institute MIHAILO PUPIN

Modeling, Fusion and Exploration of Regional Statistics and Indicators with Linked Data tools

Valentina Janev, Vuk Mijović,
Dejan Paunović, Uroš Milošević

contact: Valentina.Janev@institutepupin.com

Institute Mihajlo Pupin (IMP)

1858 -1935

- The biggest and the oldest R&D Institute in ICT area in the whole SE Europe
- Affiliated to the University of Belgrade
- Recruitment directly from University through internships/diploma work

HTTP://WWW.PUPIN.RS

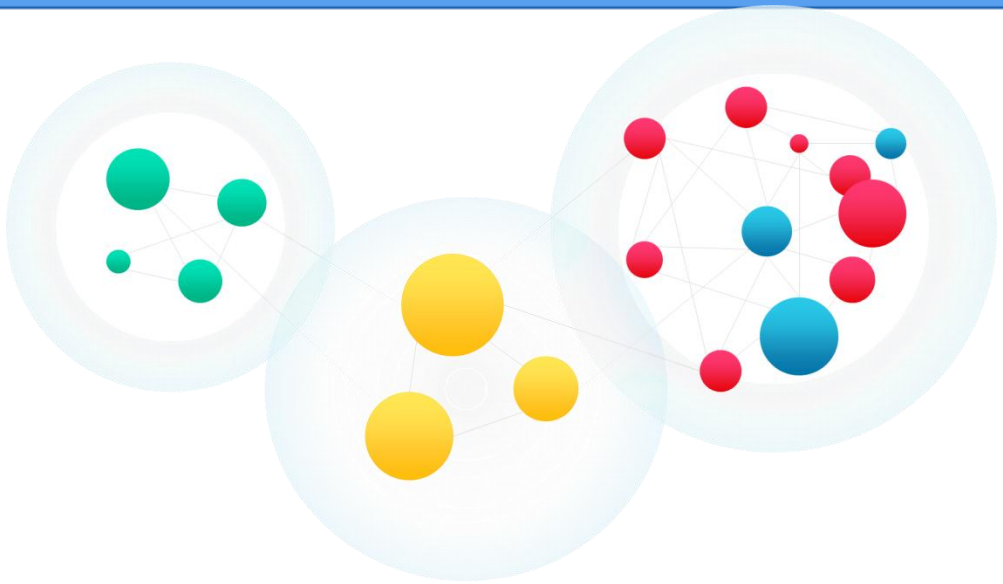
institute MIHAILO PUPIN

Overview

- Background
- Linked Data Tools (LOD2, GeoKnow)
- Serbian e-Government Case Study
- Modeling and Fusion
- Exploration and Visualization
- Concluson

Linked Open Data

- Intelligent and innovative use of large volumes of publicly available data



- Data is truly open if it is technically open (available in a machine-readable standard format, which means it can be retrieved and meaningfully processed by a computer application) and legally open (explicitly licensed in a way that permits commercial and non-commercial use and re-use without restrictions) [*World Bank Group*]

institute MIHAILO PUPIN

LOD2 - Creating Knowledge out of Interlinked Data

- Instrument: Large-Scale Integrating Project
- Objective: Intelligent Information Management
- Duration: 09/2010 – 08/2014
- Total Budget: 10,2 M€

- Consortium: 14 Partners from 11 European Countries + 1 Associated Partner from Korea

UNIVERSITÄT LEIPZIG
NUI Galway OÉ Gaillimh
OPENLINK SOFTWARE
Wolters Kluwer Deutschland
Zemanta
SEMANTIC WEB COMPANY school • consulting • projects • events • media
TenForce Pragmatic Knowledge Management
exalead connect the dots
Freie Universität Berlin
KAIST
CWI
INSTYTUT INFORMATYKI GOSPODARCZEJ
VŠE V PRAZE
institute MIHAILO PUPIN

GeoKnow - Making the Web an Exploratory for Geospatial Knowledge

- Instrument: CollaborativeProject (STReP)
- Objective: Intelligent Information Management
- Duration: 12/2012 – 11/2015
- Total Budget: **4,4 M** €

- Consortium: 7 Partners from 5 European Countries

UNIVERSITÄT LEIPZIG

The Stack

The Linked Data Stack comprises a number of tools for managing the life-cycle of Linked Data. The life-cycle comprises in particular the stages:

- Extraction of RDF from text, XML and SQL
- Querying and Exploration using SPARQL
- Authoring of Linked Data using a Semantic Wiki
- Semi-automatic link discovery between Linked Data sources
- Knowledge-base Enrichment and Repair

http://stack.linkeddata.org/
http://stack.lod2.eu/

Components in the Stack

Are you interested in using some of the Linked Data stack tools? Check the list of components available in the stack. Most of the tools that are part of the stack have online demos, give them a try!

View details »

