

LDVM implementation

Jiří Helmich, Jakub Klímek

Department of Software Engineering

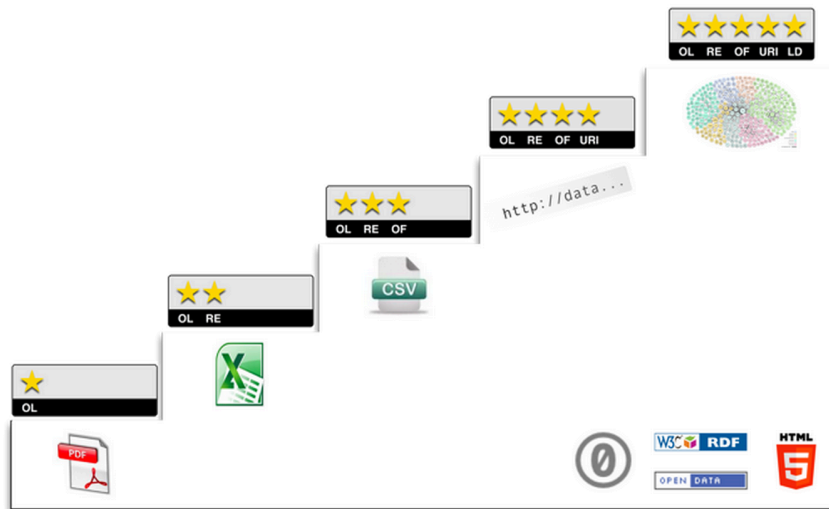
Faculty of Mathematics and Physics

Charles University

Linked Data

- ▶ Tim Berners-Lee: The next web
 - ▶ http://www.ted.com/talks/tim_berners_lee_on_the_next_web
 - ▶ “I want you to put your data on the Web!”
 - ▶ Web of data
- ▶ Use **HTTP URIs** to **denote** things so that these things can be **dereferenced** by people and user agents.
- ▶ Provide useful information about the thing when its URI is dereferenced (**RDF, SPARQL**).
- ▶ Include **links** to other related things when publishing data on the Web.

Linked Open Data

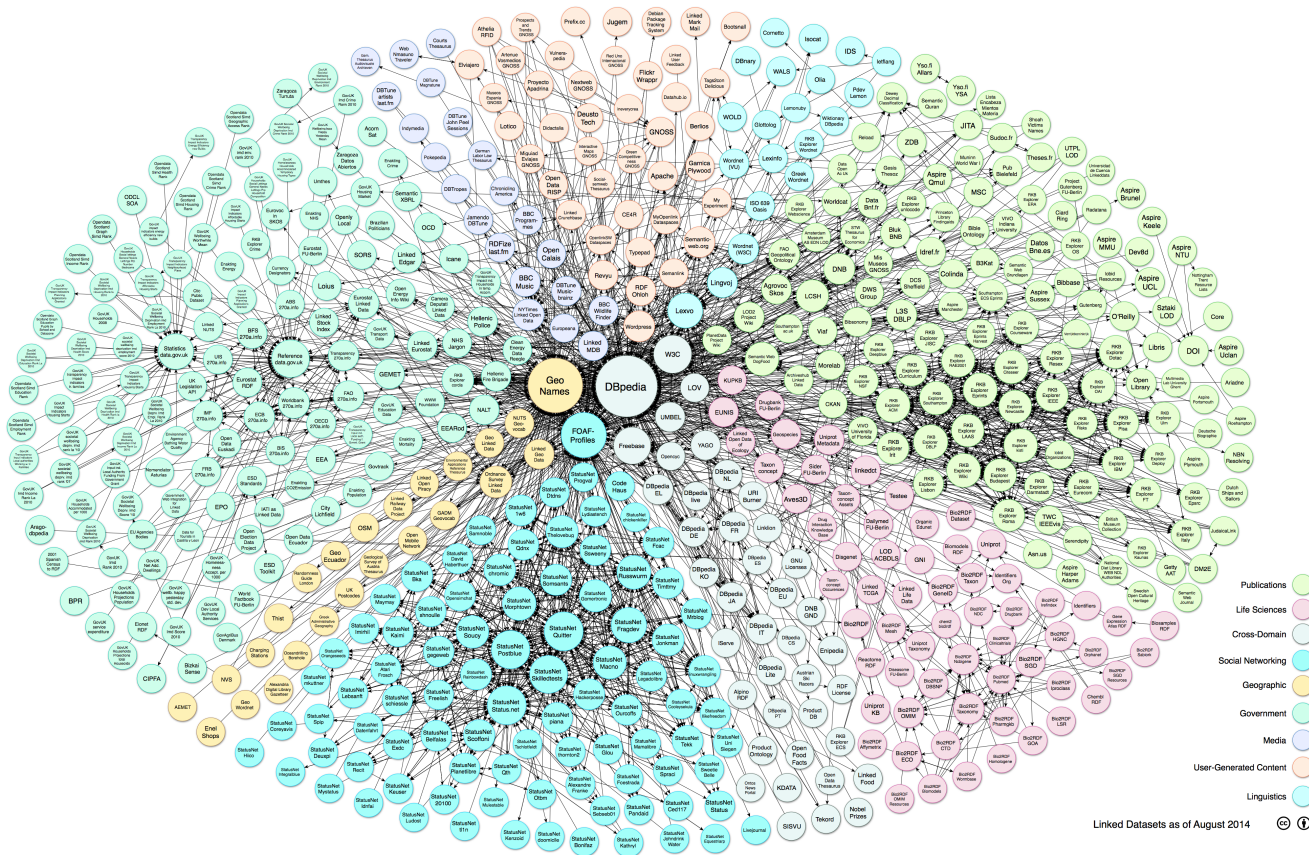


It's often very hard to obtain any data, especially governmental. That's the first star. The second one reminds me of a way how public contracts were published in my hometown, Liberec - printed, scanned and the image placed into a PDF file.

