

Insert here your thesis' task.



CZECH TECHNICAL UNIVERSITY IN PRAGUE  
FACULTY OF INFORMATION TECHNOLOGY  
DEPARTMENT OF SOFTWARE ENGINEERING



Bachelor's thesis

## Open data of Prešov

*Juraj Šprlák*

Supervisor: Mgr. Jan Starý, Ph.D.

22nd April 2018



---

## Acknowledgements

I would like to thank to my



---

# Declaration

I hereby declare that the presented thesis is my own work and that I have cited all sources of information in accordance with the Guideline for adhering to ethical principles when elaborating an academic final thesis.

I acknowledge that my thesis is subject to the rights and obligations stipulated by the Act No. 121/2000 Coll., the Copyright Act, as amended. In accordance with Article 46(6) of the Act, I hereby grant a nonexclusive authorization (license) to utilize this thesis, including any and all computer programs incorporated therein or attached thereto and all corresponding documentation (hereinafter collectively referred to as the “Work”), to any and all persons that wish to utilize the Work. Such persons are entitled to use the Work in any way (including for-profit purposes) that does not detract from its value. This authorization is not limited in terms of time, location and quantity. However, all persons that makes use of the above license shall be obliged to grant a license at least in the same scope as defined above with respect to each and every work that is created (wholly or in part) based on the Work, by modifying the Work, by combining the Work with another work, by including the Work in a collection of works or by adapting the Work (including translation), and at the same time make available the source code of such work at least in a way and scope that are comparable to the way and scope in which the source code of the Work is made available.

In Prague on 22nd April 2018

.....

Czech Technical University in Prague

Faculty of Information Technology

© 2018 Juraj Šprlák. All rights reserved.

*This thesis is school work as defined by Copyright Act of the Czech Republic. It has been submitted at Czech Technical University in Prague, Faculty of Information Technology. The thesis is protected by the Copyright Act and its usage without author's permission is prohibited (with exceptions defined by the Copyright Act).*

### **Citation of this thesis**

Šprlák, Juraj. *Open data of Prešov*. Bachelor's thesis. Czech Technical University in Prague, Faculty of Information Technology, 2018.



---

## Abstrakt

V několika větách shrňte obsah a přínos této práce v češtině. Po přečtení abstraktu by se čtenář měl mít čtenář dost informací pro rozhodnutí, zda chce Vaši práci číst.

**Klíčová slova**    Otvorené dáta, štátna správa, webová aplikácia, mesto Prešov, Java

---

## Abstract

Sem doplňte ekvivalent abstraktu Vaší práce v angličtině.

**Keywords**    Open data, state service, web application, Prešov, Java



---

# Contents

<b>Introduction</b>	<b>1</b>
<b>1 Open data</b>	<b>3</b>
1.1 Definition . . . . .	3
1.2 Requirements . . . . .	3
1.3 Open data in the world . . . . .	8
<b>2 Open data in Slovakia</b>	<b>11</b>
2.1 Political support . . . . .	11
2.2 Slovak open data portal - data.gov.sk . . . . .	12
2.3 Other state projects . . . . .	13
2.4 Business register . . . . .	14
2.5 Non-state initiatives . . . . .	14
<b>3 Analyses of the Prešov open data</b>	<b>17</b>
3.1 Prešov contracts, grants, invoices and orders . . . . .	17
<b>4 Implementation</b>	<b>21</b>
4.1 Data management . . . . .	21
4.2 Web application . . . . .	24
4.3 Deployment . . . . .	26
<b>Conclusion</b>	<b>27</b>
<b>Bibliography</b>	<b>29</b>
<b>5 Acronyms</b>	<b>31</b>
<b>6 Obsah příloženého CD</b>	<b>33</b>



---

## List of Figures

1.1	5-star data openness rating scheme. [6]	7
1.2	LMI for all data providing scheme. [9]	9
2.1	Graph created using visualization tool at foaf.sk.	16
4.1	Diagram describing retrieving the data from egov.presov.sk and transforming it to initialize database.	22
4.2	Diagram describing web application structure.	26



---

# Introduction

Impact of information is great nowadays. Internet helps us in everyday's life to get the information about almost everything almost immediately and to make decisions. Neither the public affairs did not escape this phenomenon. The society is getting more and more interested in transparency of state government. The idea of Open data was created to satisfy this demand. They can provide the evidence that public money is being well spent and proclaimed policies are being implemented. Publishing state economy data, public procurements, tax declarations, even fortune of politicians or private companies and other data on public web servers is a common practice in many developed countries.

Thanks mainly to European Union and other international institutions, but also enthusiasts and engaged citizens, more and more projects targetting electronisation, systematization and publishing the data are being started also in Slovakia. The idea to publish in machine-readable form all that is not secret or personal data is still young, though, and often meets unwillingness or incompetence of public institutions. As a result, quite often the data published by Slovak state and municipality governments, proclaimed to be open, are useless due to their non-machine-readable formats, incompleteness or lack of interconnectivity. Despite these problems, some progress has been made in this area in recent years and the number of datasets published by state institutions grows every day. Thanks to that there is a growing potential for software developers to create applications for manipulating, interconnecting, analysing and presenting these data and thus turning them into useful information. This is one of the goals of this thesis.

## Open data of Prešov

Looking at datasets published on the official website of Slovak government - [data.gov.sk](http://data.gov.sk) one easily discovers, that the most active of all the cities and towns in Slovakia in publishing the data is Prešov. Not only by number of published datasets, but also by quality of the data this city in eastern Slovakia shares

the most from its governing with its citizens. These data include all past contracts, invoices, orders and grants of the municipality. That is why these data were chosen for the purpose of this bachelor thesis.

### **Thesis Goals**

The main goal of this thesis is to analyse, interconnect and present the open data of Prešov municipality to demonstrate their value and usefulness for the citizens. Focus is on interconnecting them by logical ties not only to each other, but also to other published open data in Slovakia, e.g., Slovak Business Register. This requires analyses of all published open data in Slovakia. The final product of the thesis is a web application to present the analysed and interconnected data. The final goal is to present the application to Prešov city officials and try to convince them to publish and run it on their web servers, so it can serve the public. The thesis also explains the general concept of open data and all the attributes, that such data need to have.



---

# Open data

This chapter is devoted to the problematics of open data. It provides knowledge about what the open data are, what are their main attributes, explains these attributes and informs about open data situation in Europe and some of the important organizations concerned with open data and open government.