How to Start

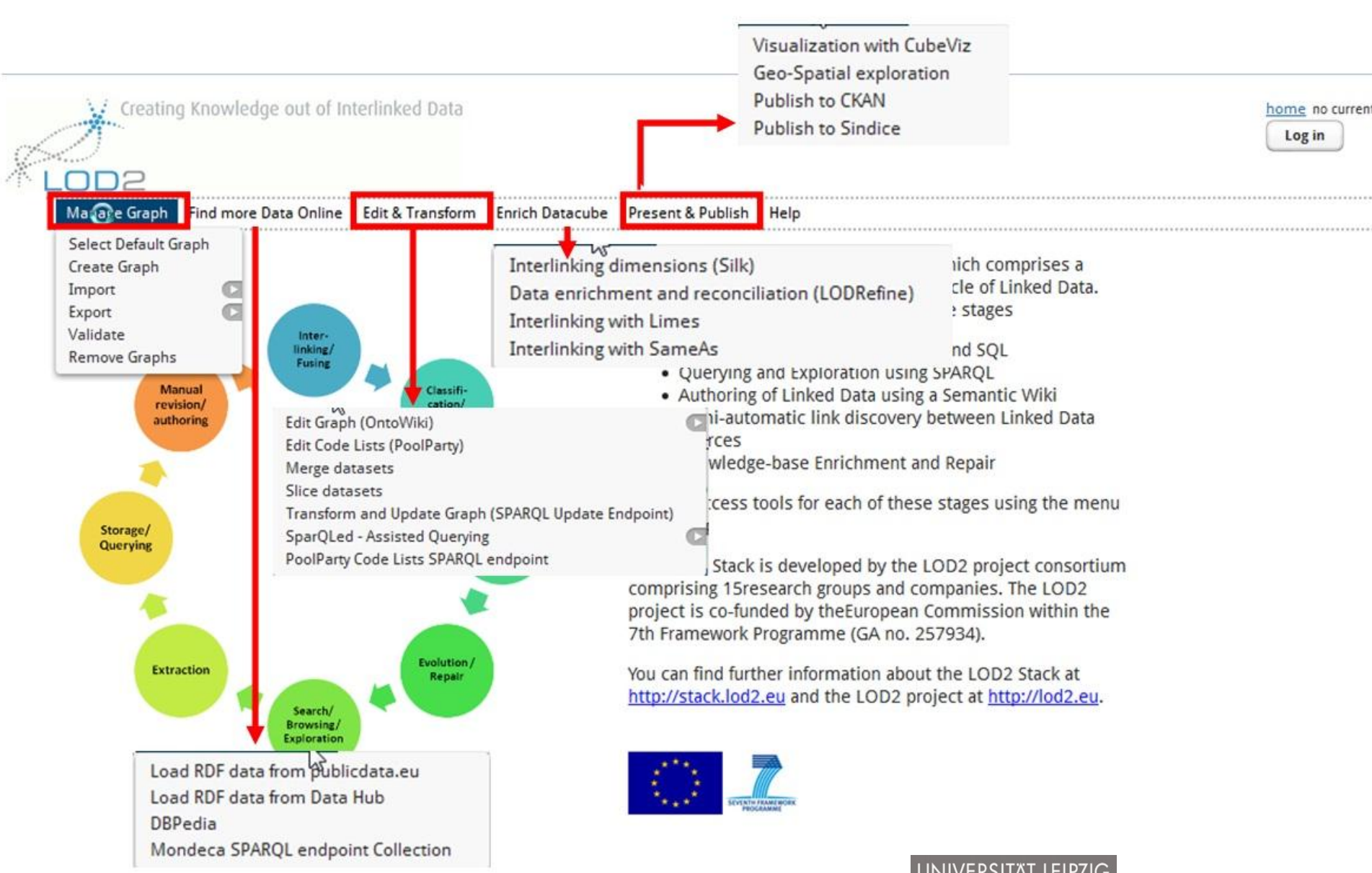
The Linked Data Stack is a collection of tools that give support to the Linked Data lifecyfle. These tools are created under different research founding programs and third parties, and you can use these tools in your Linked Data creation process.

View details »

How to Contribute

Have you created that missing tool that supports the Linked Data lifecycle?. You can contribute to the Linked Data stack by adding your tool to our repository and reach targeted users from out community.

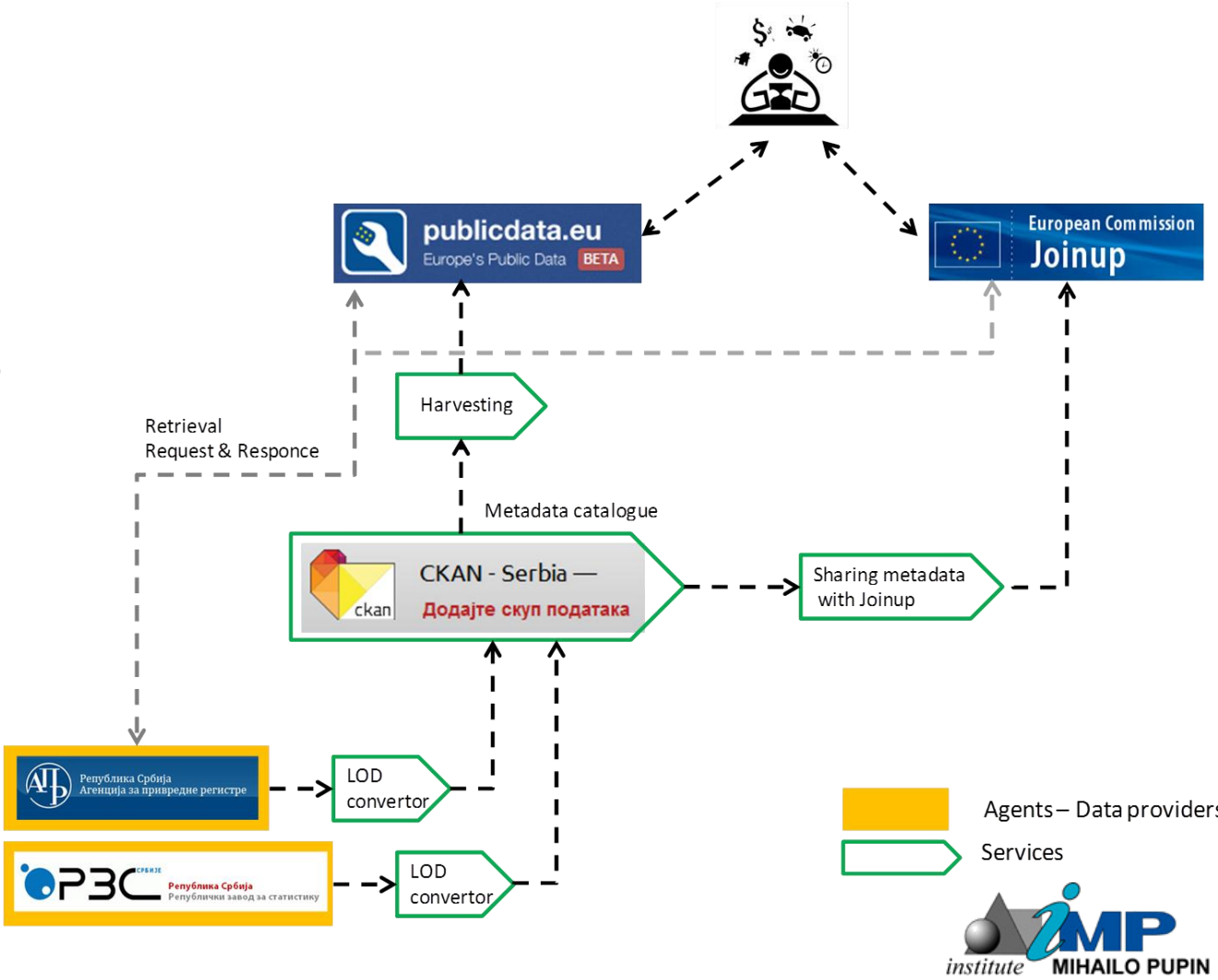
View details »


