

Open Data and Data Analysis

Big Data and Artificial Intelligence
Approach to Open Data

Matej Kovačič

matej.kovacic@ijs.si

Jožef Stefan Institute

Centre for Knowledge Transfer
in Information Technologies

Artificial Intelligence Laboratory

Open data

Open data are data that are freely available to everyone to use and republish as they wish, without restrictions from copyright, patents or other mechanisms of control.

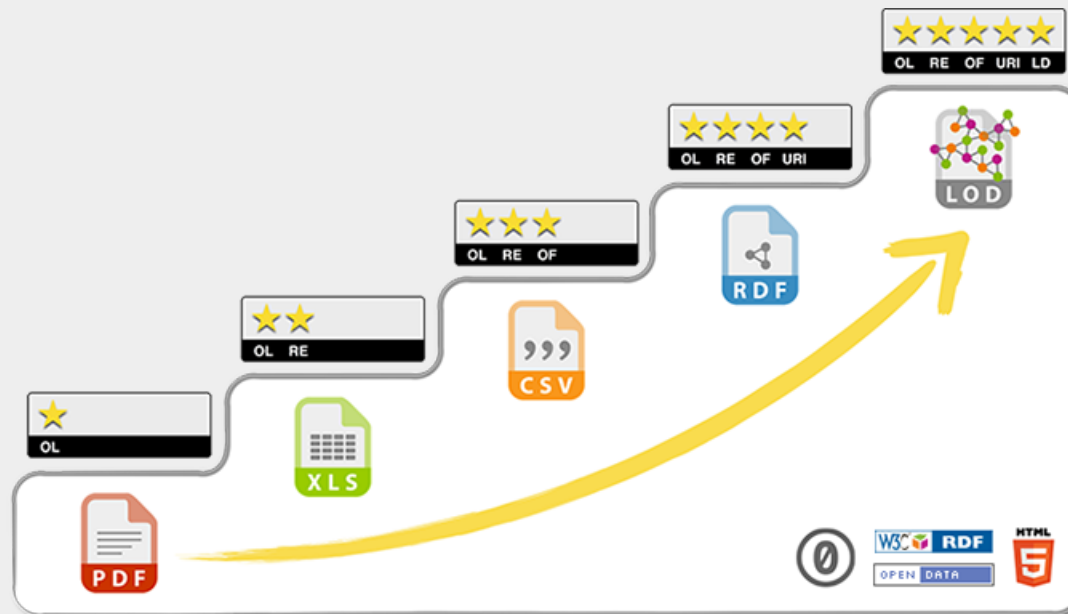
Elements of open data:

- Availability and access: available as a whole, free or for reasonable reproduction costs, preferably downloadable from internet, available in a convenient and modifiable form.
- Re-use and redistribution: terms of use permit re-use and redistribution.
- Universal participation: for anyone for any purpose (no copyright, patents, other control mechanisms)

Formally, open data has a licence saying it is open data. Sometimes licence requires attribution and share-alike.

Open data

Tim Berners-Lee, the inventor of the Web and Linked Data initiator, suggested a 5-star deployment scheme for Open Data.



Source: <http://5stardata.info/>

1. Make your data available on the Web (in whatever format) under an open license.
2. Make it available as structured data (e.g., Excel instead of image scan of a table).
3. Make it available in a non-proprietary open format (e.g., CSV instead of Excel).
4. Use URIs to denote things, so that people can point at your data.
5. Link your data to other data to provide context.

Public information

Public information is all information originating from the field of work of the public sector bodies and occurring in the form of a document, a case, a dossier, a register, a record or other documentary material (hereinafter referred to as “the document”) drawn up by the body, by the body in cooperation with other body, or acquired from other persons.

Natural and legal persons can under certain conditions use this information also for commercial or non commercial reuse.

Some sources of open data

- Open data in science:
https://en.wikipedia.org/wiki/Open_data#In_science
- Open data in government:
https://en.wikipedia.org/wiki/Open_data#In_Government
- EU Open data portals:
<https://ec.europa.eu/digital-single-market/en/open-data-portals>
- Open data by American Government:
<https://www.data.gov/>
- Slovenian National Interoperability Framework Portal:
<http://nio.gov.si/nio/vstopna.nio?lang=en>
- Slovenian community Opendata.si: <https://opendata.si/>

Big Data

Big data is data sets that are so large or complex that traditional data processing application software is inadequate to deal with them.

Big data challenges include capturing data, data storage, data analysis, search, sharing, transfer, visualization, querying, updating and information privacy.

Lately, the term "big data" tends to refer to the use of predictive analytics, user behavior analytics, or certain other advanced data analytics methods that extract value from data, and seldom to a particular size of data set.

Example of Big Data analysis: <http://eventregistry.org/>

Big Data

Big data are usually raw, “found” or “organic data”. Usually in real time.

Machine (for example from sensors) or user generated.

In social sciences: data about behavior in the natural environment and in real time.

Problems with big data for research purposes:

- Secondary data, intended for another primary use;
- Sometimes not structured or they are not well defined;
- Data quality? Cheap to collect, but expensive to clean and process;
- Ethical concerns: no informed consent policies, privacy issues.

Some examples of open data and its
analysis

Slovenian business registry

Ajpes (The Agency of the Republic of Slovenia for Public Legal Records and Related Services) provided:

- ePRS Slovenian Business Register (central public database on all business and legal entities, their subsidiaries, and other organizations' segments located in Slovenia (public and private institutes)), including deleted companies;
- eRTR - Register of legal entity's bank accounts.

The screenshot displays the ePRS Slovenian Business Register website. The browser address bar shows www.ajpes.si/?language=english. The page features a navigation menu with links for **START-UPS**, **REGISTERS**, and **CREDIT RATING**. The main heading is **AJPES - an indispensable primary source of business information**. Below this, there is a descriptive paragraph about the agency's role and services. A search form titled **ENTER SEARCH CRITERIA** is visible, with the instruction: "Data can be accessed by entering the search criteria in one or more fields, or by selecting the search criterion from the drop-down menu. The search criterion can be any of the following: data, part of the data (word), beginning of the word." The search form includes fields for: Identification number, Tax Number, Name, Street name and number, City, Municipality, Post office code, Main activity, Institutional sector, Legal organizational form, and Rank. The Rank dropdown menu is currently set to "active company". Below the search form, there is a "Search" button, an "Advanced search" button, and a "Clear form" button. The page also includes a "We recommend" section and a footer with the ePRS logo and additional information about the register.

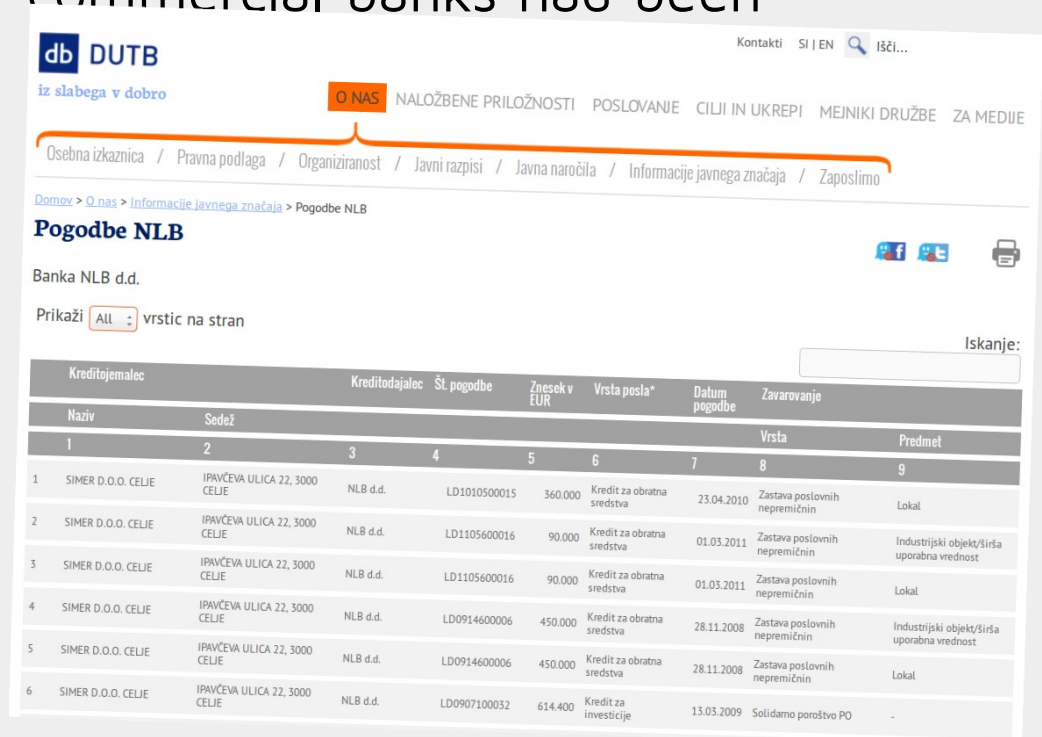
Some other basic registers...

- *Registry of budget users*: list and classification of all direct and indirect Slovenian budget users.
- *Standard Classification of Activities*: the obligatory national standard used for recording, collecting, processing, analysing, mediating and disseminating data connected with activity).
- *CPV (Common procurement vocabulary)*: CPV establishes a single classification system for public procurement aimed at standardising the references used by contracting authorities and entities to describe the subject of procurement contracts.
- ...

Bad investments transferred on *Bank Assets Management Company*

The Bank Assets Management Company (BAMC) was established in March 2013 as a government-owned company with the task of facilitating the restructuring of banks with systemic importance that were facing severe solvency and liquidity problems.

Non-performing assets of commercial banks had been transferred to BAMC.

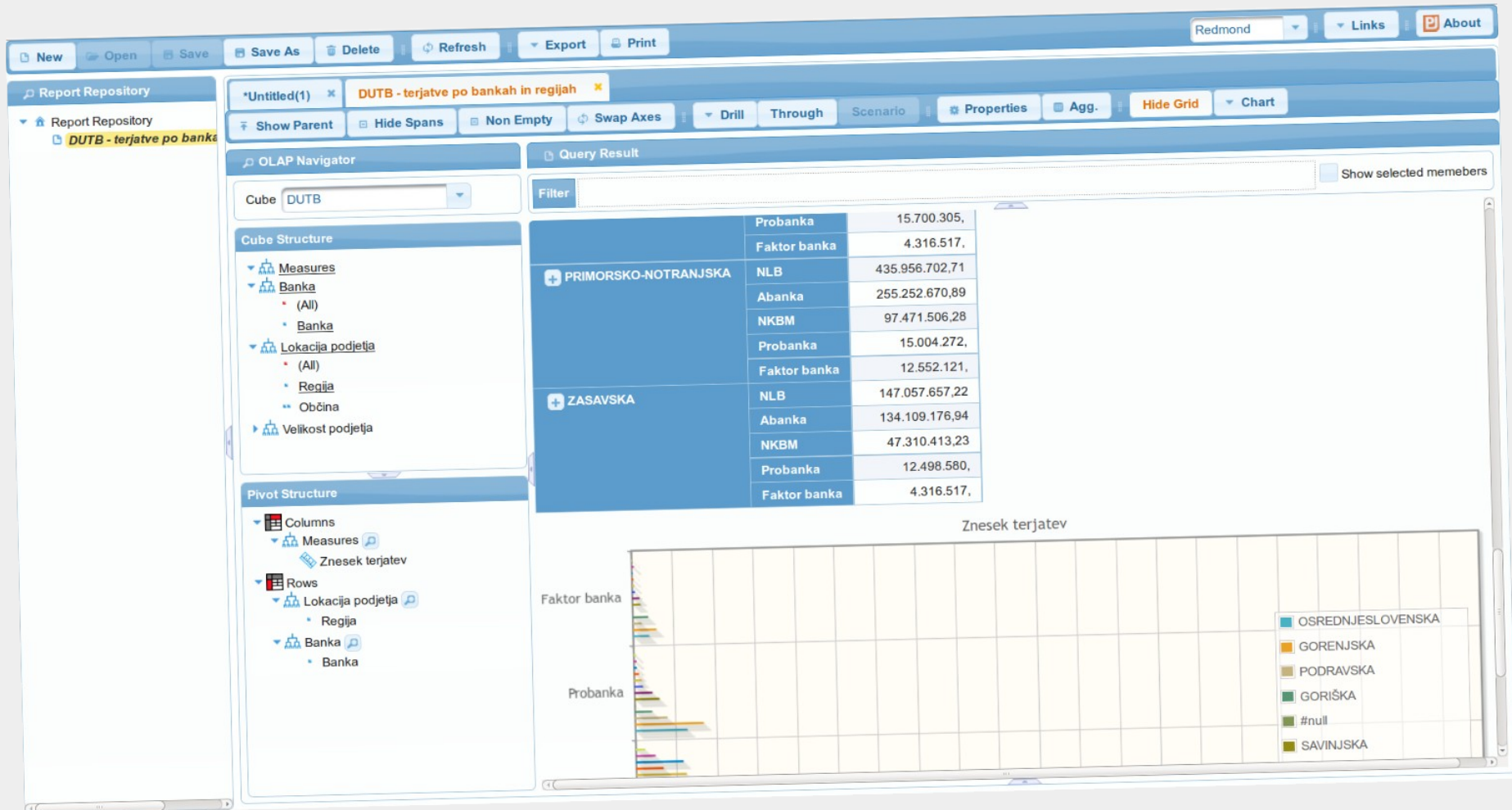


The screenshot shows the website of DUTB (Družba za upravljanje vrednostima) with a navigation menu and a search bar. The main content area displays a table titled "Pogodbe NLB" (Agreements NLB) for Banka NLB d.d. The table lists six agreements with the following columns: Kreditojemalec (Creditor), Kreditodajalec (Lender), Št. pogodbe (No. of agreements), Znesek v EUR (Amount in EUR), Vrsta posla* (Type of business), Datum pogodbe (Date of agreement), and Zavarovanje (Insurance). The table is filtered to show 6 results.

Kreditojemalec	Kreditodajalec	Št. pogodbe	Znesek v EUR	Vrsta posla*	Datum pogodbe	Zavarovanje			
Naziv	Sedež								
1	2	3	4	5	6	7	8	9	
1	SIMER D.O.O. CELJE	IPAVČEVA ULICA 22, 3000 CELJE	NLB d.d.	LD1010500015	360.000	Kredit za obratna sredstva	23.04.2010	Zastava poslovnih nepremičnin	Lokal
2	SIMER D.O.O. CELJE	IPAVČEVA ULICA 22, 3000 CELJE	NLB d.d.	LD1105600016	90.000	Kredit za obratna sredstva	01.03.2011	Zastava poslovnih nepremičnin	Industrijski objekt/sirša uporabna vrednost
3	SIMER D.O.O. CELJE	IPAVČEVA ULICA 22, 3000 CELJE	NLB d.d.	LD1105600016	90.000	Kredit za obratna sredstva	01.03.2011	Zastava poslovnih nepremičnin	Lokal
4	SIMER D.O.O. CELJE	IPAVČEVA ULICA 22, 3000 CELJE	NLB d.d.	LD0914600006	450.000	Kredit za obratna sredstva	28.11.2008	Zastava poslovnih nepremičnin	Industrijski objekt/sirša uporabna vrednost
5	SIMER D.O.O. CELJE	IPAVČEVA ULICA 22, 3000 CELJE	NLB d.d.	LD0914600006	450.000	Kredit za obratna sredstva	28.11.2008	Zastava poslovnih nepremičnin	Lokal
6	SIMER D.O.O. CELJE	IPAVČEVA ULICA 22, 3000 CELJE	NLB d.d.	LD0907100032	614.400	Kredit za investicije	13.05.2009	Solidarno poroštvo PO	.

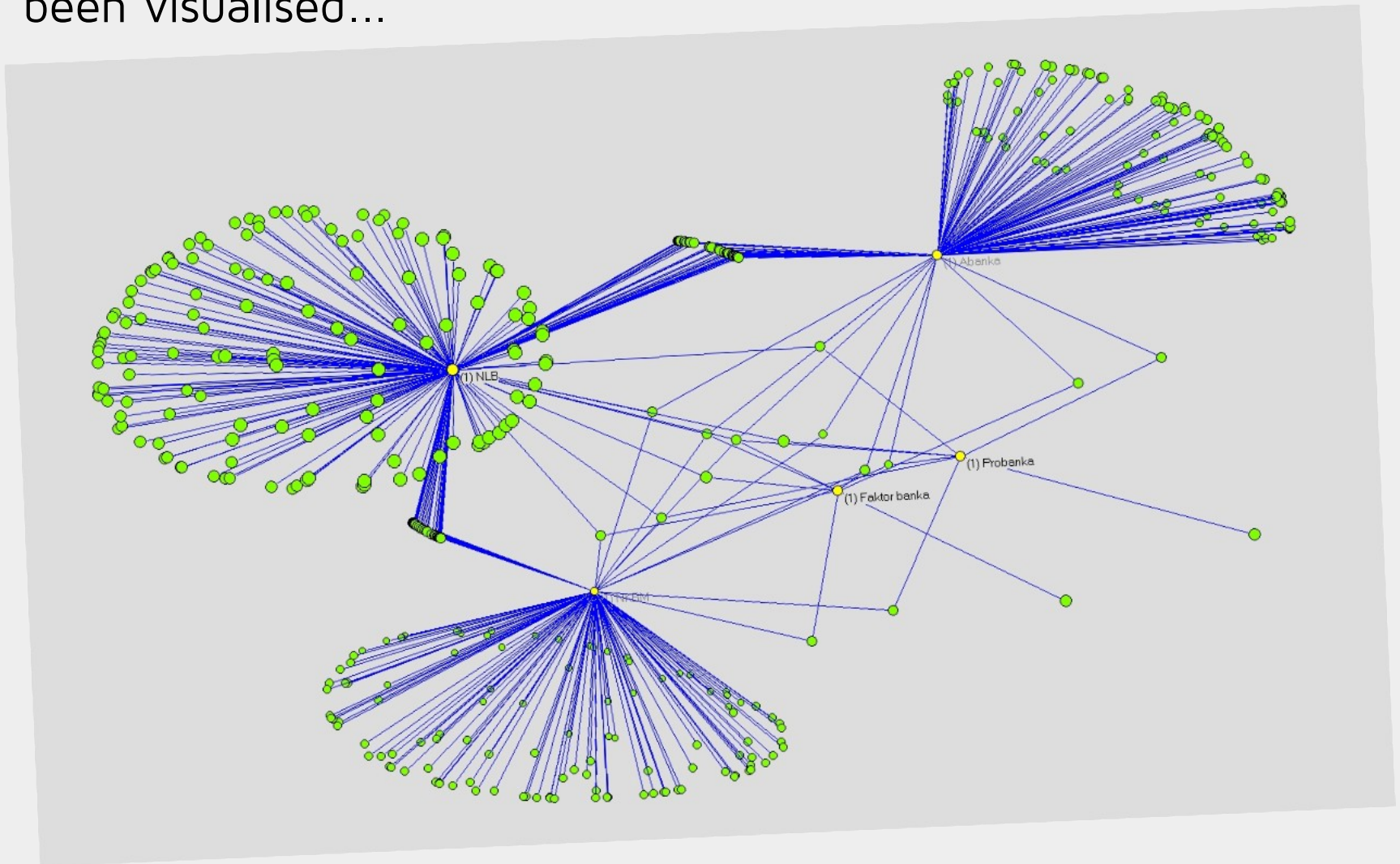
Bad investments transferred on *Bank Assets Management Company*

Data has been imported into the database and linked to business registry...



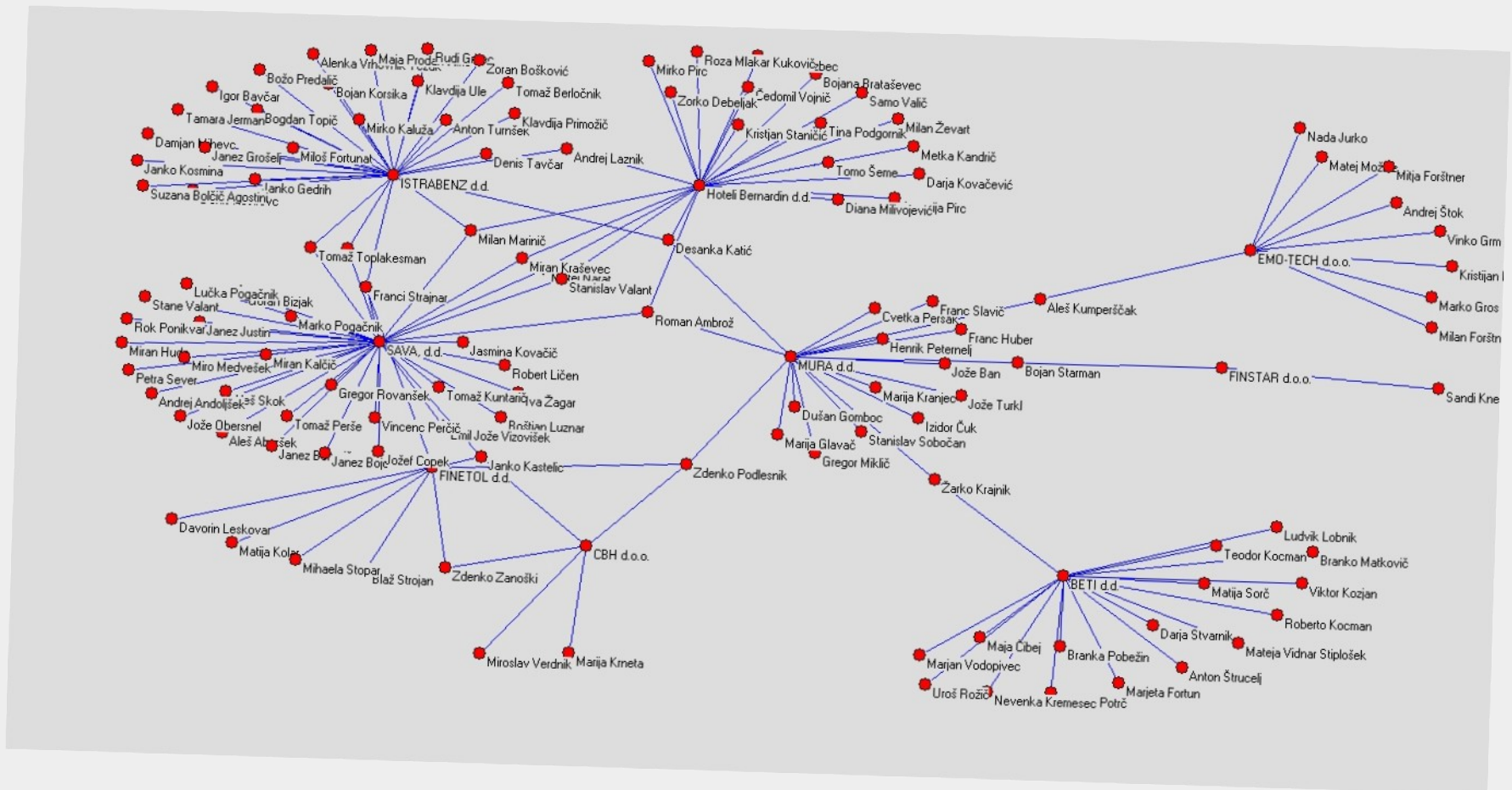
Bad investments transferred on *Bank Assets Management Company*

... multiple bad loans to companies transferred to BAMC has been visualised...



