving public access to a significant part of public procurement contracts. The problem is that these data are often difficult for many users to consult and understand, in addition to being time consuming. In this paper, we present *EU Contract Hub*, a platform for efficient exploration and analysis of large amounts of data on public procurement in Europe. The tool incorporates an innovative ingestion pipeline that unifies and enriches the information of more than a million public contracts extracted from various formats into a single common structure. This unification of information enables the efficient analysis of large amounts of public procurement data, especially related to healthcare and its impact during the COVID-19 period, within the framework of the European project '*Procure*: Public procurement assessment in the healthcare sector'. *EU Contract Hub* can be accessed through the platform page at https://procure.linkeddata.es/.

Keywords

Public procurement consultation, Public contract management, Digital statistical tool

1. Introduction

Every day, public institutions from all the European Union member states purchase goods and services of various kinds. This process, known as public procurement, has to be highly rigorous and transparent, given the impact that it has to the economical state of each country and the European Union as a whole. Regarding transparency, the European Union has various initiatives to bring procurement information closer to the public. *'Supplement to the Official Journal'* of the EU and its online version *Tenders Electronic Daily (TED)* [1], for example, is a platform where public procurement contracts with a high economical value across the EU are published to ensure transparency and open competition.

One clear example of the relevance of procurement is the task carried out during the COVID-19 crisis, where the purchase of medical goods and services played a major role in combating the pandemic. *PROCURE* [2] is a EU funded project that aims to assess the impact the COVID-19 pandemic had on health procurement organisations and practices from 13 participating EU countries. In the context of this project, we present *EU Contract Hub*, an open platform that facilitates the exploration and analysis of procurement data, making this information accessible to interested institutions and the general public. Leveraging large amounts of heterogeneous data can be challenging, so we defined an ingestion work that unifies contracts in different formats into a single common structure to facilitate the joint consultation of these documents. With this, we aim to simplify the public procurement consultation process, adapting public consultation data to a more user-friendly representation for non-experts and using large language models to further enrich this information. We enable an easier cross-reference of

EKAW 2024: EKAW 2024 Workshops, Tutorials, Posters and Demos, 24th International Conference on Knowledge Engineering and Knowledge Management (EKAW 2024), November 26–28, 2024, Amsterdam, The Netherlands.

*Corresponding author.

†These authors contributed equally.

virginia.ramon@upm.es (V. Ramón-Ferrer); alvaro.fontecha@upm.es (Á. Fontecha); carlos.badenes@upm.es (C. Badenes-Olmedo); oscar.corcho@upm.es (O. Corcho)

0009-0009-7197-1676 (V. Ramón-Ferrer); 0009-0009-2077-7558 (Á. Fontecha); 0000-0002-2753-9917 (C. Badenes-Olmedo); 0000-0002-9260-0753 (O. Corcho)

information along different data sources by homogenising both the data format and the language. This platform can be openly accessed through its online page at https://procure.linkeddata.es/.