Creating Knowledge out of Interlinked Data
LOD2
home
Log in
Manage Graph
Find more Data Online
Edit & Transform
Enrich Datacube
Present & Publish
Help
Visualization with CubeViz
Geo-Spatial exploration
Publish to CKAN
Publish to Sindice
Select Default Graph
Create Graph
Import
Export
Validate
Remove Graphs
Interlinking dimensions (Silk)
Data enrichment and reconciliation (LODRefine)
Interlinking with Limes
Interlinking with SameAs
Inter-linking/ Fusing
Classifi-cation/
Manual revision/ authoring
Storage/ Querying
Extraction
Evolution / Repair
Search/ Browsing/ Exploration
Edit Graph (OntoWiki)
Edit Code Lists (PoolParty)
Merge datasets
Slice datasets
Transform and Update Graph (SPARQL Update Endpoint)
SparQLed - Assisted Querying
PoolParty Code Lists SPARQL endpoint
Load RDF data from publicdata.eu
Load RDF data from Data Hub
DBPedia
Mondeca SPARQL endpoint Collection
ich comprises a
cle of Linked Data.
stages
nd SQL
Querying and Exploration using SPARQL
Authoring of Linked Data using a Semantic Wiki
i-automatic link discovery between Linked Data
rces
wledge-base Enrichment and Repair
cess tools for each of these stages using the menu
Stack is developed by the LOD2 project consortium comprising 15research groups and companies. The LOD2 project is co-funded by theEuropean Commission within the 7th Framework Programme (GA no. 257934).
You can find further information about the LOD2 Stack at http://stack.lod2.eu and the LOD2 project at http://lod2.eu.
UNIVERSITÄT LEIPZIG
TenForce
The Pragmatic Company
AKSW
institute
MIHAILO PUPIN

Summary: Serbian e-Government Case Study

Goals

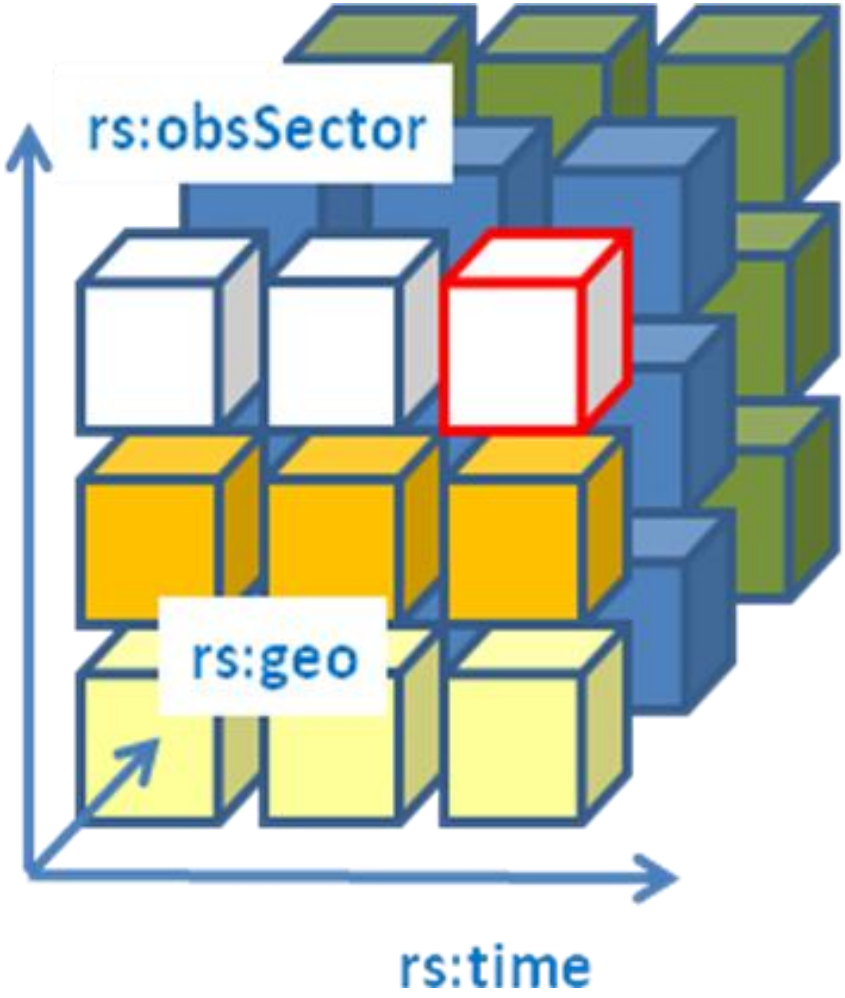
- Contribute to modernization of statistical services
- Define infrastructure for publishing and Integration of public sector information into the LOD cloud
- Provide an integrated collection professional tools for statistical LOD publishing (1) and consuming (2)