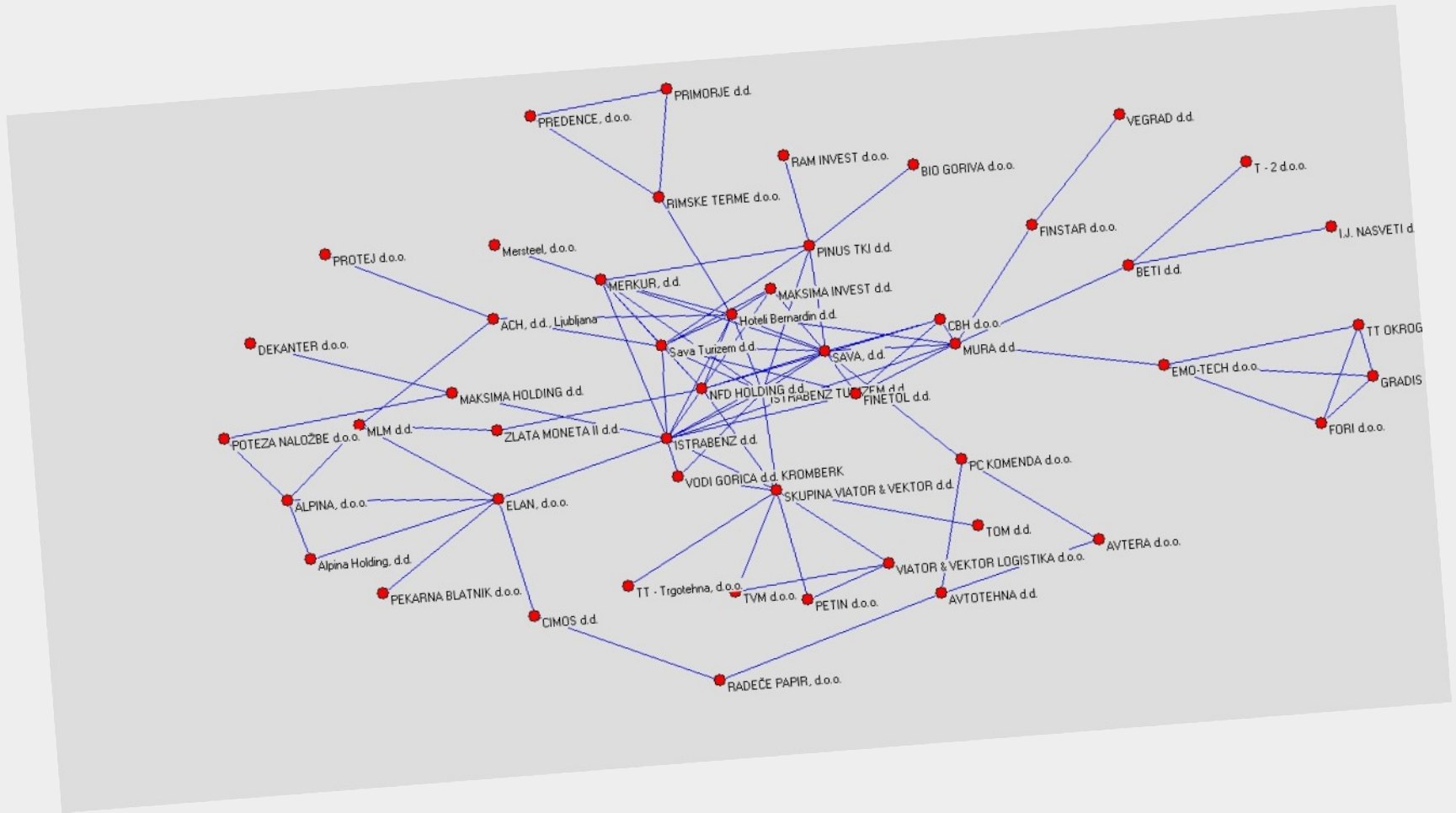
Bad investments transferred on *Bank Assets Management Company*

... the network of connected persons has been created (persons who are or have been directors and/or members of supervisory boards of those companies).



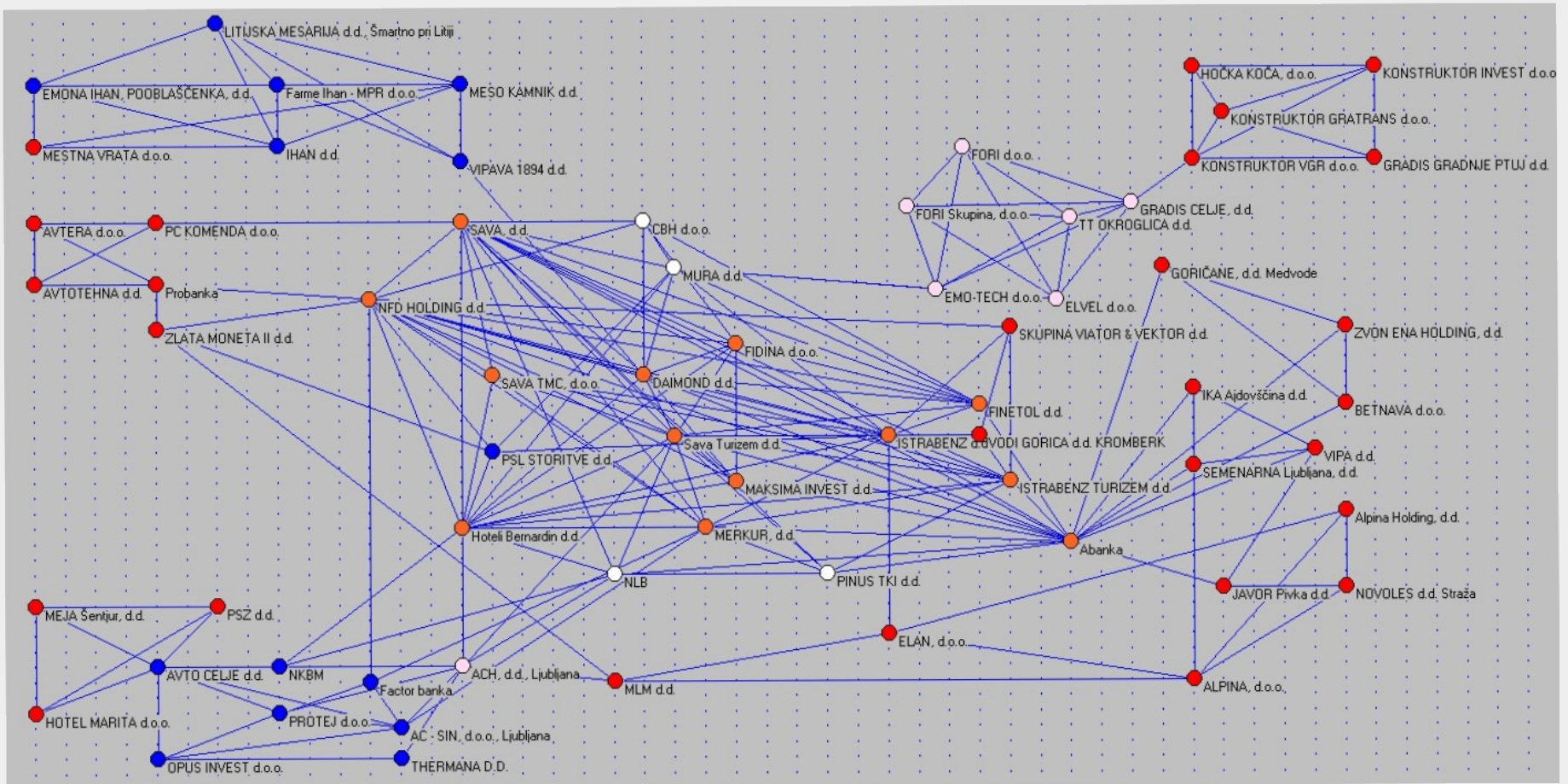
Bad investments transferred on *Bank Assets Management Company*

The network shows (management) connections among corporations...



Bad investments transferred on *Bank Assets Management Company*

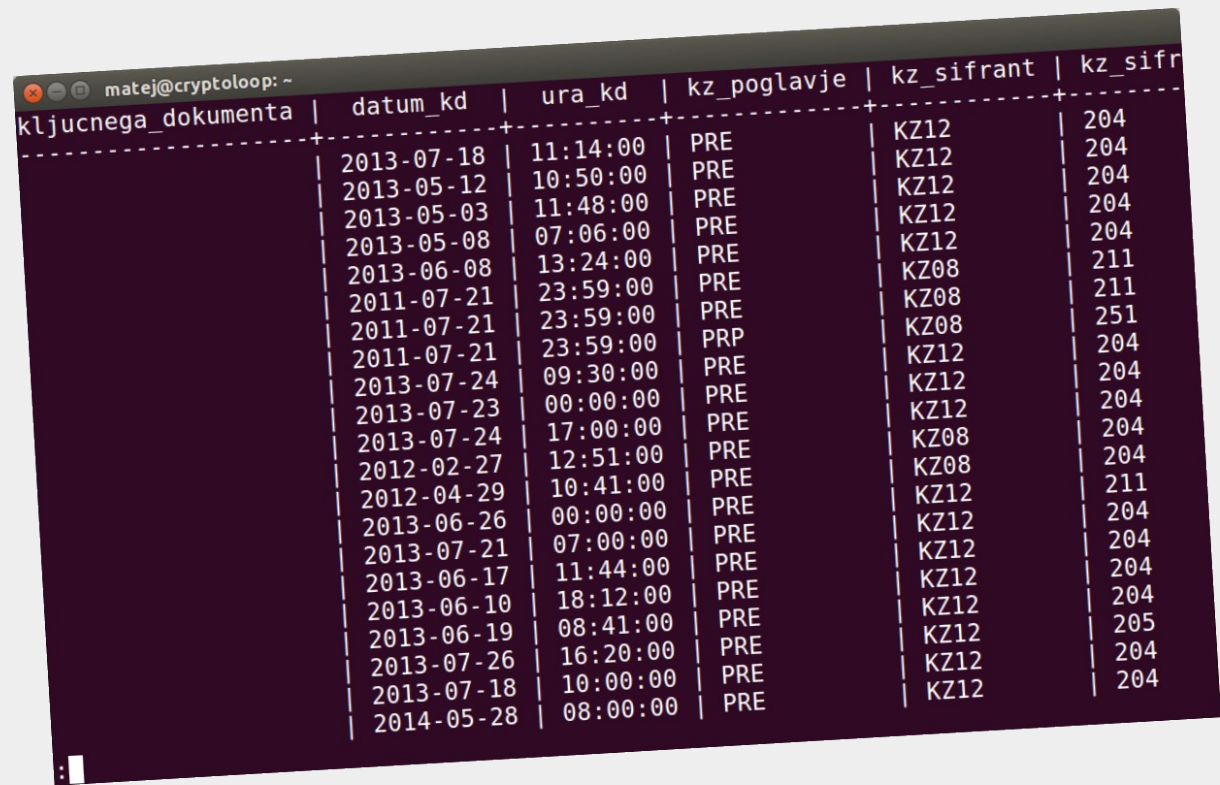
The biggest component of the network contains approx. 33% of all companies transferred to BAMC. Those companies has been granted approx. 61% of bad loans.



The incidence of crime [police database of all committed offenses]

The Institute of Criminology at the Faculty of Law has obtained access to police records of all committed offenses...

Database contains description of committed offense, geolocation data, description of involved persons, description of method of offense, ID of persons, etc.

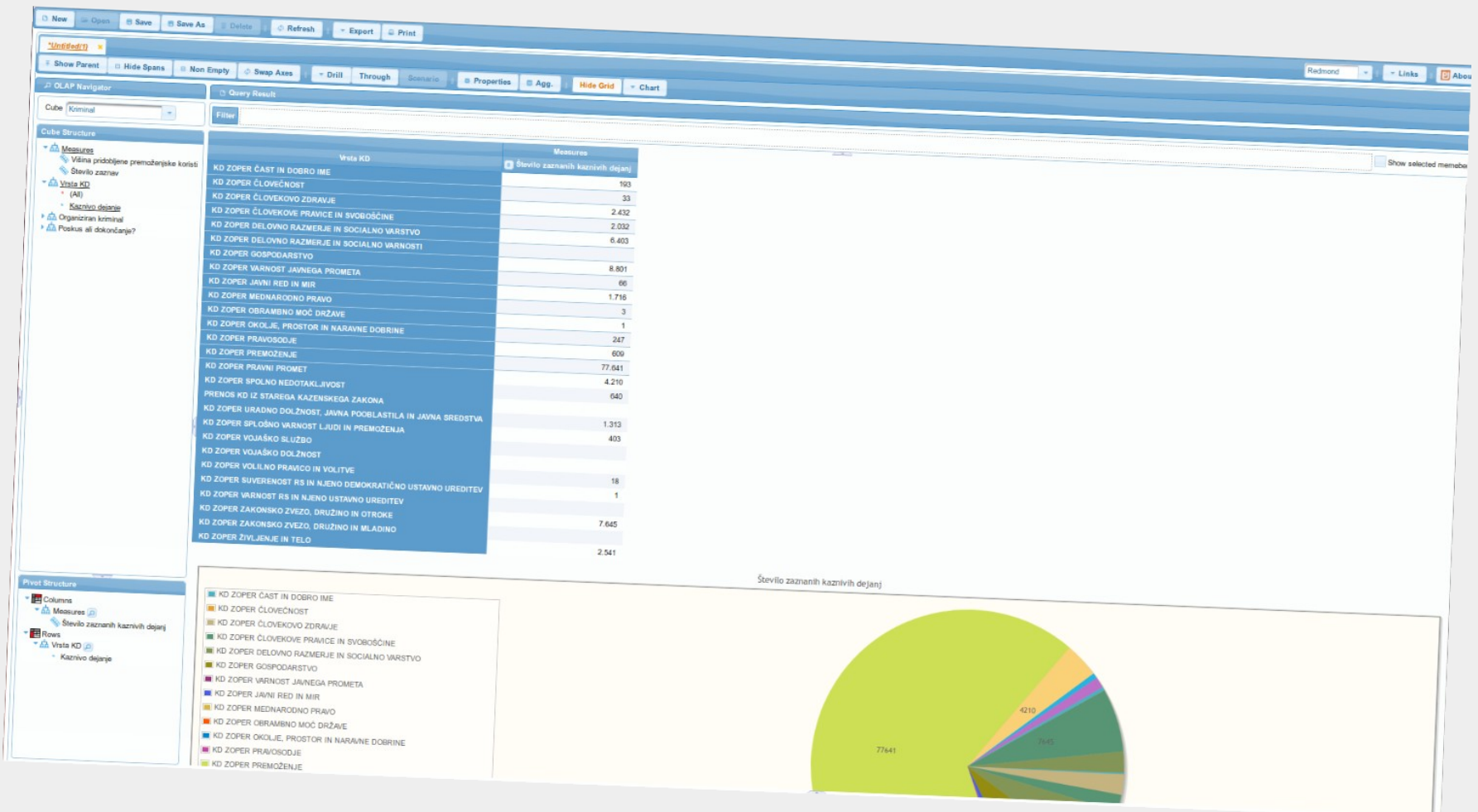


```
matej@cryptoloop: ~  
ključnega dokumenta | datum_kd | ura_kd | kz_poglavje | kz_sifrant | kz_sifr
```

ključnega dokumenta	datum_kd	ura_kd	kz_poglavje	kz_sifrant	kz_sifr
	2013-07-18	11:14:00	PRE	KZ12	204
	2013-05-12	10:50:00	PRE	KZ12	204
	2013-05-03	11:48:00	PRE	KZ12	204
	2013-05-08	07:06:00	PRE	KZ12	204
	2013-06-08	13:24:00	PRE	KZ08	211
	2011-07-21	23:59:00	PRE	KZ08	211
	2011-07-21	23:59:00	PRP	KZ08	251
	2011-07-21	23:59:00	PRE	KZ12	204
	2013-07-24	09:30:00	PRE	KZ12	204
	2013-07-23	00:00:00	PRE	KZ12	204
	2013-07-24	17:00:00	PRE	KZ08	204
	2012-02-27	12:51:00	PRE	KZ08	204
	2012-04-29	10:41:00	PRE	KZ12	211
	2013-06-26	00:00:00	PRE	KZ12	204
	2013-07-21	07:00:00	PRE	KZ12	204
	2013-06-17	11:44:00	PRE	KZ12	204
	2013-06-10	18:12:00	PRE	KZ12	204
	2013-06-19	08:41:00	PRE	KZ12	205
	2013-07-26	16:20:00	PRE	KZ12	204
	2013-07-18	10:00:00	PRE	KZ12	204
	2014-05-28	08:00:00	PRE	KZ12	204

The incidence of crime

Data were imported into the database and simple OLAP analytic has been programmed...



The incidence of crime

Data could also be easily imported in PSPP (software for advanced statistical analysis).

The screenshot displays the GNU PSPP interface. The main window shows a data table with columns: Case, id_kd, enota_ki, id_dosjeja, id_zaznave, vrsta_zaz, st_1_zakljucnega_c, datum_1, sifra_1_z, datum_ki, ura_kd, kz_pogla, kz_sifran, kz_odsta, kz_tocka, kz_alinej, krim_ozn, krim_ozn, and krim_oz. A secondary window titled '*kriminal.sps — PSPPIRE Urejevalnik sintakse' contains the following SQL query:

```
GET DATA /TYPE=PSQL  
/CONNECT='host=127.0.0.1 port=5432 dbname= user= password='  
/SQL='select * from kriminal_dogodki.'
```

A red arrow points to the SQL query window. Another window titled 'Rezultati_analize — PSPPIRE Pregledovalnik rezultatov analize' displays statistical results:

T-TEST
T-TEST /VARIABLES=visina_premozenjske_koristi
/GROUPS=organiziran_kriminal("0" "1") /MISSING=ANALYSIS
/CRITERIA=CIN(0.95).

Group Statistics		organiziran_kriminal	N	Povprečje	Standardni odklon	Povprečje std. napake:
visina_premozenjske_koristi	0	116513	443	2171,76	210282,22	616,05
	1	0	1	1635,14	17942,08	852,45

Independent Samples Test

	Leveneov test enakosti varianc	t-test for Equality of Means						95% Confidence Interval of the Difference		
		F	Stat. značilnost	t	prost.st.	Sig. (2-tailed)	Povprečna razlika	Std. Error Difference	spodnja	zgornja
visina_premozenjske_koristi	Equal variances assumed Enakost varianc ni predpostavljena	,01	,909	,05	116954,00	,957	536,62	9990,99	-19045,56	20118,81
				,51	1023,18	,610	536,62	1051,76	-1527,23	2600,47

At the bottom, there are buttons for 'Pregled podatkov', 'Pregled spremenljivk', '... izključeni', 'Uteži izključene', and 'No Split'.

The incidence of crime

Some other possible analysis:

- visualisation of crime locations;
- time series;
- analysis of recidivism;
- clustering of criminal offences;
- crime predictions;
- ...

Data about lobbyists and lobbying contacts

We have received data about registered lobbying contacts from the *Commission for the Prevention of Corruption*. Unfortunately, data were not provided in machine readable form...

REPUBLIKA SLOVENIJA
KOMISIJA ZA PREPREČEVANJE KORUPCIJE
Dolžnik št. 1000 / Ljubljana

REGISTER LOBISTOV

KOMISIJA ZA PREPREČEVANJE KORUPCIJE
INTEGRITETA, ODGOVORNOST, VLADAVINA PRAVA

PRUJAVA STIKA Z OSEBO/LOBISTOM, KI SE NI DOŁAN VPISATI V REGISTER

Skladno z 71. členom Zakona o integriteti in preprečevanju korupcije (ZInPK, Uradni list RS, št. 45/2010, 26/2011 in 43/2011) prijavljam stik z lobistom, ki se ni dolžan vpisati v register lobistov.

Čerți odstavek 58. člena ZInPK določa, da se osebi/lobistu, ki lobira za interesno organizacijo, v kateri je zaposlen, za ta nemerni ni dolžan vpisati v register. Enako velja za zakonitega zastopnika, ali izvoljenega predstavnika interesne organizacije.

1. OSERNI PODATKI LOBIRANCA

1.1 Priimek, ime: mag. Vasja Klavora, Franc Jurša, Anton Urh, Matjaž Zanoškar

1.2 Naziv in naslov organa v katerem ste zaposleni (institucija, hiša št., naselje, pošta, kraj): Državni zbor RS, PS DeSUS, Šubičeva 4, 1000 Ljubljana

2. PODATKI LOBISTA:

Utemeljite in vzpostavite določeno kodo več možnosti!

Lobist je vpisan v register lobistov

Lobist lobira za interesno organizacijo, v kateri je zaposlen

Lobist je zakoniti zastopnik interesne organizacije za katero lobira

Lobist je izvoljeni predstavnik interesne organizacije za katero lobira

2.1 Priimek in ime: Leo Oblak (Radio Antena), Rina Klinar (Radio Triglav), Ivan Seljak (Radio Cerčno)

2.2 Prebivališče (ulica, hiša št., naselje, pošta, kraj):

2.3 Naziv in naslov interesne organizacije, ali druge organizacije za katero lobist lobira (institucija, hiša št., naselje, pošta, kraj): Radio Antena d.o.o., Stegne 11b, 1000 Ljubljana; Radio Triglav d.o.o., Trig Toneta Čufarja 4, 4270 Jesenica; Radio Cerčno d.o.o., Ličarjeva ulica 7, 5282 Cerčno

2.4 Namen in cilj, zaradi katerega lobist lobira: predstavitev predloga novele Zakona o medijih, ki je bila poslana vsem poslancem na elektronski naslov in tudi v vednost KPK Da Ne

2.5 Ali se je lobist identifičiral in pokazal pooblastilo interesne organizacije za lobiranje v določeni zadevi? Da Ne

Komisija za preprečevanje korupcije
Dunajska 56, 1000 Ljubljana
tel: 01/400-57-10
fax: 01/400-04-72

Poslanska skupina Socialnih demokratov (SD)

Številca o lobiranju: 24
KOMTO LOBIRANJE: /
ME ZAKONITEGA LOBIRANJA: 30

ITNO PREPOVEDANO LOBIRANJE:
projekti prijav ni mogoče šteti lobističnih stikov. Z isto prijavo je namreč lahko prijavljen več lobističnih stikov, projekti prijav so zajete tudi prijave, ki po določbah ZInPK nimajo naravo lobističnega stika.

