TheyBuyForYou
Visualising and analysing public procurement data

Matej Kovačič

Matej Posinković

(CC) 2021

Jožef Stefan Institute
Ljubljana, Slovenia

UNESCO
United Nations
Educational, Scientific and
Cultural Organization

IRCAI
International Research Centre
on Artificial Intelligence
under the auspices of UNESCO

http://tbfy.ijs.si
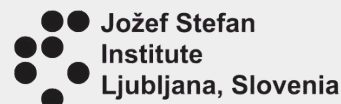
About - general

The TheyBuyForYou project started in January 2018 and has completed at the end of December 2020.

TheyBuyForYou consortium consisted of 10 leading companies, universities, research centres, government departments and local authorities in the UK, Norway, Italy, Spain and Slovenia.

TheyBuyForYou has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 780247.

About - aim

Project's aim was to make procurement in Europe more accessible. That included:

- collecting and normalizing tenders and awards data;

- building a European procurement Knowledge Graph containing millions of detailed contract and tender data;

- Making data available in standardized OCDS format;

- developing several tools for analysis and visualisation of collected data;

- testing tools upon several business cases.

TheyBuyForYou Platform

TheyBuyForYou Platform
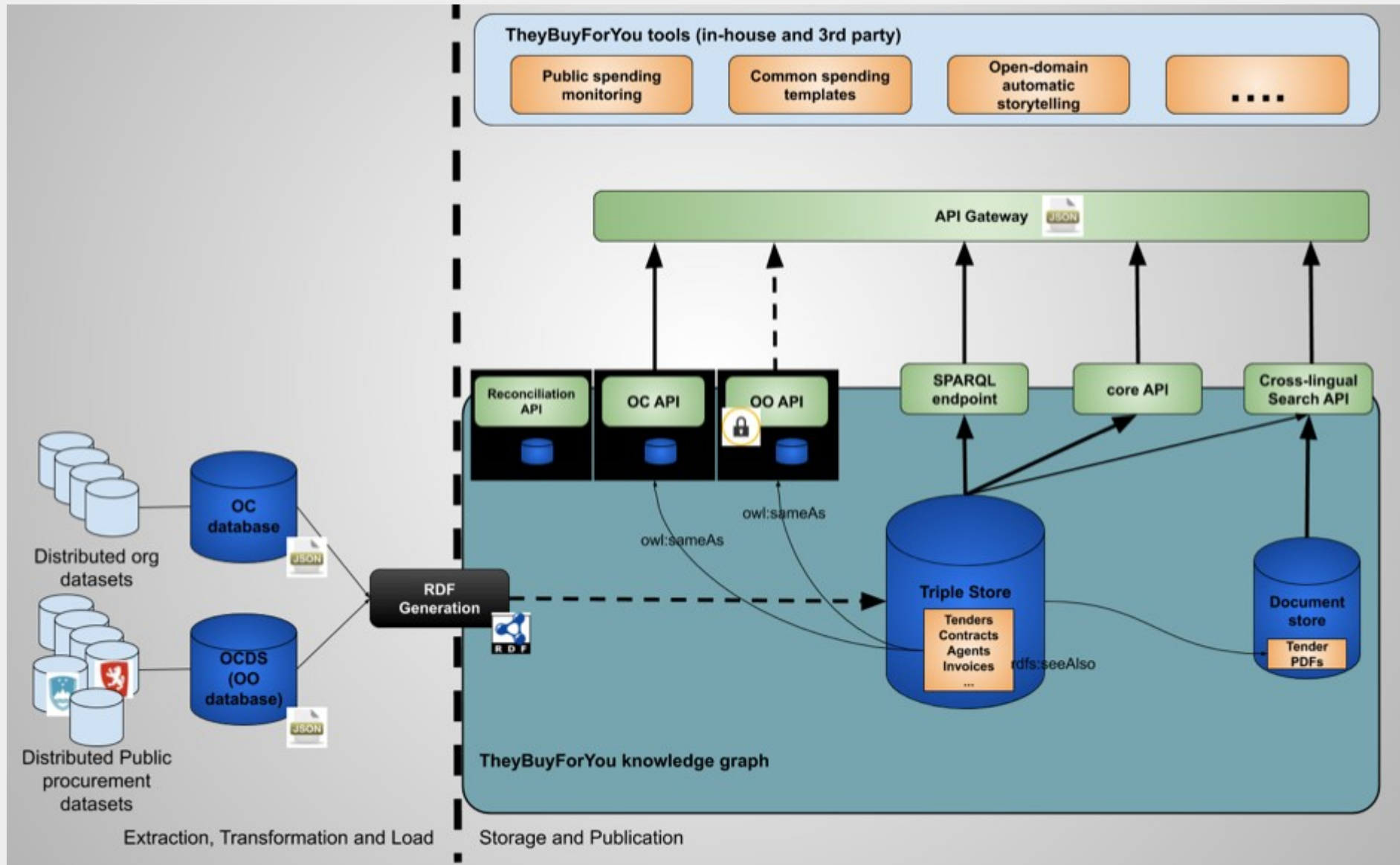
TheyBuyForYou Platform (https://tbfy.github.io/platform/):

