# Detailed Case Study

# Moldova: Policy and Business Advice and Support in Legislative Drafting for eProcurement Reforms

# Redeveloped OCDS-based visualisation tool and OCDS data feed to handle available data sources, including data quality assessment and evaluation of data consolidation architecture

# **Big data and data analytics**

### Open Contracting Data Standard based data visualisation

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Lead Organisation**: The Ministry of Finance of the Republic of Moldova; European Bank for Reconstruction and Development (EBRD) |  | **Location:** Republic of Moldova |
|  | **Problem Statement:** **The Moldova Contract Data Visualisation website** was developed and launched in September 2017. During the implementation of its’ end-to-end eProcurement system (MTender) Moldova had **several pilots** and **simultaneously a working legacy eProcurement system**. To provide visibility and transparency numbers on the whole procurement environment of the country, EBRD decided to support the **redevelopment of the existing data visualisation website** so that all data sources could be integrated at the basic level in order to enable visualisation of data and main Key Performance Indicators (KPIs) for the countrywide digital procurement. | | |
|  | **Description:** Data from the sources available of public procurement information in Moldova was integrated using the Open Contracting Data Standard (**OCDS) data structure** and reused for an **open-source data visualisation tool** previously developed by the World Bank. **Data quality analysis** was conducted to assess the practicality of **the data consolidation for the development of full-scale analytical infrastructure**. This provides transparency and visibility of the procurement information by using a single point of data visualisation and the ability to download integrated procurement data from a single point of access. | | |
|  | **Lessons learnt: 1.** The **underlying quality of the data** is a key determinant of a success – impacting the costs of the data transformation and the accuracy of the public procurement statistics generated; **2**. **Availability of data and documentation** leads to more fast and accurate data operation, saving time and money; **3.** **Data consolidation** is a much more complicated process even while operating with standardised data because of semantical, logical and structural differences of the data sources. | | |
|  | **Cost: €18 040** (the main factor impacting the cost of equivalent projects is the quality of the underlying data). |  | **Impact**: Enables public procurement policy-makers to make **data-driven decisions**; Provides access point to integrated procurement data; Ensures transparency of public spending and data visibility. |
|  | **Human resources:** Project implemented by **EBRD and IT/consulting contractors**, providing expertise in business intelligence software, web applications, and data analysis. |  | **Project timeline**: Sep 2019 – Nov 2019 |
|  | **Project status:** Fully deployed |
|  | **Email:** [NiewiadE@ebrd.com](mailto:NiewiadE@ebrd.com) |  | **Website:** <http://opencontracting.date.gov.md> <http://opencontracting.eprocurement.systems> |

#### Context and problem statement

A Consultant was appointed by EBRD under the technical cooperation project “Moldova: Policy and Business Advice and Support in Legislative Drafting for eProcurement Reforms” to redevelop the existing Moldova OCDS Contract Data Visualisation website.

The Republic of Moldova has several existing electronic public procurement data sources which formed the basis for the project:

* **The Moldova Contract Data Visualisation website** developed by the Government of Moldova's Public Procurement Agency under the Ministry of Finance in partnership with the eGovernment Center and with support from the Open Contracting Partnership;
* **DB-1** (<https://public.api.mepps.openprocurement.net/api/0/tenders>) – source of below the threshold procurement data;
* **DB-2** (<https://public.mtender.gov.md/tenders>)–source of above the threshold procurement data;
* **Registrul de Stat al Achizitiilor Publice** (SIA RSAP).

Because of the development of the eProcurement system, **the Moldova Contract Data Visualisation website had to be redeveloped** in order to report on consolidated contracting data from the new electronic public procurement system MTender which brings together procurement data from two dedicated data sources (DB-1 and DB-2), as well as the historical data from the old Sistemului informational automatizat "Registrul de Stat al Achizitiilor Publice" (SIA RSAP).

During the redevelopment process, **the data quality analysis had to be conducted** to ensure the quality and consistency, reliability and compatibility of data obtained from the different data sources.