Spatio-Temporal Analysis tool for Linked Data

- Technical challenges:
  - use emerging W3C standards (RDF Data Cube vocabulary, SKOS vocabulary OWL-Time ontology) in the field of statistical and temporal geospatial information management
  - implement mechanisms for efficient search and visualization of data at different levels of granularity (space, time)
  - support different analysis / visualization options (bar charts, run charts, spatial analysis on a geographical maps)
- Testing and Validation
  - **Use case:**
    - Serbian Business Registers Agency – Registries
    - Statistical Office of the Republic of Serbia – Dissemination DB

RDF Data Cube RDF vocabulary

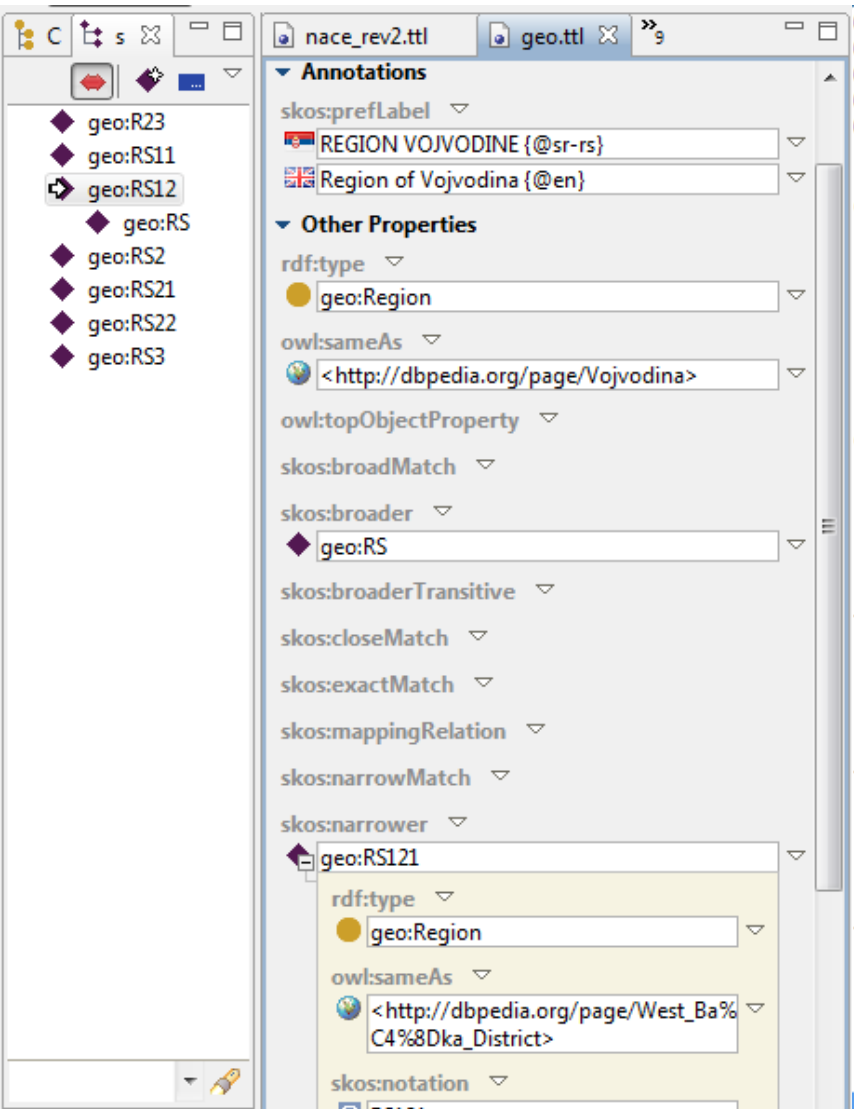


- published by the W3C Government Linked Data Working Group , focused purely on the publication of multi-dimensional data
- compatible with the cube model that underlies SDMX (Statistical Data and Metadata eXchange), an ISO standard for exchanging and sharing statistical data and metadata among organizations
- **W3C Recommendation 16 January 2014**