LDVM IMPLEMENTATION

3

So, we have Linked Data...



DBpedia is still the center of the Linked Data universe. You can see that the density of links is higher in its neighbourhood. Anyway, it's great to see a vast amount of Linked Data! But having it published is not enough...

LDVM IMPLEMENTATION

4

... even Czech LOD cloud.

► **OPENDATA cz** since 2012

► CUNI

► CTU

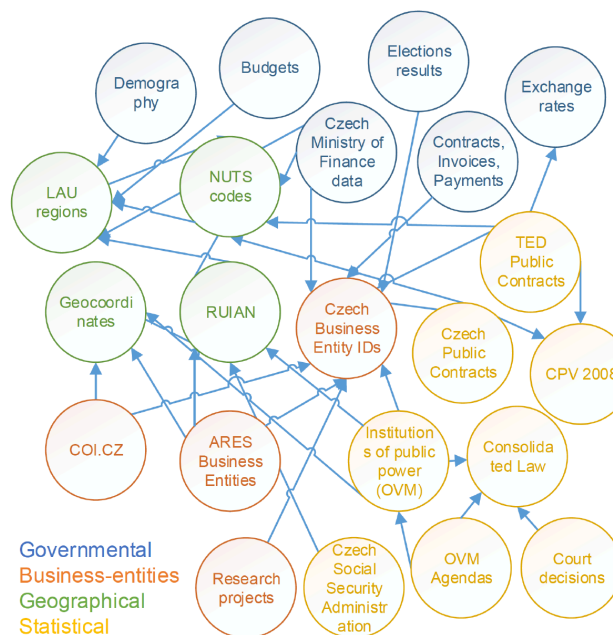
► VŠE

► 600M triples

► + 600M triples - RUIAN

► OpenLink Virtuoso 7

► <http://linked.opendata.cz/sparql>



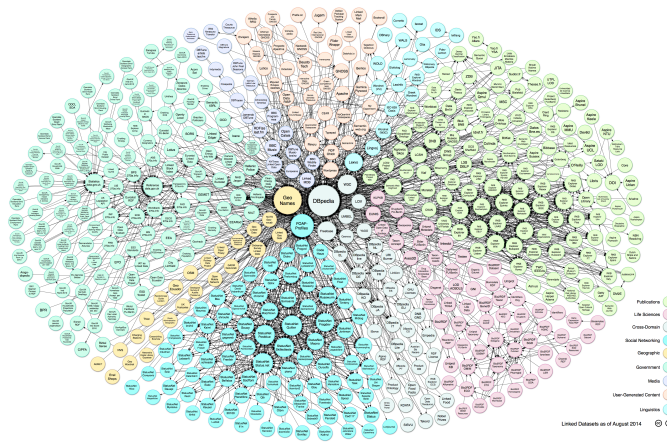
So, I published a vast amount of data ...

- ▶ What about getting something in return?
 - ▶ Consuming data
 - ▶ Presenting it in a standard way (charts, maps, ...)
- ▶ Linked Data
 - ▶ Integration with other datasets
 - ▶ ~~reasoning~~

But it's still hard to tell, whether you can use some dataset or directly combine it with yours. You don't know, what a node in LOD represents. You don't know how to visualize it, how to use it. Even if those dataset are similar and cover the same topic, they may use different schemas to describe those concepts (e.g. schema.org duplicates some vocabularies). Reasoning is forgotten, not implemented by many tools.

... but!

- ▶ What's in there?
- ▶ How can I use some dataset to enrich mine?
 - ▶ Is it even compatible with my dataset?
- ▶ Can I use some of those to prepare a cool visualization for my boss?

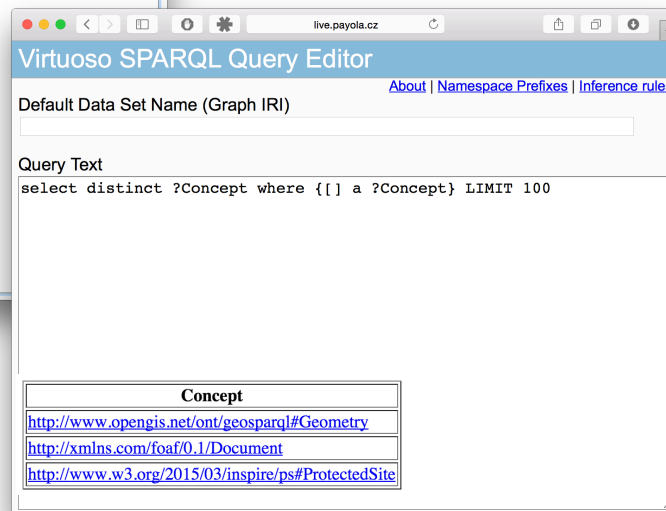
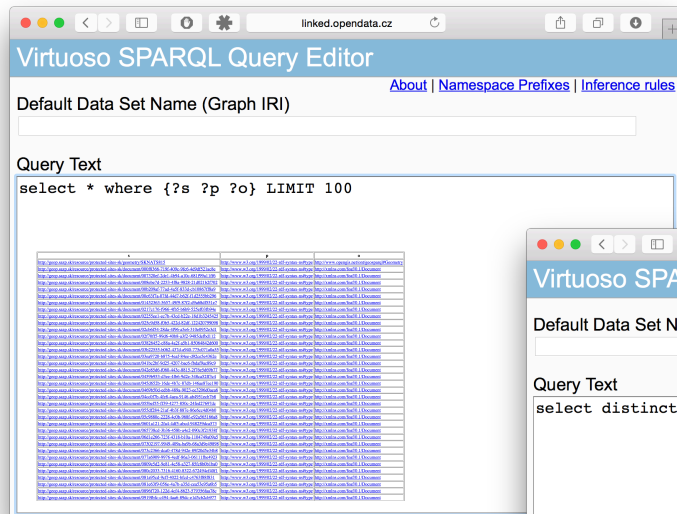


Catalogues - CKAN

- ▶ A good way to start
 - ▶ datahub.io: “finance”
 - ▶ *The Public Finances Databank is a compilation of published data covering the main aspects of the Government Finances including receipts, expenditure, borrowing and debt.*
 - ▶ linked.opendata.cz
- ▶ Too high-level



Reality



You start asking SPARQL queries. Top concepts, top queries. That's the appropriate level, but ...