KOMISIJA ZA PREPREČEVANJE KORUPCIJE, Dunajske ceste 56, 1000 Ljubljana
Telefoni: +386 1 400 57 10, Fax: +386 1 400 04 72 e-pošta: ast@kocp.gov.si, s.j@kocp.gov.si, s.j@kocp.gov.si

stran 1 / 1

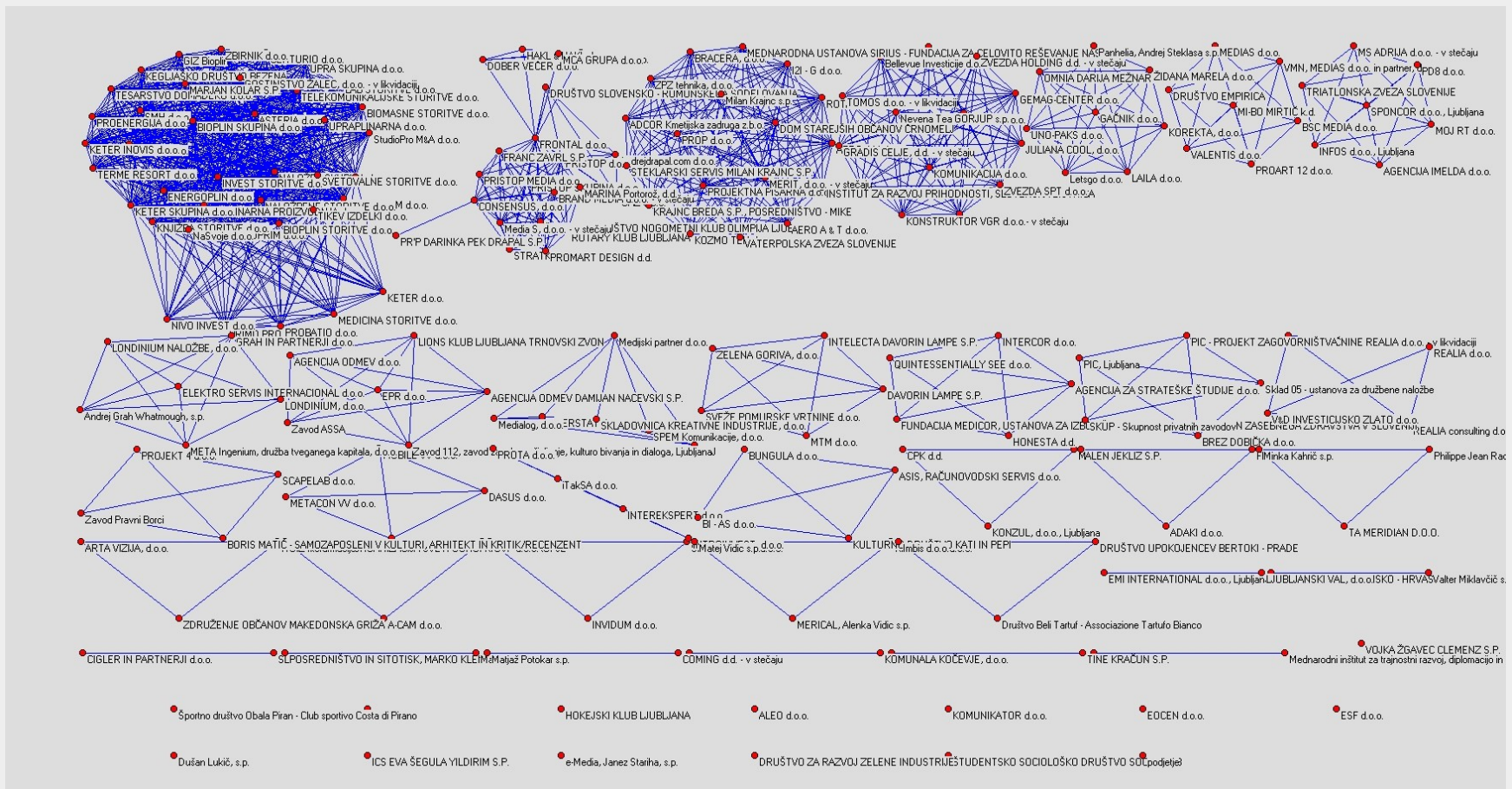
Data about lobbyists and lobbying contacts

... however, volunteers entered the data into the database...

AD	AE	AF	AG	AI	AJ	AK	AL	AM	AN
lobist_identificiran	nasprotje_interesov	darilo_pogostitev	datum_stika	datum_porocanja_stika	kraj_stika_clean	naziv_1_interesne_org_clean	naziv_2_interesne_org_clean	naziv_3_interesne_org_clean	
1	da	ne	2011-03-29	2011-03-30	Ljubljana	HESS, d.o.o.			
2	da		2011-09-15	2011-09-16	Ljubljana	GIZ RADIODIFUZNIH RADIJSKIH IN TELEVIZIJSKIH MEDIJEV			
3	da		2011-09-23	2011-09-26	Ljubljana	HSE d.o.o.	TEŠ d.o.o.		
4	da		2012-12-07	2012-12-07	Ljubljana	GZS			
5	da		2011-01-18	2011-03-14	Ljubljana	EUROLUX d.o.o.			
6	ne		2013-07-24	2013-07-25	Ljubljana	Dr. Ferk & Partner d.o.o.			
7	da	ne	2013-06-12	2013-06-14	Ljubljana	ZAVOD PIP			
8	ne	ne	2012-09-06	2012-09-06	Ljubljana	SENG d.o.o.			
9	da	ne	2011-04-19	2011-04-20	Ljubljana	FOCUS, DRUŠTVO ZA SONARAVEN RAZVOJ			
10	ne	ne	2012-10-20	2012-10-23	Ljubljana				
11	da	da	2013-09-12	2013-09-12	Ljubljana	ZDRUŽENJE LASTNIKOV NEPREMIČNIN V SLOVENIJI			
12	da	ne	2013-11-26	2013-12-13	Ljubljana				
13	da	da	2013-09-12	2013-09-12	Ljubljana	ZDRUŽENJE LASTNIKOV NEPREMIČNIN V SLOVENIJI			
14	da	da	2013-09-12	2013-09-12	Ljubljana	ZDRUŽENJE LASTNIKOV NEPREMIČNIN V SLOVENIJI			
15	da	ne	2013-09-25	2013-09-23	Ljubljana	PETROL d.d., Ljubljana			
16	ne	ne	2013-03-07	2013-03-07	Ljubljana	MEDNARODNI FORUM ZNANSTVENO RA&NOVARTIS PHARMA SERVICES INC. Podru	SERVIER PHARMA d.o.o.		
17	ne	ne	2014-06-13	2014-06-17	Ljubljana	ECRS			
18	ne	ne	2014-06-13	2014-06-17	Ljubljana	ECRS		ZDRUŽENJE SLOVENSKA UNIJA GLASBENIH USTVARJALCEV	
19	ne	ne	2012-09-18	2012-09-19	Ljubljana	ZDRUŽENJE SAZAS		HIT d.d., Nova Gorica	
20	ne	ne	2013-10-15	2013-10-16	Ljubljana	CASINO Portorož, d.d.			
21	da	ne	2011-01-12	2011-01-12	Ljubljana	UL FS			
22	ne	ne	2013-11-05	2013-11-07	Ljubljana	CONSENSUS, d.o.o.			
23	da	ne	2013-05-08	2013-05-08	Ljubljana	GZS			
24	ne	ne	2010-12-08	2010-12-08	Ljubljana	ZDRUŽENJE LASTNIKOV NEPREMIČNIN V SLOVENIJI			
25	ne	ne	2011-04-13	2011-11-15	Ljubljana	REVOZ d.d.			
26	ne	ne	2011-03-30	2011-03-31	Ljubljana	HESS, d.o.o.			
27	da	ne	2011-04-15	2011-04-19	Ljubljana	YHD - DRUŠTVO ZA TEORIJO IN KULTURO DRUŠTVO ŽAREK UPANJA, Ljubljana			
28	da	ne	2011-09-08	2011-09-08	Ljubljana	PEKO, d.d.		INFORMATIKA d.d.	
29	ne		2013-03-08	2013-03-18	Ljubljana	TELEKOM SLOVENIJE, d.d.			
30	ne	ne	2013-02-15	2013-02-19	Ljubljana	HOTEL SAVAROGAŠKA d.o.o.			
31	ne	ne	2013-01-22	2013-01-23	Ljubljana	TERMIA d.o.o.			
32	da	ne	2013-03-06	2013-03-06	Ljubljana	SI MOBIL d.d.			
33	da	ne	2013-07-25	2013-07-30	Ljubljana	NEVIJA d.o.o.			
34	da	ne	2011-08-09	2012-02-29	Ljubljana	NEMESIS d.o.o.			
35	ne	ne	2014-03-21	2014-03-28	Ljubljana	TELEKOM SLOVENIJE, d.d.			
36	ne	ne	2012-10-11	2012-10-12	Ljubljana	CASINO Portorož, d.d.		HIT d.d., Nova Gorica	
37	da	ne	2013-10-15	2013-10-16	Ljubljana	TEŠ d.o.o.			
38	da	ne	2011-09-22	2011-09-29	Ljubljana				
39	ne	ne	2013-05-31	2013-06-03	Ljubljana	KZ KRKA z.o.o.			
40	ne	ne	2014-06-04	2014-06-04	Ljubljana	DATALAB d.d.			
41	ne	ne	2013-09-17	2013-09-19	Ljubljana	ZDRUŽENJE LASTNIKOV NEPREMIČNIN V SLOVENIJI			
42	da	da	2013-09-12	2013-09-12	Ljubljana	VZAJEMNA d.v.z.			
43	da	da	2013-09-12	2013-10-29	Ljubljana				
44	ne	da							

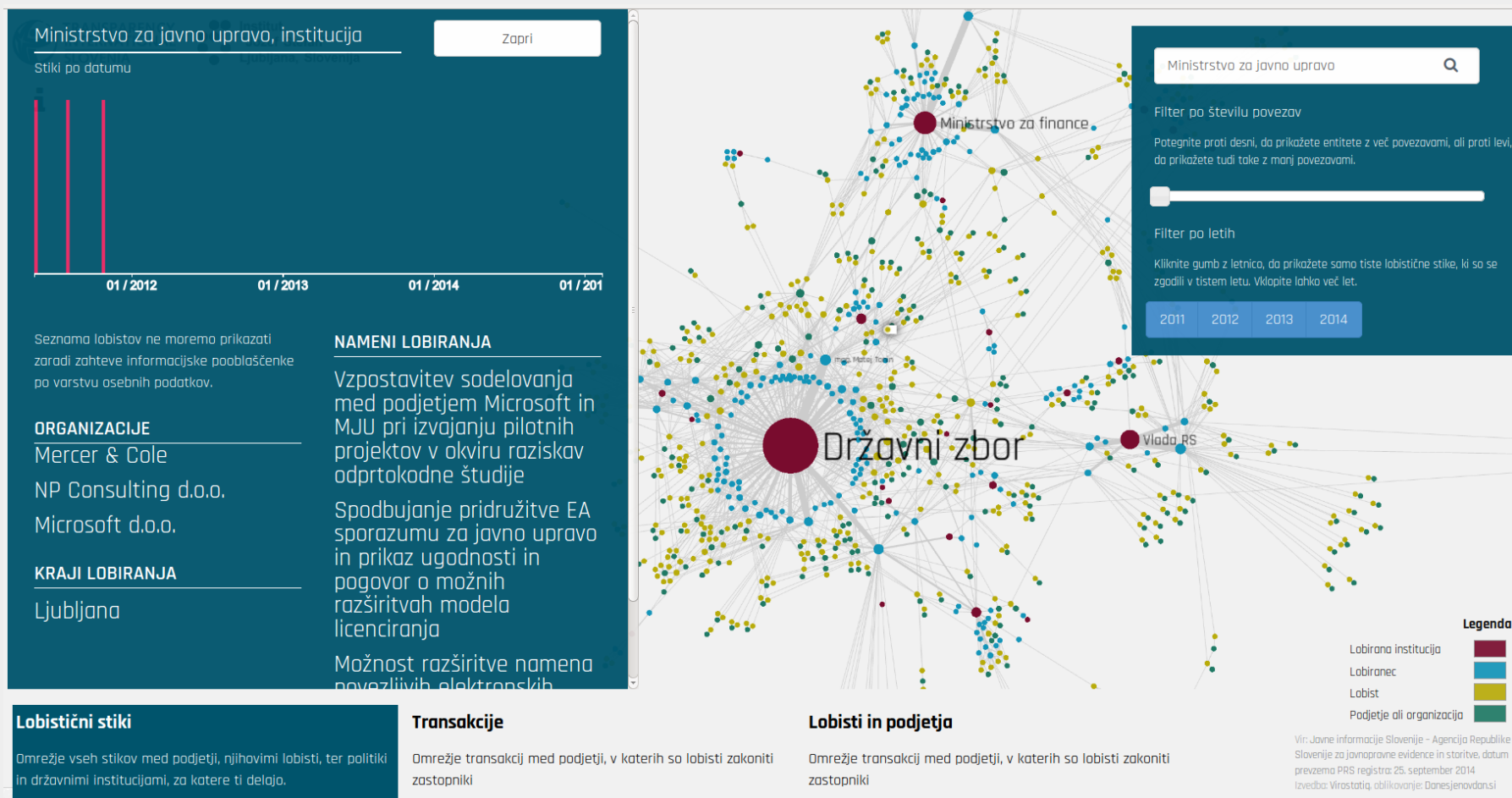
Data about lobbyists and lobbying contacts

... Database was linked to business registry and some other databases and some analysis of lobbyists and their companies has been prepared.



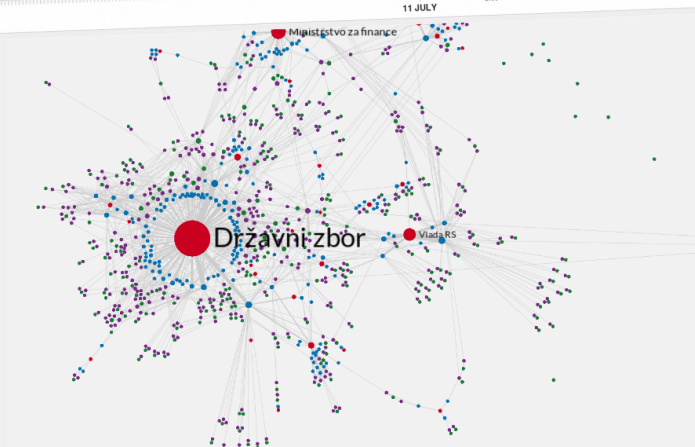
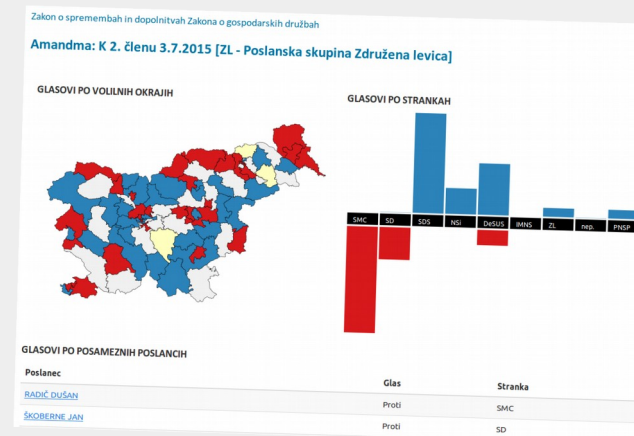
Data about lobbyists and lobbying contacts

Finally, visualisation of lobbying contacts has been created.



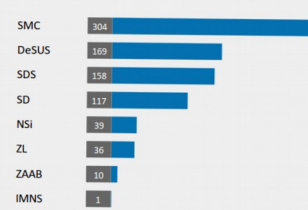
Legislative Monitor [zakonodajni-monitor.si]

Legislative Monitor is (was) an interactive tool which gives an insight into the legislation process, MEP's activities, analysis of their discussions, reported lobbying contacts, etc.



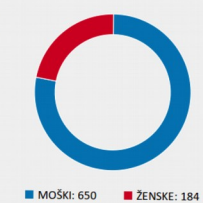
Število izjav po strankah

Pripadniki katerih strank so se v postopku sprejemanja tega akta največkrat oglašili?



Število izjav po spolu

Kolikokrat so se v postopku sprejemanja tega akta oglašili moški, in kolikokrat ženske?



Teme razprave

O katerih temah so največkrat govorili pripadniki posameznih strank?



<http://zakonodajni-monitor.si/>

Salaries in public sector

Ajpes is collecting all data about salaries of civil servants. Through access of public information act we received the database of salaries from 2010. Database also contains (anonymised) ID's of civil servants.

The image shows a screenshot of the ISPAP website and a terminal window displaying a database query result.

ISPAP - INFORMACIJSKI SISTEM ZA POSREDOVANJE

V tej rubriki je mogoče pregledovati podatke o izplačan

Osnovno iskanje

Leto obračuna plače: 2015
Mesec obračuna plače: 1
Šifra delovnega mesta: - vse -
Plačni razred: - vsi -

Išči

Dodatni pogoji (izbrati morate šifro delovnega mesta in/ali plačni razred)

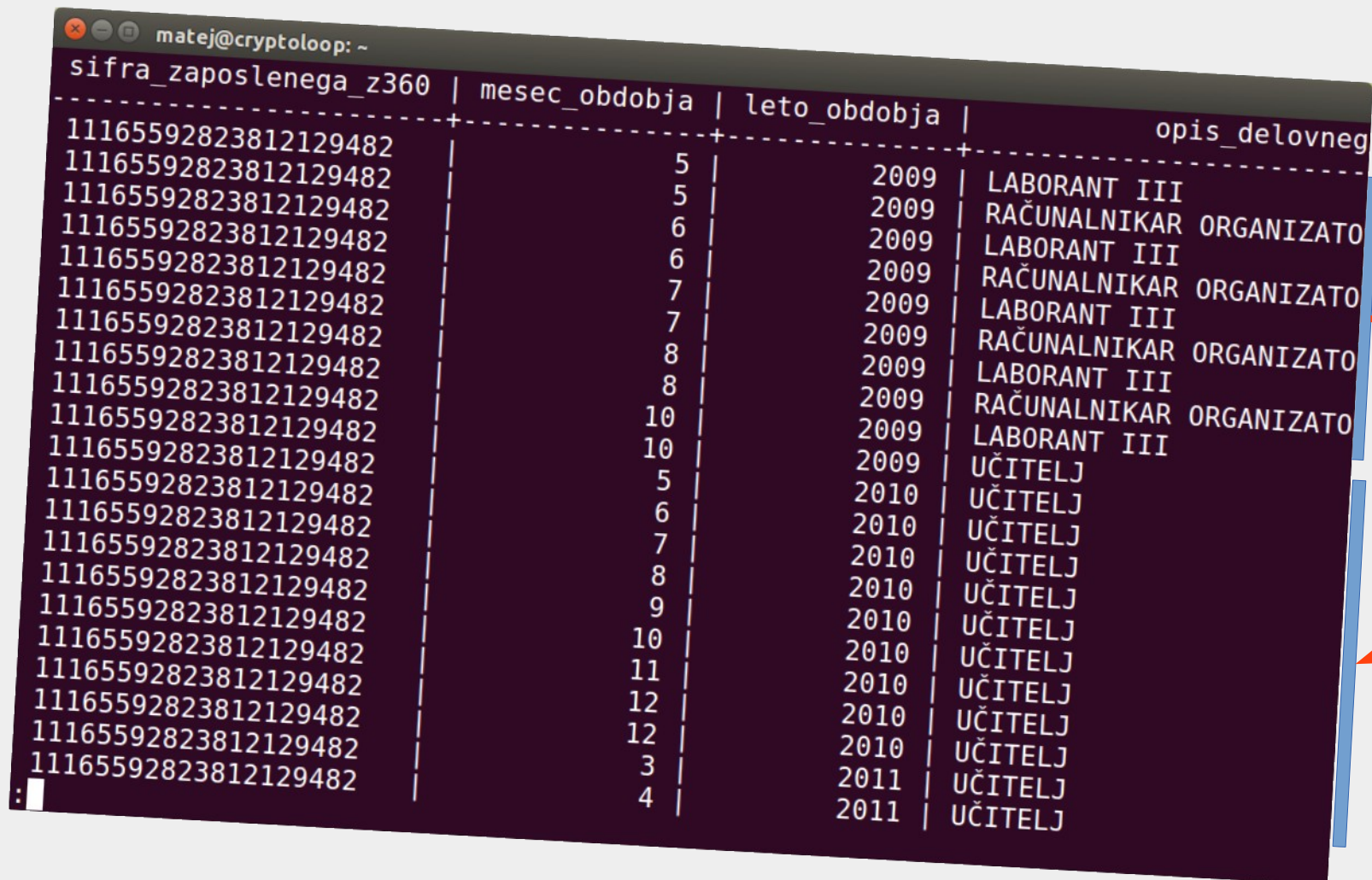
Dodatki: - ne prikaži -
Delovna uspešnost: - ne prikaži -
Devizni dodatek: - ne prikaži -

Išči

```
matej@cryptoloop: ~
60 | sifra_delovnega_mesta_z370 | opis_delovnega_mesta_z370opis | sifra_n
-----|-----|-----|-----
C027001 |          |          | 2
C027002 |          |          | 2
J016005 |          |          | 0
C027001 |          |          | 1
C027001 |          |          | 1
C027001 |          |          | 2
C025003 |          |          | 1
J017137 |          |          | 2
J027005 |          |          | 1
J034069 |          |          | 0
C027005 |          |          | 0
C027001 |          |          | 0
C027001 |          |          | 3
J027001 |          |          | 1
J016027 |          |          | 1
C027010 |          |          | 0
C027001 |          |          | 0
J035059 |          |          | 3
C027001 |          |          | 1
A037331 |          |          | 0
C027005 |          |          | 1
          |          |          | 0
          |          |          | 2
```

Salaries in public sector

From the salaries database career paths could be computed...

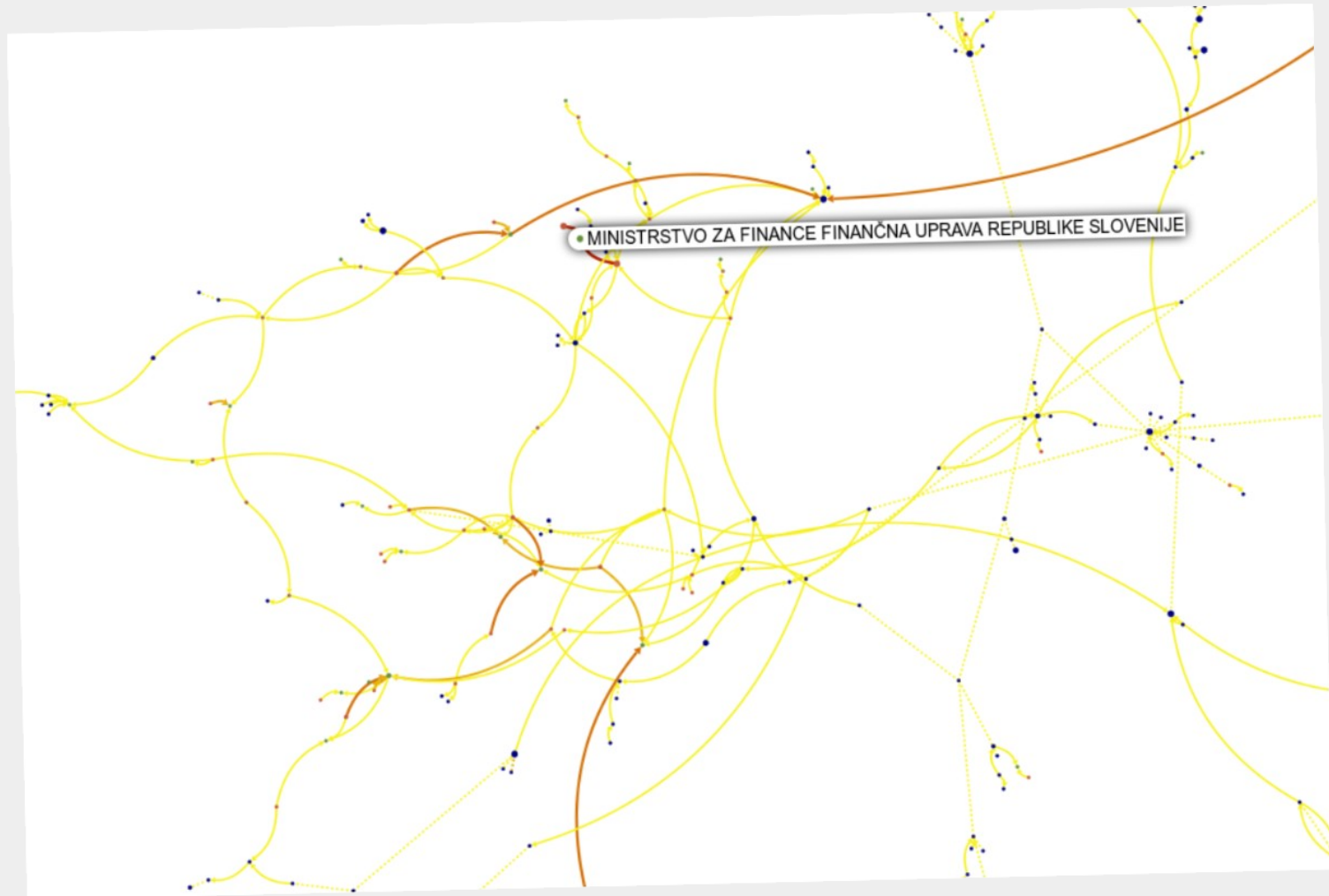


```
matej@cryptoloop: ~
sifra_zaposlenega_z360 | mesec_obdobja | leto_obdobja | opis_delovneg
-----+-----+-----+-----
11165592823812129482 | 5 | 2009 | LABORANT III
11165592823812129482 | 5 | 2009 | RAČUNALNIKAR ORGANIZATO
11165592823812129482 | 6 | 2009 | LABORANT III
11165592823812129482 | 6 | 2009 | RAČUNALNIKAR ORGANIZATO
11165592823812129482 | 7 | 2009 | LABORANT III
11165592823812129482 | 7 | 2009 | RAČUNALNIKAR ORGANIZATO
11165592823812129482 | 8 | 2009 | LABORANT III
11165592823812129482 | 8 | 2009 | RAČUNALNIKAR ORGANIZATO
11165592823812129482 | 10 | 2009 | LABORANT III
11165592823812129482 | 10 | 2009 | UČITELJ
11165592823812129482 | 5 | 2010 | UČITELJ
11165592823812129482 | 6 | 2010 | UČITELJ
11165592823812129482 | 7 | 2010 | UČITELJ
11165592823812129482 | 8 | 2010 | UČITELJ
11165592823812129482 | 9 | 2010 | UČITELJ
11165592823812129482 | 10 | 2010 | UČITELJ
11165592823812129482 | 11 | 2010 | UČITELJ
11165592823812129482 | 12 | 2010 | UČITELJ
11165592823812129482 | 12 | 2010 | UČITELJ
11165592823812129482 | 3 | 2011 | UČITELJ
11165592823812129482 | 4 | 2011 | UČITELJ
```

The image shows a terminal window with a table of salary data. The table has four columns: 'sifra_zaposlenega_z360' (employee ID), 'mesec_obdobja' (month), 'leto_obdobja' (year), and 'opis_delovneg' (job description). The data shows a progression from 'LABORANT III' in 2009 to 'UČITELJ' in 2010 and 2011. A blue vertical bar highlights the right side of the table, and a red arrow points from the right edge of the terminal window towards the right side of the slide.

