

Open contracts: Linking open public data

Michal POLÁK*, Nghia PHAM VAN*, Michal SELICKÝ*, Michal ŠTEVULIAK*,
Marek ŠTEVULIAK*, Martin FUKAS*, Ľubomír SAMOTNÝ*

*Slovak University of Technology in Bratislava
Faculty of Informatics and Information Technologies
Ilkovičova 2, 842 16 Bratislava, Slovakia
timak9.fiit@gmail.com*

The state publishes a large amount of open data about contracts, companies, entrepreneurs and many others. Since the year 2011, the state is obliged to disclose these data about contracts otherwise they would be automatically invalid. For this reason, the state began to publish contracts using its interface¹. However, the interface was not easy to work with. In order to allow easy access to this data, the project *otvorenezmluvy*² was created. The project started as a collaboration of *Aliancia Fair-play*³ and *Transparency International*⁴ with the main goal of analyzing this data via public crowdsourcing and to find suspicious contracts. This can increase the level of state management of its budget and reduce corruption.

However, after the project started, any work on the project was inactive for several years, until we started working on it within the Team Project. During this inactive time, the technologies the project was using became outdated and the state also changed its publication methods of contracts and the existing method was not functional anymore. Since the downloading of contracts is a key part of the project, it had to be repaired.

Nowadays the state is publishing even more data. None of this data is linked. Therefore, all the necessary information is not available when analyzing it. Linking this data would greatly improve its analysis and the detection of fraudulent contracts. Also, providing the people channel of a communication about contracts can be helpful for connecting people and also increase public awareness. Adding new sources of data is therefore one of our main goals.

One of the datasets that can greatly increase the amount of information that the project offers is the Register of Legal Subjects. It is one of the key registers in public administration. It allows us to obtain specific information about the companies that are part of a contract. All contracts contain the identification number of the organization. We can therefore connect that company from the contract with the data from the register. This way, we can get valuable information about the company, for example its official company name, since the names on contracts can vary. Another information we can gain is the residence or economic activity of company. Although the state provides its own API, it is preferable to use API from *slovensko.digital*⁵, which integrates multiple registers and provides a way to synchronize data.

Another dataset is the Register of Public Sector Partners⁶. This dataset contains information about business end owners. With this information, the contract can be associated with a particular person standing behind the company and gaining profit.

The Registry of Financial Statements⁷ contains data about the economic activities of subjects such as its turnover, profit, liabilities or assets. It could also increase the chance of detecting fraudulent contracts. For example, if a company signs a contract with a high value, it is expected that it has an adequate turnover and assets.

We already managed to add the Register of Legal Subjects during the winter semester and are currently working on adding another new data source and connecting them with existing ones. All these newly

* Master degree study programme in field: Intelligent Software Systems
Supervisor: Dr. Róbert Móro, Institute of Informatics, Information Systems and Software Engineering, Faculty of Informatics and Information Technologies STU in Bratislava

¹ <https://www.crz.gov.sk>

² <http://otvorenezmluvy.sk>

³ <http://fair-play.sk/>

⁴ <http://transparency.sk>

⁵ <https://ekosystem.slovensko.digital/otvoreneapi>

⁶ <https://rpvs.gov.sk/rpvs>

⁷ <http://www.registeruz.sk>

added data sources will not just increase the amount of information provided but will also allow for more accurate and easier searching of contracts or companies. So, the next step after adding new data sources is making sure to properly integrate them into the project and linking them with the original data to enhance transparency. For example, by enabling a new and more specific search method or adding new user screens.

Not all of public data is of equal quality. Anomalies when the contract value is zero, but it is worth millions in reality, are very common. The detection of these anomalies is another challenge which can increase data transparency.

Since the project started in the year 2012 and was not updated since then, the technologies it uses are outdated. The project is using Ruby on Rails, Elasticsearch and PostgreSQL as a database. Sustainability is one of the important properties of software projects. It is therefore important to update these technologies for easier future work. The technologies that the project is using evolved during the inactive period of the project and newer versions are not completely compatible with older versions. It is one of our main goals to update them. There is a great possibility that other teams within the Team Project might continue working on this project and updated technologies will allow them to focus on delivering new functionality and improving already existing ones.

The project `otvorenezmluvy.sk` will allow citizens to conveniently and comfortably read, search and evaluate contracts concluded by the state and state institutions. This will improve the state's state management of its budget, increase transparency and reduce corruption. Our main goal in this project is updating outdated technologies, adding new data sources and to provide a convenient and simple way to work with this information.