LDVM IMPLEMENTATION

9

Not very helpful.



Visualizations

- ▶ Human is a visual being
 - ▶ Best way how to involve people
- ▶ This is not visual:

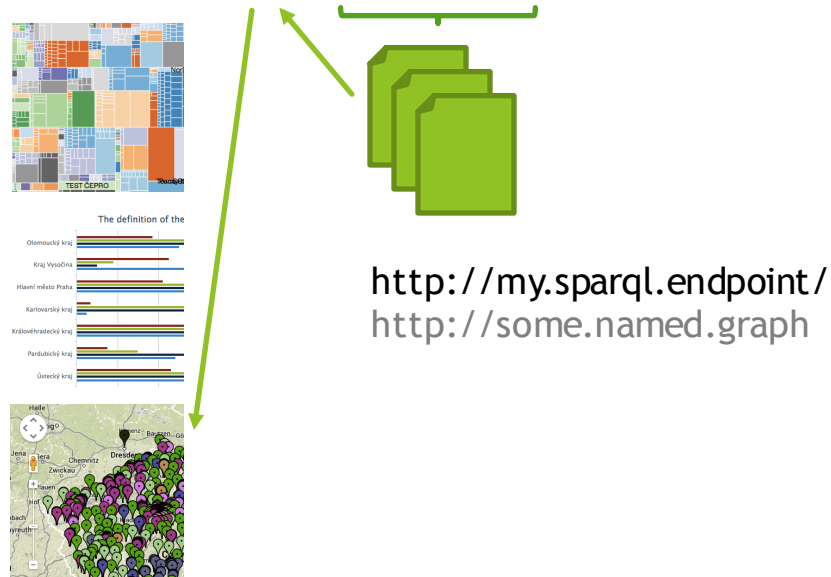
```
<http://linked.opendata.cz/resource/business-entity/CZ00088455> a gr:BusinessEntity , s:Organization ;
  adms:identifier <http://linked.opendata.cz/resource/business-entity/CZ00088455/identifier/fromOvmDetail> ,
<http://linked.opendata.cz/resource/domain/seznam.gov.cz/ovm/ovms/SprvKrknNP/identifier/fromOvmList> .

<http://linked.opendata.cz/resource/business-entity/CZ75151537> a gr:BusinessEntity , s:Organization ;
  adms:identifier <http://linked.opendata.cz/resource/business-entity/CZ75151537/identifier/fromOvmDetail> ,
<http://linked.opendata.cz/resource/domain/seznam.gov.cz/ovm/ovms/KRPUK/identifier/fromOvmList> .

<http://linked.opendata.cz/resource/domain/seznam.gov.cz/ovm/ovms/AOPrdKnCR> a gr:BusinessEntity , s:Organization , ovm:OVM ;
  gr:legalName "Agentura ochrany přírody a krajiny České republiky" ;
  s:name "Agentura ochrany přírody a krajiny České republiky" ;
  skos:notation "AOPrdKnCR" ;
  ovm:typSubjektuText "Orgán státní správy" ;
  ovm:stavSubjektu <http://linked.opendata.cz/ontology/domain/seznam.gov.cz/ovm/StavSubjektu/1> ;
  ovm:pravniForma <http://linked.opendata.cz/ontology/domain/seznam.gov.cz/ovm/PravniForma/325> ;
  ovm:primarniOvm "true"^^xsd:boolean ;
  ovm:overeniSubjektem "true"^^xsd:boolean ;
  ovm:datumOvereni "2012-02-01T15:31:57Z"^^xsd:dateTime ;
  ovm:business-entity <http://linked.opendata.cz/resource/business-entity/CZ62933591> ;
  s:address <http://linked.opendata.cz/resource/domain/seznam.gov.cz/ovm/ovms/AOPrdKnCR/address/fromOvmDetail> ,
<http://linked.opendata.cz/resource/domain/seznam.gov.cz/ovm/ovms/AOPrdKnCR/address/fromOvmList> ;
  ovm:datovaSchranka <http://linked.opendata.cz/resource/domain/seznam.gov.cz/ovm/datove-schranky/dkdkdj> ;
  s:contactPoint <http://linked.opendata.cz/resource/domain/seznam.gov.cz/ovm/ovms/AOPrdKnCR/email/fromOvmDetail/1> ,
```

Basic LDVM use-case

- Show me, how I could **VISUALIZE** my data!



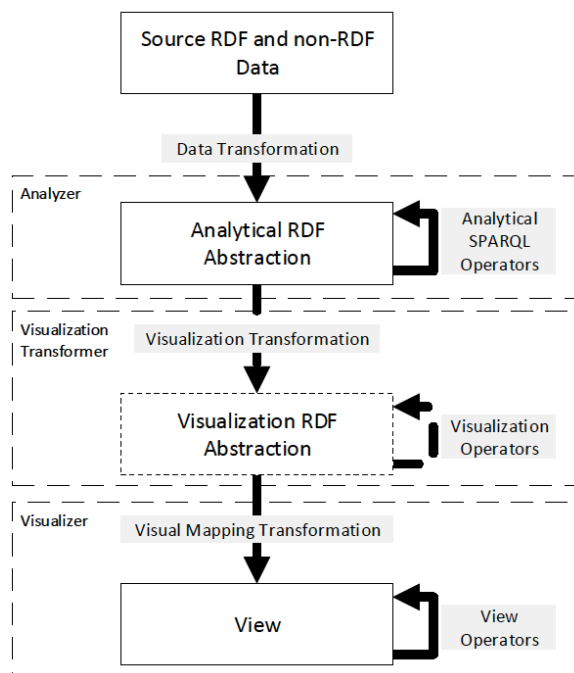
- Could I combine another datasets with mine?