2. System implementation

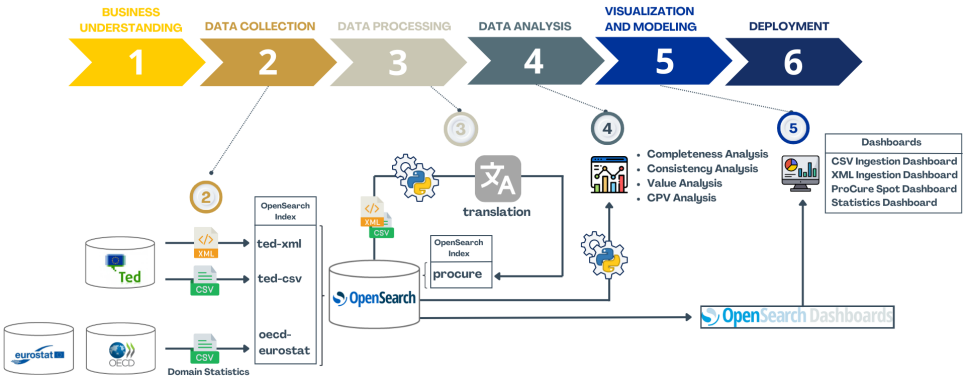


Figure 1: Data ingestion pipeline

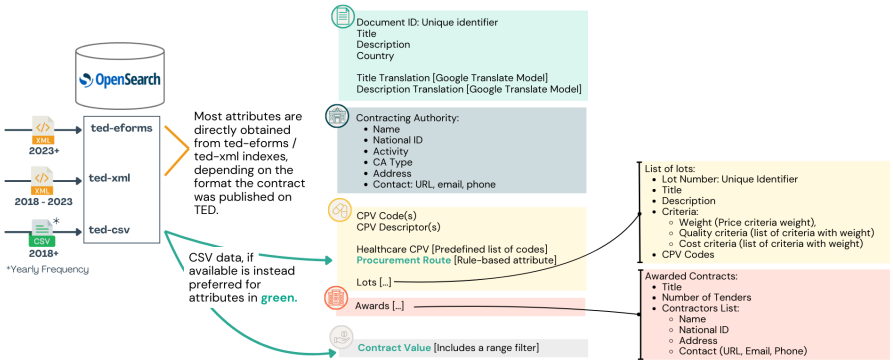


Figure 2: Data unification structure

EU Contract Hub[1] is a document-oriented platform which allows us to create data visualisations and integrate them into interactive dashboards. Currently[2], this hub stores and enables the efficient consultation of over 1.2 million public procurement contracts extracted from *Tenders Electronic Daily (TED)* [1], in addition to socioeconomic data about the consortium countries extracted from OECD [3] and EUROSTAT [4]. These sources were chosen taking into account a pre-defined series of questions regarding the public procurement environment in the healthcare sector and key players, such as the total amount of Healthcare Public contracts in value for each member country, and the organization of the national health system, such as the overall country population, that the *PROCURE* [2] members needed to be able to answer using our tool. As shown in figure 2, data is sourced and unified from various standardised forms completed by contracting authorities, which results in a wide range of quality and completeness levels across documents. Given that the same document can exist in different versions with various formats, specifically eForms, XML and CSV, and each format may contain information related to the contract that isn't present in the others, we implemented an ingestion framework that processes each formatted document and unifies the documents to gather the most

[1]Link to source code: https://github.com/procure-project/EU-Contract-Hub/
[2]This document was drafted on the 17th of September 2024, but the tool is continuously updated with new information and visualisations, so there may be changes in the future.

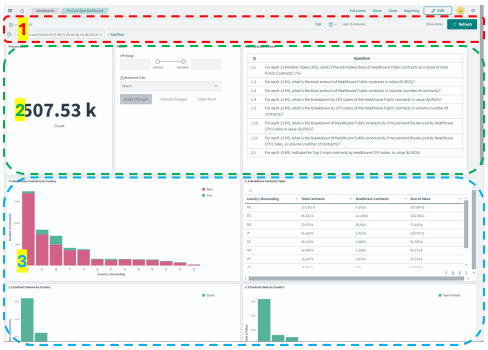
information possible in relation to a specific contract in a single clean, processed structure. When conflicts exist, meaning that the same attribute is available in various formats, we generally prioritise data extracted from eForms and XML, except some specific attributes, such as the contract value, where we prioritise CSV information. We also collaborated with procurement experts to identify information that can be inferred through a set of rules from the contract data but was not explicitly present, such as the legal mechanisms used in each contract. This inference allows us to provide more extensive information of interest to both the consortium of the project and the general public of our tool. Furthermore, though there are some search tools available at *TED* that enable the consultation of the contracts individually, our aim was to also be able to identify overall trends, which the questions defined in the matrix of the project required, and have the ability to create user-friendly visualizations for non-experts. In *EU Contract Hub* we provide said functionalities to ensure this kind of analysis is supported .

The contracts stored have information related to the contract awards and tenders, including the type and purpose of the contract, the value, and the winning bidders. Given the multilingualism present in the EU, most contracts are in the native language of the contracting country. This, as a user who wants to be able to consult contracts from all over the EU, makes it difficult to grasp a large part of the relevant information represented in the contract notices. To tackle this problem, we have machine-translated the non-English contract fields to English in order to widen the accessibility to the information. Given that we could not supervise the translation of over a million contracts and needed a model that could automatically translate at least 15 different languages[3] to English, we decided to use the Deep-Translator library [5] with the Google Translate [6] model, which we previously evaluated over two corpus extracted from *OPUS* [7], the *Opus-100 Corpus* [8] with a sample size of 2000 and the *OPUS Europarl Corpus* [9] with a sample size of 4000. This evaluation presented results of COMET [10] of 0.81 for the *Opus-100 Corpus* sample and 0.86 for the *OPUS Europarl Corpus* sample over the 16 non-English languages used in the 13 member states of the *ProCure* Project. This model was also chosen since it could automatically detect the original language used and translate it to English, saving the need to detect the language beforehand. Currently[2], about 50% of the contracts in the platform have translated attributes available, but the translation of the remaining contracts' attributes is under development, and we anticipate that the remaining translations will be available in the near future.