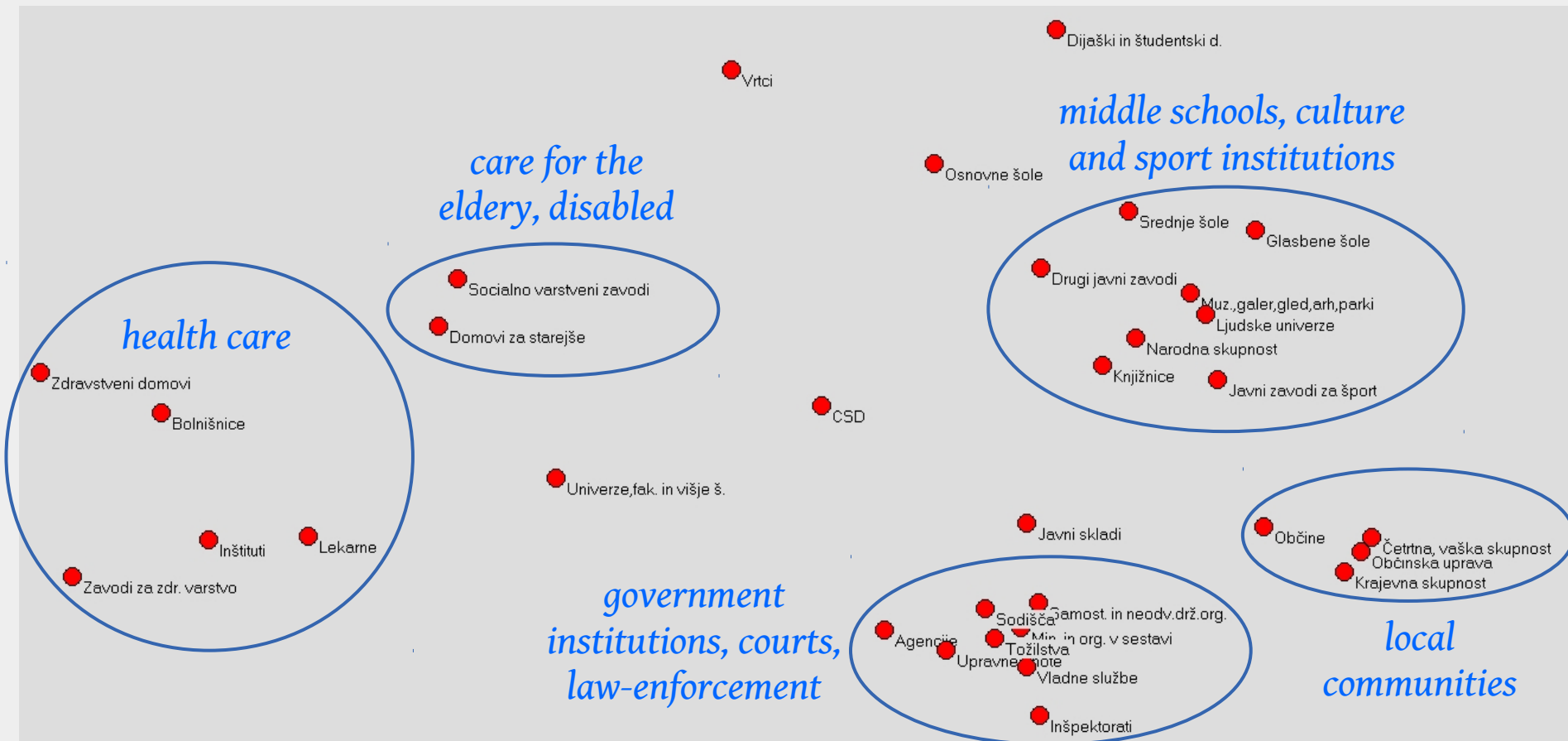
Salaries in public sector

... and from career paths migrations of workforce among budget users could be computed and visualised.



Salaries in public sector

Clustering of employees migrations among public sector bodies.



Salaries in public sector

Another possible analysis:

- the influence of leadership change on staff recruitment;
- the influence of leadership change in sick leaves and terminations of employment contracts (possible indicator of mobbing);
- rate between minimum and maximum salary in public different parts of sector;
- amount of women in different employment positions;
- visualisations of wage premiums;
- what-if scenarios for salaries reform;
- ...

Public procurement database

We have received database of all Slovenian public procurements including small value procurements (published on a portal of public procurements).

The screenshot displays the 'E-NAROČANJE' portal. At the top, there is a navigation bar with links: 'Pregled objav', 'O portalu', 'Pogoji uporabe', 'Aktualno', and 'Kontakt'. The main header includes the 'URADNI LIST' logo and the text 'REPUBLIKA SLOVENIJA, MINISTRSTVO ZA JAVNO UPRAVO, DIREKTORAT ZA JAVNO NAROČANJE'. The central content area is titled 'OBVESTILO O NAROČILU' and features a blue 'Blago' label. Below this, it specifies 'Razna živila', 'Datum objave: 12. 3. 2015', and 'Številka objave: JN1614/2015'. To the right, there is a 'Prijava za naročnike - ZJN-2, ZJNVETPS' section with input fields for 'Uporabniško ime (email):' and 'Geslo:', a 'PRIJAVA' button, and radio button options for 'Pozabljeno geslo' and 'Nov uporabnik'. The main body of the notice is organized into sections: 'ODDELEK I: NAROČNIK' and 'ODDELEK II: PREDMET NAROČILA'. Section I.1 details the contractor's name, address (OSNOVNA ŠOLA BIČEVJE, Splitska ulica 13, SI-1000 Ljubljana), and contact information. It also lists requirements for additional information, specifications, and tenders. Section II.1 is titled 'OPIS' and includes sub-sections II.1.1 (Name of goods), II.1.2 (Type of goods and location), and II.1.3 (Information on public procurement). Section II.1.4 provides information on framework agreements, with a specific field for 'Trajanje okvirnega sporazuma:'.

E-NAROČANJE

Pregled objav | O portalu | Pogoji uporabe | Aktualno | Kontakt

URADNI LIST

REPUBLIKA SLOVENIJA
MINISTRSTVO ZA JAVNO UPRAVO
DIREKTORAT ZA JAVNO NAROČANJE

OBVESTILO O NAROČILU

Blago

Razna živila
Datum objave: 12. 3. 2015
Številka objave: JN1614/2015

Prijava za naročnike – ZJN-2, ZJNVETPS

Uporabniško ime (email):
Geslo:

PRIJAVA

Pozabljeno geslo
Nov uporabnik

ODDELEK I: NAROČNIK

I.1) **IME, NASLOVI IN KONTAKTNE TOČKE**
OSNOVNA ŠOLA BIČEVJE, Splitska ulica 13, SI-1000 Ljubljana.
Tel. +386 12833013. Telefaks +386 12833013. E-pošta o-bicevje.lj@guest.arnes.si.

Dodatne informacije dobite pri
Kot pri zgoraj navedenih kontaktnih točkah.

Specifikacije in dopolnilno dokumentacijo dobite pri
Kot pri zgoraj navedenih kontaktnih točkah.

Ponudbe in prijave za sodelovanje je treba poslati na naslov
Kot pri zgoraj navedenih kontaktnih točkah.

I.2) **VRSTA NAROČNIKA**

I.3) **GLAVNA PODROČJA DEJAVNOSTI**

I.4) **ODDAJA NAROČIL V IMENU DRUGIH NAROČNIKOV**

ODDELEK II: PREDMET NAROČILA

II.1) **OPIS**

II.1.1) **Naziv naročila, ki ga je določil naročnik:**
Sukcesivne nabave konvencionalnih in ekoloških živil.

II.1.2) **Vrsta naročila in mesto gradnje, kraj dobave ali izvedbe**
Blago.
Nakup.

II.1.3) **Informacije o javnem naročilu, okvirnem sporazumu ali dinamičnem nabavnem sistemu**

II.1.4) **Informacije o okvirnem sporazumu**

Trajanje okvirnega sporazuma:

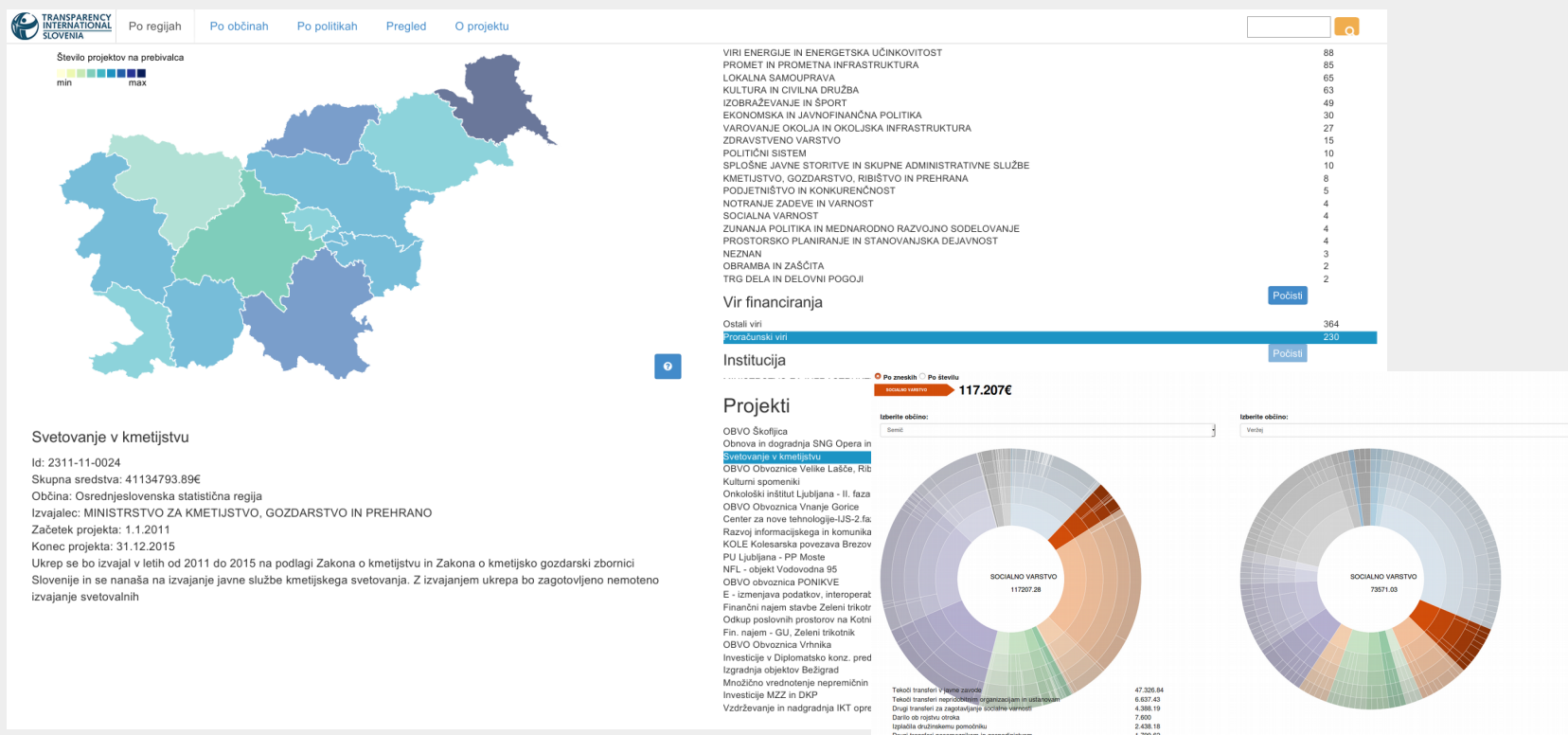
Public procurement database

Some possible analysis:

- the influence of leadership change on public procurement;
- price comparison;
- visualisation of geographical distribution of granted procurements (*“Do the local companies usually win?”*)
- detection of hidden consortium;
- ...

Visualisation of financial data

In 2016 we did a two small visualizations of open data: municipalities' budgets and development programs of Slovenian government.



<http://www.transparency.si/projekti/drzavni-projekti/>

<http://www.transparency.si/projekti/proracuni-obcin/>

Financial transactions of public sector

- Supervisor (now Erar) is an online application which enables simple browsing through the financial transactions and their graphical presentation.
- Application was developed by *Commission for the Prevention of Corruption* in 2011. Later was renamed to Erar.
- User can view all money transfers from selected budget user or all money transfers from the budget user to selected company.
- Application also shows data about public procurements, information about the business entities in Slovenia, accounting entries of payments (for direct budget users), and some other additional data (e.g.: data about tax debtors and transfers to the so called favourable tax environments (“tax havens”), etc.

Financial transactions of public sector

Data on monthly sums of funds received from individuals of public funds

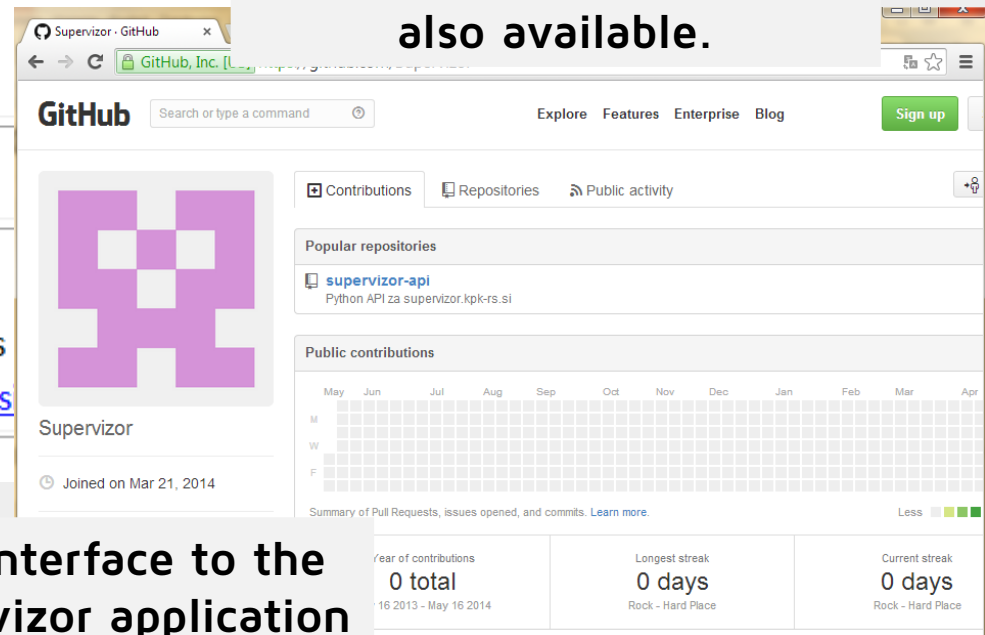
The data is located in a CSV text files (individual fields are in a format with Unix/Linux end of the line. The file has a set of individual files contain data for each year (since 2003). Files are on a monthly basis. Each file is compressed in gz format.

Data structure description

<i>field</i>	<i>field description</i>
<u>sifra_pu</u>	Budget user code.
<u>(budget user code)</u>	Codes of budget users (http://www.ujp.gov.si)

Part of Erar data is available for download in machine readable form.

Description of data is also available.



The screenshot shows a GitHub profile for 'Supervisor', which joined on Mar 21, 2014. The profile features a purple pixelated avatar. The 'Popular repositories' section lists 'supervisor-api' (Python API za supervisor.kpk-rs.si). The 'Public contributions' section shows a calendar grid for May through April, with a summary of 0 total contributions, 0 longest streak, and 0 current streak. The streaks are noted as 'Rock - Hard Place'.

API interface to the Supervisor application is also provided.

<http://erar.kpk-rs.si>

Financial transactions of public sector

With Erar database of financial transactions of public sector several analysis is possible. For instance:

- analysis of payments to media companies;
- analysis of grants paid to private companies;
- correlation of government in charge with financial flows;
- detecting interesting financial flows patterns;
- ...

The image shows a screenshot of a data analysis interface. On the left, there are two panels: 'Cube Structure' and 'Pivot Structure'. The 'Cube Structure' panel shows a hierarchy of dimensions: Measures, Obdobje, Proračunski uporabnik, and Lokacija podjetja. The 'Pivot Structure' panel shows a pivot table with columns for Measures (Nepovratna sredstva, Izplačila) and rows for Proračunski uporabnik categories.

Proračunski uporabnik	Nepovratna sredstva	Izplačila
Agencije	40.637.791,75	3.084.264.018,36
Bolnišnice	98.444,37	5.780.516.587,22
CSD	76.736,98	155.772.529,96
Dijaki in študentski d.	9.544,37	140.439.757,31
Domovi za starejše	16.792,25	1.127.886.285,47
Drugi javni zavodi	110.860.153,81	15.841.399.375,82
Glasbene šole	12.832,97	107.286.105,63
Inšpektorati	47.170.563,35	107.992.109,97
Inštituti	223.587,89	1.321.001.783,04
Javni skladi	288.873.295,73	4.795.237.348,12
Javni zavodi za šport	9.267.361,56	301.529.239,62
Knjižnice	573.293,18	299.811.871,92
Krajevna skupnost	4.949.870,64	397.359.044,7
Lekarne	315.215,66	5.435.168.755,02
Ljudske univerze	257.558,39	70.594.866,22
Min. in org. v sestavi	31.321.475.251,68	96.559.551.936,67
Muz., galer., gled., arh., parki	54.636,9	447.180.308,19
Narodna skupnost	2.229.875,11	13.486.364,29
Občine	1.172.483.413,72	23.730.934.338,98

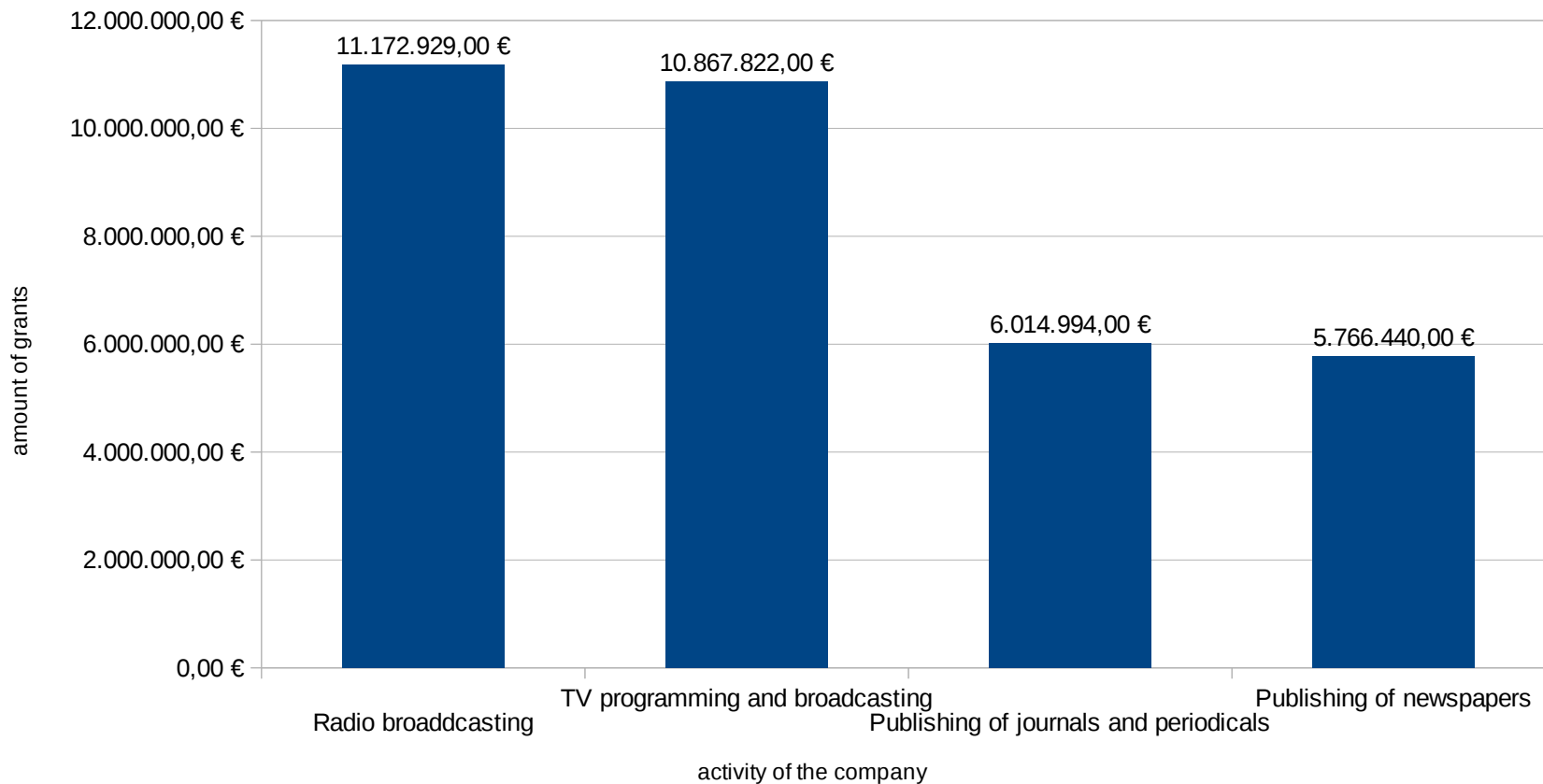
On the right, there is a list of transactions with columns: 'kacije', 'racun_v_dobro', and 'naziv_prejemnika'. The list shows various payment details and recipient names.