DEMO



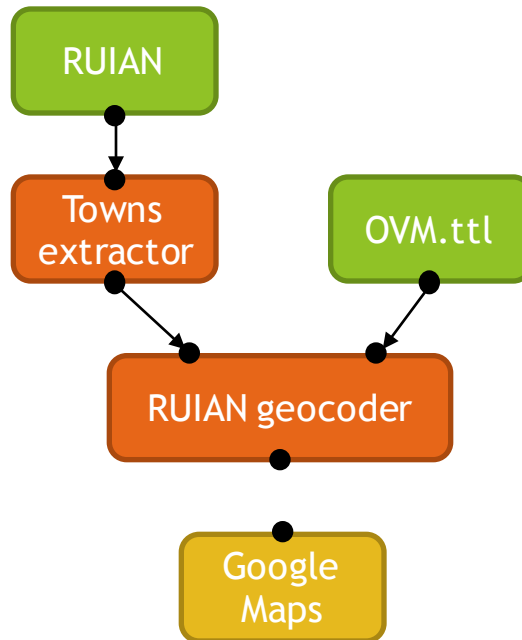
<https://goo.gl/r5Ln89>

Linked Data Visualization Model



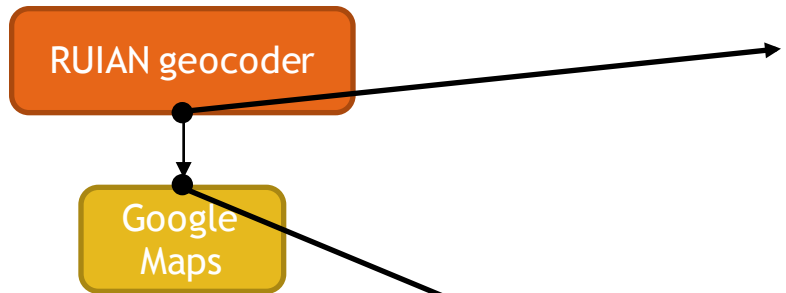
LDVM descripton, read e.g. http://events.linkedata.org/ldow2014/papers/ldow2014_paper_13.pdf

LDVM pipeline



This is the pipeline you saw in our first demo. How is it possible that we are able to discover it automatically? Let's assume that the fragment ending with RUIAN geocoder was already assembled by our discovery algorithm.

LDVM compatibility check



Output data sample

```
[] ruianlink:obec <ex1>;  
  s:geo [  
    s:longitude "14.417780" ;  
    s:latitude  "50.084552" .  
  ] .
```

Input descriptor

```
ASK {  
  ?s s:geo ?g .  
  ?g a s:GeoCoordinates ;  
    s:latitude ?lat ;  
    s:longitude ?lng .  
}
```

We use compatibility checking. Each output has an output data sample, sample RDF data that covers the possible output as well as possible, being as small as possible. Each input has a set of descriptors, SPARQL ASK queries, that describes an expected input. If descriptors match data samples, corresponding components are compatible and can be chained into a visualization pipeline.

LDVM discovery

- ▶ Iterative process
- ▶ Chaining registered LDVM components
- ▶ Starts with pipeline fragments based on available data sources