## 1.1 Definition

Open data are data that are accessible, can be used, re-used and redistributed by anyone for free (or no more than reproduction cost), subject only, at most, to the requirement to attribute and share-alike<sup>1</sup>. For the sake of interoperability (i.e. ability of diverse systems and organizations to work together) the exact definition of open data is crucial. [1]

## 1.2 Requirements

Requirements that open data need to meet are following: [2]

- **Open license**

The work must be available under an open license. Any additional terms accompanying the work (such as a terms of use, or patents held by the licensor) must not contradict the terms of the license.

- **Access**

The work shall be available as a whole and at no more than a reasonable one-time reproduction cost, preferably downloadable via the Internet without charge. Any additional information necessary

---

<sup>1</sup>Share-alike is a copyright term, describing works that require copies or adaptations of the work to be released under the same or similar licence as the original.

for license compliance (such as names of contributors required for compliance with attribution requirements) must also accompany the work.

- **Open Format**

The work must be provided in a convenient and modifiable form such that there are no unnecessary technological obstacles to the performance of the licensed rights. Specifically, data should be machine-readable, available in bulk<sup>2</sup>, and provided in an open format or, at the very least, can be processed with at least one free/libre/open-source software tool.

To understand the requirements properly, next part is devoted to explaining the most important points mentioned above.

### 1.2.1 Open licenses

Open data need to be available under an open license. A licence tells anyone the conditions, under which they can access, use and share the data. Unless the data have a licence, data may be ‘publicly available’, but users do not have permission to access, use and share it under copyright or database laws. A license is open if it:

- allows free use and redistribution (including sale) of the licensed work.
- allows creation of derivatives of the work and their distribution under the same terms
- does not impose any fee arrangement as part of its conditions
- does not discriminate against any person or group

The next part looks at some of the most common open licenses.

#### 1.2.1.1 Creative commons

Creative Commons (CC) is a non-profit organization devoted to the idea of making the creative works available for others to legally build upon and share. Among other activities it provides free copyright licenses to make a simple and standardized way to give the public permission to share and use creative work on conditions of choice. [3]

---

<sup>2</sup>Data is available in bulk if the whole dataset can be downloaded easily. It is considered non-bulk if the citizens are limited to getting parts of the dataset through an online interface.

**CC Attribution-ShareAlike** lets people remix, tweak, and build upon the licensed work even for commercial purposes, as long as they credit the owner and license their new creations under the identical terms. This is the license used by Wikipedia, and is recommended for materials that would benefit from incorporating content from Wikipedia and similarly licensed projects.

**CC0** is the "no copyright reserved" option - it means giving up all copyright rights that the owner of the work holds and dedicating those rights to the public domain under copyright law, to the extent allowed by law.

**CC Attribution** allows re-distribution and re-use of a licensed work on the condition that the creator is appropriately credited.

### 1.2.1.2 GNU Free Documentation License

The GNU Free Documentation License is a copyleft license - which means that derivative works must be made available under the same or a similar license. The creators of the license recommend that it is used "principally for works whose purpose is instruction or reference."

### 1.2.1.3 Open government licenses

Some governments have chosen to develop their own licenses for the data they publish. The good examples are France <sup>3</sup> and Britain <sup>4</sup>.

## 1.2.2 Data file formats

File format is the digital base in which the information is stored. For the open data to be easily useable, the correct file format is crucial. Choosing the format depends mainly on the character of the data (e.g. structured data for statistics, geodata for geographic data displayed in map, etc.). The formats in which information is published can either be "open" or "closed". If a file format is "closed", this may be either because the file format is proprietary and the specification is not publicly available, or because the file format is proprietary and even though the specification has been made public, re-use is limited. Releasing information in a closed file format can cause significant obstacles to reusing the information encoded in it, forcing those who wish to use the information to buy the necessary software. Open format is a file format with a freely available published specification which places no restrictions upon its use. The preference from the open government data perspective therefore is that information be released in open file formats which are machine-readable.

---

<sup>3</sup>Available at <https://www.etalab.gouv.fr/licence-ouverte-open-licence>

<sup>4</sup>Available at <http://www.nationalarchives.gov.uk/doc/open-government-licence/version/3/>

## 1. OPEN DATA

---

Following are the data formats that are most-widely used for data publishing by government institutions. [5]

### **JSON**

JSON is a simple file format that is very easy for any programming language to read. Its simplicity means that it is generally easier for computers to process than others, such as XML.

### **XML**

XML is a widely used format for data exchange because it gives good opportunities to keep the structure in the data and the way files are built on, and allows developers to write parts of the documentation in with the data without interfering with the reading of them.

### **RDF**

RDF makes it possible to represent data in a form that makes it easier to combine data from multiple sources. RDF data can be stored in XML and JSON, among other serializations. RDF encourages the use of URLs as identifiers, which provides a convenient way to directly interconnect existing open data on the Web. RDF is still not widespread, but it has been a trend among Open Government initiatives, including the British and Spanish Government Linked Open Data projects.

#### **1.2.2.1 CSV, XLS/XLSX**

CSV, or Comma Separated Values can be a very useful format because it is compact and thus suitable to transfer large sets of data with the same structure. It is particularly important for the comma-separated formats that documentation of the individual fields are accurate. Furthermore it is essential that the structure of the file is respected, as a single omission of a field may disturb the reading of all remaining data in the file without any real opportunity to rectify it, because it cannot be determined how the remaining data should be interpreted. Many authorities have information left in the spreadsheet, for example Microsoft Excel. In this case, it is stored in XLS format, which is closed (and thus not suitable for open data), or XLSX format, which is partially open. This data can often be used immediately with the correct descriptions of what the different columns mean.

#### **1.2.2.2 Text formats, PDF, TXT**

Classic documents in formats like Word (closed format), ODF, OOXML, or PDF may be sufficient to show certain kinds of data - for example, relatively

stable mailing lists or equivalent. It may be cheap to exhibit in, as often it is the format the data is born in. The format gives no support to keep the structure consistent, which often means that it is difficult to enter data by automated means. Use of templates as the basis of documents is important to make it possible to pull information out of documents. Plain text documents (.txt) are very easy for computers to read. They generally exclude structural metadata from inside the document however, meaning that developers will need to create a parser that can interpret each document as it appears.