kacije	racun_v_dobro	naziv_prejemnika
	010000000200097	Nova Ljubljanska banka d.d.
	012086030644985	OSNOVNA ŠOLA PRESERJE
	012086030644985	OSNOVNA ŠOLA PRESERJE
	031081010513116	KRIM D.O.O.
	029800052180243	D PLAST ALEKSEJ DOBROVNIK S.P.,
	029800052180243	D PLAST ALEKSEJ DOBROVNIK S.P.,
	029800052180243	D PLAST ALEKSEJ DOBROVNIK S.P.,
	051008000087486	SRC.SI D.O.O.
	023730091203307	GIRUS - KRŽAN BRANKO S.P.,
	023730017615227	PGD PIROŠICA
	023730013809044	PGD SKOPICE
	031501000003824	INTEL SERVIS, D.O.O.
	031501000003824	INTEL SERVIS, D.O.O.
	051008010455719	DNEVNIK D.D.
	029700010636797	KVM-SOFTWARE, RAČUN. PROGR. INŽ., D.O.O
	330007321322535	ENERGETSKI SERVIS D.O.O.
	012096450784722	KRAJEVNA SKUPNOST BIZELJSKO
	031181000002191	SLUGA ALEŠ S.P.
	023420015567621	CADIS D.O.O.
	023400014211077	TUROPOLJE MURSKA SOBOTA D.O.O.
	012100100011634	OBCINA TIŠINA-PRORAČUN

Financial transactions of public sector

Grants to media companies

Grants from January, 1st 2003 till October, 1st 2014

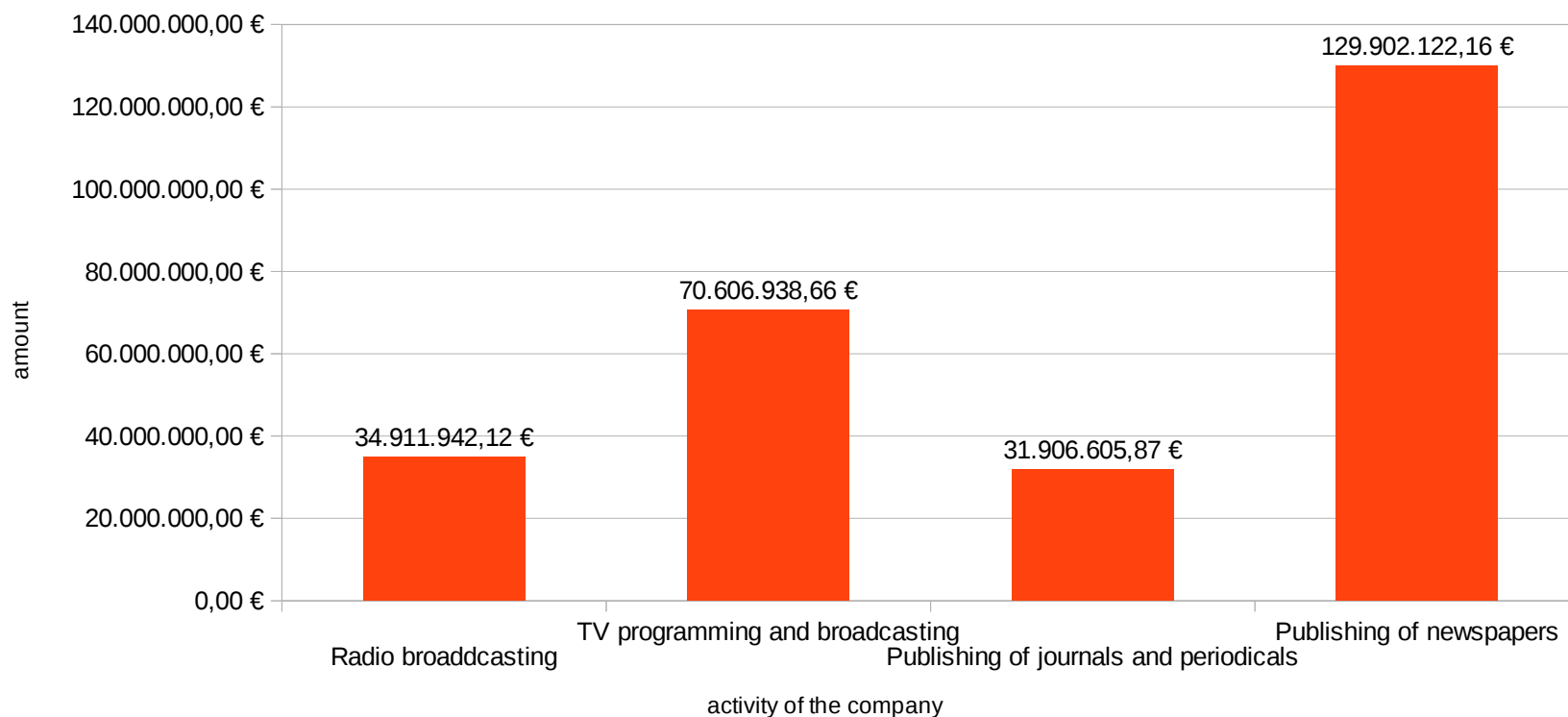


Amount of grants from Slovenian budget users to media companies. Media companies were identified from business registry by their main activity registered.

Financial transactions of public sector

All financial transactions to media companies

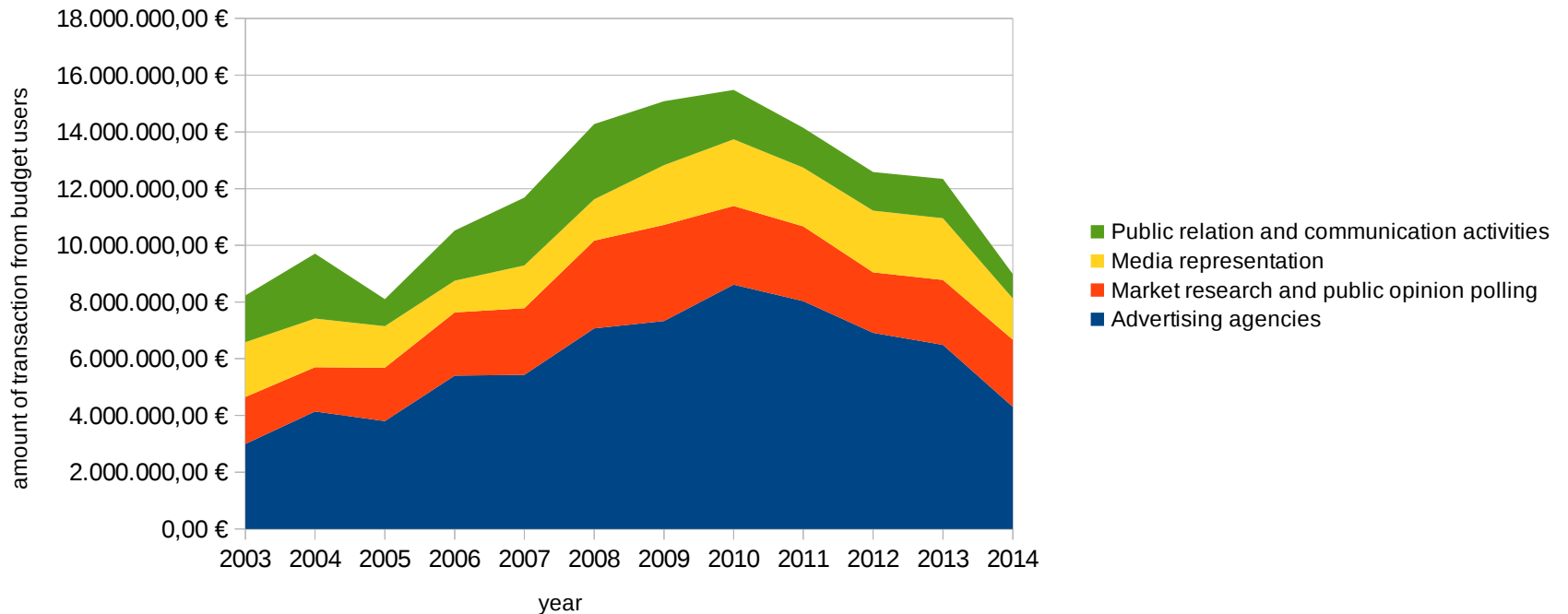
Payments from January, 1st 2003 till October, 1st 2014.



Amount of payments from Slovenian budget users to media companies. Media companies were identified from business registry by their main activity registered.

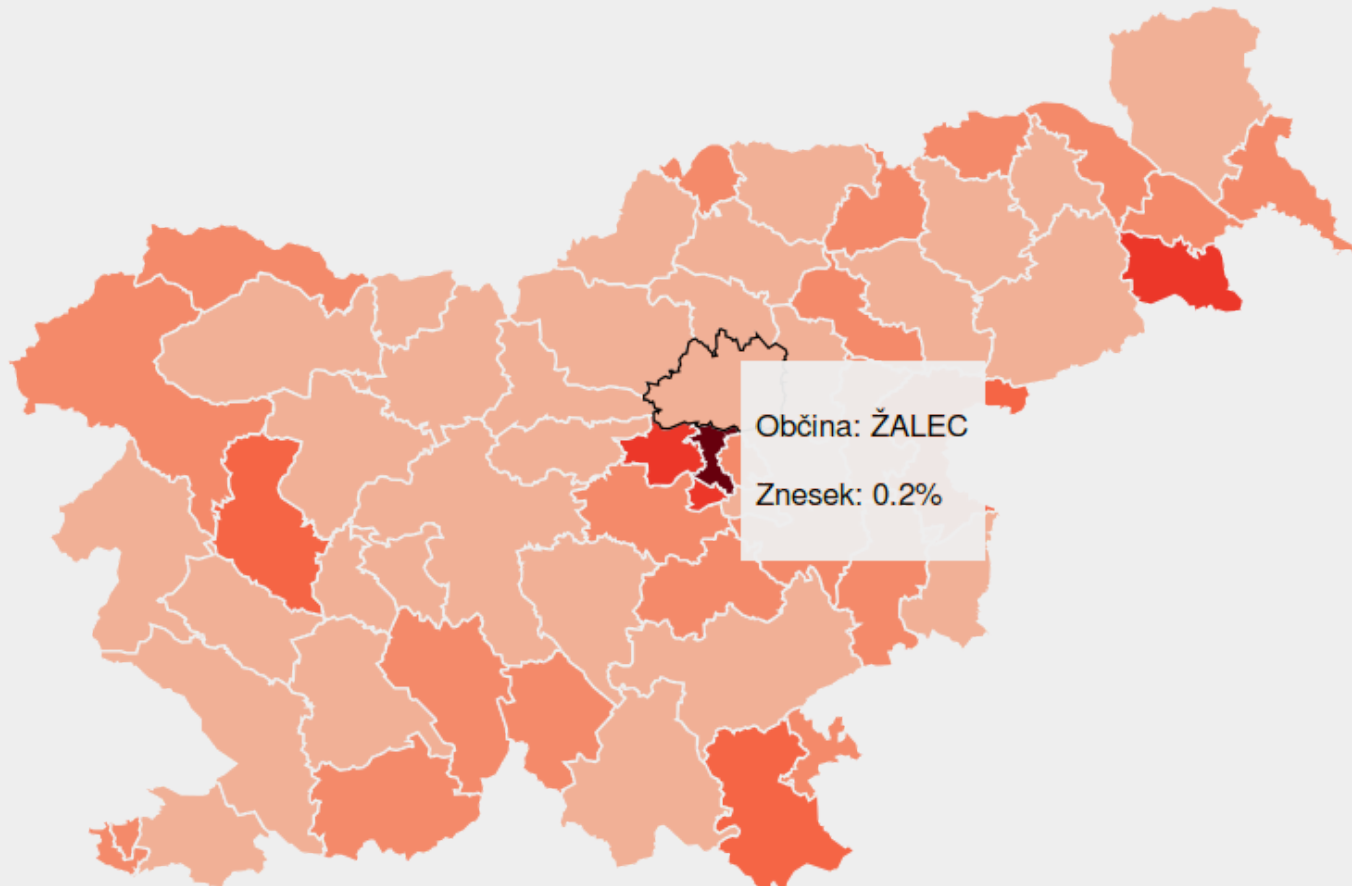
Financial transactions of public sector

Annual transactions to companies registered to...



Annual amount of payments from Slovenian budget users to companies registered for management consultancy, PR, advertising, marketing research, etc. Types of companies were identified from business registry by their main activity registered.

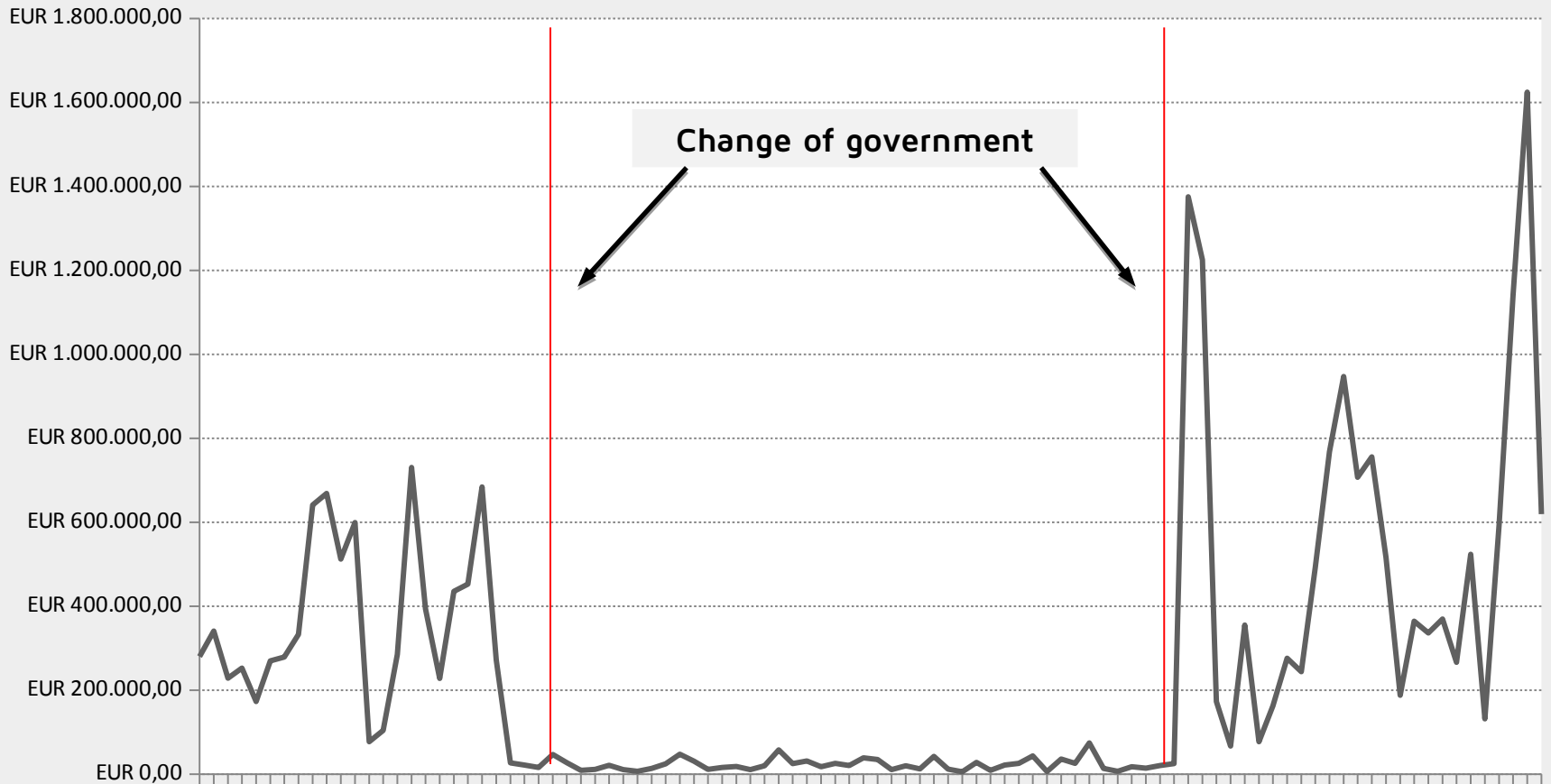
Financial transactions of public sector



Visualisation of share of grants paid to companies located in different municipalities.

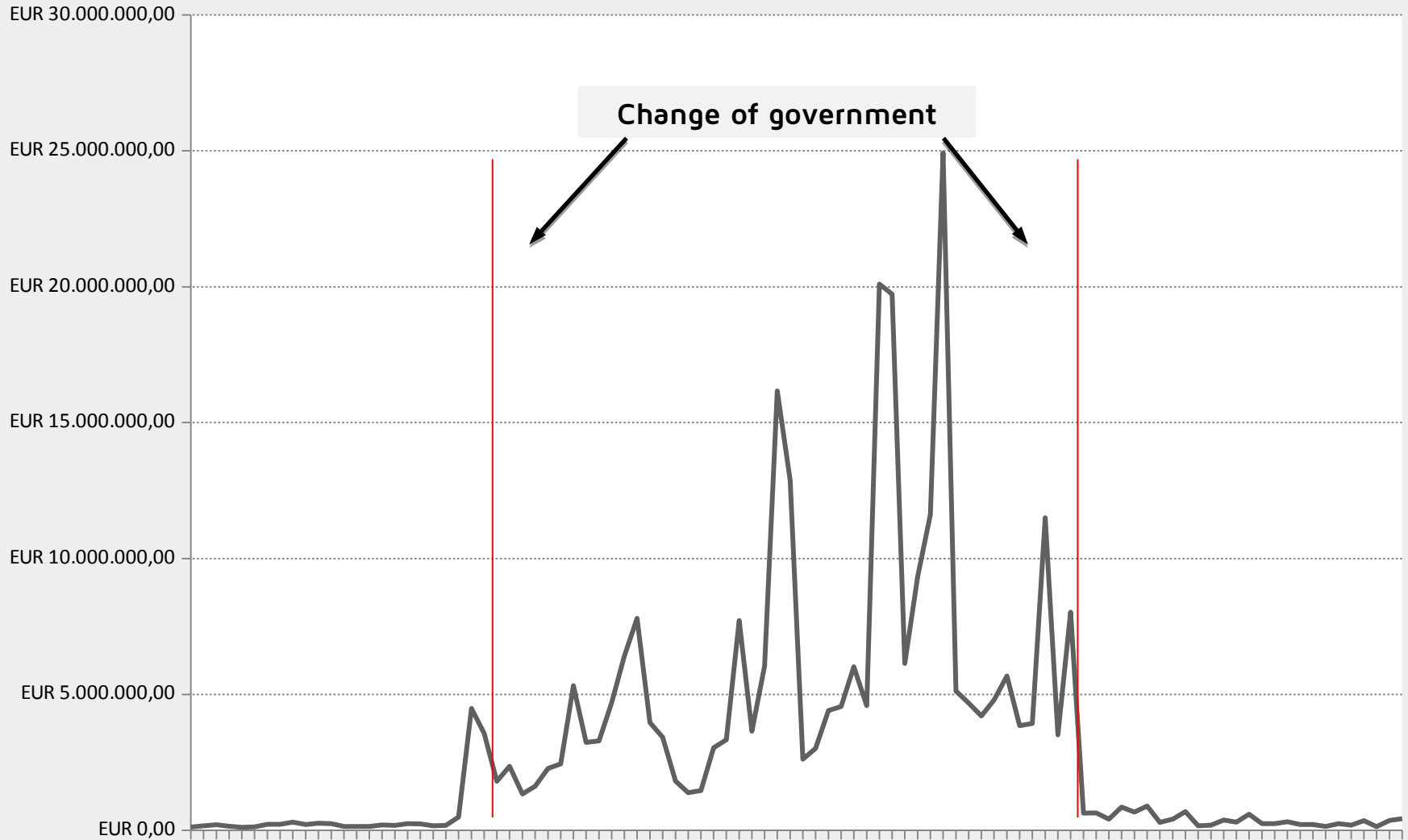
Financial flow analysis

Summary of monthly payments to 65 companies



Financial flow analysis

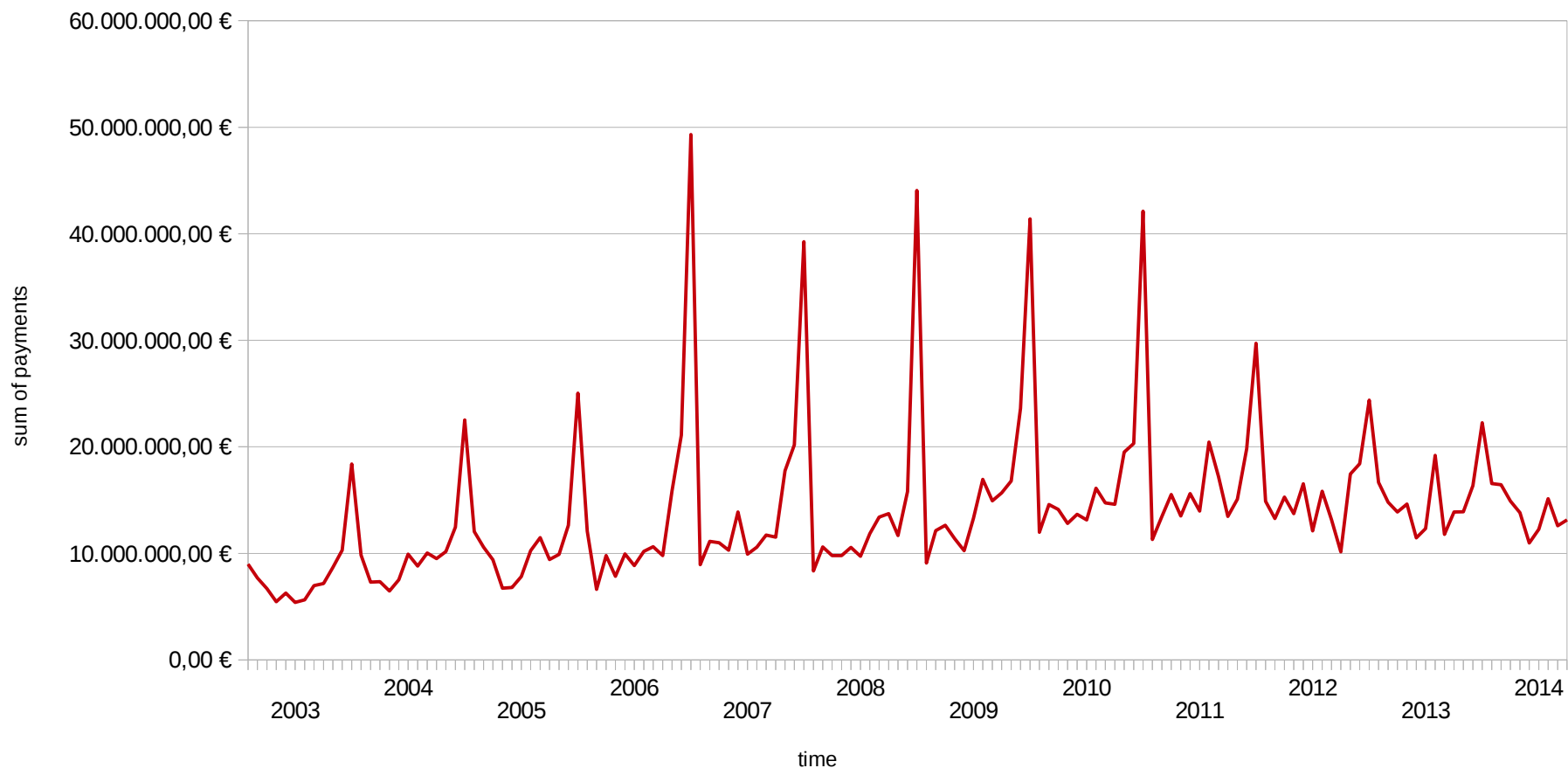
Summary monthly payments to 252 companies



Financial flow analysis

Monthly payments to companies registered for 'IT activities'

In Slovenian Business Register there are 6223 companies registered for 'IT activities', 3796 of them had business with budget users.