- Data (TBFY Knowledge Graph and Documents repository).

- Schemas (TBFY ontology that imports the OCDS ontology for procurement data and the euBusinessGraph ontology for the company data).

- Tools (Harvester, R4R for RESTful services, KG data ingestion pipeline, SPARQL GUI and OptiqueVQS).

- APIs (set of core APIs built or used in the project).

- Added-value services (services and tools).

Knowledge Graph

The TheyBuyForYou (TBFY) Knowledge Graph (KG) integrates procurement and company data. The KG covers data from January 2019 onwards. New data is onboarded every night.

As of May 2021 the KG consisted of more than 189 million triples and contained information about:

- 1.92 million tenders,

- 2.51 million awards, and

- 133 thousand companies (reconciled suppliers).

Knowledge Graph

The KG data is provided as open data under the Creative Commons BY-NC-SA 4.0 License (use, share and adapt the data for non-commercial uses).

Data (http://data.tbfy.eu/) available through:

- SPARQL endpoint;
- API;
- as data dump.

Anomaly detection platform

An online toolkit developed by JSI for exploring public spending and tender data to make public procurement more efficient and transparent (http://tbfy.ijs.si/).

Anomaly detection platform

Analysis of **financial transactions** between public entities and privately held companies.
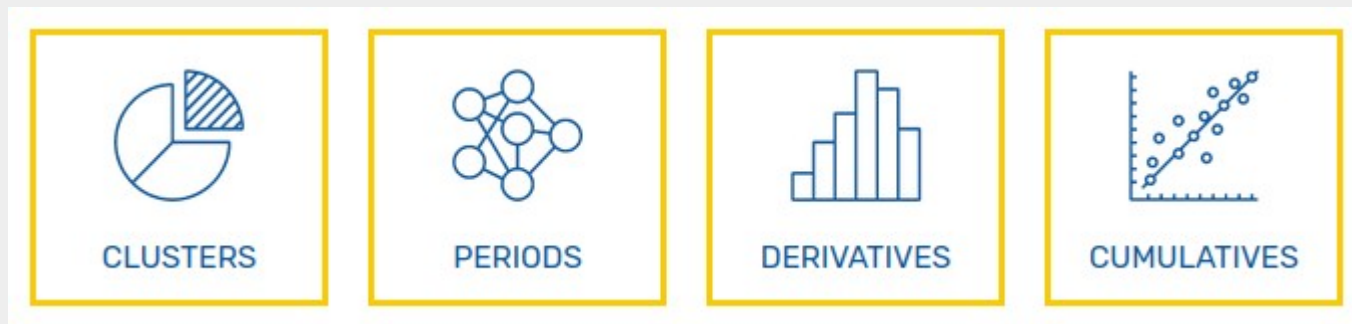

Anomaly detection and visualisation of detected anomalies in spending data could be done by several methods: Average Deviation Anomaly, Jenks Natural Breaks, Period Margin Points Cumulative, Local Extremes Detection and Time Periods Deviations. Anomalies are detected in the whole dataset or in different industry sectors only.