Most data need further processing (to save the data in the database, connect the data to other sources, etc.). As the data can be too big to manually process, manipulating with them should be automatable. The best formats to make automatization of the data more convenient are the structured ones like JSON, XML and RDF. On the contrary, images and text formats like PDF are very difficult to process. Spreadsheet formats (XLS/XLSX, CSV) are structured, but their structure is easily corrupted and manipulating data in these formats is usually less convenient.

### 1.2.3 Data openness levels

Tim Berners-Lee, the inventor of the Web and Linked Data initiator, suggested a 5-star deployment scheme for Open Data. This scheme is widely used (e.g., by British government open data) to rate given data based on their attributes like file format, structure, inteconnectivity, etc.

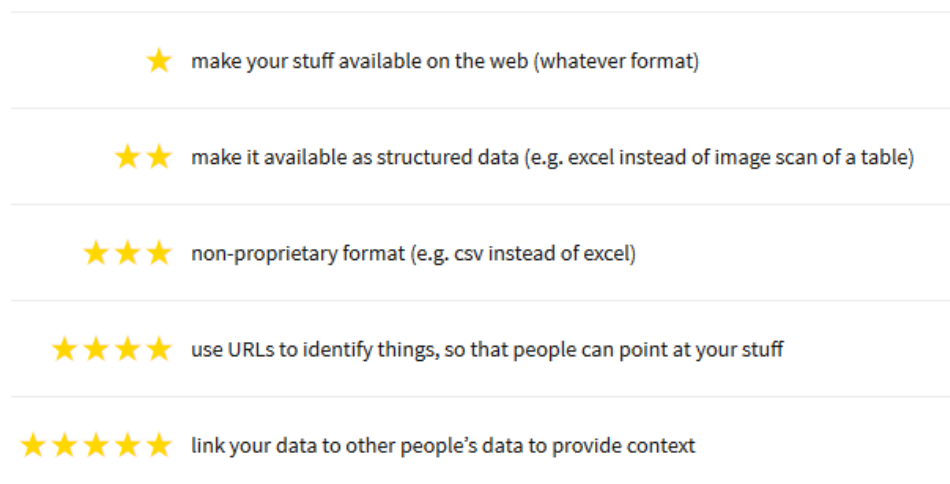


Figure 1.1: 5-star data openness rating scheme. [6]

### 1.3 Open data in the world

**Open data barometer** is a World Wide Web Foundation's global measure of how governments are publishing and using open data for accountability, innovation and social impact. [7]

Barometer ranks governments on:

- readiness for open data initiatives.
- implementation of open data programmes.
- impact that open data have on business, politics and civil society.

The ranking made by Open data Barometer in 2016 is following:

1. Great Britain
2. Canada
3. France
4. USA
5. South Korea
29. Slovakia
31. Czech Republic

#### 1.3.1 Great Britain's open data

Great Britain is the global leader in opening government data. It scored 100/100 in the overall assesment of the prevalence of open data initiatives of Open data barometer. There are several reasons why the UK's open data are ranked so well. One of them is its 'Open Data Readiness', which is the policy framework for the release of open data. It includes having national guidelines that are applicable to specific sectors, and these guiding principles ensure a smooth release of open data that can benefit anyone that wishes to access it.

Government's Digital Service Standard provides guidelines on data publishing. A number of regional or city initiatives are integrated into the data.gov.uk portal, too, such as Open Manchester, Leeds Data Mill, Birmingham Open Data, Bristol City Council Open Data, to mention just a few.

The UK Government Licensing Framework provides a policy and legal framework for licensing the use and re-use of public sector information across the public sector. It sets out best practice, standardises the licensing principles for government information, mandates the Open Government Licence (OGL) as the default licence for Crown bodies and recommends OGL for other public sector bodies (see 1.2.1.3).

Following in the UK's footsteps in open data guidelines would be a good place to start for other countries hoping to improve in this area. [8]

One of the interesting applications of UK's open data is the **LMI For All** project. It is an online data portal, developed by the UK Commission for Employment and Skills, which brings together existing national sources of high quality labour market information (LMI) that can inform and help people make right choices about their careers. The online portal includes LMI that can help answering the questions people ask when thinking about their careers, like how much people earn or what type of persons do what kind of jobs, etc. It includes information from the Annual Survey of Hours and Earnings, the Labour Force Survey, the Employer Skills Survey and Working Futures. All of its data are made freely available via an Application Programming Interface (API) for use in websites and applications. Other interesting applications built using UK's open data can be seen at [data.gov.uk/apps](http://data.gov.uk/apps). These include Car Insurance Comparison Application, All GB Railway Stations, Inheritance Tax Calculator UK and more.



Figure 1.2: LMI for all data providing scheme. [9]

### 1.3.2 Open Government Partnership

Open Government Partnership (OGP) is the international governments initiative, which was launched in 2011 by 8 founding governments (Brasil, Indonesia, Mexico, Norway, Philippines, South Africa, United Kingdom and United States). The focus of the initiative is the improvement of government by promoting transparency, effectiveness and responsibility, harnessing new technologies and setting up institutions, which help citizens to participate in governing. To become a member of OGP, participating countries must commit to uphold the principles of open and transparent government by endorsing the Open Government Declaration. In total there are over 70 OGP participating countries and 15 subnational governments and together they have made over 2,500 commitments to make their governments more open and accountable. [17]

### 1.3.3 European Union initiatives

**European Union Open Data Portal**, set up in 2012, is the open data portal of the European Union containing datasets that are collected and published by the European Institutions and other EU bodies. It provides a standardized catalogue of EU open data and a list of applications and web tools reusing these data. [10]

**European Data Portal** is an initiative of the European Commission, and is part of the Digital Single Market. It was created to gather Public Sector Information of the 28 European Member States and the four EFTA (Norway, Iceland, Switzerland and Liechtenstein) countries. Information regarding the provision of data and the benefits of re-using data is also included. Going beyond the harvesting of metadata, the strategic objective of the European Data Portal is to improve accessibility and increase the value of Open Data. Within the Portal, sections are dedicated to searching datasets, providing data, using data, training and library. It also contains ratings of open data of every involved country. [11]