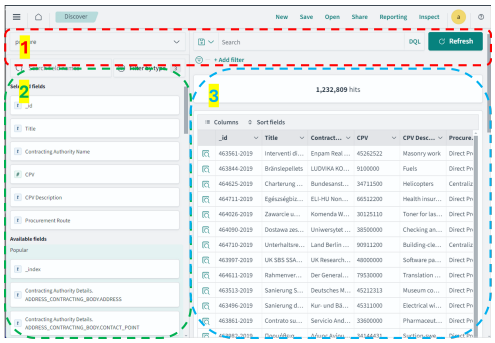
This platform has been developed following the needs of the end users, especially the consortium of the *ProCure* project. In addition, we are also constantly receiving feedback from consortium members and applying the necessary changes to further improve the tool. The two main requirements that we have based the development are that the contracts can be consulted individually, by applying filters based on the existing attributes of each contract, and globally, by creating data visualisations that enables the user to analyse the tendencies present in them. To do this, the platform is divided into two main consulting interfaces, namely *1) Dashboard* and *2) Discover*:

1) Dashboard. This interface enables graphical visualisation of the data to extract statistics and tendencies. As shown in figure 3a, we have created three main areas in the visualisation: The first highlighted area is where the data filters are to be entered if needed. These filters can be entered by hand, using *Documentum Query Language (DQL)*, or using the "Add Filter" option, where a filter dialogue appears to enable the user to define the filter through selection of options. The second area corresponds to the general information display, where the number of contracts taken into account for the visualisations below is displayed together with a small window where we can filter the data directly by *Common Procurement Vocabulary (CPV)* values and the list of questions defined by the consortium. Finally, the third highlighted area displays the different graphics and tables needed to answer the consortium questions according to the filters entered. The visualisations on the dashboard are currently

[3]List of languages: German, Dutch, French, Greek, Hungarian, Italian, Maltese, Portuguese, Romanian, Slovak, Spanish, Catalan, Galician, Basque and Swedish.



(a) Dashboard



(b) Discover

Figure 3: Screenshots of EU Contract Hub interfaces

designed to support natural language questions defined by the ProCure project consortium in the context of healthcare procurement, but they can be modified and other visualisations can be created if needed.

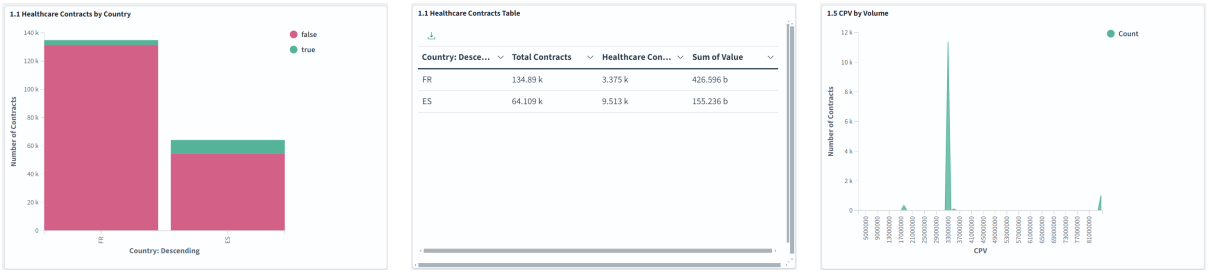
2) Discover. This interface enables the consultation of individual contracts through different attributes, such as the presence of the word ”mask” in one of the attributes or the existence of a specific value of a specific attribute. This search platform enables us to consult over a million contracts individually in a matter of seconds. As shown in figure 3b, there are three main areas in the visualisation. First we have the filter panel, where we can enter the requirements of our search by attributes, such as defining a specific country of contracts or a range of CPV values, for example. Same as with the dashboard, filters can be entered both by hand, using *Documentum Query Language (DQL)*, or using the ”Add Filter” option. The second highlighted area shows the attribute selection, where the election of displayed attributes of the contracts is defined according to the users’ necessities. The third and last highlighted area displays the results of the search with the atributtes previously selected and the filters, if defined, applied. The contracts displayed there can be then consulted individually by clicking on the magnifying glass icon next to the desired contract, which will open a side window where the document details will be fully shown.

3. EU Contract Hub

During the demo we are going to demonstrate two use cases: *a)* Extraction of statistics and visualisations related to healthcare contracts from Spain and France with a value equal or higher than 100.000€ and *b)* The search for information on a specific contracts, namely the procurement route of contracts made by the Spanish institution ”Universidad Politécnica de Madrid” about protective gear.

Case a) To do this search, we go to the ”Dashboard”[4] interface of the platform (on the side menu of the platform, *OpenSearch dashboards →Dashboards*). We stated that the requirements of our search are contracts from Spain and France with a value equal or higher than 100.000€, so we define a *DQL* query stating this filters: *”(Country: ES or Country: FR) and Value >= 100000”*. The graphics and tables displayed in the dashboard will be adapted following the filters defined. For example, as we can see in figure 4, we can see the proportion of contracts that are healthcare related against the ones that are not for the chosen countries (figure 4a). We can also find the numerical value of the volume of contracts that are healthcare related against the total number of contracts and the total monetary value of these (figure 4b) or the distribution of the healthcare contracts by their CPV value (figure 4c).