Analysis of financial transactions
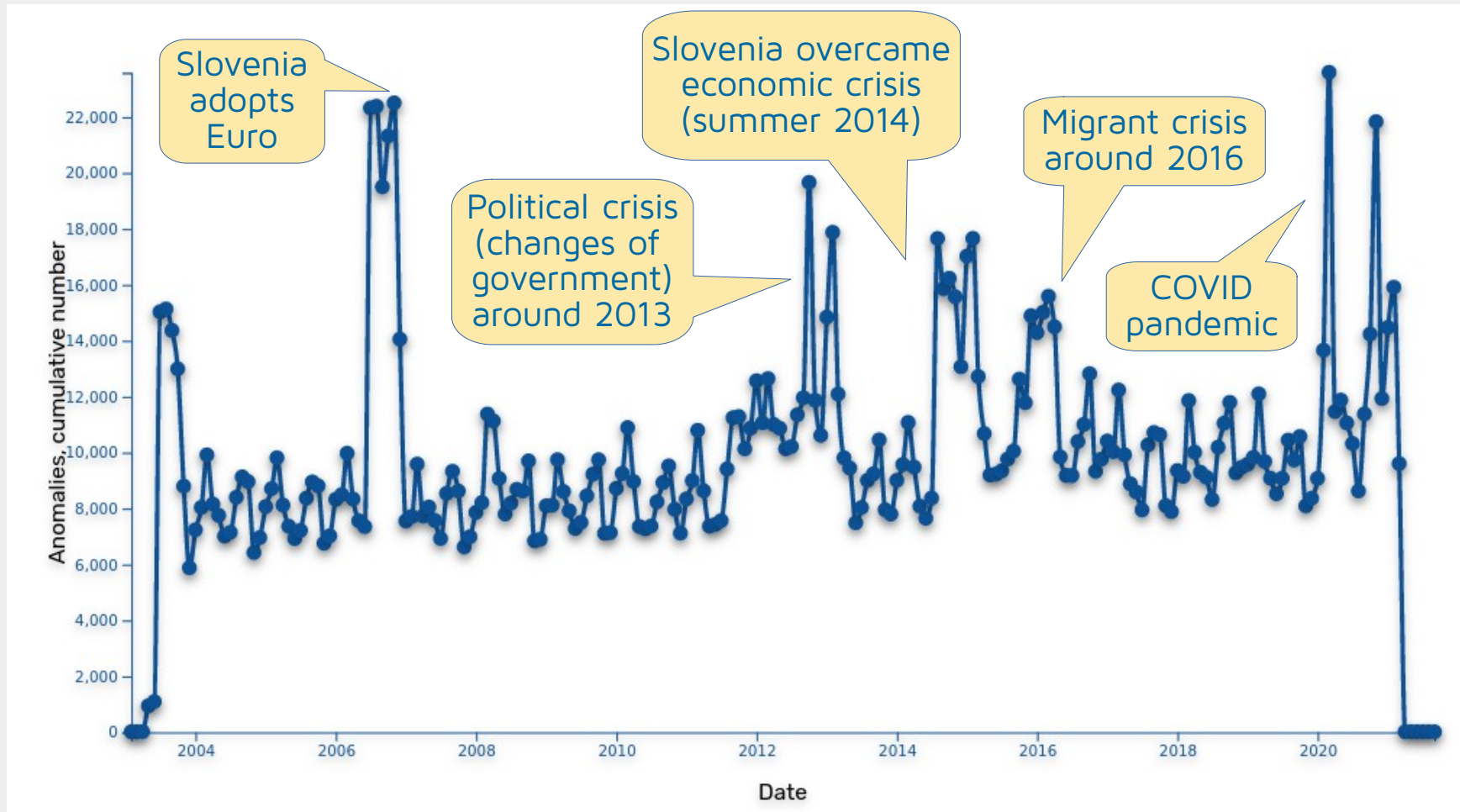
Financial transaction is defined as a base relation between two entities (public sector entity and business entity). We detect when relation started or ended and accumulate starting/ending periods on a timeline. Based on that, we identify deviations and list entities as part of identified extremes (periods method).