---

## Open data in Slovakia

Slovakia ranked 29th in the above mentioned Open data barometer for 2016. It moved 7 places up the ranking since 2015. European Data portal publishes factsheets, which summarize readinnes, challenges and shortcomings of open data in each member country. According to 2017 factsheet dedicated to Slovakia, main possitives of Slovak open data are: presence of action plan of Open Government for 2017-2019; presence of national guidelines on data publication; data licensing 100% free of charge and big amount of regional data initiatives. On the contrary, the main negatives/barriers are: less then 25% of the data uploaded automatically; lack of integrated regional portals; passive approach of Open Data holders, mainly by municipalities and lack of promotion of prescribed technical standards. [12]

### 2.1 Political support

A lot of positive steps were made towards open government in Slovakia during the short government of Iveta Radičová (2010-2012). Most notable ones were making publishing of contracts on the internet mandatory (see 2.3.1) and introduction of electronical auction for public procurements. On 20th September 2011, the Government of the Slovak Republic joined the Open Government Partnership and in 2012 approved the first action plan for the Open Government Initiative for years 2012-2013, which created an open data portal. In June 2016, a new office – Deputy Prime Minister’s office for Investments and Informatization was set up, with one of the main tasks being creation and implementation of uniform state policy in the field of informatization of the society. In March 2017, Open Government Partnership National Action Plan for years 2017-2019 was adopted by the Slovak government. This plan addresses open data (by formulating new commitments on publicly available APIs and open source software), open education, public policy making and open justice.

### 2.2 Slovak open data portal - data.gov.sk

Official government open data portal - data.gov.sk was launched in 2013 as part of the Open Government Initiative, which aims to improve governance through increasing transparency, efficiency and accountability. Declared goals of the portal are publishing data and metadata (data describing data like column descriptors, etc.) in machine-readable form using open standards and licenses. The website allows users to search datasets by organization that published them, license, file formats and tags. It also supports full-text search by name and description of dataset and even by their location on the map. Following are some statistics from the portal (March 2018):

- 4399 unique visitors on average per month
- 30% of the visitors foreign
- most datasets available in CSV format
- 99% of the datasets machine-readable
- 66 organizations publishing data
- 1425 datasets available

The most active organizations by number of published datasets:

Organization	Number of datasets
Slovak Statistical Office	625
Public Procurement Office	64
The city of Prešov	58
Ministry of Interior of the Slovak republic	57
Central Agricultural Control and Testing Institute	43

The most commonly used file formats (many datasets are published in more than one file format):

File format	Number of datasets
CSV	877
XML	336
HTML	218
XLSX	185
JSON	11

The most commonly used licenses (see 1.2.1):

License	Number of datasets
Creative Commons Attribution Share-Alike	737
Creative Commons CCZero	87
Creative Commons Attribution	24
Other non-commercial	14
GNU Free documentation license	7

The most common tags:

Tag	Number of datasets
Statistics	607
Key tables	190
Elections	177
Key Indicators	98
Parliament	63

## 2.3 Other state projects

Beside above-described official Slovak open data portal, several other big state-run projects are present. The most important ones are Central Register of Contracts and Business register.

### 2.3.1 Central Register of Contracts

Started on 1st January 2011, Central Register of Contracts (available at [crz.gov.sk](http://crz.gov.sk)) is a website containing contracts (with their appendices and attachments) concluded by so-called obliged entities - Government Office, ministries, central government authorities, public bodies and their subordinate organizations (contributors, budgetary organizations, etc.) since the beginning of 2011. As of March 2018, the database contains 1 578 659 contracts. All contracts concluded prior to 2011 are available at older address [zmluvy.gov.sk](http://zmluvy.gov.sk). Operator of both websites is the Government Office of the Slovak Republic. The register has proven to be successful. The procurement costs of public orders fell by almost one third in the very first year of the register's operation due to reduced financial mismanagement. Iveta Radičová, former Slovak Prime Minister during whose government the project was started, described its success by these words:

“Following the first year of the register’s operation, savings in public finances amounted to an average of 30 percent of costs. There were also some areas – for example, the transport sector – where the savings reached 50 percent. The Office of the Government

alone saved EUR 3.5 million in the first year of the register's operation. The publication of a contract also prevented an overpriced purchase of flowers for EUR 10 000 at the Ministry of Education." [15]

### 2.4 Business register

The Business Register (available at [orsr.sk](http://orsr.sk)) is a public registry that contains particular information specified by the law about individual entrepreneurs, companies and other legal entities. The registry is owned and maintained by the Ministry of Justice of Slovakia. It can be searched by business name, identification number, registered seat, registration number or name of a person. The register provides information from January 2001 onwards.

### 2.5 Non-state initiatives

In addition to official projects of Slovak republic and municipalities, there are several civic associations, non-profit organizations and private companies dedicated to state open data. Some of them are trying to improve the quality and usefulness of the available data, others interconnect them with other sources and present them in an attractive form. Most often their main goal is to bring more transparency to public spending and state government. Next section looks at the most important ones.

#### 2.5.1 Slovensko.digital

Slovensko.digital is a civic association aimed at improving the quality of state digital services in Slovakia. Their goals are summarized on their website:

"We want the digital services of the state to be simple, meaningful and have a normal price. We need to increase the transparency and efficiency of spending of public resources as well as the participation of the professional public in the electronisation of public administration."

In accordance with these goals, slovensko.digital set up the platform for all IT specialists, who want to join the discussion about state IT projects. This civic association's members also developed a methodology that focuses on more than 50 criteria of project success and use it to systematically rate state IT projects and alert the public to the risky ones in one of their project called Red Flags. Thanks to this, slovensko.digital was able to stop a couple of senseless state projects, saving Slovak citizens hundreds of millions of euros. They've also started several projects as part of [ekosystem.slovensko.digital](http://ekosystem.slovensko.digital), some of which provide APIs to public data sources (otherwise only accessible

on websites) like Register of legal entities, Central register of contracts, Public procurement journal, Business journal, Registry of financial statements and more.

### 2.5.2 Finstat

Finstat was started in 2013 as a web portal, which helps people to freely assess financial health of Slovak companies. It connects data from various data sources: ministries, state insurance companies, Business register, Statistical office, Slovak Chamber of executors, Trade Journal, Register of entrepreneurs, Register of financial statements, Register of bankrupts, Czech statistical office and many more and by their processing and analysing provides complex information about Slovak companies, individual sectors and groups of entrepreneurs. The portal also helps media and non-profit organizations to analyse companies and investigating corruption cases by development of analyses and making the data of the portal accessible for them. A lot of its features are for free and some features, like APIs access and datasets in CSV are parts of paid packages.