The problem statement for this project was designed according to the requirement to redevelop the existing Moldova OCDS Contract Data Visualisation tool. The input data for the redevelopment was as follows:

* A repository on the GitHub platform of the old version of **the Moldova OCDS Contract Data Visualisation tool**;
* **DB-1**and**DB-2** OCDS Application Programming Interfaces (API) as well as their procurement data mapping rules;
* **The web-interface**.

The most common issues impacting the development process were the following:

* More than one data source (multi-platform systems);
* Data redundancy and/or insufficiency;
* "Dirty data" which included but is not limited to omissions, abnormal values, duplicates and contradictions.

The following list of tasks was formulated for the purpose of the assignment to redevelop the old version of the Moldova OCDS Contract Data Visualisation tool:

* Review the source code from the GitHub repository of the old version of the Moldova OCDS Contract Data Visualisation tool, which was designed by another technical team. Consider whether it can be adjusted to receive new data from the DB-1 and DB-2 databases along with the old ones, thus operating in the multi-platform environment;
* Map the data from new data sources to the OCDS format that is used on the portal to display the correct data;
* Develop a data import module that aims to extract data from the DB-1 and DB-2 OCDS APIs, transform it to the required OCDS format used on the web portal and further export these OCDS formatted data as a JSON or comma-separated values (CSV) file;
* Deploy a web portal prototype based on the multi-platform system to the new server.

#### Objectives and vision

The purpose of the technical cooperation project was to modernise and update the existing public procurement analytical tools as well as to consolidate data from old and new eProcurement data source to provide a single point of procurement data visualisation and the ability to download integrated procurement data from a single point of access, **increasing transparency, visibility and availability of the procurement information and driving improved public procurement policies and decisions**. The TC project aimed to:

* Provide improved public procurement statistics to guide decision making;
* To increase the accessibility to all available electronic public procurement data of the country;
* Provide transparency on public spending for citizens.

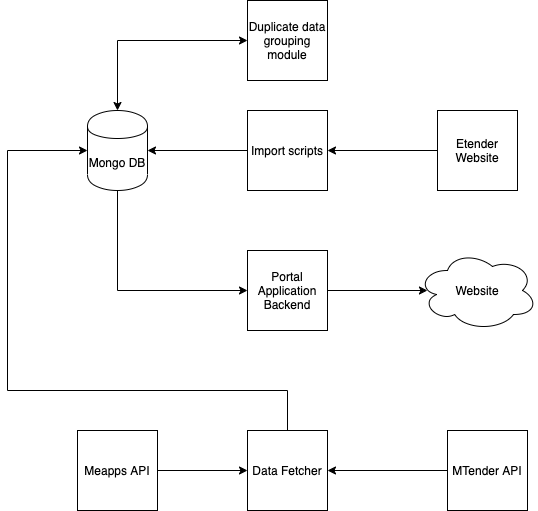
The vision to achieve these objectives contains two primary points:

* **Redevelopment of the Moldova OCDS Contract Data Visualisation tool** in order to help public procurement policy-makers to make data-driven decisions and ensure transparency of public spending and data visibility;
* **Development of a single access point,** including actions on **data quality investigation**, providing recommendations on how to reduce or eliminate data quality issues and providing the initial data compliance with the required data quality as well as suggestions on how to operate with data from multi-platform systems, **in order to have an ability to download integrated data on countrywide procurement processes**.

#### Technological solution and implementation

The solution developed **fetches data from DB-1 and DB-2 APIs and stores them in the dedicated Mongo database (DB)** which is built for further data representation in the OCDS format. Another module of the solution **gathers historical data and also stores it in the Mongo DB**. **The module for duplicated data grouping** **checks** whether the names of Contracting Authorities are different **and removes any**. **Duplicates**. The **portal application backend** has been set up so that **access to this data directly is available through an API and visual representations of data and main KPIs for the countrywide digital procurement are available** viathe redeveloped Moldova OCDS Contract Data Visualisation tool**.** The structure of the redeveloped Moldova OCDS Contract Data Visualisation tool is presented in Figure 1.