RUIAN

OVM.ttl

Google
Maps

Towns
extractor

RUIAN geocoder

LDVM discovery - before iteration 1

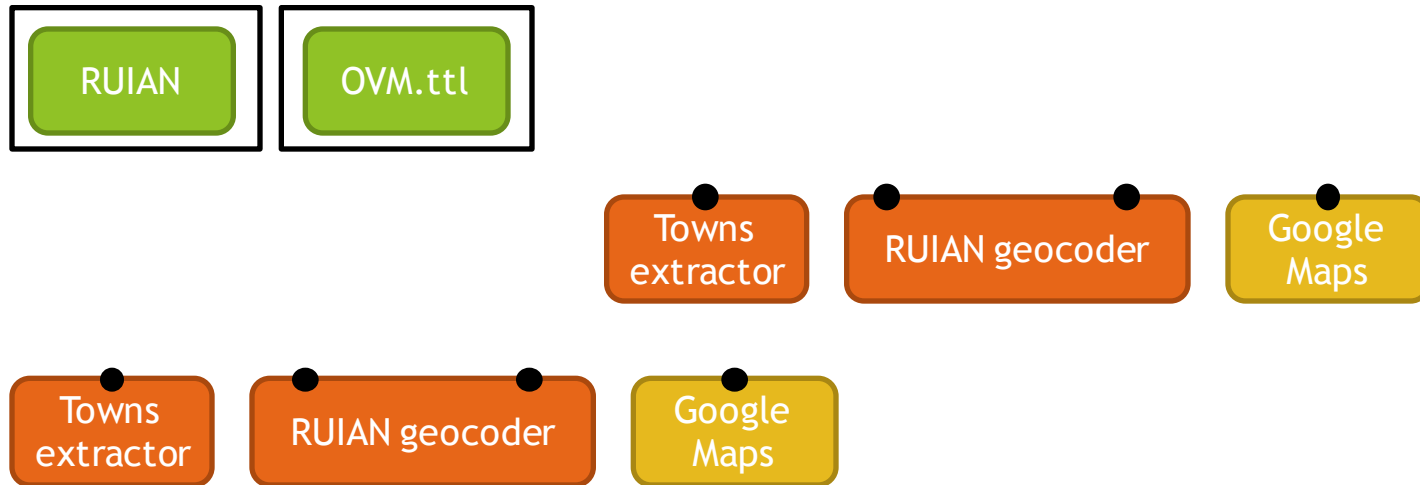
Pipeline fragments



Available components



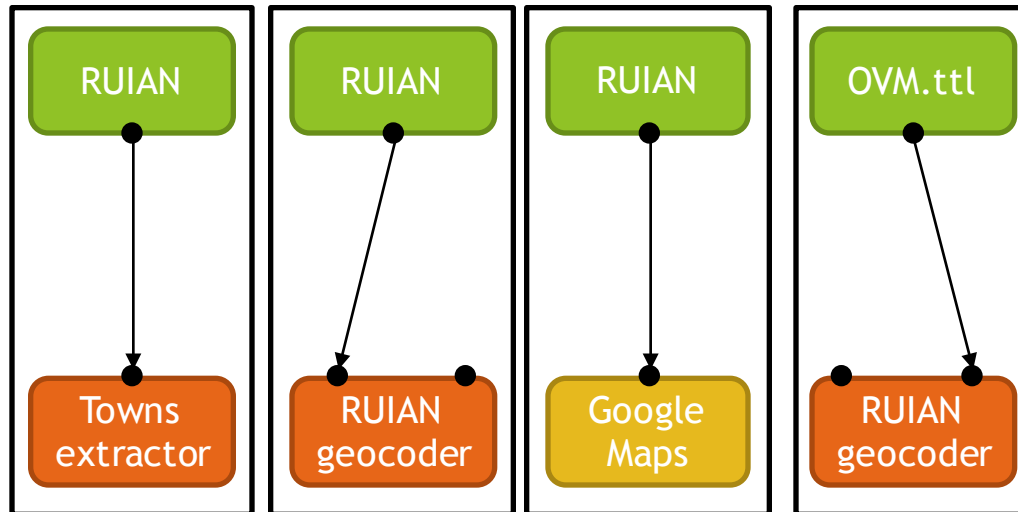
LDVM discovery - iteration 1



For every input, every descriptor query is asked against every available datasource. If the datasource answers TRUE, it's possible for the component to use that datasource via one of its inputs. Original slides contains animation where some inputs turn green and some red, based on compatibility checking.

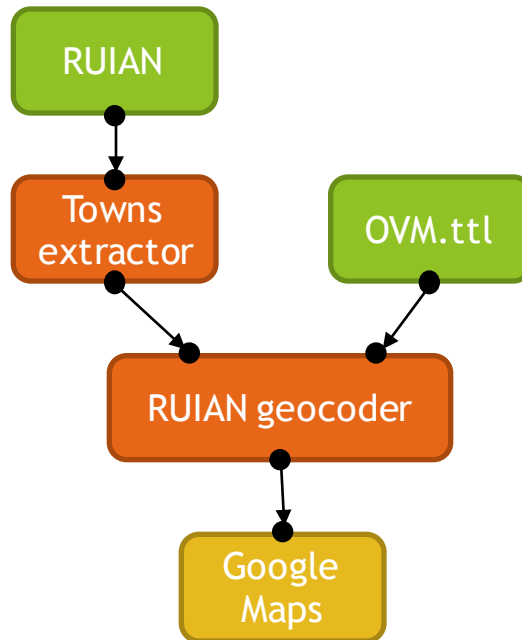
LDVM discovery - after iteration 1

Pipeline fragments



Fragments created in iteration one. Those ending with visualizers cannot be extended, those are no longer fragments, those are complete pipelines. No. 2 and 4 are merged into a single fragment, of course. When no fragments are added during an iteration, discovery terminates.

LDVM discovery - after iteration 4

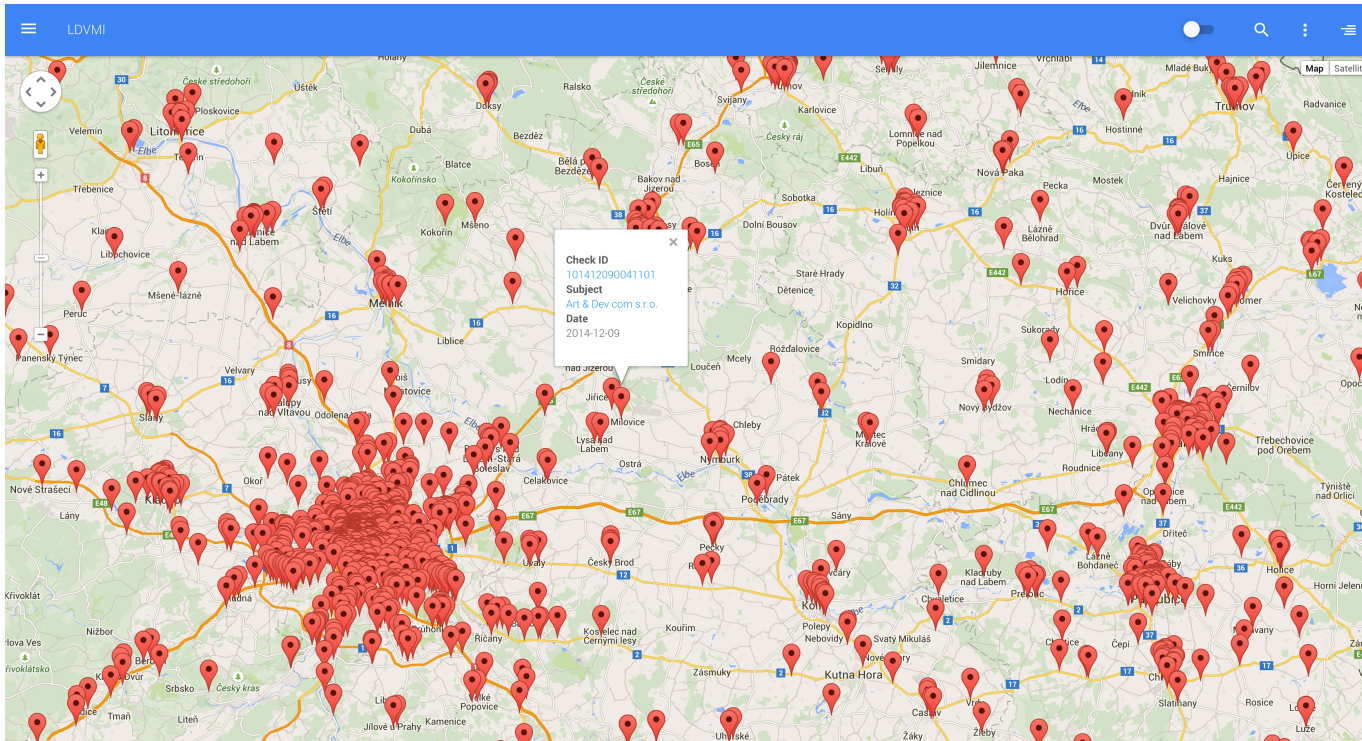


DEMO



<https://goo.gl/CEiZpR>

LDVM features



To display a marker on a map is just one of the visualizer's features. It is also able to provide facets, display labels and descriptions. Those features could be described formally by means of LDVM and assigned their own descriptors.