### 2.5.3 Uvostat

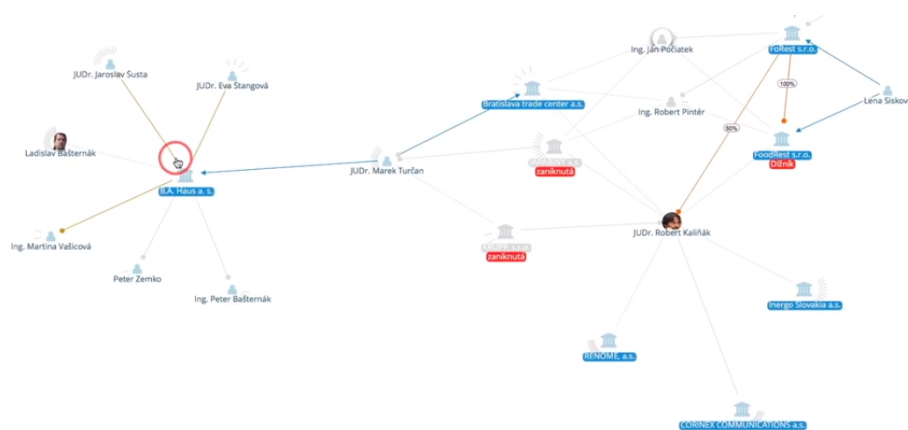
Uvostat is a project that processes public procurements data, analyses and presents them in the form of statistics and interactive graphs. The statistics include most valuable procurements, most procurements by company or by contracts concluded. The website enables searching in the list of suppliers and even fulltext search in names and contents of the procurements. It also provides noticator for tenders for new positions in state offices, which can be used by people looking for jobs to inform them, that a new position which fits their profession and location is being opened. The interesting fact about the project is that it has been developed by only two people.

### 2.5.4 Otvorené zmluvy (Open contracts)

Otvorené zmluvy is an investigative project of Slovak anti-corruption initiative Aliance Fair-play and Transparency International, which helps citizens to read, search and assess contracts concluded by state and state institutions, which are being published by law since 2011. It offers automated analyses and offers general public to join discussions (crowdsourcing), trying to draw attention to suspicious contracts.

### 2.5.5 Datanest

Datanest was also created by Aliance Fair-play as a source of information about management of Slovak public money as part of their anti-corruption initiatives. It serves mainly journalists, analytics, watchdog organizations,



## Analyses of the Prešov open data

The city of Prešov is the only municipality in Slovakia, which is in top 10 organizations by number of published datasets on data.gov.sk (see 2.2). Prešov is also the only municipality in Slovakia, that provides open geodata. At ITAPA<sup>5</sup> 2017 awards, the city's webgis geodata<sup>6</sup> project was awarded 3rd place. Mariana Hurná, Prešov's main IT coordinator was awarded as Best IT specialist of Slovak municipalities in years 2008, 2009 and 2012. The city maintains egov.presov.sk information system portal to enable its citizens access to useful information about city data. Published here are data about street names, members of municipal government, members of municipal police, camera systems locations, demographic structure of citizens of the city, drawing of the municipality budget, register of petitions, tax debtors and more. Prešov is by all means a very good example of municipality that works towards greater transparency by processing the data it collects from the citizens. Unlike most other Slovak municipalities, its employees don't only do the necessary minimum to comply with the law, but also do the additional work to provide useful data applications and progress to its citizens.

### 3.1 Prešov contracts, grants, invoices and orders

Slovak municipalities are obliged to publish data about disclosed orders, contracts and invoices of institutions like City Office, Municipal Police, Registry office and all budgetary organizations of which they are founders. This is mandatory by law 211/2000<sup>7</sup> in Slovak collection of laws. Additionally, Prešov municipality publishes all applications for grants to comply with its internal

---

<sup>5</sup>ITAPA events (Information Technologies and Public Administration) are the most important and well-known IT events in Slovakia since 2002. [18]

<sup>6</sup>data about geographic locations, that can be displayed in a map

<sup>7</sup>Available at <http://www.zakonypreludi.sk/zz/2000-211>

### 3. ANALYSES OF THE PREŠOV OPEN DATA

---

binding regulations. Prešov publishes this data on its website [egov.presov.sk](http://egov.presov.sk), as well as on national open data portal [data.gov.sk](http://data.gov.sk), in various formats (XML, JSON, XLSX, and PDF). Next sections look at the data that are used in practical part of these bachelor thesis - contracts, invoices, grants and orders. They provide basic statistics of each of these datasets and describes their attributes. All the statistics were collected from [egov.sk](http://egov.sk) on April 8th 2018.

#### 3.1.1 Contracts

Total published contracts	6027
Oldest contract published on	August 2nd 2011
Latest contract published on	April 6th 2011

Attribute	Data type
id (central number)	Integer
type	String
contracting parties	String
subject	String
total price	Integer
currency	String
date of signature	Date
date of release	Date
name (detailed type)	String
parties' national identifiers	String
parties' addresses	String
notes	String
main/appendix indicator	String
effective date	Date

#### 3.1.2 Invoices

Total published invoices	28511
Oldest invoice published on	May 26nd 2011
Latest invoice published on	April 6th 2018



### 3.1. Prešov contracts, grants, invoices and orders

Attribute	Data type
id (internal number)	Integer
name of the organization	String
supplier	String
payment purpose description	String
total price	Integer
currency	String
id of contract	Integer
id of order	Integer
release date	Date
supplier national identifier	String
supplier address	String
payout status	String
office	String
invoice number	Integer
final/deposit indicator	String
status	String

#### 3.1.3 Orders

Total published orders	28511
Oldest orders published on	May 26nd 2011
Latest orders published on	April 6th 2018

Attribute	Data type
id	Integer
supplier	String
order text	String
total value	Integer
currency	String
creation date	Date
release date	Date
supplier national identifier	Integer
supplier address	String
office	String
name of approver	String
function of approver	String
date of issue	Date
release notes	String
status	String
contract id	Integer
year of contract	Integer

### 3. ANALYSES OF THE PREŠOV OPEN DATA

---

#### 3.1.4 Grant applications

Total published grant applications	825
Oldest grant app. published in	2012
Latest grant app. published in	2018

Attribute	Data type
year of dotation	Integer
type	String
status	String
event	String
approved amount	Integer
currency	String
applicant	String
applicant's address	Integer
applicant's city	String
applicant's national identifier	Integer

---

# Implementation

As mentioned in previous chapters, the main output of this bachelor thesis is a website built on top of economic open data of Prešov. This website not only presents these data and enables searching them, but also provides added value by connecting them to other national open data sources. Next sections look at how this web application works, how it manages and presents the data and what technologies it uses.