Analysis of financial transactions
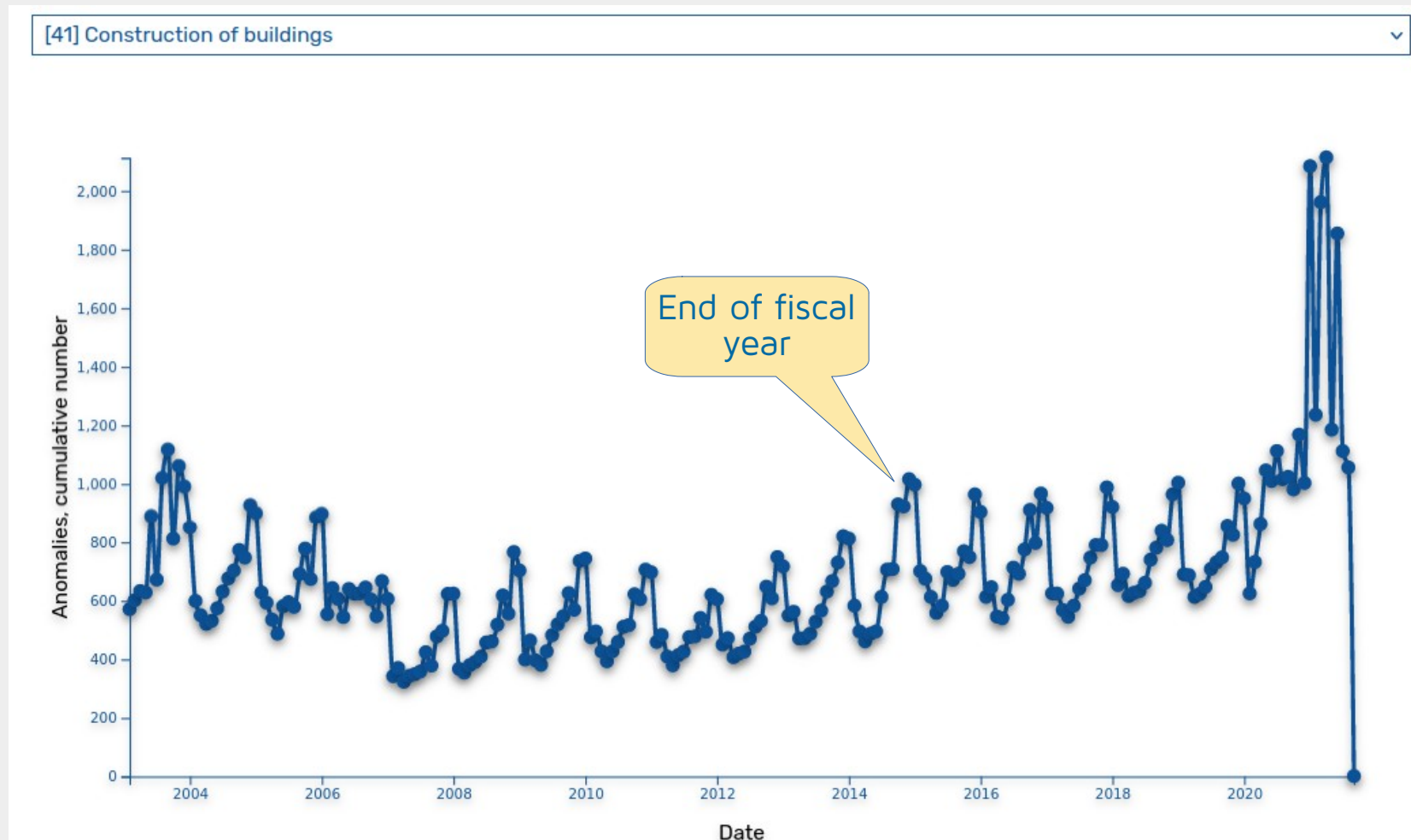
Derivatives method analyses the biggest changes within two entities and transactional relation in a given period. If a change is identified as an anomaly, it is added to the cumulative anomaly graph. The purpose of the method is to identify the companies manifesting biggest changes in transaction relations.