institute MIHAILO PUPIN

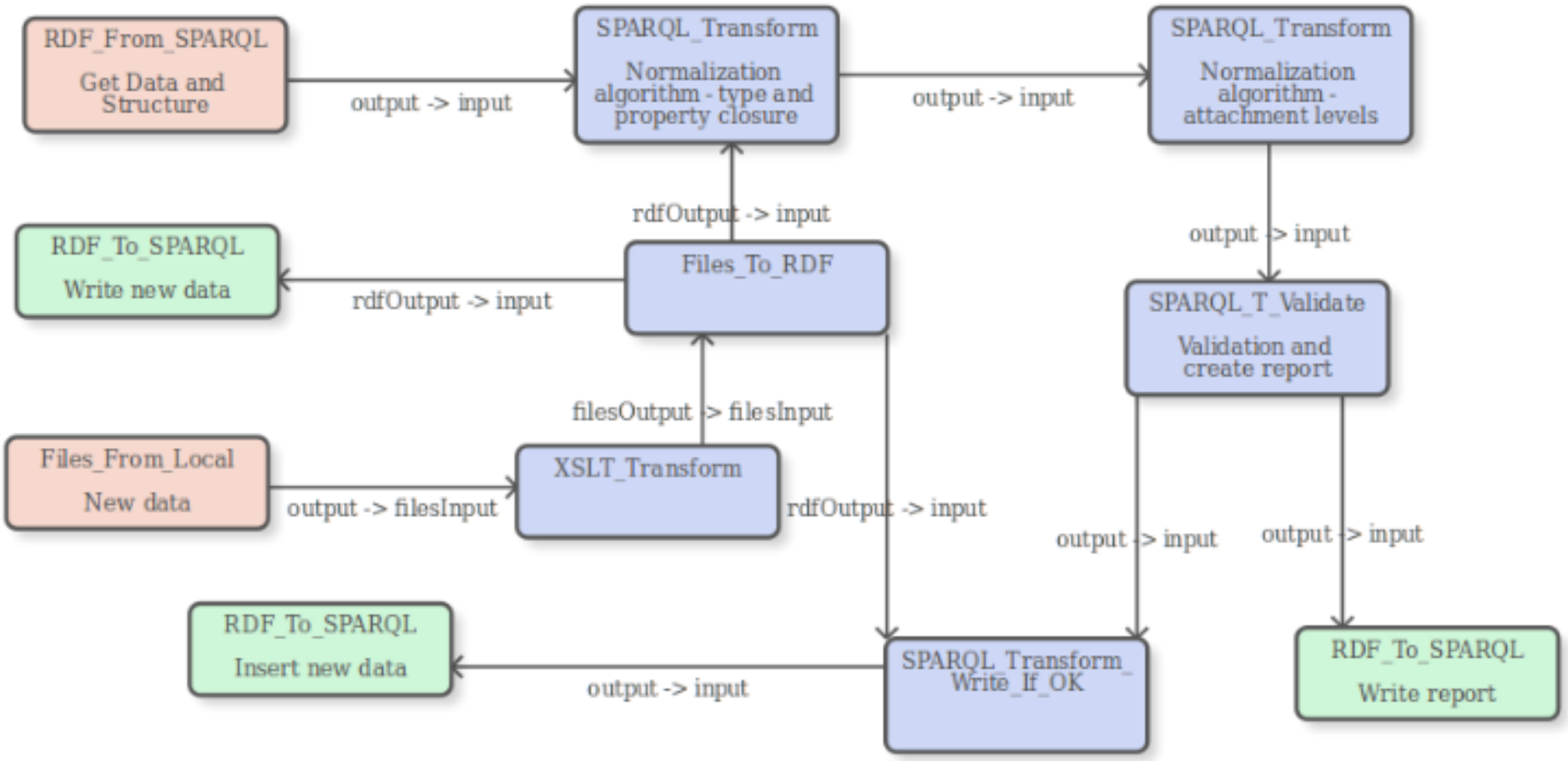
SKOS Vocabulary for Modeling

Hierarchical code lists




```
@prefix rdf:    <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix rdfs:   <http://www.w3.org/2000/01/rdf-schema#> .
@prefix skos:   <http://www.w3.org/2004/02/skos/core#> .
@prefix geo:    <http://elpo.stat.gov.rs/lod2/RS-DIC/geo/> .
@prefix time:   <http://elpo.stat.gov.rs/lod2/RS-DIC/time/> .


geo:RS21
        rdf:type geo:Region ;
        owl:sameAs      <http://dbpedia.org/page/%C5%A0umadija_and_
        skos:broadergeo:RS ;
        skos:narrower geo:RS212 , geo:RS216 , geo:RS211 , geo:RS215
        skos:narrower geo:RS218 , geo:RS214 , geo:RS217 ;
        skos:notation "RS21"^^xsd:string ;
        skos:prefLabel "Region of Sumadija and Western Serbia"@en ,

geo:RS218
        rdf:type geo:Region ;
        owl:sameAs <http://dbpedia.org/page/%C5%A0umadija_District>
        skos:prefLabel "ŠUMADIJSKI OKRUG"@sr-rs, "Šumadijska area"@
        skos:notation "RS218"^^xsd:string .


time:Y1980M1
        rdf:type time:P1M ;
        skos:broader time:Y1980Q1 ;
        skos:notation "Y1980M1"^^xsd:string ;
        skos:prefLabel "1980/january"@en .

time:Y2010Q1
        rdf:type  time:P3M ;
        skos:prefLabel "2010/I quarter"@en ;
        skos:notation "Y2010Q1"^^xsd:string ;
        skos:broader time:Y2010 .

time:Y2011
        rdf:type  time:P1Y ;
        skos:prefLabel "2011"@en ;
        skos:notation "Y2011"^^xsd:string .
```

Fusion, Validation, Interlinking…



UnifiedViews

http://www.semantic-web.at/de/unifiedviews

institute MIHAILO PUPIN

Exploration and Visualization (CubeViz)

Data Selection

Select a dataset

GDP usage - Exports of..