Other Supervisor/Erar datasets

Some other datasets:

- impositions of sanctions for offences and inspector decisions;
- list of tax debtors (published on *Tax Administration's* website);
- database of financial transfers to the so-called favourable tax environments ("tax havens"), published on the *Office for Money Laundering Prevention* website;
- accounting entries of payments for direct budget users from MFERAC database (is already included in Erar database, but not published in machine readable form yet);

Other Supervisor/Erar datasets

Some other datasets:

- list of restrictions of business due to conflict of interest, published by *Commission for Prevention of Corruption*;
- database of fiscal stimulus - “*Government for the development of Slovenia*”;
- financial transactions from European funds;
- adopted legislation;
- election candidates;
- list of registered media companies;
- ...

Other Supervisor/Erar datasets



Visualisation of tax debtors (author Gašper Žejn, <<http://www.kiberpipa.org/~hruske/stuff/durs.html>>).

Example: Detection of restriction
on business activities violations



Restrictions on business activities

Article 35 of the Integrity and Prevention of Corruption Act
(Restrictions on business activities and the consequences of violations)

(1) A public sector body or organisation which is committed to conducting a public procurement procedure in accordance with the regulations on public procurement or which carries out the procedure for granting concessions or other forms of public-private partnership, may not order goods, services or construction works, enter into public-private partnerships or grant special and exclusive rights to entities in which the official who holds office in the body or organization concerned or in cases where the official's family member has the following role:

- participating as a manager, management member or legal representative; or
- has more than a 5% level of participation in the founders' rights, management or capital, either by direct participation or through the participation of other legal persons.

Restrictions on business activities

Article 35 of the Integrity and Prevention of Corruption Act
(Restrictions on business activities and the consequences of violations)

...

(5) Officials shall communicate the name, registration number and head office of those entities with which they or their family members have a relationship, as specified in paragraph ti of this Article, to the body in which they hold office within one month after taking office and then no later than within eight days of any change occurring. The body shall submit the list of entities referred to in the preceding sentence of this Article to the Commission no later than within ti5 days of receipt of information on or a notification of changes regarding the entities. The Commission shall publish the list of entities referred to in the first sentence of this paragraph on its website every month.

...

List of restrictions

Komisija za preprečevanje korupcije - Mozilla Firefox

https://registri.kpk-rs.si/registri/omejitve_poslovanja/seznam/#

KOMISIJA ZA PREPREČEVANJE KORUPCIJE

Začetna stran > Omejitve poslovanja > Seznam veljavnih omejitev poslovanja

Hitro iskanje:

Počisti vse

Strojno čitljiv seznam v formatu JSON
Strojno čitljiv seznam v formatu CSV

Organizacija	Matična	Davčna	Šifra PU	Naslov organizacije	Pošta organizacije	Omejitev do kmetijskega gospodarstva ali do poslovnega subjekta	Naziv poslovnega subjekta	Matična
UPRavno sodišče Republike Slovenije	1190989000	79787975	46116	FAJFARJEVA ULICA 033	1000 Ljubljana	ps	FUNDACIJA DR. BRUNO BRESCHI, USTANOVA ZA OHRANJANJE STAREJŠIH DEL SLOVENSKEGA SLOVSTVA	1421549000
UPRavno sodišče Republike Slovenije	1190989000	79787975	46116	FAJFARJEVA ULICA 033	1000 Ljubljana	ps	DANCEQUAKE, proizvodnja, storitve, trgovina, d.o.o.	1702319000
UPRavno sodišče Republike Slovenije	1190989000	79787975	46116	FAJFARJEVA ULICA 033	1000 Ljubljana	ps	CELJSKA MOHORJEVA DRUŽBA, založništvo, trgovina in storitve, d.o.o.	2028484000
UPRavno sodišče Republike Slovenije	1190989000	79787975	46116	FAJFARJEVA ULICA 033	1000 Ljubljana	ps	BARTENJEV MITJA - ODVETNIK	2367939000

Supervizor/Erar database in machine readable form

The image displays two overlapping browser windows. The top window shows the Supervizor website, which is a web application for monitoring corruption. The bottom window shows a PDF document titled 'Opis_kopije_podatkov_Supervizor.pdf'.

Supervizor - Mozilla Firefox
supervizor.kpk-rs.si/podatki/

Nazaj na KPK | Projekt Transparentnost | Navodila

KOMISIJA ZA PREPREČEVANJE KORUPCIJE

SUPERVIZOR
SPLETNA APLIKACIJA ZA SPREMLJANJE

Podatki

Na tej strani so na voljo podatki o mesečnih seštevkih sredstev.

OPOZORILO: podatki se lahko brez opozorila da čiščenje podatkov o plačah in odtegljajih o

- 2003 - [transakcije](#)
- 2004 - [transakcije](#)
- 2005 - [transakcije](#)
- 2006 - [transakcije](#)
- 2007 - [transakcije](#)
- 2008 - [transakcije](#)
- 2009 - [transakcije](#)

Opis_kopije_podatkov_Supervizor.pdf - Mozilla Firefox
https://www.kpk-rs.si/upload/datoteke/Opis_kopije_podatkov...

Page: 1 of 4

Podatki o mesečnih seštevkih prejetih sredstev po posameznih proračunskih uporabnikih in prejemnikih javnih sredstev


Podatki se nahajajo v CSV tekstovnih datotekah (posamezna polja so ločena s podpičjem) v Unicode (UTF-8) formatu z Unix/Linux koncem vrstice. Datoteka ima vzglavje z imeni podatkovnih polj. V datotekah se nahajajo podatki za vsako leto posebej (od leta 2003 dalje), za zadnje leto pa za vsak mesec posebej. Posamezna datoteka je kompresirana v .gz obliko.

Opis strukture podatkov

Polje	Opis polja
sifra_pu	Šifra proračunskega uporabnika. Šifre proračunskih uporabnikov se nahajajo v Registru proračunskih uporabnikov (http://www.ujp.gov.si/dokumenti/dokument.asp?id=354). Za nekatere ukinjene organe so šifre proračunskih uporabnikov dodane ročno. Podatki so v celoštevilskem formatu (<i>integer</i>).

Overview of reported restrictions

*Integrity And
Prevention of
Corruption Act was
enacted in 2003*



	K	L	M	O	P	Q
	PS_DAVCNA	MID	PS_NASLOV	DATUM_OD	DATUM_DO	
1						
2	76594912		IDRIJSKA CESTA 013	1970-01-01		
3	33191093		PREŠERNOVA CESTA 008	1979-10-01		
4	27699242		VRAZOVA ULICA 005	1980-01-01		
5		100243908		1981-01-01		
6	17142776		LOKEV 165	1981-06-01		
7	68256183		ULICA STANETA ROZMANA 016	1981-07-01		
8		100200859		1219 1984-08-01		
9	34423974		VAŠE 011 C	1215 Medvode	1986-09-23	
10		100268238		3224 dobjje pri planini	1988-07-26	
11	74735705		CESTA NA BOKALCE 040	1000 Ljubljana	1989-06-30	
12	48079421		NA POKLONU 007	5281 Spodnja Idrija	1989-08-01	2013-02-23
13	97905445		OPATJE SELO 003	5291 Miren	1989-10-17	
14	17454352		AHLINOVA ULICA 001	1291 Škofljica	1989-12-18	
15		100294298		3326 Topolšica	1990-01-01	
16		100314697		3325 Šoštanj	1990-01-01	
17	12455458		EFENKOVA CESTA 061	3320 Velenje	1990-01-01	
18	46871535		KOČEVARJEVA ULICA 004	8000 Novo mesto	1990-01-13	2014-12-29
19	84774533		HOLCERJEVA ULICA 027	2204 Miklavž na Dravskem polju	1990-03-01	
20	80857175		PANONSKA ULICA 001	9250 Gornja Radgona	1990-04-09	
21	91824036		KIDRIČEVA ULICA 035	5270 Ajdovščina	1990-09-10	
22	81121903		SAMOVA ULICA 012 A	1000 Ljubljana	1990-10-11	
23	91612900		TEPANJE 063	3210 Slovenske Konjice	1990-11-14	
24	92302025		GREGORČIČEVA ULICA 015	5280 Idrija	1990-11-20	
25	69846103		Letališka cesta 032	1000 Ljubljana	1991-01-09	
26	24058807		GRADAC 086	8332 Gradac	1991-02-12	
27	91201055		SMOLINCI 088	2236 Cerkevjak	1991-04-01	
28	10488332		Ulica XIV. divizije 014	3000 Celje	1991-04-09	
29	64679691		DOBJJE PRI PLANINI 032	3224 Dobje pri Planini	1991-05-07	
30	82306567		PENOJE 010	3215 Loče	1991-05-28	
31	52087964		LIPTOVSKA ULICA 028	3210 Slovenske Konjice	1991-06-12	
32	16466543		LINHARTOVA CESTA 001	1000 Ljubljana	1991-06-17	
33	72903295		LEVSTIKOVA ULICA 001	6310 Izola - Isola	1991-08-22	
34	25213800		Meža 114	2370 Dravograd	1991-09-19	

Overview of reported restrictions

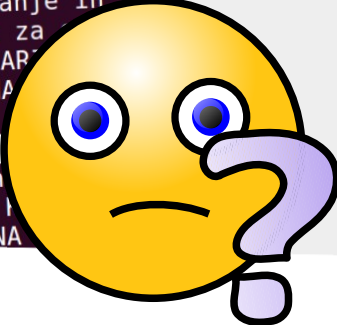
matej@cryptoloop: ~

organ	
OBČINA IDRIJA	Center za idrijsko dediščino
OBČINA LENDAVA - LENDVA KÖZSEG	DVOJEZIČNA OSNOVNA ŠOLA LENDAVA
MESTNA OBČINA MARIBOR	JAVNI MEDOBČINSKI SKLAD MARIBOR
OBČINA ČRNOMELJ	JAVNI ZAVOD ZA KOLPA
OBČINA ROGATEC	JAVNI ZAVOD ZA RIZEM IN RAZVOJ OBČINE KOPER ENTE
MESTNA OBČINA KOPER	JAVNI ZAVOD ZA RIZEM IN RAZVOJ OBČINE KOPER ENTE
OBČINA ŠKOFJA LOKA	KNJIŽNICA OBČINE ŠKOFJA LOKA
OBČINA ŠMARJE PRI JELŠAH	KNJIŽNICA OBČINE ŠMARJE PRI JELŠAH
OBČINA ZAGORJE OB SAVI	KRAJEVNA SKUPNOST TOLDOVA
OBČINA ZAGORJE OB SAVI	KRAJEVNA SKUPNOST TOLDOVA
OBČINA ČRNOMELJ	KRAJEVNA SKUPNOST TOLDOVA
OBČINA ZAGORJE OB SAVI	KRAJEVNA SKUPNOST TOLDOVA
MESTNA OBČINA MARIBOR	KULTURNI CENTER - DOM ZAGORJE OB SAVI
OBČINA ZAGORJE OB SAVI	MARIJSKA RAZVOJNA AGENCIJA p.o.
OBČINA PODLEHNIK	MLADINSKI CENTER ZAGORJE OB SAVI
OBČINA KIDRIČEVO	OBČINA PODLEHNIK
OBČINA DOBJE	OSNOVNA ŠOLA CIRKOVCE
OBČINA KOSTANJEVICA NA KRKI	OSNOVNA ŠOLA DOBJE
OBČINA ODRANCI	OSNOVNA ŠOLA JOŽETA GORJUPA KOSTANJEVICA NA KRKI
MESTNA OBČINA VELENJE	OSNOVNA ŠOLA ODRANCI
OBČINA ŠTORE	OSNOVNA ŠOLA ŠALEK VELENJE
	OSNOVNA ŠOLA ŠTORE - VRTEC LIPA ŠTORE



Do they have restriction of business with themselves???

Overview of reported restrictions



matej@cryptoloop: ~

```
OBČINA ODRANCI
MESTNA OBČINA VELENJE
OBČINA ŠTORE
MESTNA OBČINA KOPER
OKROŽNO SODIŠČE V MARIBORU
MESTNA OBČINA NOVO MESTO
OBČINA ČRENŠOVCI
OBČINA ŠKOFJA LOKA
OBČINA IVANČNA GORICA
MESTNA OBČINA MARIBOR
OBČINA LENDAVA - LENDVA KÖZSEG
OBČINA PODCETRTEK
MESTNA OBČINA VELENJE
OBČINA ROGATEC
OBČINA KOČEVJE
OBČINA ŠKOFJA LOKA
OBČINA LENDAVA - LENDVA KÖZSEG
OBČINA SLOVENSKE KONJICE
OBČINA ŠMARJE PRI JELŠAH
MESTNA OBČINA VELENJE
OBČINA ŽALEC
(42 rows)
(END)
```

OSNOVNA ŠOLA ODRANCI
OSNOVNA ŠOLA ŠALEK VELENJE
OSNOVNA ŠOLA ŠTORE - VRTEC LIPA ŠTORE
POKRAJINSKI MUZEJ KOPER, MUSEO REGIONALE DI CA
RADIOTELEVIZIJA SLOVENIJA javni zavod, Ljubljana
RAZVOJNI CENTER NOVO MESTO, Svetovanje in razvoj
SAUBERMACHER - KOMUNALA - podjetje za
SMUČARSKO TURISTIČNI CENTER STC STAR
SREDNJA ŠOLA JOSIPA JURČIČA IVANČNA
ŠPORTNI OBJEKTI MARIBOR

TURIZEM LENDAVA d.o.o.
TURIZEM PODCETRTEK, gospodarsko in
VETERINA ŠOŠTANJ d.o.o., Šoštanj, M
ŠPORTNI OBJEKTI MARIBOR

PRIS - Poslovni register SLOVENIJE

Podatki o subjektu **TURIZEM LENDAVA d.o.o.**, matična številka: 2080036000 [Preverite boniteto eS.BON](#)
[Več v FI-PO](#)

osnovni podatki o subjektu	družbeniki in poslovni deleži	osebe, pooblaščenec za zastopanje	člani organa nadzora	skupščinski sklepi	razno	sprememba družbene pogodbe / statuta
--	---	---	--------------------------------------	------------------------------------	-----------------------	--

DRUŽBENIKI

zap. št. družbenika:	492897
identifikacijska številka:	5874645000 - matična številka (pravna oseba je vpisana v PRS)
firma:	OBČINA LENDAVA - LENDVA KÖZSEG
Naslov:	Lendava, Glavna ulica 20, 9220 Lendava - Lendava
vrsta odgovornosti za obveznosti družbe:	ne odgovarja
datum vstopa:	27.01.2005
zap. št. družbenika:	576081
identifikacijska številka:	2220385000 - matična številka (pravna oseba je vpisana v PRS)
firma:	ZVEZA DRUŠTEV ZA MEDKULTURNI DIALOG MURANIA - MURANIA-EGYESÜLETEK SZÖVETÉGE A KULTÚRÁK KÖZÖTTI PÁRBESZÉDÉRT
Naslov:	Lendava, Bánffyjev trg 1, 9220 Lendava - Lendava
vrsta odgovornosti za obveznosti družbe:	ne odgovarja
obseg odgovornosti:	/
datum vstopa:	20.01.2009
zap. št. družbenika:	740816
identifikacijska številka:	5301971000 - matična številka (pravna oseba je vpisana v PRS)

Company established by the municipality

Cleaning the data

For data cleaning we need *Registry of budget users* (available on web) and *Business registry of Slovenia* (accessible through Access to Public Information Act)...

In first step we...

- remove entities which are in fact budget users;
- remove business entities established by budget users;
- *for other business entities add type, for associations we add type of association;*
- remove »*associations working in public interest*« (firemen associations, hunting and fishing associations, handicapped persons associations);
- remove political parties.

Cleaning the data

```
omejitve_poslovanja.sql (~/.Namizje/datoteke/open_data/omejitve_poslovanja) - gedit
Odpri Shrani Razveljavi
omejitve_poslovanja.sql x Supervisor_psql.txt x
# zacasna tabela
select organ, kpk_omejitve_poslovanja.sifra_pu, tip_omejitve, ps_naziv, ps_davcna, datum_od, datum_do,
datum_transakcije, znesek_transakcije, namen into temp_krsitve_omejitev from kpk_omejitve_poslovanja,
supervisor_transakcije where (kpk_omejitve_poslovanja.sifra_pu = supervisor_transakcije.sifra_pu) and
(kpk_omejitve_poslovanja.ps_davcna = supervisor_transakcije.davcna_stevilka::text) and (datum_od <
datum_transakcije) and (datum_transakcije < datum_do) and (extract(year from datum_transakcije) >= 2014);

# ali je prejemnik PU?
alter table temp_krsitve_omejitev add column ps_je_pu boolean;
update temp_krsitve_omejitev set ps_je_pu = FALSE;
update temp_krsitve_omejitev set ps_je_pu = TRUE from rpu where (ps_davcna = rpu.davcna::text);

# tip prejemnika
alter table temp_krsitve_omejitev add column ps_oblika text;
update temp_krsitve_omejitev set ps_oblika = kategorija from rpu where (ps_je_pu = TRUE) and (ps_davcna =
rpu.davcna::text);
update temp_krsitve_omejitev set ps_oblika = oblika from prs_enota_rs where (ps_je_pu = FALSE) and
(ps_davcna = prs_enota_rs.davcna::text);
update temp_krsitve_omejitev set ps_oblika = drustvo from prs_enota_rs where (ps_je_pu = FALSE) and
(ps_oblika ~* 'društvo') and (ps_davcna = prs_enota_rs.davcna::text);
update temp_krsitve_omejitev set ps_oblika = trim(ps_oblika);

# ali je ustanovitelj prejemnika PU
alter table temp_krsitve_omejitev add column ps_ustanovil_pu boolean;
update temp_krsitve_omejitev set ps_ustanovil_pu = FALSE;
update temp_krsitve_omejitev set ps_ustanovil_pu = TRUE from prs_enota_rs, prs_ustanovitelji, rpu where
(ps_davcna = prs_enota_rs.davcna::text) and (prs_enota_rs.enota = prs_ustanovitelji.enota) and
(p_subjekt_maticna = rpu.maticna);
```


Results

- Number of all public bodies reported restrictions of business: **294**
- Number of all public bodies, where financial transactions were found in 2014 and 2015: **154**
- Number of public bodies where there is suspicion of violation (cleaned data): **148**

- Number of suspicious transactions in 2014: **2.912.193,03 EUR**
- Number of suspicious transactions till end of April 2015: **1.658.190,47 EUR**

- **Finally we have 391 pairs public body – company on the list.** Can we remove more data?
 - all remaining associations and clubs?
 - all companies that contain word »*public*« in their name, even if they are not registered as public companies?

Additional cleaning of data

prejemki	st_transakcij	organ	
63744.8000	734	MESTNA OBČINA MARIBOR	MARIBORSKI VODOVOD javno podjetje d.d.
111916.4800	110	OKROŽNO SODIŠČE V KOPRU	BANKA KOPER d.d. Dvojezična firma: BANKA KOPER d.d. - s.p.a
5972.1400	75	MESTNA OBČINA NOVA GORICA	GEN-I, trgovanje in prodaja električne energije, d.o.o.
1107889.0400	72	MESTNA OBČINA LJUBLJANA	WALDORFSKA ŠOLA LJUBLJANA
34322.9900	62	OBČINA BRDA	VINSKA KLET "GORIŠKA BRDA" druga, kmetijstvo, proizvodnja
140529.6400	33	MESTNA OBČINA KRANJ	ZAVOD CARNICA - Zavod za kulturo in turizem
16074.8600	32	OBČINA POSTOJNA	ZVEZA DRUŠTEV MLADINSKI CENTER POSTOJNA
8815.0800	30	OKRAJNO SODIŠČE V LJUBLJANI	Odvetniška pisarna Leber Miklič, Rozman d.o.o.
6291.4900	30	OBČINA SLOVENSKA BISTRICA	KADRING - kadrovsko in poslovno svetovanje, d.o.o.
6166.9200	30	OBČINA PIVKA	NEMEZA, finančni inženjering in storitve d.o.o.
12580.6100	28	OBČINA VOJNIK	MPRINT servis in prodaja računalniške opreme, d.o.o.
10139.1800	28	MESTNA OBČINA MURSKA SOBOTA	DRUŠTVO UPOKOJENCEV MURSKA SOBOTA
12073.6700	25	OKROŽNO SODIŠČE V LJUBLJANI	OREL LADO - ODVETNIŠTVO
35952.0200	24	MESTNA OBČINA MARIBOR	DRUŠTVO ZA POMOČ SAMOPOMOČ NA PODROČJU ZASVOJENOSTI ZDRA
23404.4200	24	OBČINA ČRNOMELJ	MLADINSKI CENTER BIT angleški prevod firme: YOUTH CENTER BI
10809.2000	24	OBČINA SLOVENSKE KONJICE	EKOTIP - KEROMAN KOLAR proizvodnja in storitve, d.n.o. S
37560.1800	23	OBČINA ŽIRI	M SORA, trgovina in proizvodnja, d.d.
15747.6800	23	OBČINA SLOVENSKA BISTRICA	KULTURNO IZOPRAVEVALNO DRUŠTVO KERAMIKOV IN LONČARJEV PODRA

[Preverite boniteto eS.BON](#)

[Več v FI-PO](#)

Podatki o subjektu MARIBORSKI VODOVOD javno podjetje d.d., matična številka: 5067880000

osnovni podatki o subjektu	druženiki in poslovni deleži	osebe, pooblaščene za zastopanje	člani organa nadzora	skupščinski sklepi	razno	sprememba družbene pogodbe / statuta
--	--	--	--------------------------------------	------------------------------------	-----------------------	--

OSNOVNI PODATKI O SUBJEKTU

status subjekta:	vpisan
datum vpisa subjekta v sodni register:	23.04.1991
matična številka:	5067880000
davčna številka:	SI 68041527
firma:	MARIBORSKI VODOVOD javno podjetje d.d.
skrajšana firma:	MARIBORSKI VODOVOD d.d.
sedež:	Maribor
poslovni naslov:	Jadranska cesta 24, 2000 Maribor
pravnoorganizacijska oblika:	Delniška družba d.d.
število delnic:	285.489
vrsta organa nadzora:	Nadzorni svet

Type of company: Joint-stock company (d.d.)

However from the name of the company we can deduce that it is public company

Results

- Transactions of remained business entities:
3.510.827,56 EUR.