Anomaly detection platform

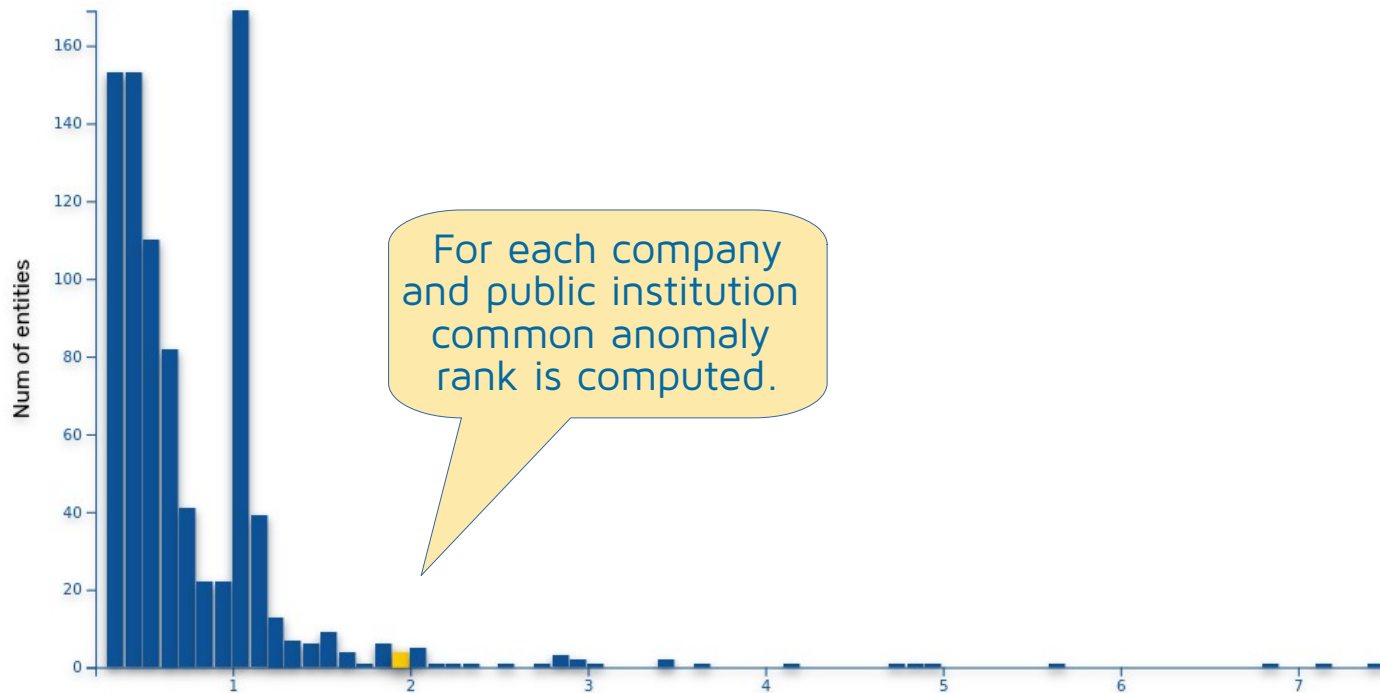Analysis of **public procurement data** in OCDS format.

Analysis of public procurement data could be done by several approaches: supervised, unsupervised and statistical analysis.

Unsupervised analysis is based on k-Means method. Supervised analysis is based on a decision tree analysis, and is used to get additional insights into the public procurement decision-making process. Statistical analysis is done in order to pursue a more intuitive and defined-in-advance goals. StreamStory tool is used to uncover, visualize and explain the inner structure within the data.

Analisys of public procurements

Analisys of public procurements

PHARMACO Informacijski inženiring d.o.o.
Bukovci 127
2281 Markovci
Slovenija

ID — 5744954000
Tax — 86193210
Size — 5-9 employees

Tenders won — 6
Funds received — 487.795,00 €
Tenders issued — 0
Funds issued — 0,00 €
Anomalies found — 5

External links
- Ajpes
- Erar

Tenders won (6) | Tenders issued (0) | **Anomalies (5)**

Number of competitive offers distribution

PHARMACO Informacijski inženiring d.o.o. is on average winning tenders within a **LESS competitive** environment as an average company does. In comparison to other companies, PHARMACO Informacijski inženiring d.o.o. can be found here having a score of -93.9.

Current company's competition for won tenders:		Average competition for companies tender winners:		Average competition for tenders winners (CPV = 72):	
No competitor	100%	No competitor	45%	No competitor	82%
One competitor	0%	One competitor	19%	One competitor	12%
2+ competitors	0%	2+ competitors	37%	2+ competitors	6%

Web platform allows users to review detected anomalies. Anomalies are also explained in natural language. In this case company competition is compared to average competition and average competition within selected CPV.

Analisys of public procurements

Several types of anomalies are detected. Here we can see dependence of contracting authorities from the selected company.