Select Unit and Measurement

Observed Value

Configure the Dimensions

Observed Indicator
2 of 2 Selected

Times
4 of 13 Selected

Unit Measure
1 of 2 Selected

Link | Show Visualization

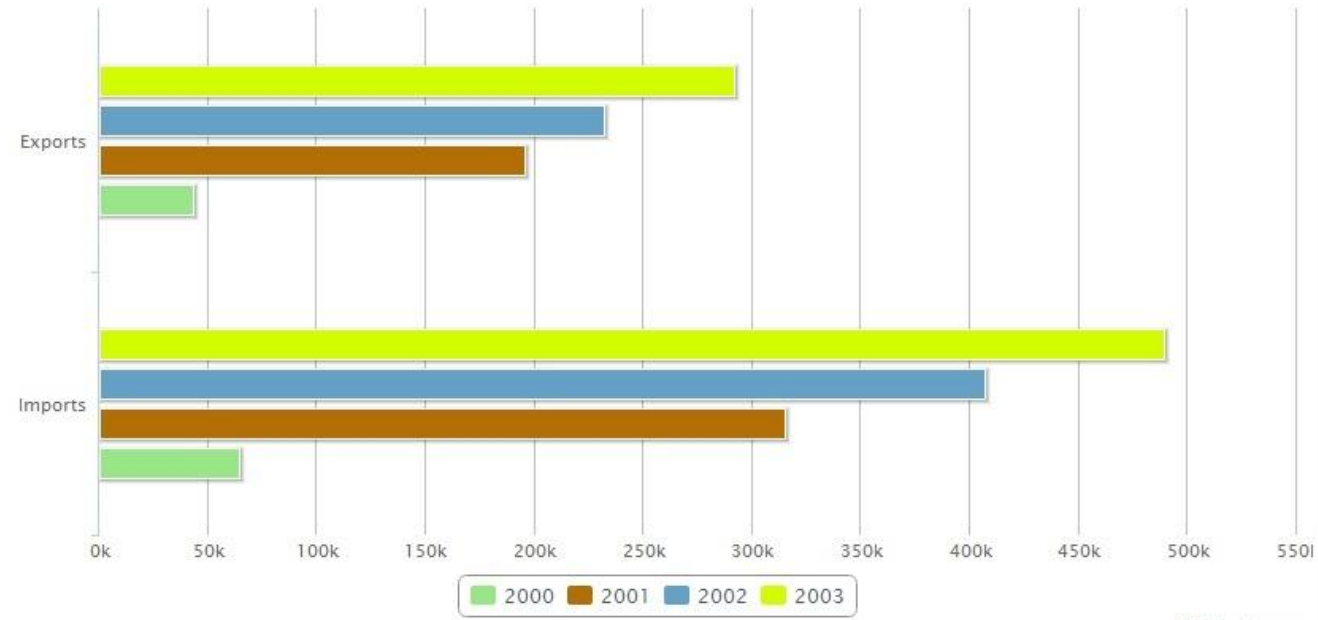
Knowledge Bases

Edit View

http://elpo.stat.gov.rs/test/cvmod/noDataSe
http://elpo.stat.gov.rs/lod2/RS-DATA/National_accounts/dsd/
RS_NA_Gross_value_added
OntoWiki System Configuration

Visualization for http://elpo.stat.gov.rs/lod2/RS-DATA/National_accounts/RS_NA_GDP_usage

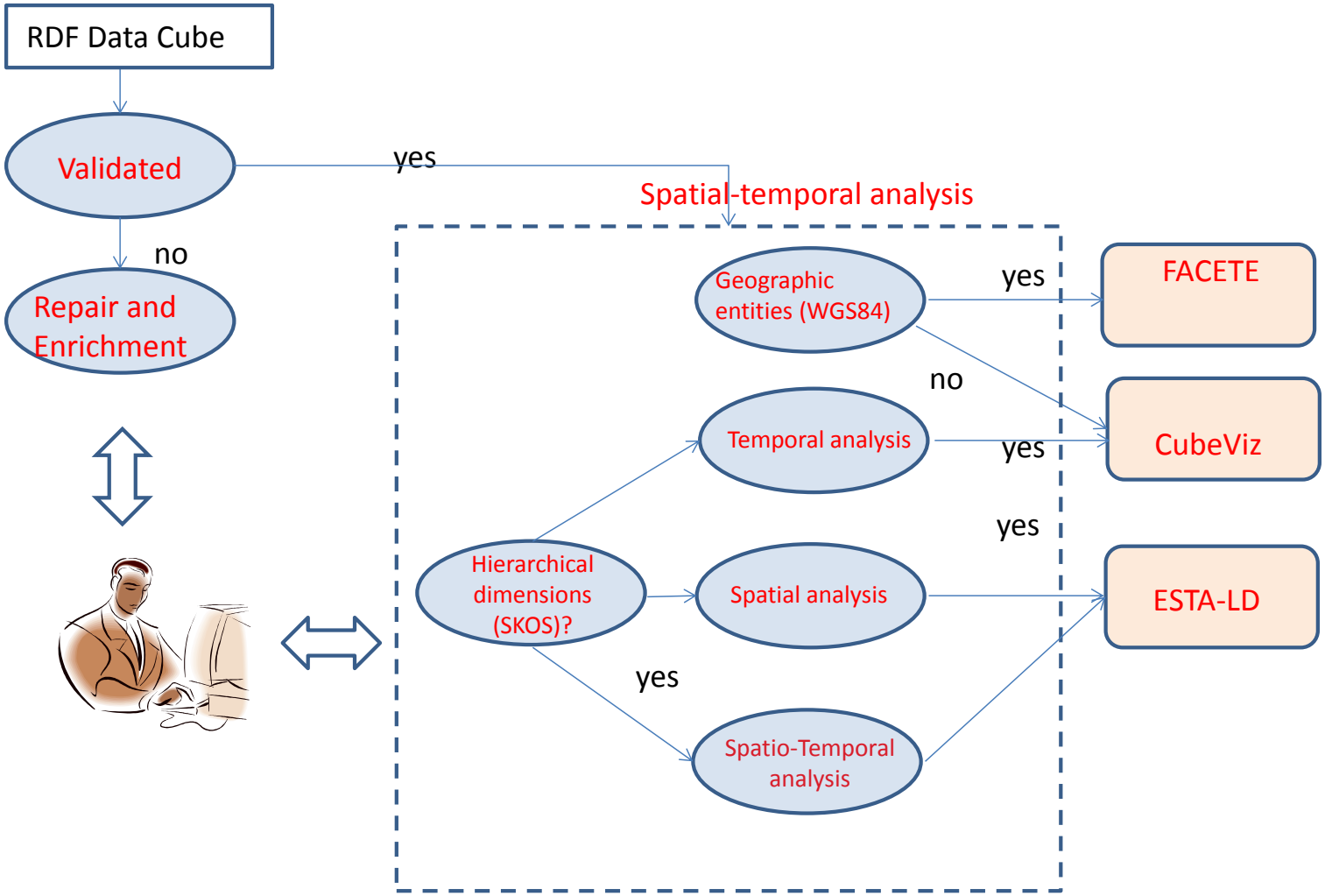
Visualization Of http://elpo.stat.gov.rs/lod2/RS-DATA/National_accounts/GDP_usage