1	A	B	C	D	E
organ	datum transakcije	znesek transakcije	namen	ps_naziv	
2	OBČINA SEŽANA	2014-01-31	12.507,67 €	13300298 30.12.2013	HORTIKULTURA SEŽANA vrtnarstvo, gozdarstvo in komunalne storitve d.o.o.
3	OBČINA ZREČE	2014-03-21	2.354,07 €	2014/1000061 8-14 31.01.2014 TRANS	TRANSLES, proizvodno trgovsko podjetje, d.o.o.
4	OBČINA ZREČE	2015-01-30	2.154,96 €	2014/1001953 68-14 30.11.2014 TRAN	TRANSLES, proizvodno trgovsko podjetje, d.o.o.
5	OBČINA IVANČNA GORICA	2014-09-19	4.453,00 €	2014/9002336 NAKUP ZEMLJIŠČA DEDNI	KMETIJSKA ZADRUGA STIČNA zadruga z omejeno odgovornostjo v Ivančni Gorici
6	OBČINA ZREČE	2015-03-27	2.154,96 €	2015/1000111 2 03.02.2015 TRANSLES	TRANSLES, proizvodno trgovsko podjetje, d.o.o.
7	MESTNA OBČINA LJUBLJANA	2014-01-03	49.282,20 €	2160329 REFUNDACIJE-DOTACIJE	WALDORFSKA ŠOLA LJUBLJANA
8	MESTNA OBČINA LJUBLJANA	2014-01-06	8.696,07 €	2163349 STORITVE	WALDORFSKA ŠOLA LJUBLJANA
9	MESTNA OBČINA LJUBLJANA	2014-01-22	5.877,24 €	2207004 REFUNDACIJE-DOTACIJE	WALDORFSKA ŠOLA LJUBLJANA
10	MESTNA OBČINA LJUBLJANA	2014-02-06	8.857,96 €	2248560 STORITVE	WALDORFSKA ŠOLA LJUBLJANA
11	MESTNA OBČINA LJUBLJANA	2014-02-17	49.537,61 €	2269922 REFUNDACIJE-DOTACIJE	WALDORFSKA ŠOLA LJUBLJANA
12	MESTNA OBČINA LJUBLJANA	2014-02-17	9.752,28 €	2269925 REFUNDACIJE-DOTACIJE	WALDORFSKA ŠOLA LJUBLJANA
13	MESTNA OBČINA LJUBLJANA	2014-03-03	6.740,45 €	2305635 REFUNDACIJE-DOTACIJE	WALDORFSKA ŠOLA LJUBLJANA
14	MESTNA OBČINA LJUBLJANA	2014-03-04	49.282,20 €	2306335 REFUNDACIJE-DOTACIJE	WALDORFSKA ŠOLA LJUBLJANA
15	MESTNA OBČINA LJUBLJANA	2014-03-06	8.791,88 €	2315394 STORITVE	WALDORFSKA ŠOLA LJUBLJANA
16	MESTNA OBČINA LJUBLJANA	2014-03-19	49.198,05 €	2341036 REFUNDACIJE-DOTACIJE	WALDORFSKA ŠOLA LJUBLJANA
17	MESTNA OBČINA LJUBLJANA	2014-04-03	9.279,25 €	2387160 STORITVE	WALDORFSKA ŠOLA LJUBLJANA
18	MESTNA OBČINA LJUBLJANA	2014-04-30	8.325,17 €	2457799 REFUNDACIJE-DOTACIJE	WALDORFSKA ŠOLA LJUBLJANA
19	MESTNA OBČINA LJUBLJANA	2014-05-07	49.198,05 €	2473606 REFUNDACIJE-DOTACIJE	WALDORFSKA ŠOLA LJUBLJANA
20	MESTNA OBČINA LJUBLJANA	2014-05-23	9.361,70 €	2502348 STORITVE	WALDORFSKA ŠOLA LJUBLJANA
21	MESTNA OBČINA LJUBLJANA	2014-05-30	5.000,87 €	2532449 REFUNDACIJE-DOTACIJE	WALDORFSKA ŠOLA LJUBLJANA
22	MESTNA OBČINA LJUBLJANA	2014-05-30	2.381,45 €	2532513 REFUNDACIJE-DOTACIJE	WALDORFSKA ŠOLA LJUBLJANA
23	MESTNA OBČINA LJUBLJANA	2014-06-05	13.073,14 €	2545523 STORITVE	WALDORFSKA ŠOLA LJUBLJANA
24	MESTNA OBČINA LJUBLJANA	2014-06-06	53.002,89 €	2548969 REFUNDACIJE-DOTACIJE	WALDORFSKA ŠOLA LJUBLJANA
25	MESTNA OBČINA LJUBLJANA	2014-06-23	49.198,05 €	2578646 REFUNDACIJE-DOTACIJE	WALDORFSKA ŠOLA LJUBLJANA
26	MESTNA OBČINA LJUBLJANA	2014-07-01	4.318,10 €	2611321 REFUNDACIJE-DOTACIJE	WALDORFSKA ŠOLA LJUBLJANA
27	MESTNA OBČINA LJUBLJANA	2014-07-09	11.873,58 €	2632201 STORITVE	WALDORFSKA ŠOLA LJUBLJANA
28	MESTNA OBČINA LJUBLJANA	2014-07-31	11.127,33 €	2686412 STORITVE	WALDORFSKA ŠOLA LJUBLJANA
29	MESTNA OBČINA LJUBLJANA	2014-08-08	49.198,05 €	2707840 REFUNDACIJE-DOTACIJE	WALDORFSKA ŠOLA LJUBLJANA
30	MESTNA OBČINA LJUBLJANA	2014-08-08	2.987,93 €	2707841 REFUNDACIJE-DOTACIJE	WALDORFSKA ŠOLA LJUBLJANA
31	MESTNA OBČINA LJUBLJANA	2014-08-19	49.198,05 €	2723661 REFUNDACIJE-DOTACIJE	WALDORFSKA ŠOLA LJUBLJANA
32	MESTNA OBČINA LJUBLJANA	2014-09-01	2.313,97 €	2757916 REFUNDACIJE-DOTACIJE	WALDORFSKA ŠOLA LJUBLJANA
33	MESTNA OBČINA LJUBLJANA	2014-09-02	10.933,57 €	2758347 STORITVE	WALDORFSKA ŠOLA LJUBLJANA
34	MESTNA OBČINA LJUBLJANA	2014-09-17	49.765,95 €	2792281 REFUNDACIJE-DOTACIJE	WALDORFSKA ŠOLA LJUBLJANA
35	MESTNA OBČINA LJUBLJANA	2014-09-30	2.189,67 €	2831570 REFUNDACIJE-DOTACIJE	WALDORFSKA ŠOLA LJUBLJANA

Overview of purposes of transactions

431.8800		OKROŽNO SODIŠČE V LJUBLJANI	GRAD, računalniški programi in inženiring d.o.o.
431.8800		OKROŽNO SODIŠČE V LJUBLJANI	GRAD, računalniški programi in inženiring d.o.o.
431.8800		OKROŽNO SODIŠČE V LJUBLJANI	GRAD, računalniški programi in inženiring d.o.o.
431.8800		OKROŽNO SODIŠČE V LJUBLJANI	GRAD, računalniški programi in inženiring d.o.o.
431.8800		OKROŽNO SODIŠČE V LJUBLJANI	GRAD, računalniški programi in inženiring d.o.o.
431.8800		OKROŽNO SODIŠČE V LJUBLJANI	GRAD, računalniški programi in inženiring d.o.o.
431.8800		OKROŽNO SODIŠČE V LJUBLJANI	GRAD, računalniški programi in inženiring d.o.o.
431.8800		OKROŽNO SODIŠČE V LJUBLJANI	GRAD, računalniški programi in inženiring d.o.o.
431.8800		OKROŽNO SODIŠČE V LJUBLJANI	GRAD, računalniški programi in inženiring d.o.o.
431.8800		OKROŽNO SODIŠČE V LJUBLJANI	GRAD, računalniški programi in inženiring d.o.o.
458.6000		OKROŽNO SODIŠČE V LJUBLJANI	TRGOVINA, POSREDNISTVO IN ZASTUPSTVO DRAFT MI
441.7900		OKROŽNO SODIŠČE V LJUBLJANI	OREL LADO - ODVETNIK
447.7900		OKROŽNO SODIŠČE V KOPRU	BANKA KOPER d.d. Dvojezična firma: BANKA KOPE
451.4000		OKROŽNO SODIŠČE V MARIBORU	ŠOŠTERIČ FRANCI - ODVETNIK
478.8500		JANI	Odvetniška družba Fašun, Melihen, Milač, Stro
494.9000		JANI	ISTRABENZ PLINI, plini in plinske tehnologije
531.0200		JANI	Odvetniška družba Fašun, Melihen, Milač, Stro
554.6100		JANI	OREL LADO - ODVETNIK
609.0000		JANI	Odvetniška družba Fašun, Melihen, Milač, Stro
632.5700		JANI	Odvetniška družba Fašun, Melihen, Milač, Stro
638.7500		JANI	BANKA KOPER d.d. Dvojezična firma: BANKA KOPE
643.5500		JANI	Odvetniška pisarna Leben, Miklič, Rozman d.o.
653.9200		JANI	OREL LADO - ODVETNIK
707.4800		JANI	OREL LADO - ODVETNIK
772.7700		OKROŽNO SODIŠČE V MARIBORU	ŠOŠTERIČ FRANCI - ODVETNIK
799.5000		OKROŽNO SODIŠČE V KOPRU	BANKA KOPER d.d. Dvojezična firma: BANKA KOPE
800.0000		OKROŽNO SODIŠČE V KOPRU	BANKA KOPER d.d. Dvojezična firma: BANKA KOPE
847.9000		OKROŽNO SODIŠČE V MARIBORU	ŠOŠTERIČ FRANCI - ODVETNIK
854.2400		OKRAJNO SODIŠČE V LJUBLJANI	Odvetniška pisarna Leben, Miklič, Rozman d.o.
866.0800		OKROŽNO SODIŠČE V LJUBLJANI	OREL LADO - ODVETNIK
868.0300		OKRAJNO SODIŠČE V LJUBLJANI	Odvetniška pisarna Leben, Miklič, Rozman d.o.
1207.6800		OKROŽNO SODIŠČE V LJUBLJANI	OREL LADO - ODVETNIK
1212.4700		OKROŽNO SODIŠČE V LJUBLJANI	MERKUR - trgovina in storitve, d.d.
1591.8000		OKROŽNO SODIŠČE V LJUBLJANI	OREL LADO - ODVETNIK
3956.0000	VL 2011/187093 VRAČILO VARŠČINE	Okrajno sodišče v Šmarju pri Jelšah	JAGROS trgovina, proizvodnja in storitve, d.o
4119.8000	VL 2011/187093 VRAČILO VARŠČINE	Okrajno sodišče v Šmarju pri Jelšah	JAGROS trgovina, proizvodnja in storitve, d.o
5482.4000	2986445 STORITVE-F.O.	OKROŽNO SODIŠČE V MARIBORU	ŠOŠTERIČ FRANCI - ODVETNIK
18037.5200	VL 2011/187093 VRAČILO VARŠČINE	Okrajno sodišče v Šmarju pri Jelšah	JAGROS trgovina, proizvodnja in storitve, d.o
20633.3400	IN 223/12 DOLŽNIK MENTOR ČERIFČAKI	OKROŽNO SODIŠČE V KOPRU	BANKA KOPER d.d. Dvojezična firma: BANKA KOPE

(245 rows)

Series of transactions with the same amount of money - it seems there is some kind of periodic payments. Payments of service?

»VRAČILO VARŠČINE« (RETURN OF JUDICIAL SECURITY) - probably there is no business.
»STORITVE F. O.« - services of natural persons?

Overview of purposes of transactions

- ŠOŠTANJ MUNICIPALITY -> NTS, navtika, trgovina in storitve, d.o.o.
 - 2014-09-30, 2.547,99 €
 - »*ASIGNACIJA*« (*ASIGNATION*)
- ŽIRI MUNICIPALITY -> M SORA, trgovina in proizvodnja, d.d.
 - 2015-02-23, 4.052,90 €
 - »*TAX REFUND*«
- ŽELEZNIKI MUNICIPALITY -> DOMEL, Elektromotorji in gospodinjski aparati, d.o.o.
 - 2014-07-30, 20.000,00 €
 - »*RN41412014042NAKUP ENOSOBNEGA*« (*THE PURCHASE OF ONE-ROOM (APARTMENT)*)

Overview of purposes of transactions

- SLOVENSKE KONJICE MUNICIPALITY -> MLADINSKI CENTER DRAVINJSKE DOLINE
 - 2014-11-18, 2.280,00 €
 - *»PF 2014/8000555 SOF.MLAD.PROGRAMA« (CO-FINANCING OF YOUTH PROGRAM)*
- LOVRENC NA POHORJU MUNICIPALITY -> PERGOLA, družba za gradbeništvo, posredovanje, trgovino in druge storitve, d.o.o.
 - 2014-07-21, 38.253,60 €
 - *»PLAČILO RAČUNA PF 2011/D000180 004« (BILL PAYMENT)*
- MESTNA OBČINA CELJE, CELJE -> NT&RC Podjetje za časopisno in radijsko dejavnost d.o.o. Celje
 - 2015-04-10, 3.008,55 €
 - *»PLAČ.RAČ.-OPRAV.STORITVE« (BILL PAYMENT – SERVICES PROVIDED)*

Overview of purposes of transactions

- BRDA MUNICIPALITY -> VINSKA KLET "GORIŠKA BRDA" zadruga, kmetijstvo, proizvodnja vina in trgovine z.o.o. Dobrovo
 - 2014-10-06, 3.518,84 €
 - »*PL.RAČ. ZA VINO*« (*BILL PAYMET FOR VINE*)
- KRŠKO MUNICIPALITY -> AVTOBUSNI PREVOZI BOŽIDAR BRAČUN S.P.
 - 2015-04-01, 2.787,60 €
 - »*PREVOZI ŠOL. OTROK 2-2015 2015/300*« (*TRANSPORTATION OF SCHOOL CHILDREN*)
- BLED MUNICIPALITY -> PROMOTIVA, ORGANIZACIJA DOGODKOV, ANTOLIN GREGA S.P.
 - 2014-11-13, 3.000,00 €
 - »*RN?101-2014-OTROŠKI ANIMACIJSKI NA*« (*CHILDRENS ENTERTAINMENT*)

Overview of purposes of transactions

- ŽALEC MUNICIPALITY -> DRUŠTVO SVETNIKOV SDS ŽALEC:
 - 2014-10-27: 622,60 €
 - 2014-09-22: 877,20 €
 - 2014-09-02: 877,20 €
 - 2014-07-11: 877,20 €
 - 2014-06-24: 877,20 €
 - 2014-05-29: 877,20 €
 - 2014-04-22: 877,20 €
 - 2014-03-18: 2.631,60 €
- OKROŽNO SODIŠČE V LJUBLJANI -> GRAD, računalniški programi in inženiring d.d.:
 - 2015-04-13: 417,24 €
 - 2015-03-19: 417,24 €
 - 2015-02-25: 417,24 €
 - 2015-01-12: 431,88 €
 - 2014-12-10: 431,88 €
 - 2014-12-01: 431,88 €
 - 2014-10-13: 431,88 €
 - 2014-09-15: 431,88 €
 - 2014-08-14: 431,88 €
 - 2014-07-14: 431,88 €
 - 2014-06-27: 431,88 €
 - 2014-05-15: 431,88 €
 - 2014-04-14: 431,88 €
 - 2014-03-17: 431,88 €
 - 2014-02-14: 431,88 €
 - 2014-01-15: 431,88 €

Example: Text analysis of Slovenian legislation



Text analysis of Slovenian legislation

- Slovenian legislation (full text) was obtained in HTML format and imported into the database with some additional metadata.
- Data were exported to “named-line documents format”: first word is the name of the document (act), after space/tab there is a full text of the document till the newline:

```
\COPY (select replace[naslov_zakona, ' ', '_'],  
regexp_replace[vsebina_zakona, E'[\n\r\u2028]+', ' ', 'g' ] from  
zakonodaja_zakoni, zakonodaja_besedila where [zakonodaja_zakoni.sop  
= zakonodaja_besedila.sop] and [id ~* 'zak']) to  
'slovenian_acts.txt' delimiter E'\t';
```

- Slovenian characters (čšž) were removed:

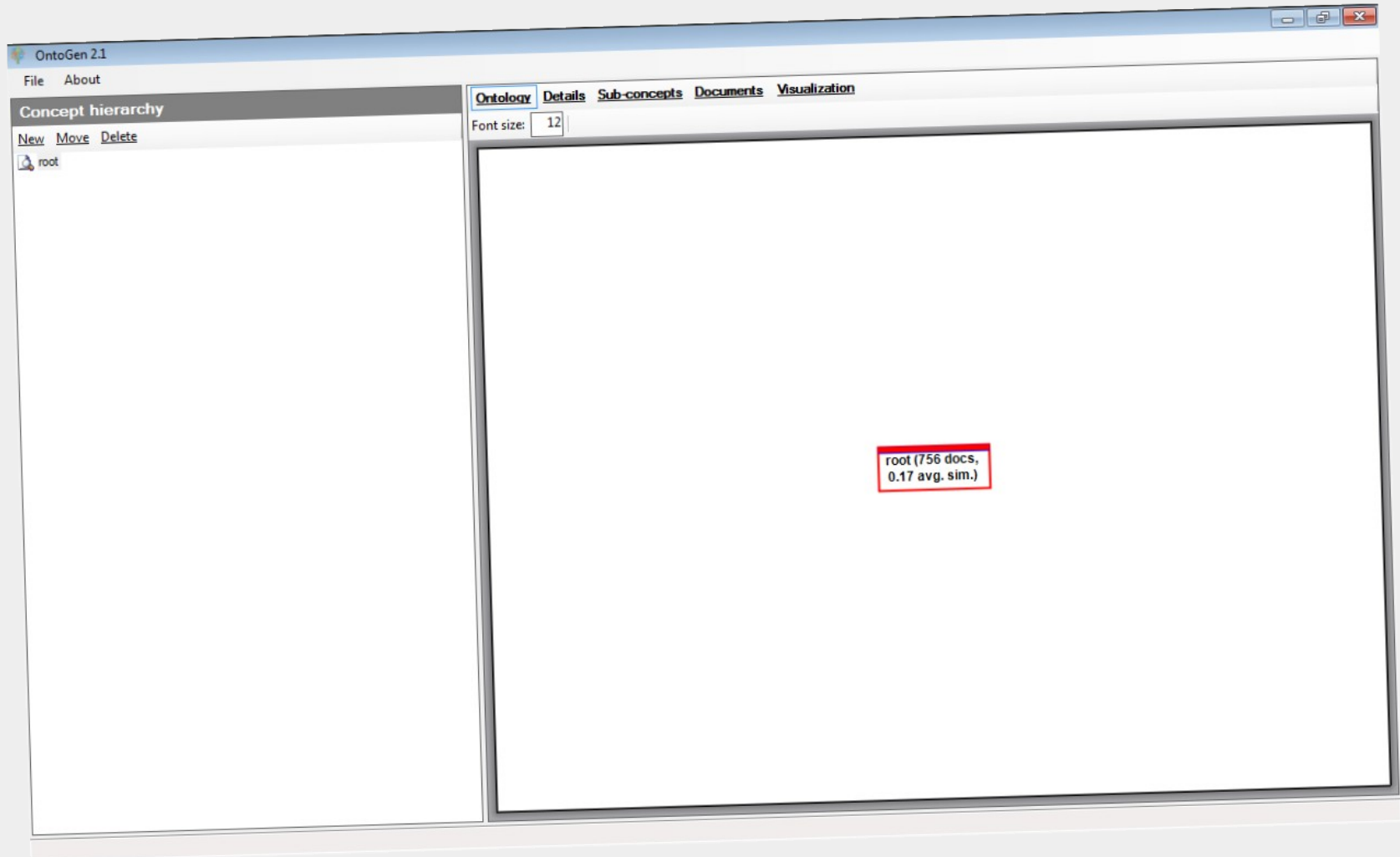
```
sed 's/\č/c/g' slovenian_acts.txt | sed 's/\š/s/g' | sed 's/\ž/z/g'  
| sed 's/\Č/C/g' | sed 's/\Š/S/g' | sed 's/\Ž/Z/g' >  
slovenian_acts_csz.txt
```

- OntoGen application is quite old and unfortunately does not use UTF-8 encoding, so file needs to be encoded in CP1250:

```
iconv -c -f utf-8 -t cp1250 slovenian_acts_csz.txt >  
slovenian_acts_win.txt
```

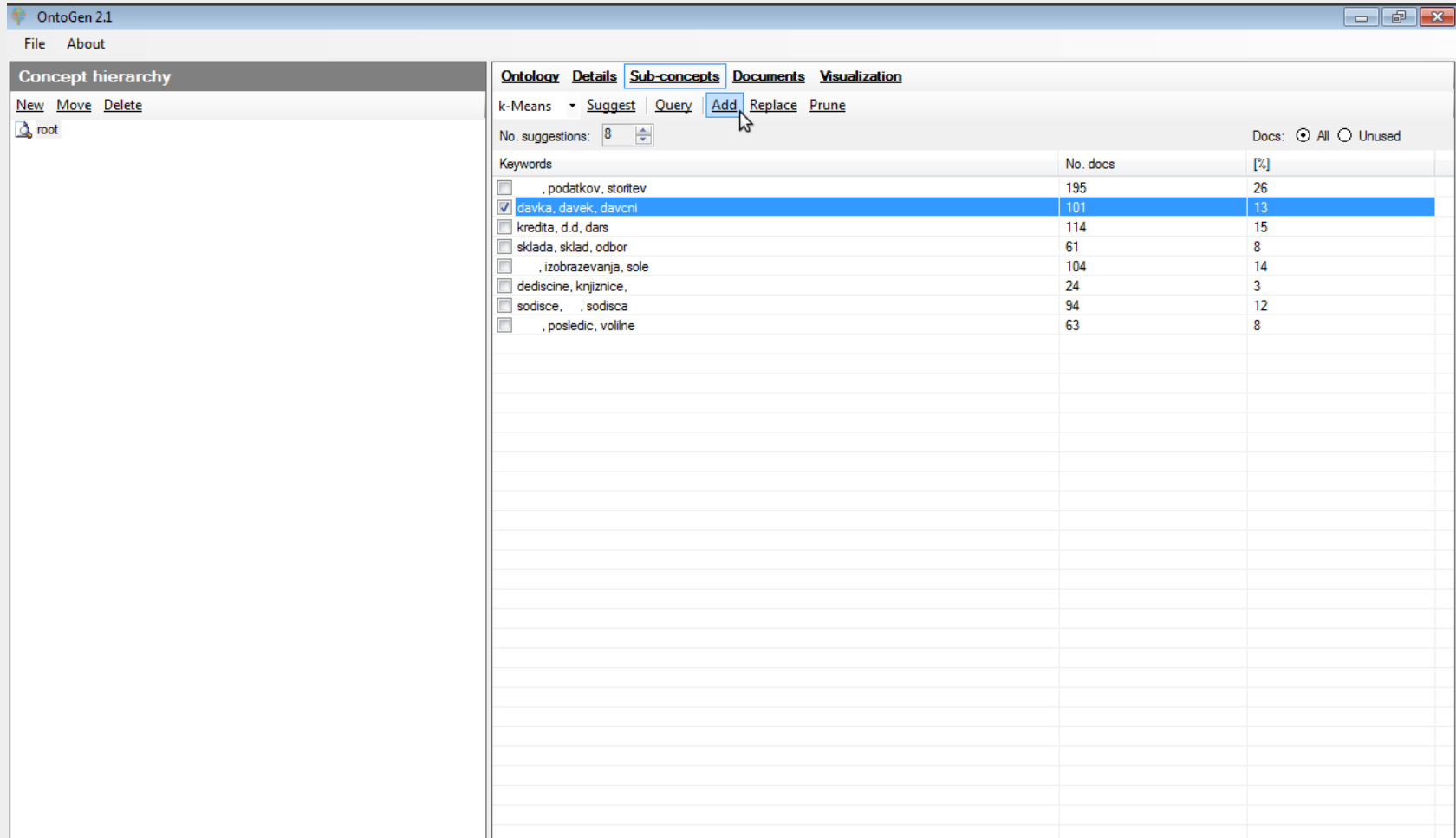
Text analysis of Slovenian legislation

Slovenian legislation is then imported into OntoGen...



Text analysis of Slovenian legislation

OntoGen suggests us 8 subconcepts. We add subconcept named "tax"...



The screenshot shows the OntoGen 2.1 application window. The 'Sub-concepts' tab is active, displaying a table of suggestions. The table has three columns: 'Keywords', 'No. docs', and '[%]'. The row for 'davka, davek, davčni' is highlighted in blue, indicating it is the selected subconcept. The 'Add' button is highlighted with a mouse cursor.

Keywords	No. docs	[%]
<input type="checkbox"/> , podatkov, storitev	195	26
<input checked="" type="checkbox"/> davka, davek, davčni	101	13
<input type="checkbox"/> kredita, d.d, dars	114	15
<input type="checkbox"/> sklada, sklad, odbor	61	8
<input type="checkbox"/> , izobrazevanja, sole	104	14
<input type="checkbox"/> dediscine, knjiznice,	24	3
<input type="checkbox"/> sodisce, . sodisca	94	12
<input type="checkbox"/> , posledic, voline	63	8

Text analysis of Slovenian legislation

For this subconcept we can preview documents included into this concept. We can also add or remove documents...