Anomaly detection methods
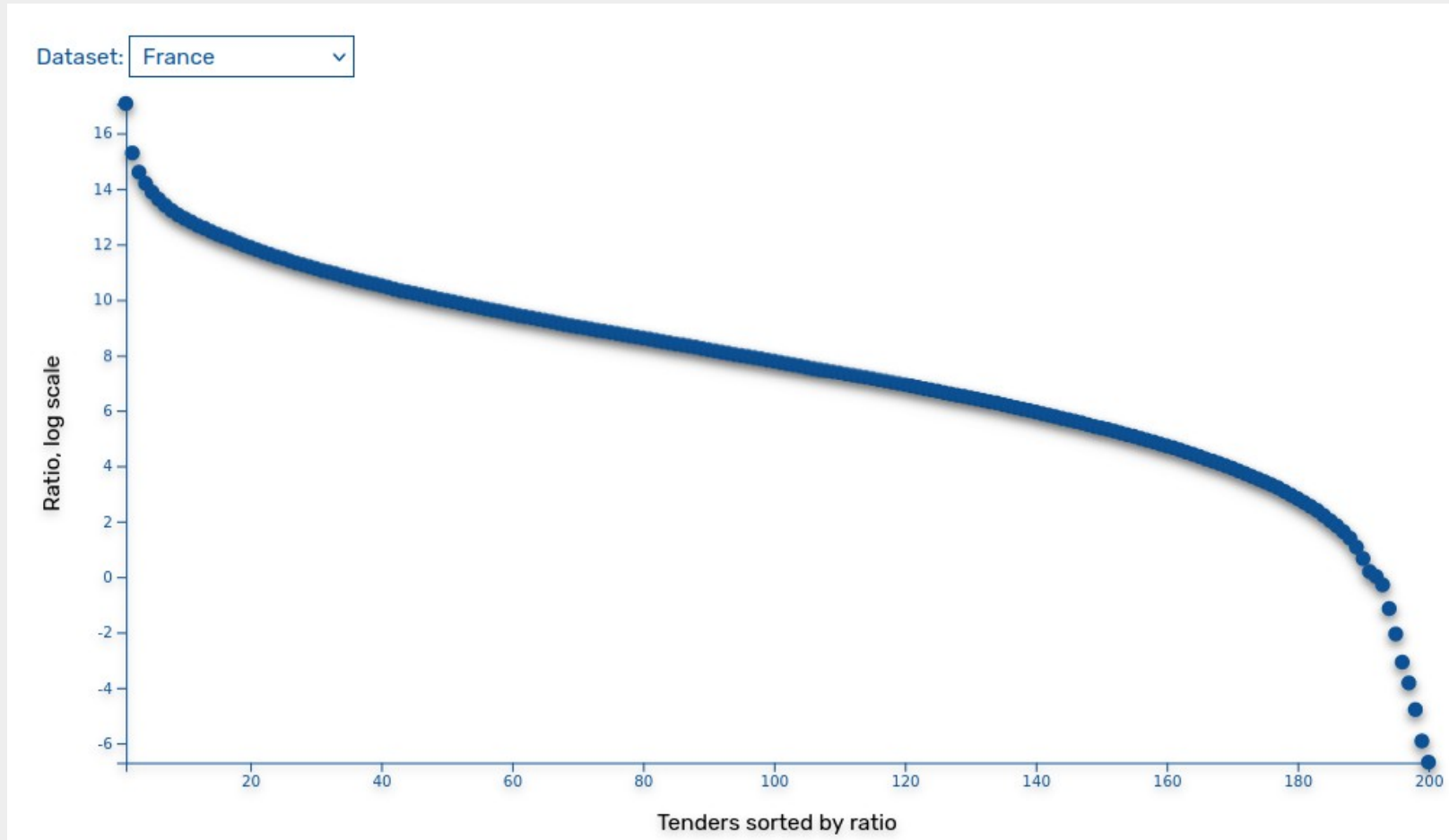
Implemented statistical analysis method (Ratios method) is showing a visual presentation of interdependence between tender value and number of employees of economic operator. Our tool can analyse any data in OCDS format.