Show Information About | Selected Configuration | Retrieved Data

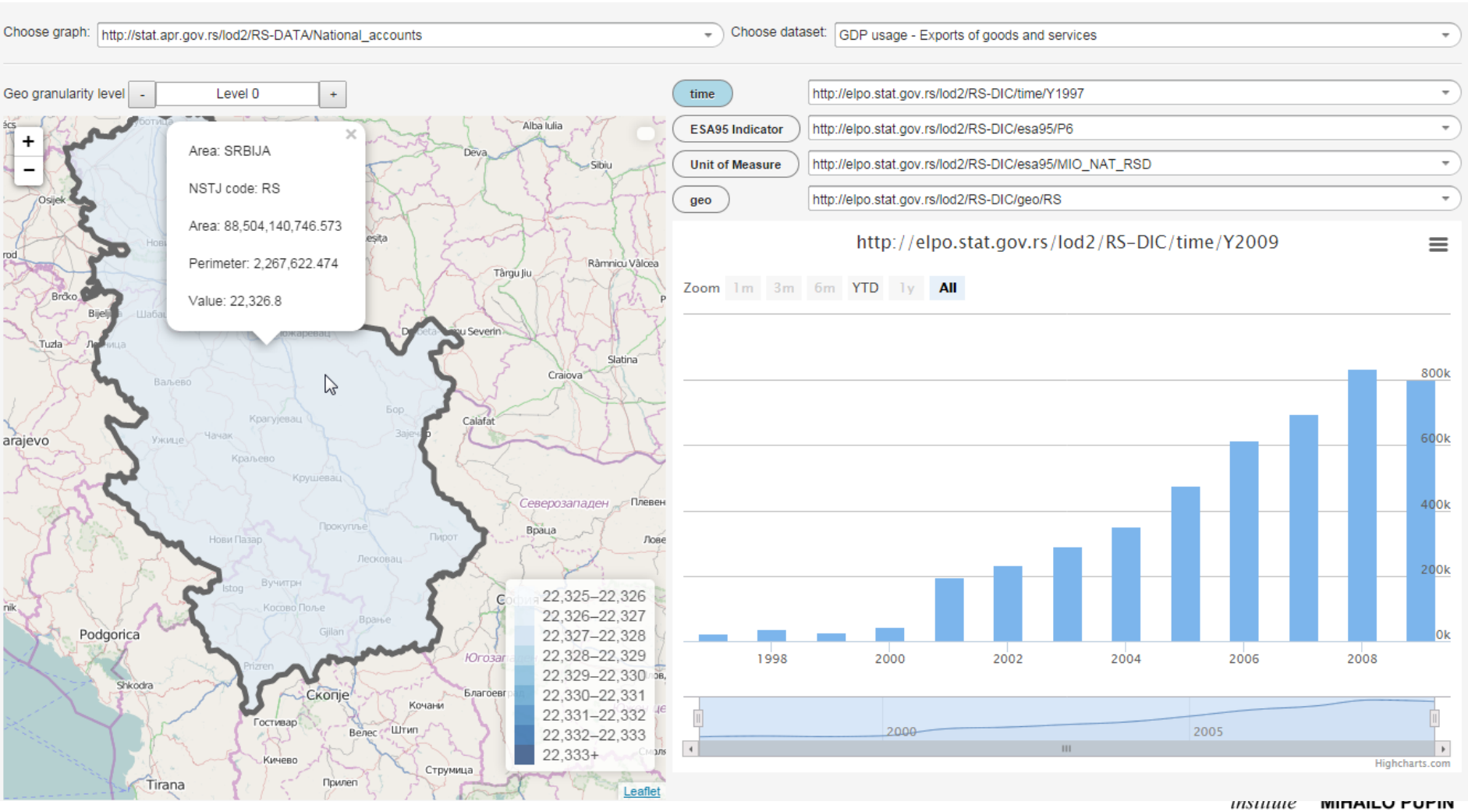
Export as

Guided analysis of spatio-temporal data