## 4.1 Data management

### 4.1.1 Data from Prešov web server

As described in chapter 3, the main concern of this web application are data about Prešov's contracts, invoices, orders and applications for grants. These data are being updated daily on the internet and can be downloaded in various formats, including XML and JSON (1.2.2). JSON format is more modern alternative to XML and is easier and faster to parse, less memory-consuming and more intuitive to read and that's why the data from Prešov's web servers were downloaded in this format for the purpose of this application. To enable fast searching and accessing these data, they need to be stored in the database. *PostgreSQL* was chosen as the database management system. To fill the database with the JSON data, the conversion from JSON to SQL statements is needed. In this case, this is done by script written in *Python* programming language. This script takes JSON file as an input, along with the parameter indicating what kind of data are being inserted ("i" for invoices, "c" for contracts, etc.), parses the JSON file, transforms each JSON object into SQL INSERT statement and outputs the statement. The result is a series of SQL statements which can be stored as a SQL script and used to fill 4 database tables - Contracts, Invoices, Grants and Orders. The process of getting the data from the web server and transforming it to fill the database is described in the diagram 4.1.

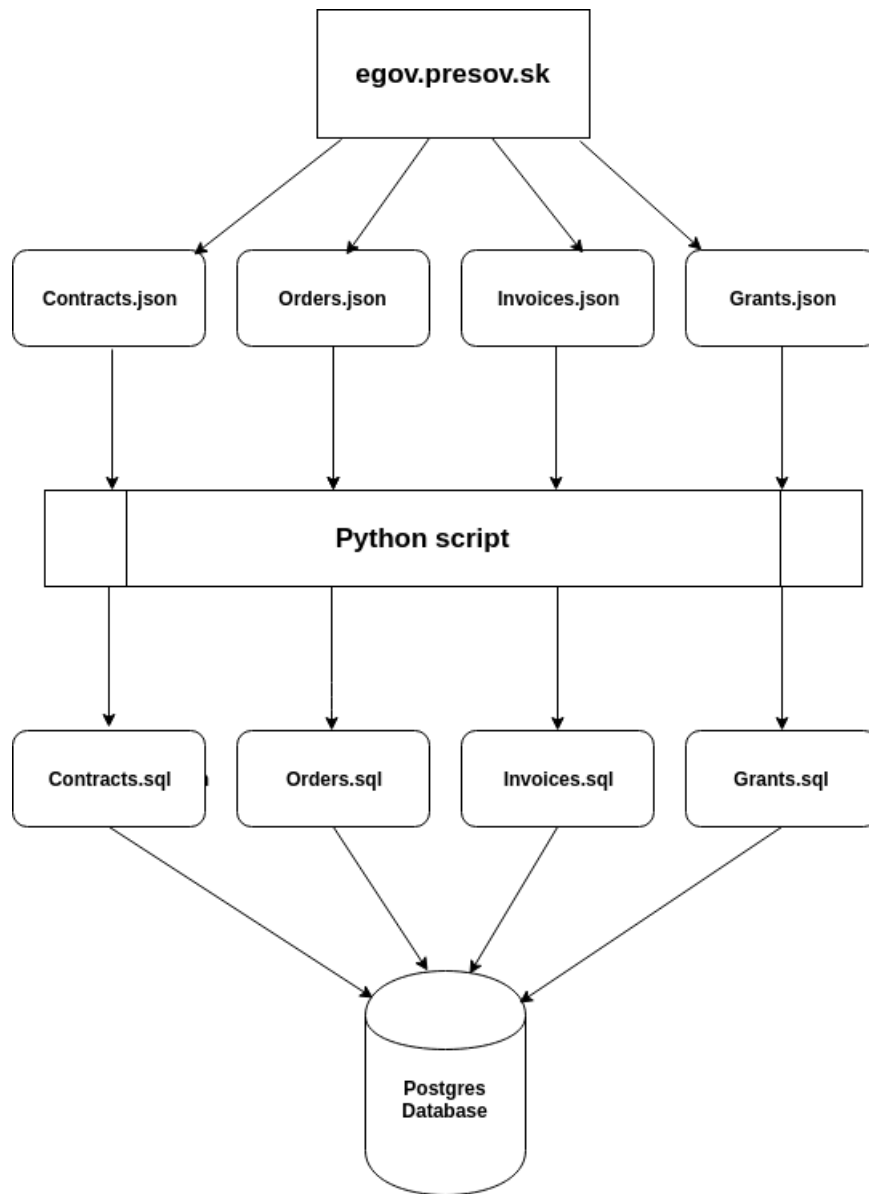


Figure 4.1: Diagram describing retrieving the data from `egov.presov.sk` and transforming it to initialize database.

### 4.1.2 Data provided by `slovensko.digital`

The data retrieved from Prešov's website contain national identifiers (IČO - organization identification number) of contract parties, grant applicants and suppliers (see 3.1). Through these identifiers, a given legal entity can be searched in Slovak Business Register (see 2.4). The problem of getting the data from this register, which does not provide any API was targeted and

solved by slovensko.digital by providing free API to this register as part of their ekosystem.slovensko.digital project (see 2.5.1). Through this API it is possible to retrieve information about legal entity with given IČO in JSON format.