LDVM features

```
?s s:geo ?g .  
?g a s:GeoCoordinates ;  
    s:latitude ?lat ;  
    s:longitude ?lng .
```

```
?s s:geo ?g ;  
    ?p ?o .  
?o a skos:Concept  
.
```

```
?o s:name ?n .
```

```
?o s:description ?d .
```

A dataset must contain some coordinates. If it does not, the maps visualizer is not a good candidate. Therefore that feature is a mandatory one. Facets are generated using skos:Concept, but that's an optional feature. The same holds for labels and descriptions. Based on feature coverage, we will be able to rank both datasets and visualization pipelines.

LDVM vocabulary

- ▶ Define your component or pipeline in RDF
- ▶ <http://lov.okfn.org/dataset/lov/vocabs/ldvm>
- ▶ <https://github.com/payola/ldvm>

LDVMi

- ▶ <http://ldvm.opendata.cz>
- ▶ <https://github.com/payola/LDVMi>



Running instance of our tool, formerly known as Payola. Currently supporting just our components. Work in progress ;-)

LDVMI

The INSPIRE Protection Classification

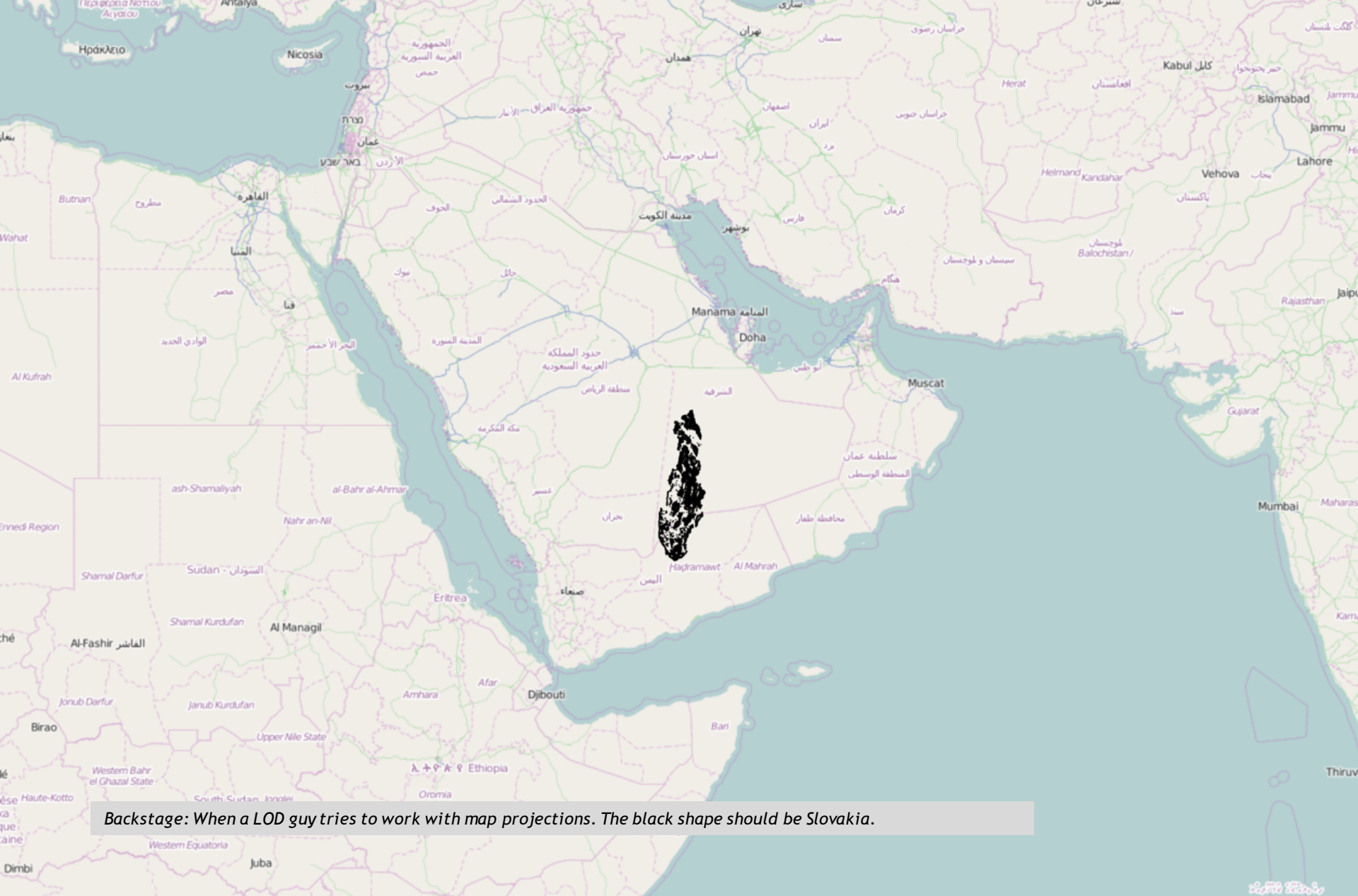
Enumeration values expressed in SKOS

<input type="checkbox"/>	landscape	(0)
<input checked="" type="checkbox"/>	ecological	(1645)
<input type="checkbox"/>	archaeological	(0)
<input type="checkbox"/>	geological	(0)
<input type="checkbox"/>	cultural	(0)
<input type="checkbox"/>	environment	(1645)
<input type="checkbox"/>	nature conservation	(1645)