***Figure 1:*** *The structure of the redeveloped Moldova OCDS Contract Data Visualisation tool*



##### **Development of a single access point, providing initial data for the new Moldova OCDS Contract Data Visualisation tool**

At the start of the redevelopment process, **the technical team pulled the source code from the original repository**. As **the database was not initialised** when the application starts, the technical team had to find a backup copy of the DB in the source code. Then, **a number of issues were also faced** that certainly could cause a delay in completing the assignment. The project team discovered these issues, identified ways to resolve them and fixed those issues where possible. The issues and actions performed to eliminate them:

* The source code of the portal didn't allow viewing data after the 2017 year, thus all "2017" values in the code were removed and updated to a dynamically calculated variable based on data for the last year in the database.
* The Etender import scripts failed while scraping data from the website, so the exception handlers for errors were added for the cases when the source formatting differs from what the scraping tool expects.
* There was no documentation on what data is stored in the collections and used for each query on the website so all the necessary data from the website had to be manually mapped to the data from the APIs.

The technical team successfully **found the scripts, fixed the issues, and initialised the DB** enabling the **data from the Etender to be pulled to the Mongo DB**.

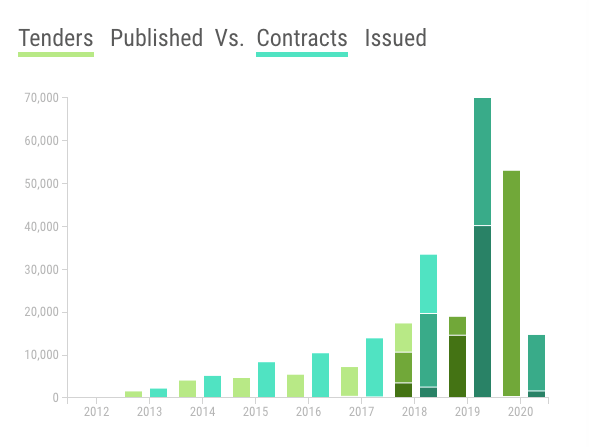
After that, **the technical team performed several steps as follows**:

* Developed the Import-Module to grep[[1]](#footnote-2) data from the DB-1 and DB-2 APIs;
* Developed the fetched data grouping module to remove duplicate data;
* Created the adjusted DB (MongoDB) with collections from different sources.

##### **Redevelopment of the Moldova OCDS Contract Data Visualisation tool**

As for the project goals, the visual part of the website was redeveloped to enable public procurement policy-makers to implement **data-driven decisions** and **ensure transparency of public spending and data visibility.** As a result, the old version of the Moldova OCDS Contract Data Visualisation tool was reviewed and updated by the Consultant’s analytical and technical teams. Several filters have been added or updated, and a new visual layout has been created (Figure 2). A **prototype of a multi-platform system web portal of the Moldova OCDS Contract Data Visualisation tool was created** **and deployed** to the new server; all the data required is stored in the version of the OCDS format of the website.

***Figure 2:*** *New visual layout with data distribution by different data sources*



Steps performed in order to redevelop the visual part of the Moldova OCDS Contract Data Visualisation tool:

* Dynamic variables were added in order to take the last year in the database. This allowed the system to take the date of the last tender ID, set the year of the procedure as the last one we have in the system and dynamically update it on the website.
* The visual part of the website was reviewed and updated.

##### **Data quality evaluation**

The value and reliability of knowledge obtained as a result of data analysis depend not only on the effectiveness of the analytical methods and algorithms used but also on how the initial data for analysis is selected and prepared and the data quality investigation and corresponding actions were provided by the analytical team. For a better understanding and interpretation of the data quality analysis results, a **dedicated Business Intelligence (BI) tool was developed**.

The dedicated BI tool **operates with data about tenderers and Contracting Authorities from the DB1 and DB-2**. In order to carry out the data quality evaluation, the following actions were undertaken:

* **Data preparation** – defining the scope of procurement data for analysis as well as creating specific queries to collect data from the data sources;
* **Data processing and cleansing** – to fetch the most adequate and suitable results for analysis;
* **Data quality analysis** – identification of the data quality issues, the examination of the reasons for their occurrence and the impact on the correct calculation of procurement KPIs; creation of the dedicated BI application.