Anomaly detection methods

A high deviation indicates a high deviation between tender size and the number of economic operator employees. It is not common for large tenders to be won by companies with extremely low number of employees. While it is not necessarily in breach of the public procurement legislation, it does warrant a closer look.

Dataset: si-ministry   Sample size: 8788   CPV:
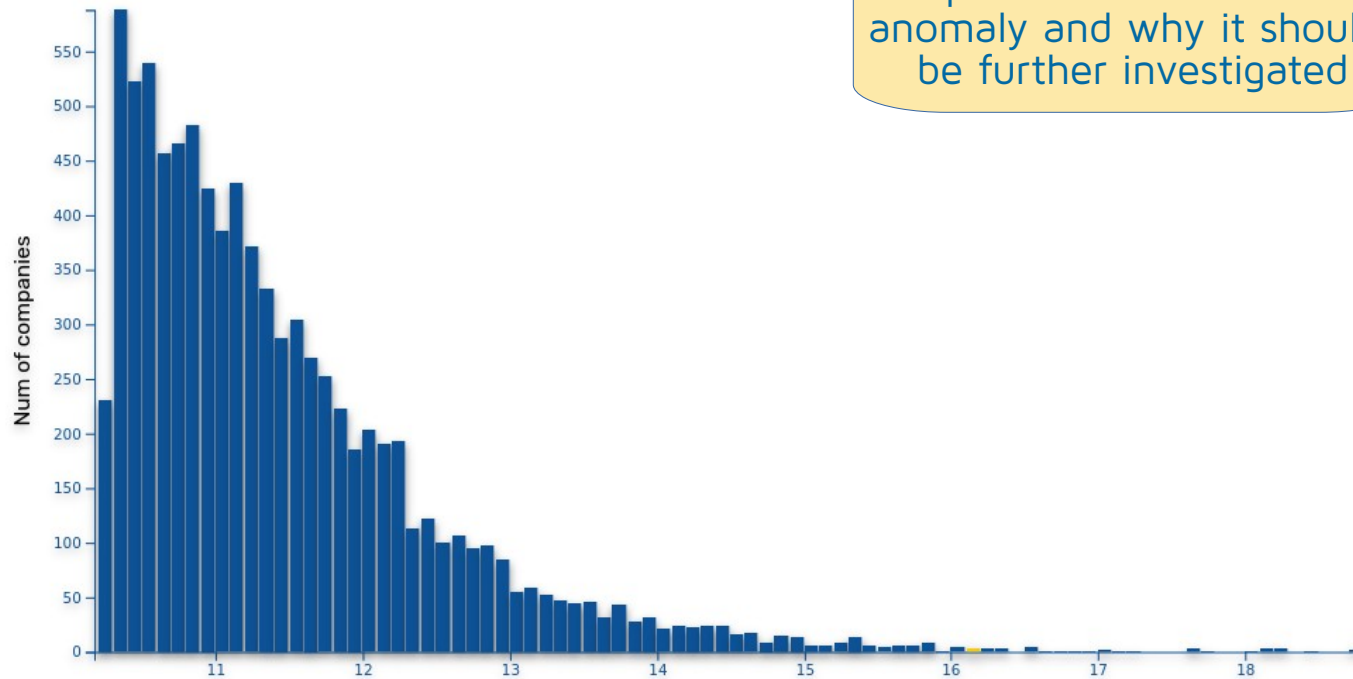
Explanation of detected anomaly and why it should be further investigated

Anomaly measure: higher values infers higher anomaly

Contracting authority:	Economic operator:	Tender amount	Economic operator employees
MINISTRSTVO ZA INFRASTRUKTURO, DIREKCIJA REPUBLIKE SLOVENIJE ZA INFRASTRUKTURO [id:5300177000]	MEGALIT inženirski biro d.o.o. [id:8241767000]	10.614.858,24 €	1