[4]https://procure.linkeddata.es/



(a) Healthcare Contracts by Country (b) Healthcare Contracts Table (c) Healthcare contracts by CPV

Figure 4: Screenshot of case a) *(Dashboard)* demo results

Case b) To do this search, we are going back to the ”Discover” interface of the platform (on the side menu of the platform, *OpenSearch dashboards →Discover*). We stated that the requirements of our search are that the Contracting Authority is the Universidad Politécnica de Madrid and that the product asked for is protective gear, so we define a *DQL* query stating this filters: *”Contracting Authority Name: ”Universidad Politécnica de Madrid” and CPV Description: ”Protective gear””*. Given that we wanted to get the procurement route of these contracts, in the attribute selection area we select ”Procurement Route”, in addition to ”_id”, ”Title” and ”CPV”, to have a little more context over the resulting contracts. As shown in figure 5, we can see that this search gives us two resulting contracts, which both state that they follow a ”Direct Procurement” route.

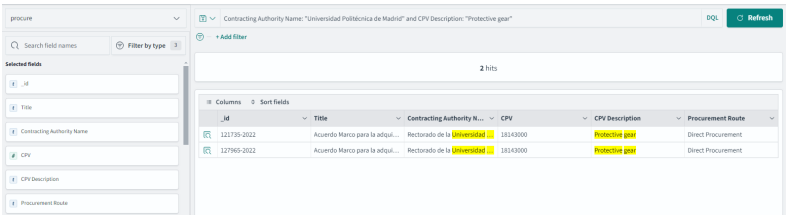


Figure 5: Screenshot of case b) *(Discover)* demo

4. Conclusions and future work

In this paper we presented *EU Contract Hub*, an open platform for the efficient exploration, analysis, unification, and enrichment of over a million public procurement contracts. We also presented the ingestion framework created to unify various contract formats into a single, unified structure that enables the efficient consultation of both individual and global information regarding public procurement.

As future work, we will continue to study the feedback received by the users and implement the changes needed to improve our tool, in addition to the finalisation of the contract translations. We are also currently developing the partial and/or total verbalisation of contracts using , which will enable us to give users a textual summary of the contract in addition to the table-formatted data available currently. Our intention is to be able to create a Retrieval Augmented Generation system using all this data where contract consultation using natural language is enabled.

Acknowledgments

This work is supported by the European Union under the project ProCure (Grant Agreement number 101128437).

References

[1] Supplement to the official journal of the european union, https://ted.europa.eu/TED/main/HomePage.do, 2024. Accessed: 2024-09-16.

[2] Procure: Public procurement assessment in the healthcare sector, https://www.projectprocure.eu/, 2024. Accessed: 2024-09-16.

[3] Organisation for Economic Co-operation and Development, OECD Statistics Database, 2024. URL: https://stats.oecd.org, accessed: 2024-09-13.

[4] Eurostat, Eurostat Database, 2024. URL: https://ec.europa.eu/eurostat, accessed: 2024-09-13.

[5] N. Baccouri, Deep-translator, https://github.com/nidhaloff/deep-translator, 2023. Accessed: 2024-09-16.

[6] Google, Google translate, https://translate.google.com, 2024. Accessed: 2024-09-16.

[7] J. Tiedemann, Parallel data, tools and interfaces in OPUS, in: N. Calzolari, K. Choukri, T. Declerck, M. U. Doğan, B. Maegaard, J. Mariani, A. Moreno, J. Odijk, S. Piperidis (Eds.), Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC'12), European Language Resources Association (ELRA), Istanbul, Turkey, 2012, pp. 2214–2218. URL: http://www.lrec-conf.org/proceedings/lrec2012/pdf/463_Paper.pdf.

[8] B. Zhang, P. Williams, I. Titov, R. Sennrich, Improving massively multilingual neural machine translation and zero-shot translation, in: D. Jurafsky, J. Chai, N. Schluter, J. Tetreault (Eds.), Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, Association for Computational Linguistics, Online, 2020, pp. 1628–1639. URL: https://aclanthology.org/2020.acl-main.148. doi:10.18653/v1/2020.acl-main.148.

[9] P. Koehn, Europarl: A parallel corpus for statistical machine translation, in: Proceedings of Machine Translation Summit X: Papers, Phuket, Thailand, 2005, pp. 79–86. URL: https://aclanthology.org/2005.mtsummit-papers.11.

[10] R. Rei, C. Stewart, A. C. Farinha, A. Lavie, COMET: A neural framework for MT evaluation, in: Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP), Association for Computational Linguistics, Online, 2020, pp. 2685–2702. URL: https://www.aclweb.org/anthology/2020.emnlp-main.213.