The data obtained was compared to the data from the official catalogues and several data quality issues were detected:

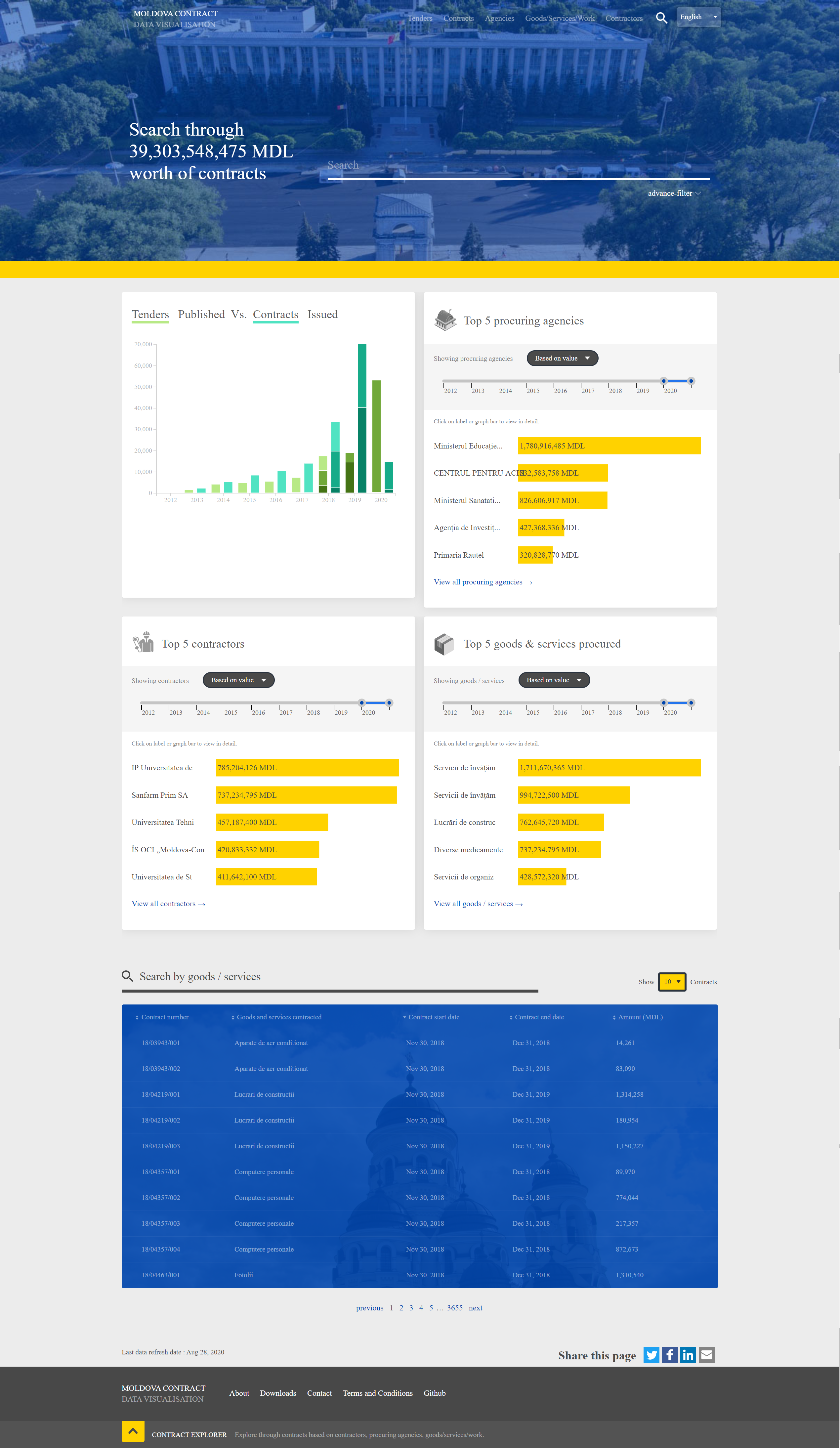
* Invalid or missing identifier for a tenderer;
* Invalid or missing legal name for a tenderer;
* Invalid or missing country for registration of a tenderer.

Unfortunately, the register of Contracting Authorities wasn’t publicly available and data quality issues in the Contracting Authority data couldn’t be determined.