#### 4.1.2.1 Data retrievable by IČO

To be able to use the above-mentioned API to retrieve the full data about legal entity by its IČO, the user needs to use an intermediary API, which provide consolidated basic data about the legal entity and also serves as a gateway to other services which provide more detailed information. To use this premium intermediary API, an *Authorization token* is required, which needs to be included in the API's URL request header. This token is provided by slovensko.digital upon email request, in which one should describe the purpose of the data usage.<sup>8</sup> The result of successful request contains national and tax (vatin) identifiers, name of the subject, address attributes, dates of establishment and termination, registration office attributes, two system columns describing dates of creation and modification of the entry and finally the URLs to APIs providing more detailed information - full data from Business Register and Register of Accounting Entities. Following is the example of JSON data returned by this API for legal entity with IČO 36421928, which is slovensko.digital:

```
{
  "id": 1358414,
  "cin": 50158635,
  "tin": 2120264674,
  "vatin": null,
  "name": "Slovensko. Digital",
  "formatted_address": "Staré Grunty 6207/12, 841 04 Bratislava –  
mestská časť Karlova Ves",
  "street": "Staré Grunty",
  "reg_number": 6207,
  "building_number": "12",
  "postal_code": "841 04",
  "municipality": "Bratislava – mestská časť Karlova Ves",
  "country": "Slovenská republika",
  "created_at": "2016-06-09T12:26:10.521084Z",
  "updated_at": "2017-06-30T02:10:45.967257Z",
  "established_on": "2016-01-29",
  "terminated_on": null,
  "vatin_paragraph": null,
  "registration_office": "MV SR",
  "registration_number": "VVS/1-900/90-48099",
  "formatted_street": "Staré Grunty 6207/12",
  "street_number": "6207/12",
  "rpo_organizations": [
    {
      "id": 6562824,
```

---

<sup>8</sup> The complete guideline to using the API and retrieving the authorization token is available here: <https://ekosystem.slovensko.digital/premiovie-api#access-token>

## 4. IMPLEMENTATION

---

```
        "url": "https://datahub.ekosystem.slovensko.digital/api  
              /data/rpo/organizations/6562824"  
      },  
    ],  
    "ruz_accounting_entities": [  
      {  
        "id": 1617045,  
        "url": "https://datahub.ekosystem.slovensko.digital/api  
              /data/ruz/accounting_entities/1617045"  
      }  
    ]  
  }  
}
```

### 4.1.2.2 Full Business Register data

Using the URL contained in the above-mentioned data under *rpo\_organizations* attribute, one can obtain full data available for the given entity in the Business Register. These include: list of addresses, legal forms, organization names, alternate names, predecessors, successors, registration numbers and statutes (with name, address and statutory type). Included are also entity's economic data like economic activity types, money deposits put into the organization (with name of the depositor and the amount), company equities (investment, paid and approved amounts), shares (with identifiers of shareholder and share price) and stakeholders (name, address and type). All of these entries contain dates of effectiveness from and to, which can be used to construct the history of the legal entity.

### 4.1.2.3 Register of Accounting Entities data

The URL under *ruz\_accounting\_entities* attribute points to an API which provides access to data collected from Register of Accounting Entities. This API, however, is only in alpha version and thus only provides limited information like the size of an organization (by number of employees), ownership type, economic activity, municipality and district.

## 4.2 Web application

### 4.2.1 Used technologies

The web application was implemented using *Java Platform, Enterprise Edition* (Java EE) technology. *Java Servlet Technology* was used together with *JavaServer Pages* (JSP) to generate web pages based on user request. The application was deployed to *WildFly* application server. *Apache Maven* was used for building the application and dependency management, and *Hibernate ORM* for persistence layer.

#### **4.2.1.1 Java EE**

Java Platform, Enterprise Edition is a collection of Java APIs that software developers can use to write server-side applications. Its architecture provides services that simplify some common challenges faced by developers building web applications like client requests handling or working with database.

#### **4.2.1.2 Java Servlet Technology and JSP**

Java Servlet Technology was created as a way to provide dynamic, user-oriented web content. A servlet is a Java class used to respond to various, most often HTTP requests. JavaServer Pages technology is also used to create web content based on user request. Unlike Java servlet, JaveServer page is not a Java class, but a HTML-like document supporting Java scriptlets, which are pieces of Java-code.

#### **4.2.1.3 Hibernate**

Hibernate ORM (Object-relational mapping) is a free Java tool, which provides a framework for mapping an object-oriented domain model to a relational database. Its primary feature is mapping from Java classes to database tables and Java data types to SQL data types. Hibernate also supports data querying by generating SQL calls, relieving the developer from manual object conversion.

#### **4.2.1.4 WildFly**

WildFly, formerly known as JBoss, is an free open-source application server implementing Java EE specification. Being implemented in Java, it can run on multiple platforms.

#### **4.2.1.5 Maven**

Maven is a build automation tool used mostly in Java projects. It addresses two important parts of software development - describing, how software is build and its dependencies. Is is configured by XML file called POM (Project Object Model). Maven comes with pre-defined targets for performing tasks as code compilation and packaging. It dynamicly downloads libraries and plugins from either its own central repository or ones defined in POM file.

### **4.2.2 Architecture**

The architecture of the web application is a classic Java Servlet/JSP architecture. The HTTP request is handled by the application server, which finds the correct Java servlet. The servlet then uses application Java classes to do the business logic like retrieving and processing the data from the database or

## 4. IMPLEMENTATION

---

through slovensko.digital's API (this includes parsing the JSON data). After doing the necessary operations the servlet forwards the request to appropriate JSP document, which handles the UI part. This architecture is described in the diagram 4.2.