The screenshot displays the OntoGen 2.1 application window. On the left, the 'Concept hierarchy' shows a tree structure with 'root' and a sub-concept 'davka, davek, davčni'. The main area is divided into several panes. The top pane, 'Documents', shows a list of documents with their similarity scores. The selected document, 'Zakon o finančni upravi', is highlighted in blue. Below this list, a 'Keywords for selected documents' section shows a list of terms. The right pane displays the text of the selected document, 'Zakon o finančni upravi', which includes a warning and the beginning of the law's text. At the bottom, a graph shows a curve representing the document's position in the hierarchy, with a vertical line indicating the selected document's position.

Concept hierarchy

- root
 - davka, davek, davčni

Documents

Document	Similarity
<input checked="" type="checkbox"/> Zakon o davku od iger na srečo -- ZAKON O DAVKU OD IGE...	0,400
<input checked="" type="checkbox"/> Zakon o izvrševanju proračunov Republike Slovenije za leti 201...	0,399
<input checked="" type="checkbox"/> Zakon o posebnem prometnem davku od posebnih iger na sreč...	0,394
<input checked="" type="checkbox"/> Zakon o dodatnem davku od dohodkov članov poslovodstev in...	0,348
<input checked="" type="checkbox"/> Zakon o davku od prometa zavarovalnih poslov -- Opozorilo: B...	0,347
<input checked="" type="checkbox"/> Zakon o davku na motorna vozila -- ZAKON O DAVKIH NA M...	0,341
<input checked="" type="checkbox"/> Zakon o davku na dobitke pri klasičnih igrah na srečo -- Opozo...	0,338
<input checked="" type="checkbox"/> Zakon o denacionalizaciji -- Opozorilo: Besedilo osnovnega pre...	0,327
<input checked="" type="checkbox"/> Zakon o izvrševanju proračunov Republike Slovenije za leti 201...	0,325
<input checked="" type="checkbox"/> Zakon o davčni službi -- ZAKON O DAVČNI SLUŽBI (ZDS-1) 1...	0,325
<input checked="" type="checkbox"/> Zakon o davku od dobitka od odsvojitve izvedenih finančnih in...	0,320
<input checked="" type="checkbox"/> Zakon o izvrševanju proračunov Republike Slovenije za leti 201...	0,319
<input checked="" type="checkbox"/> Zakon o davku na vodna plovila -- Opozorilo: Besedilo osnovn...	0,317
<input checked="" type="checkbox"/> Zakon o davku na bilanco vsoto bank -- ZAKON O DAVKU N...	0,304
<input checked="" type="checkbox"/> Zakon o ureditvi določenih vprašanj v zvezi z graditvijo avtoce...	0,293
<input checked="" type="checkbox"/> Zakon o finančni upravi -- Opozorilo: Besedilo osnovnega pred...	0,290
<input checked="" type="checkbox"/> Zakon o davku na dodano vrednost -- ZAKON O DAVKU NA ...	0,279
<input checked="" type="checkbox"/> Zakon o ureditvi določenih vprašanj -- _v_zvezi_z_graditvijo_z...	0,278
<input checked="" type="checkbox"/> Zakon o stavbnih zemljiščih -- Opozorilo: Besedilo osnovnega p...	0,276
<input checked="" type="checkbox"/> Zakon o urejanju trga dela -- Opozorilo: Besedilo osnovnega pr...	0,274
<input checked="" type="checkbox"/> Zakon o posebnem davku na določene prejemke -- Opozorilo: ...	0,272

Keywords for selected documents:

davka, davek, davčni, zavezanec, nepremičnine, nepremičnin, stavbe, davčni zavezanec, dohodka

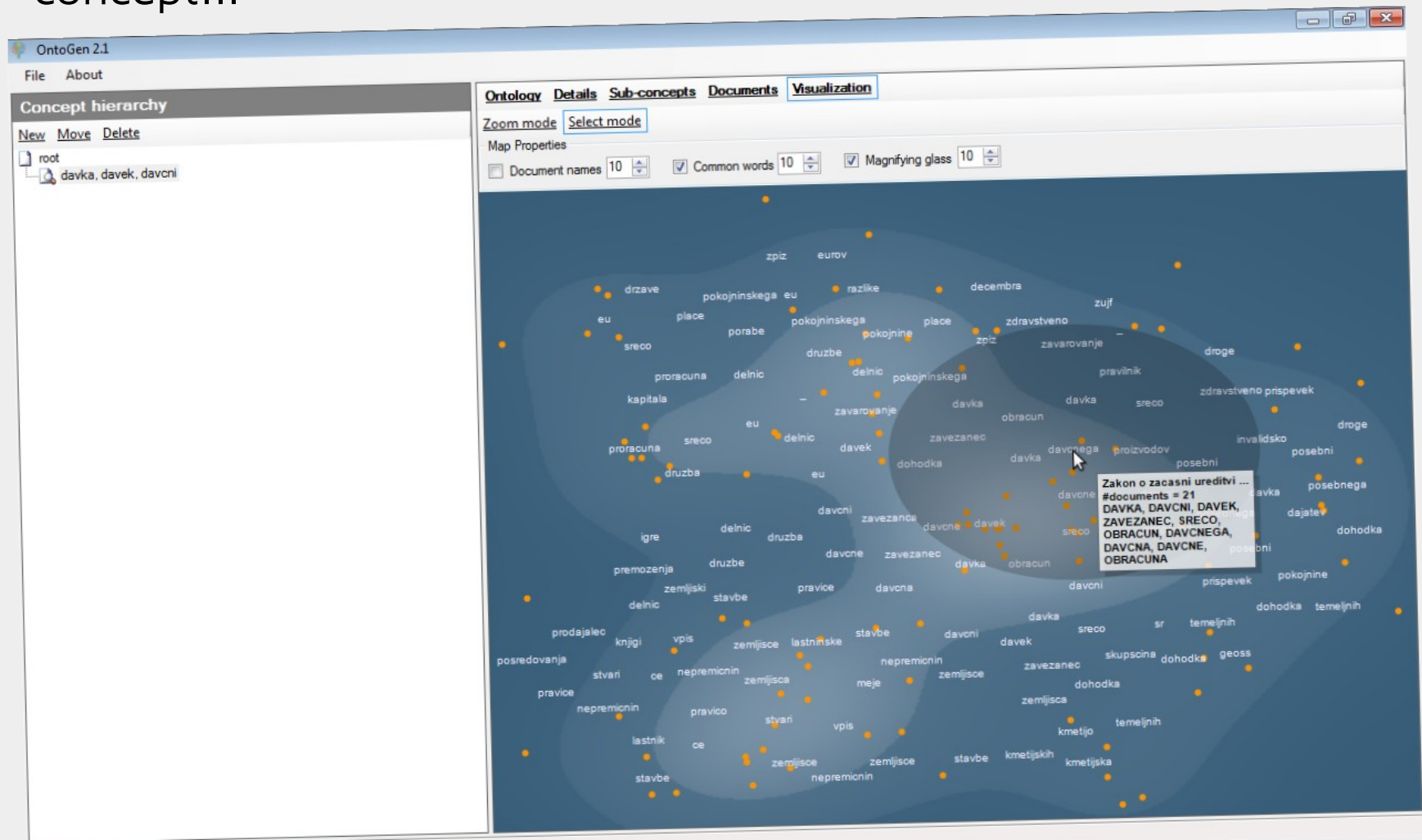
Zakon o finančni upravi

Opozorilo: Besedilo osnovnega predpisa
ZAKON O FINANČNI UPRAVI (ZFU) I. SPLOSNE
DOLOČBE 1. člen (vsebina zakona) Ta zakon ureja načela delovanja, organizacijo, naloge in pooblastila finančne uprave, pravice ter posebnosti delovno-pravnih razmerij javnih uslužbencev in javnih uslužbenk finančne uprave (v nadaljnjem besedilu: uslužbenki finančne uprave) in druga vprašanja, povezana z delovanjem finančne uprave. 2. člen (opredelitev pojmov) Pojmi, uporabljeni v tem zakonu, pomenijo: 1. obvezna dajatev vključuje pojem davka, kot je opredeljen z zakonom, ki ureja davčni postopek, uvozne in izvozne dajatve, predpisane s predpisi Evropske unije (v nadaljnjem besedilu: EU), prispevke za socialno varnost in druge dajatve, ki so prihodek ene od blagajn javnega financiranja ali proračuna EU in so s posebnim predpisom predpisane kot splošno obvezne ter za pobiranje katerih je pristojna finančna uprava. Obvezna dajatev vključuje tudi koncesijsko dajatev, ce je za finančni nadzor v zvezi z njo v skladu s posebnim predpisom pristojna finančna

Document name: Zakon o davku od dobitka od odsvojitve izvedenih finančnih instrumentov

Text analysis of Slovenian legislation

Ontogen creates visualisation of the documents within the same concept...



Text analysis of Slovenian legislation

We can view concept details...

The screenshot displays the OntoGen 2.1 application window. On the left, the 'Concept hierarchy' pane shows a tree structure with 'root' and a sub-entry 'davka, davek, davčni'. The main area is titled 'Ontology' and has several tabs: 'Details', 'Sub-concepts', 'Documents', and 'Visualization'. The 'Details' tab is active, showing the following information:

- Name: davka, davek, davčni
- Top keywords: davka, davek, davčni, zavezanec, nepremicnine, nepremicnin, . stavbe, davčni zavezanec, dohodka
- Central documents: Zakon_o_davcnem_postopku, Zakon_o_davkih_obcanov, Zakon_o_davku_na_nepremicnine, Zakon_o_davku_na_promet_nepremicnin, Zakon_o_davku_na_dediscine_in_darila, Zakon_za_uravnotezenje_javnih_financ, Zakon_o_davku_od_sreck, Zakon_o_dohodnini, Zakon_o_davku_od_dohodkov_pravnih_oseb, Zakon_o_davku_na_financne_storitve
- SVM Keywords: davka, davek, davčni, nepremicnine, zavezanec, nepremicnin, stavbe, pokojninskega, zemljišca, dohodka
- All documents: 101
- Unused documents: 101
- Avg. similarity: 0,244

A red arrow points from a text box to the SVM Keywords field. The text box contains the following text:

Keywords, which separate the selected concept from other concepts most

The Windows taskbar at the bottom shows the system tray with the date 14.9.2015 and time 14:28, and the system tray icons including SL, network, volume, and power. The taskbar also shows several application icons and the 'Right Ctrl' button.

Text analysis of Slovenian legislation

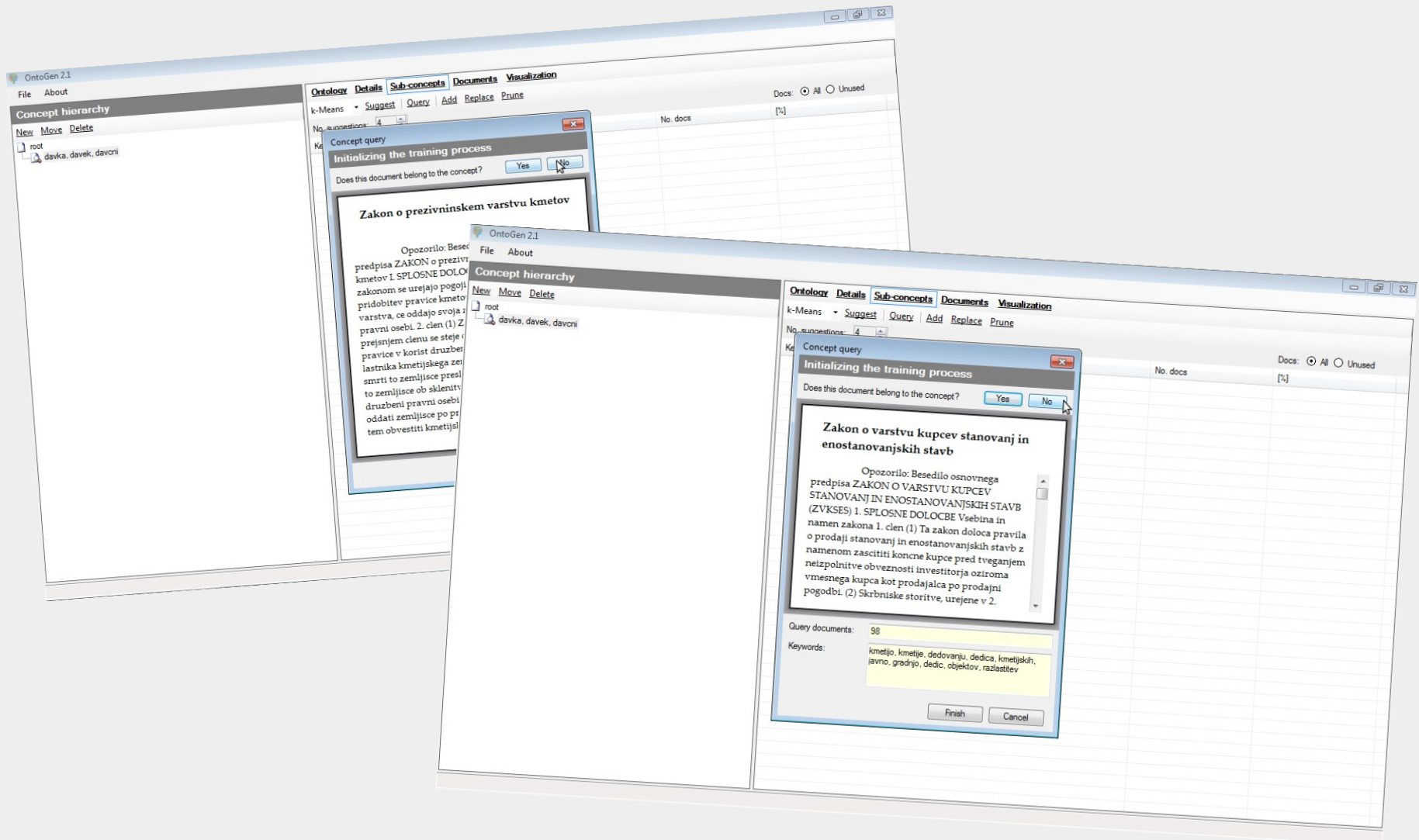
We can zoom in the concept...

The screenshot shows the OntoGen 2.1 application window. On the left, the 'Concept hierarchy' panel shows a tree structure starting with 'root' and a sub-item 'davka, davek, davčni'. The main area is divided into several sections:

- Navigation:** 'File About' menu and tabs for 'Ontology', 'Details', 'Sub-concepts', 'Documents', and 'Visualization'.
- Search:** A 'Zoom mode' section with a 'Name' field containing 'zemljisca, gradnjo, stavbnih' and 'Add'/'Prune' buttons. Below it, a 'Keywords' field contains 'zemljisca, gradnjo, stavbnih, razlastitev, nepremicnino, zemljisce, objektov, javno, zemljisci, zemljiscu'.
- Map Properties:** Checkboxes for 'Document names', 'Common words', and 'Magnifying glass', each with a value of '10'.
- Visualization:** A word cloud of Slovenian legislation titles. A tooltip is displayed over the text 'Zakon o stavbnih zemljiščih', showing the following information:
#documents = 1
ZEMLJIŠCA, STAVBNEGA,
ZEMLJIŠCE, STAVBNIH,
GRADITEV, OBJEKTOV,
ZEMLJISCU, STAVBNO,
UREJANJA, SKLAD

Text analysis of Slovenian legislation

We can add new (sub)concept by learning...

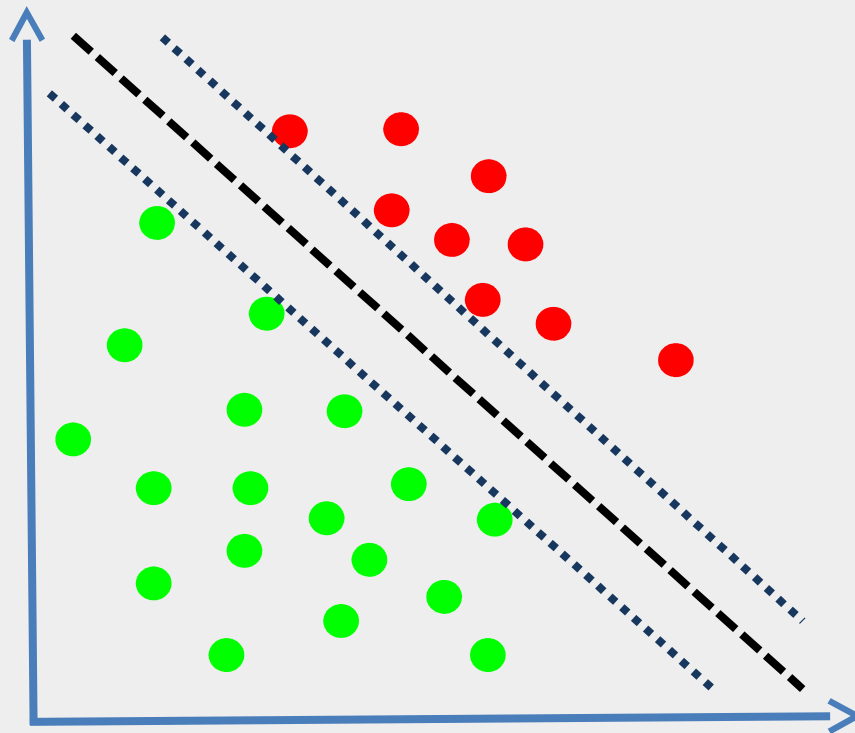


Example: Anonymisation of court decisions



Anonymisation of court decisions

Classification of words in two classes with the method of **support vectors**.



Linear distinguishable example

Word is represented as m -dimensional vector of several **features** (in our case $m \sim 600$).

Method tries to find an optimal **hyperplane**, which segregates points (words represented by vectors) in this (mapped) place.

Example: Geolocation of photographs



Geolocation of photographs

- Photographs taken from mobile phones have embed several metadata, including GPS location (usually in WGS 84 format).
- So we grab all data from the JPG images in a directory:

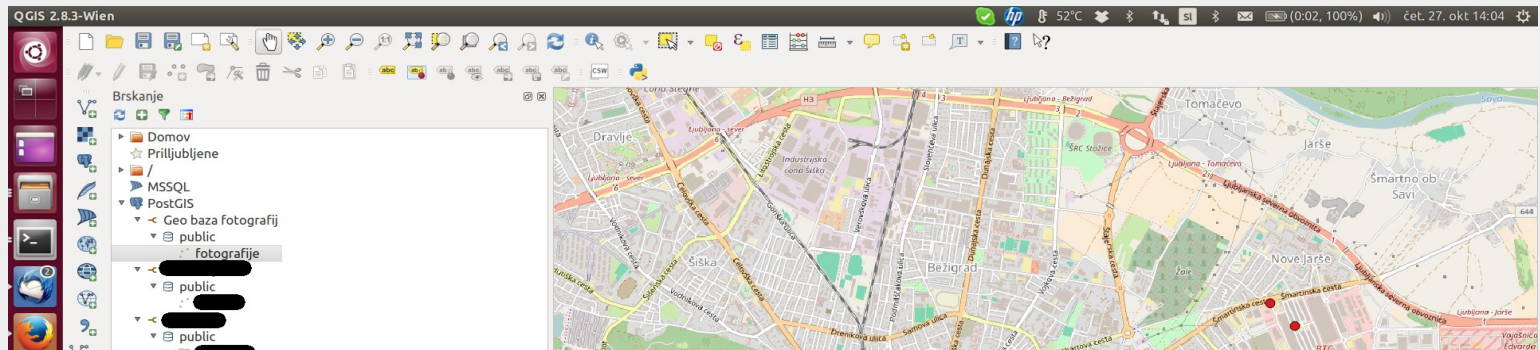
```
for file in *.jpg; do exiftool -c "%.6f" $file | grep "GPS  
Position" | awk -v filename="$file" '{print filename, $4, $6}' | tr  
[:blank:] \t >> fotografije.csv; done;
```
- ... import it into a PostgreSQL database:

```
create table fotografije (filename text, latitude numeric,  
longitude numeric);  
\COPY fotografije from 'fotografije.csv' delimiter E'\t';
```
- ...and add geolocation (postGIS) info:

```
select AddGeometryColumn('fotografije', 'geolokacija', 4326,  
'POINT', 2);  
update fotografije set geolokacija =  
ST_SetSrid(ST_GeomFromText('POINT(' || longitude || ' ' || latitude  
|| ')'), 4326);
```


Geolocation of photographs

...and visualise data with opensource QGIS!

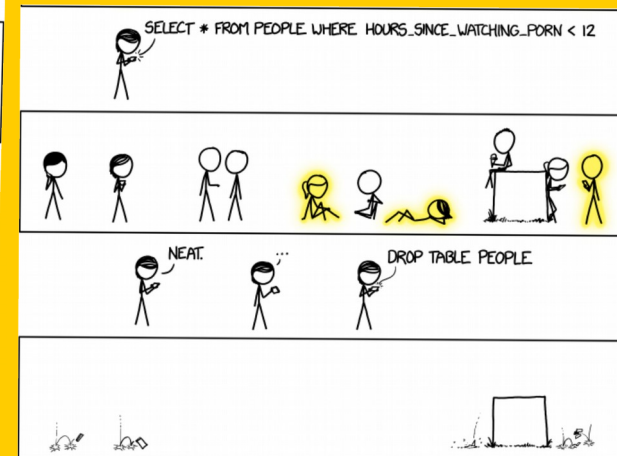
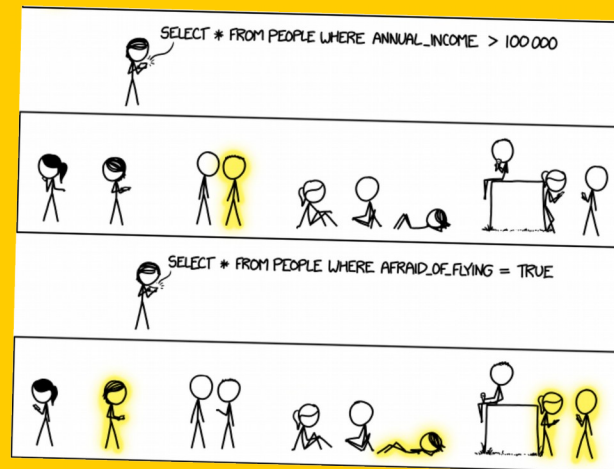
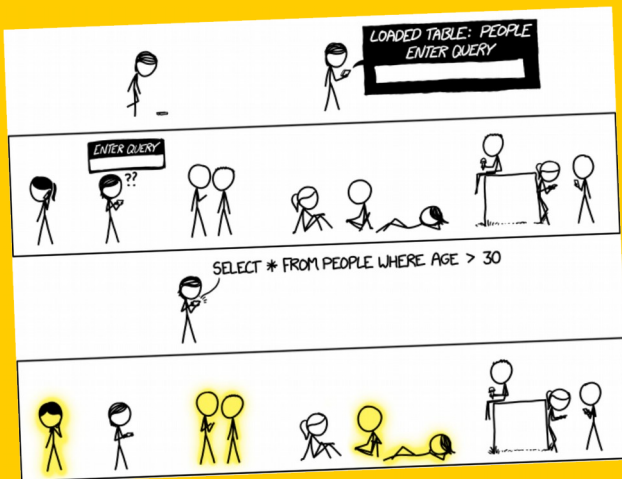


With PostGIS we can also do some nice locational analysis:

```
select ime_obcine, count(*) as num_of_photos from fotografije, obcine where (geolokacija_slo && geom) and {ST_Contains(geom, geolokacija_slo)} group by ime_obcine;
```

ime_obcine	num_of_photos
LJUBLJANA	86
KRANJ	5

SELECT *questions* FROM *audience*;



Cartoon: (CC) xkcd.com

Matej Kovačič
matej.kovacic@ijs.si

Centre for Knowledge Transfer in Information Technologies
<http://ct3.ijs.si/>

Artificial Intelligence Lab
<http://ailab.ijs.si>