#### Results and future expectations

As a result of the project, **a prototype of a multi-platform system web portal of the Moldova OCDS Contract Data Visualisation tool has been created and deployed**. It provides transparency and visibility of the procurement information by using a single point of data visualisation and the ability to download integrated procurement data from a single point of access.

***Figure 3:*** *Prototype redeveloped Moldova OCDS Contract Data Visualisation tool*



In terms of the goals set for the project, the redeveloped tool:

* **Provides transparency on public spending** – through the open online Moldova OCDS Contract Data Visualisation tool;
* **Ensures the accessibility to the Moldova procurement data** – through the developed single access point to all available electronic public procurement data of the country
* Differences between the data sources and approaches to the publication of data in the OCDS format led to the decision to consolidate the multi-platform system data at the database level. A dedicated OCDS database was created and organised the storage of all data in the special collections to consolidate the multi-platform system data. We consider this approach to be the most suitable in similar situations;
* **Provides support for public administration regarding public procurement decisions** – due to the improved public procurement statistics presented on the redeveloped website.

For future development, when the amount of data has increased, the queries for the tables should be updated. The current request is time-consuming and needs to be updated, as well as a pagination method, which can sometimes cause system freezing due to a memory leak.

Also, it is better to have similar public procurement data within one data source to avoid data intersection, duplication and ambiguous data.

Three issues were determined among the key information on tenderers during the data quality evaluation. The main reason for all the issues is wrong, unverified data input. The following steps can be taken to eliminate these issues:

* **Add validation to existing input data sources available on platforms**: check out the number of spaces entered, the format of identifiers, empty input, etc.;
* **Integrate available official registers and catalogues with the existing input data sources**: it means that the tenderer can be authorised in the system only with the identifier and legal name contained in the official register and a drop-down selection menu provided to enable the selection of the proper values from the official catalogues.

Another issue was that the registry of Contracting Authorities data isn't publicly available, and it is recommended to resolve this issue is to **make all official registers and catalogues available for public use**.

#### Costs and requirements

| **Description** | **Total, EUR** |
| --- | --- |
| Mapping schemas development, requirements explained to developers, setting up the development | 2340 € |
| Extractor for DB2 | 2340 € |
| Extractor for DB1 | 1170 € |
| Extractor for eTander | 780 € |
| Main page interface, missing data issues, testing, bugs fixing | 5460 € |
| Project Management for developers | 1950 € |
| Preparation of the documentation on visualisation tool redevelopment | 1,000 € |
| Data quality analysis of the available OCDS data sources | 3,000 € |
| **TOTAL** | **18,040 €** |

#### Challenges and lessons learnt

**Standardisation** of published public procurement data can **lead to better transparency, efficiency, the effectiveness of public procurement procedures and increase the international trading potential**. However, the OCDS standard (used in Moldova) allows different structures for the data that are to be published. It is very convenient for depicting local public procurement processes, but it leads to difficulties in operating with, and understanding, the published data and its structure. It also leads to technical issues e.g. when it is necessary to get data from the release packages of one data source and record packages of another data source, the same import scripts cannot be used, and they have to be adjusted for each data source.

Another difficulty is associated with the **semantic meaning that each variable is endowed with** – each publisher has to provide a unified semantic meaning for each variable published. Also, the issue with a different process representation was encountered: in one data source contracting process ends on the contracting stage, while in another data source, it ends on the awarding stage.

All these issues should be addressed and resolved to avoid issues with **data consolidation and integration** in the future.

In addition, it **is essential to provide** analytical tools, analysts, government organisations, Non-Governmental Organisations and other stakeholders interested in public procurement with **the data of high quality**. As a result of data preparation, processing and cleansing, as well as data quality assessment, the data quality issues were detected and investigated. The reasons for data quality issues occurring were determined and addressed, and the possible impact of data quality issues on analytics was defined. Our analytical team developed recommendations for the elimination of the issues identified. Moreover, for a better understanding and interpretation of the data quality analysis results, a dedicated BI tool was developed.

1. Grep is a command-line utility for searching plain-text data sets for lines that match a regular expression. Its name comes from the ed command g/re/p (Globally search for a Regular expression and Print matching lines) [↑](#footnote-ref-2)