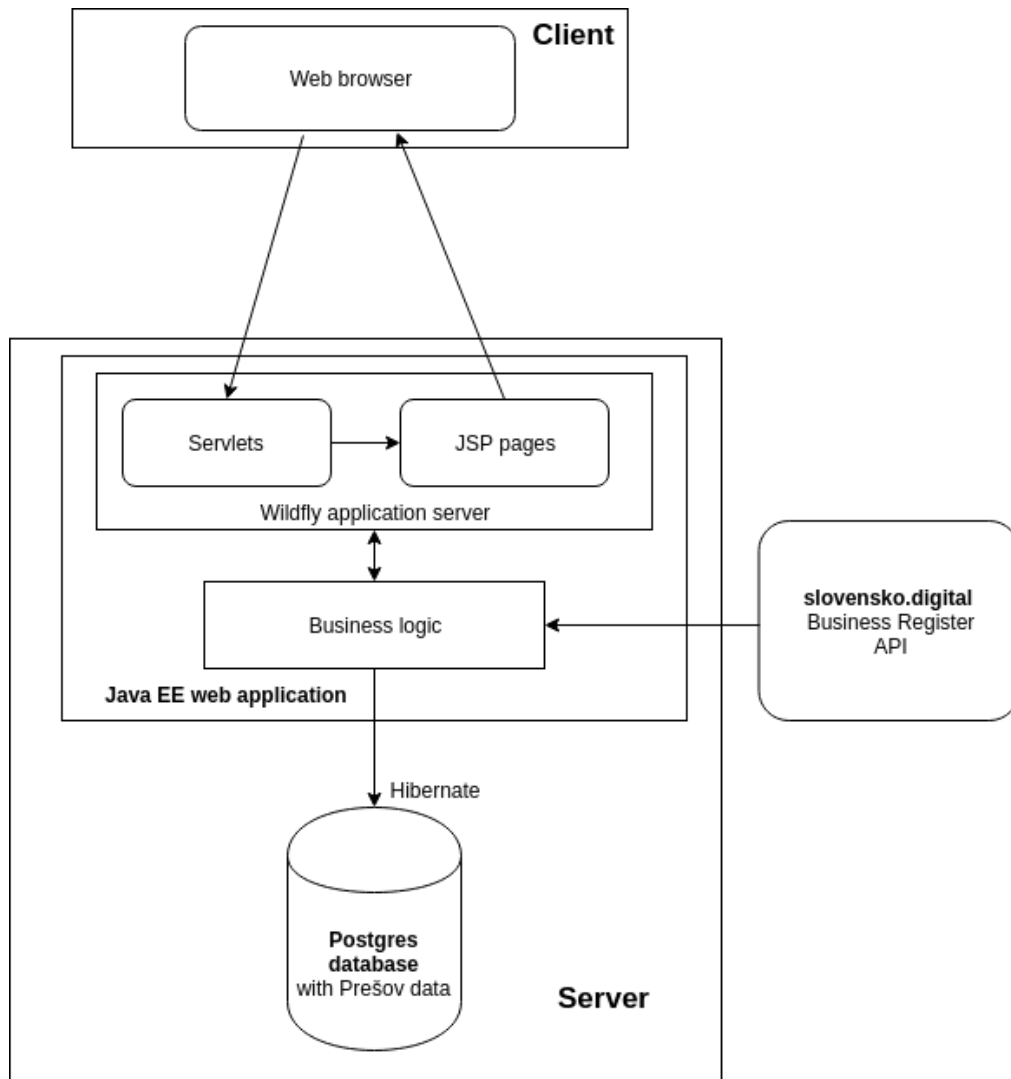


Figure 4.2: Diagram describing web application structure.

### 4.3 Deployment



---

## Conclusion



---

# Bibliography

- [1] *What is Open Data?* [online]. [cit. 2018-02-28]. Available from: <http://opendatahandbook.org/guide/en/what-is-open-data/>
- [2] Open Definition. *Open Definition 2.0* [online]. [cit. 2018-03-02]. Available from: <http://opendefinition.org/od/2.0/en/>
- [3] *Creative Commons licenses* [online]. [cit. 2018-19-04]. Available from: <https://creativecommons.org/licenses/>
- [4] European Data Portal. *What is open data?* [online]. [cit. 2018-03-02]. Available from: <https://www.europeandataportal.eu/elearning/en/module1>
- [5] *File formats* [online]. [cit. 2018-03-02]. Available from: <http://opendatahandbook.org/guide/en/appendices/file-formats/>
- [6] *5 star data* [online]. [cit. 2018-23-03]. Available from: <http://5stardata.info/en/>
- [7] *Open data barometer* [online]. [cit. 2018-25-03]. Available from: <https://opendatabarometer.org/>
- [8] *Government Computing* [online]. [cit. 2018-25-03]. Available from: <http://central-government.governmentcomputing.com/features/who-is-leading-open-data-in-europe-walking-the-open-data-talk-5653451>
- [9] *LMI for all* [online]. [cit. 2018-25-03]. Available from: <http://www.lmiforall.org.uk/about-lmi-for-all/>
- [10] *European Union Open Data Portal* [online]. [cit. 2018-13-04]. Available from: <http://data.europa.eu/euodp/en/home/>

## BIBLIOGRAPHY

---

- [11] *European Data Portal* [online]. [cit. 2018-25-03]. Available from: <https://www.europeandataportal.eu/en/what-we-do/our-activities>
- [12] *European Data Portal Factsheet* [online]. [cit. 2018-26-03]. Available from: [https://www.europeandataportal.eu/sites/default/files/country-factsheet\\_slovakia\\_2017.pdf](https://www.europeandataportal.eu/sites/default/files/country-factsheet_slovakia_2017.pdf)
- [13] *Action plan of Slovak Open data 2017-2019* [online]. [cit. 2018-26-03]. Available from: <http://www.rokovania.sk/Rokovanie.aspx/BodRokovaniaDetail?idMaterial=26262>
- [14] *Central Registry of Contracts* [online]. [cit. 2018-30-03]. Available from: <http://www.vlada.gov.sk/od-1-januara-2011-zacal-fungovat-centralny-register-zmluv/>
- [15] *Joinup - Slovakian online central register contracts - crs* [online]. [cit. 2018-30-03]. Available from: <https://joinup.ec.europa.eu/document/slovakian-online-central-register-contracts-crs>
- [16] *Otvorené zmluvy* [online]. [cit. 2018-04-04]. Available from: <http://www.otvorenezmluvy.sk/o-projekte>
- [17] *Open Government Partnership* [online]. [cit. 2018-06-04]. Available from: <https://www.opengovpartnership.org/about/about-ogp>
- [18] *ITAPA events* [online]. [cit. 2018-19-04]. Available from: <http://www.itapa.sk/o-itapa-podujatiach/>

## Acronyms

<b>IT</b>	Information Technology
<b>XML</b>	Extensible markup language
<b>JSON</b>	JavaScript Object Notation
<b>RDF</b>	Resource Description Framework
<b>CSV</b>	Comma Separated Values
<b>PDF</b>	Portable Document Format
<b>LMI</b>	Labour market information
<b>URL</b>	Uniform Resource Locator
<b>Java EE</b>	Java Platform, Enterprise Edition
<b>JSP</b>	JavaServer Pages
<b>HTTP</b>	Hypertext Transfer Protocol
<b>UI</b>	User Interface



## Obsah přiloženého CD

	readme.txt .....	stručný popis obsahu CD
	exe .....	adresář se spustitelnou formou implementace
	src	
	impl.....	zdrojové kódy implementace
	thesis .....	zdrojová forma práce ve formátu L <sup>A</sup> T <sub>E</sub> X
	text .....	text práce
	thesis.pdf .....	text práce ve formátu PDF
	thesis.ps .....	text práce ve formátu PS