REFRESH

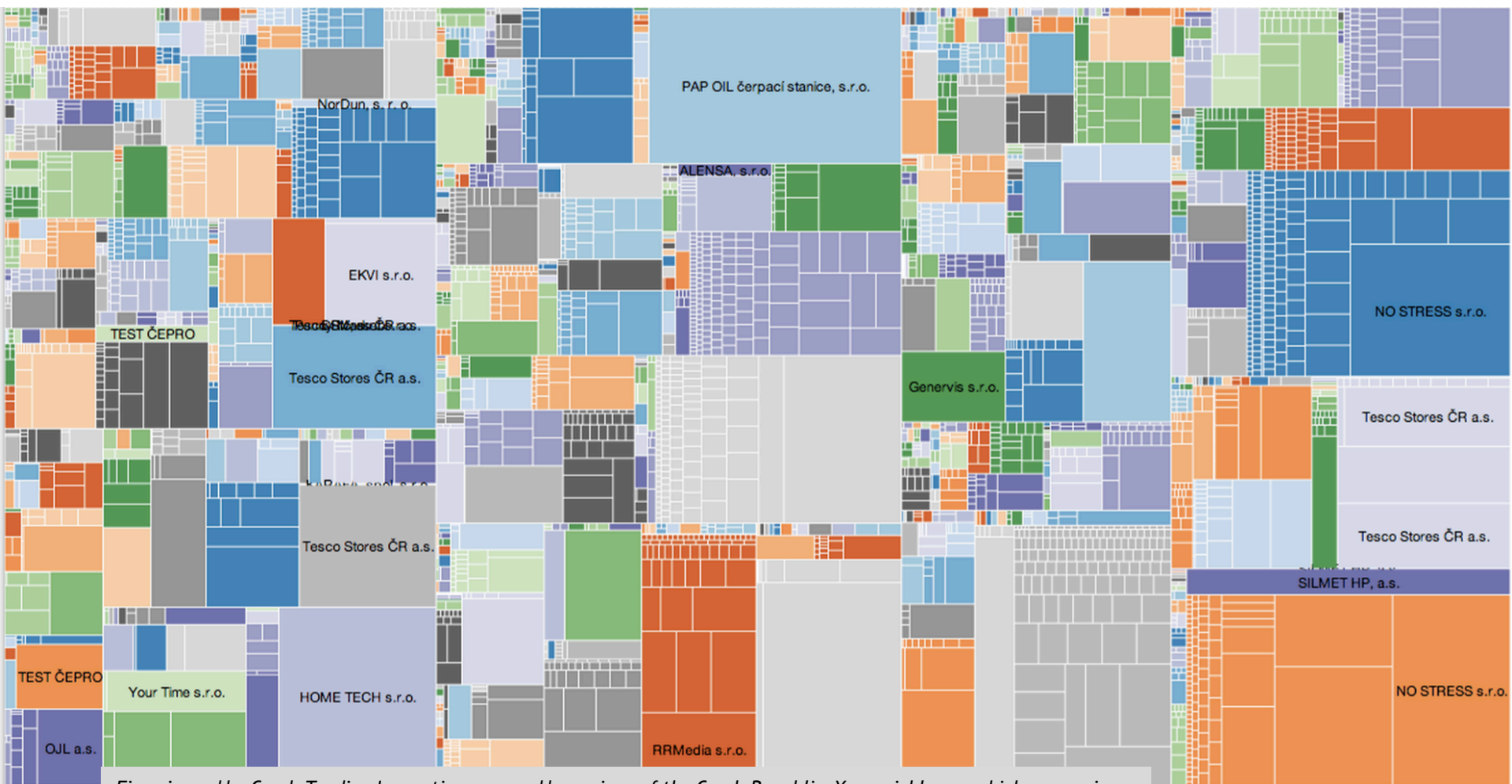
Lates achievements. Note that the filtering is not enough in this case, we'll have to introduce something more clever.

LDVM IMPLEMENTATION



Backstage: When a LOD guy tries to work with map projections. The black shape should be Slovakia.

LDVM IMPLEMENTATION

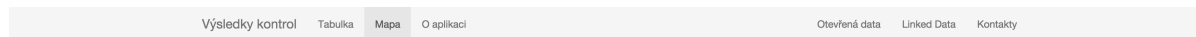


Fines issued by Czech Trading Inspection grouped by regions of the Czech Republic. You quickly see which companies had to pay the biggest fines. No stress...