Anomaly detection methods
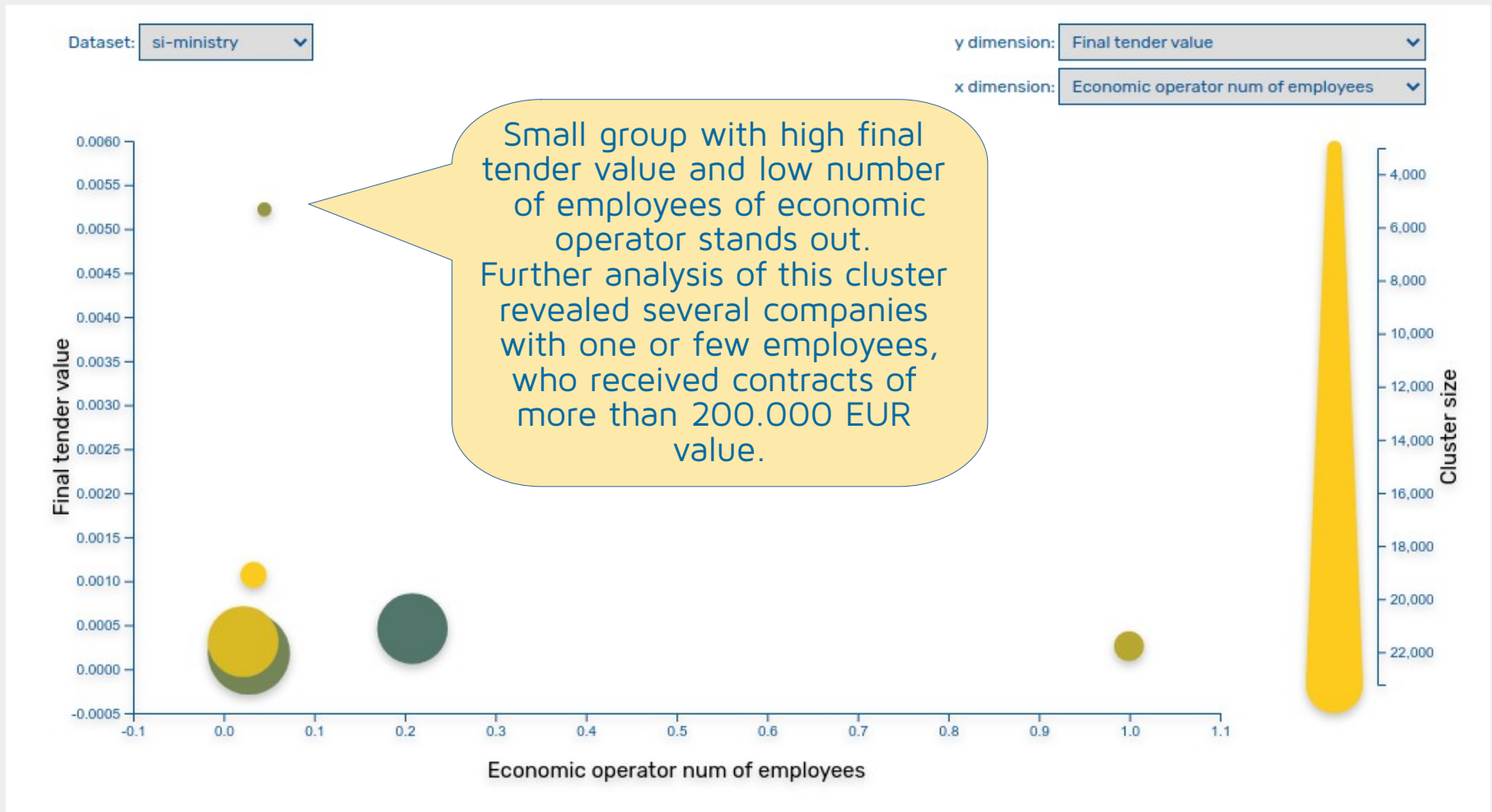
Visualisation of Clusters method. This method is looking for previously undetected patterns in a data, usually those, we are not aware of.

Anomaly detection platform

Platform is accessible at: http://tbfy.ijs.si/.

Platform provides daily exports of the Slovenian procurement data in OCDS format. Data have been made available thanks to the Slovenian ministry for public administration.

Uses procurement data from Knowledge Graph and Slovenian financial transactions data from Erar (https://erar.si/static-data/).

Source code is available at Github: https://github.com/TBFY/anomaly-detection-tool/.

Questions?

Matej Kovačič

Matej Posinković

Jožef Stefan Institute
Ljubljana, Slovenia

UNESCO
United Nations
Educational, Scientific and
Cultural Organization

IRCAI
International Research Centre
on Artificial Intelligence
under the auspices of UNESCO