LDVM IMPLEMENTATION

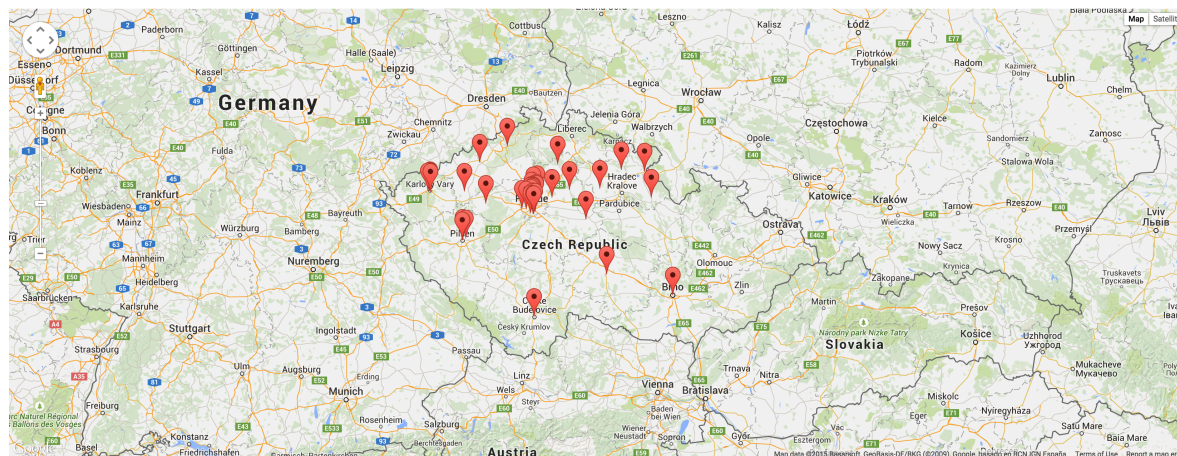
Advertisement

► <http://vysledkykontrol.cz>

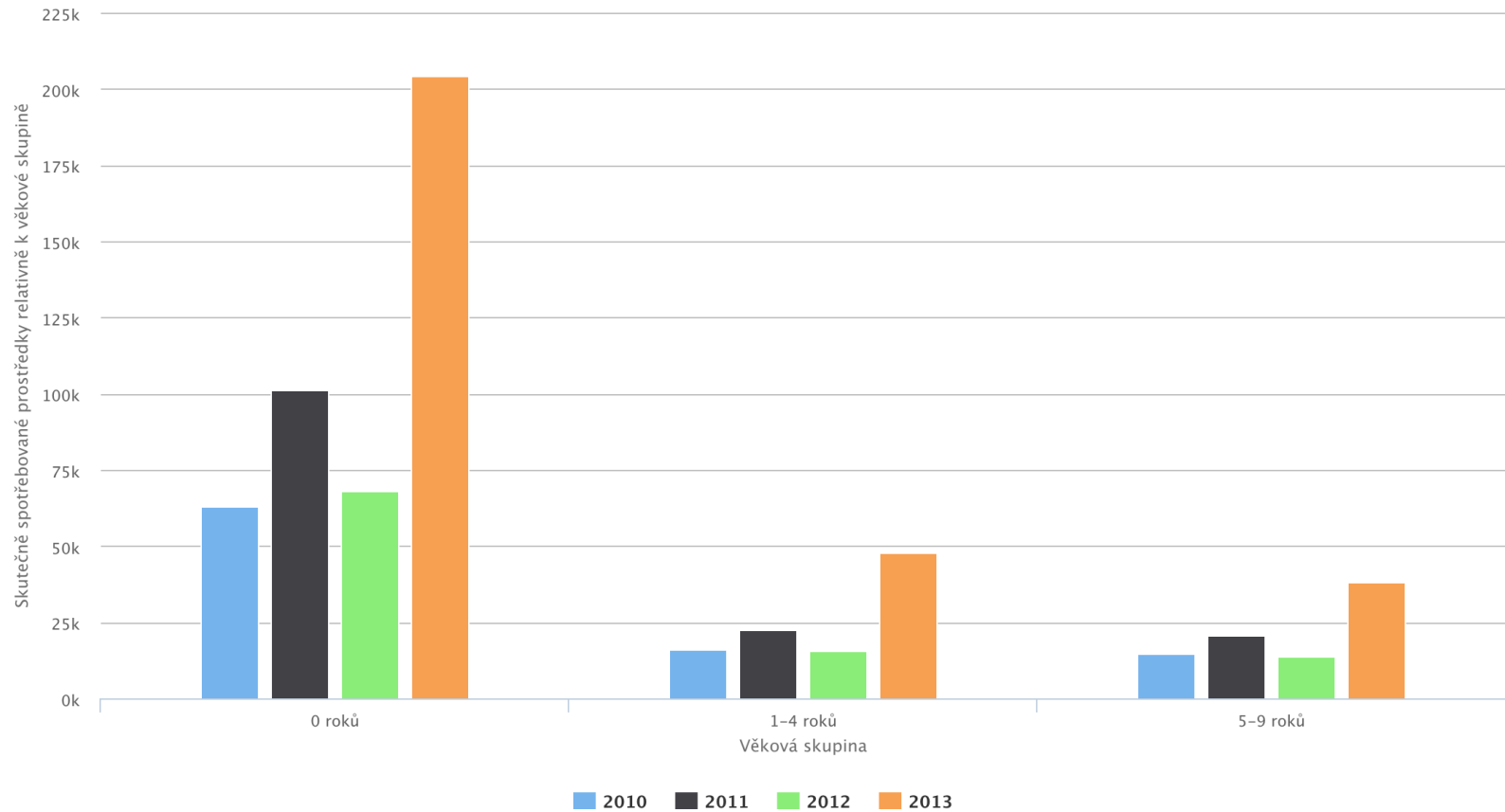


Zde můžete prohledávat otevřená propojená data o kontrolách provedených kontrolními orgány ČR. Tato stránka je v beta stádiu.

Zobrazují 50 kontrol z celkových 57024 v aktuální mapové oblasti. Pokud chcete zobrazit všechny kontroly, zasoďte tak, aby jich v oblasti bylo méně než 50.



Výdaje na předškolní a základní vzdělávání v okresech ČR v porovnání s velikostí dětské populace dle Ministerstva financí ČR a Českého statistického úřadu



DataCube values fixed to get this chart:

- Rozpočtová položka
 - 3111
- Okres
 - Semily

The most advanced visualizer we currently have is the DataCube visualizer. Bosses love charts.

LDVM IMPLEMENTATION

Available components

- ▶ Datasources
 - ▶ Easy to create
- ▶ Analyzers
 - ▶ Just SPARQL queries right now
- ▶ Transformers
 - ▶ Easy to implement, SPARQL queries
 - ▶ Hard to check compatibility
- ▶ Visualizers
 - ▶ Google Maps, OpenLayers, Treemap, Sunburst, PackLayout, DataCube
- ▶ Do your own!
 - ▶ Even now, you can implement your own and extend our library.

Thank you!