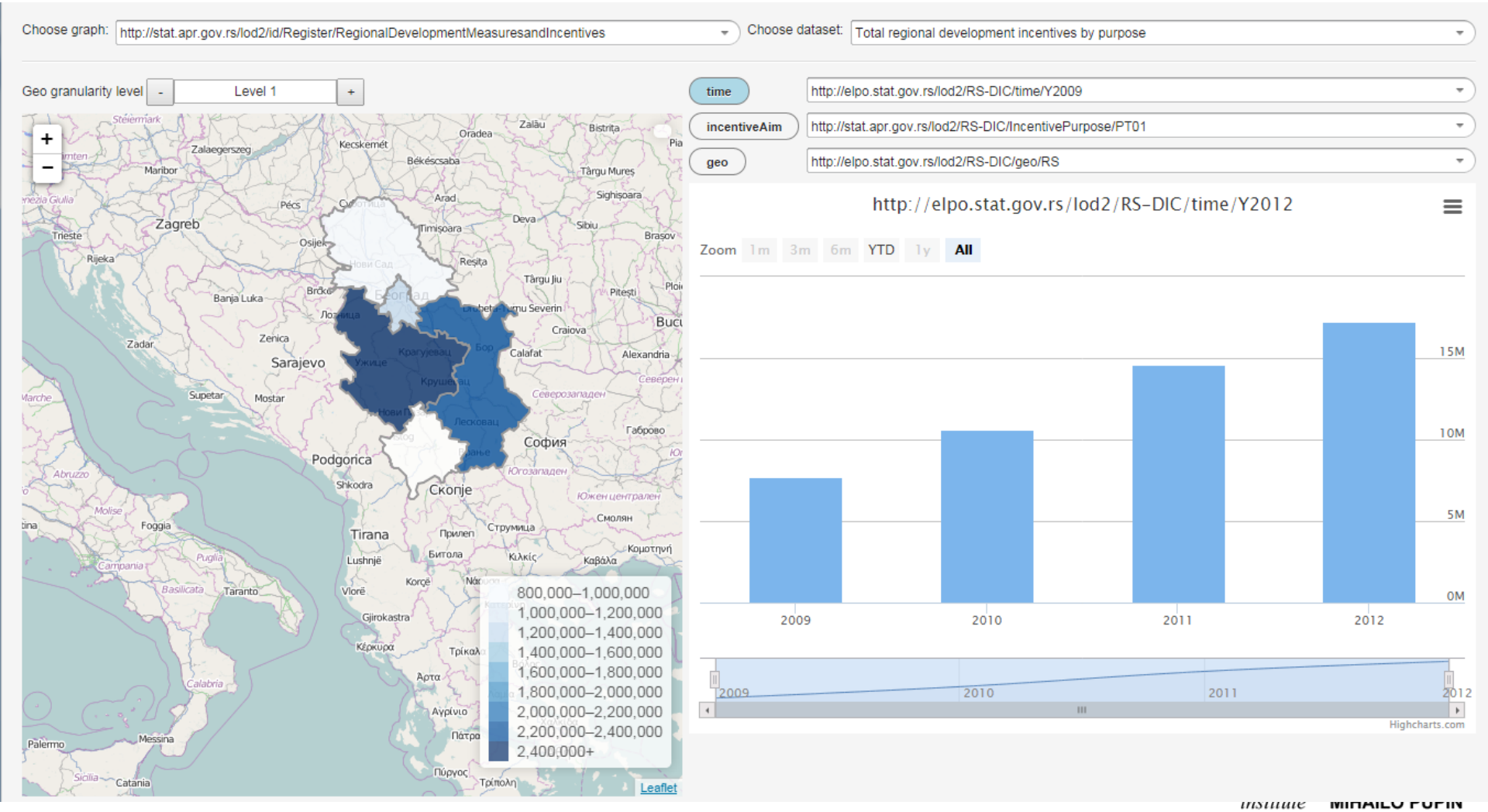


institute MIHAILO PUPIN

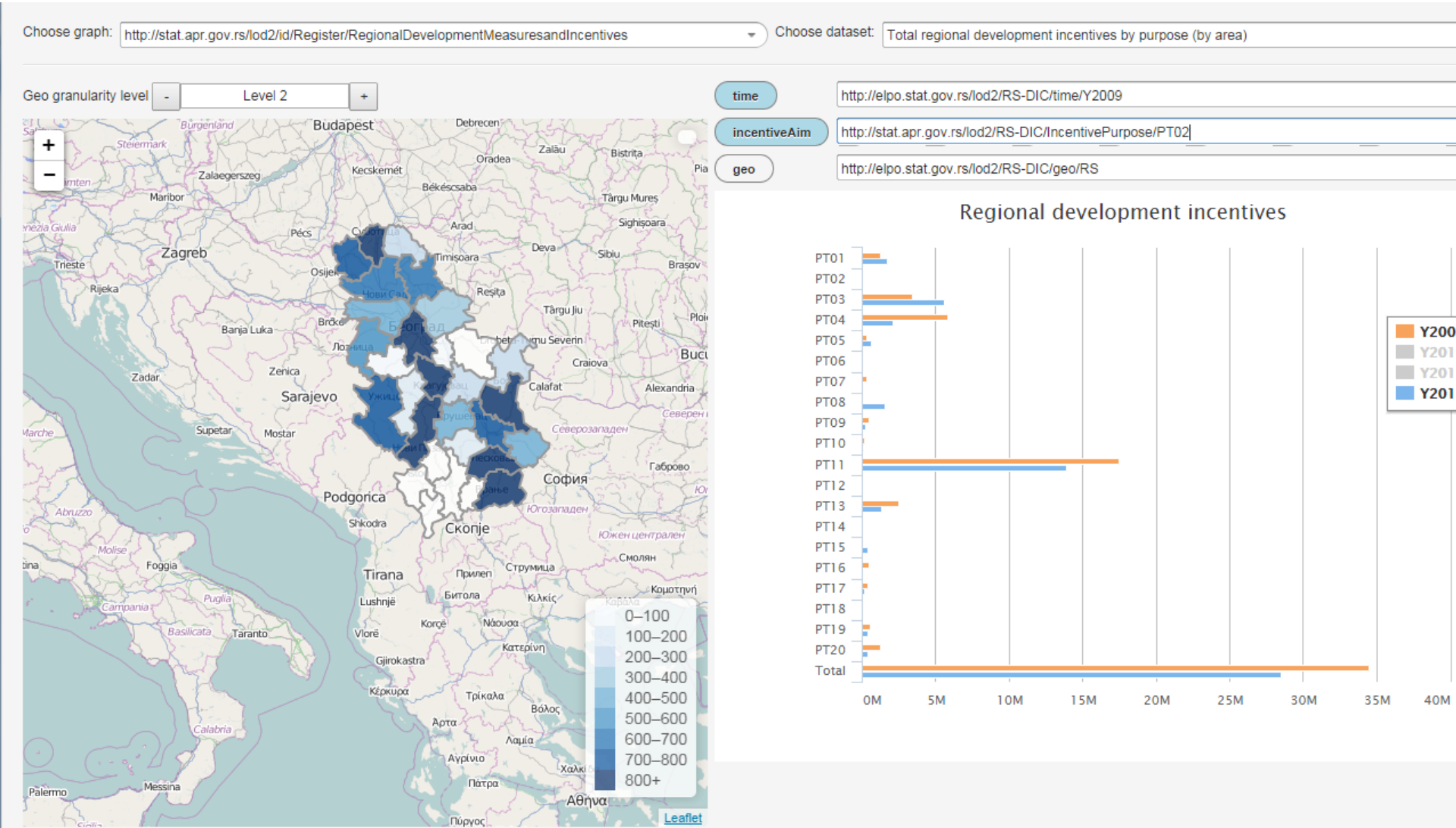
Exploration and Visualization



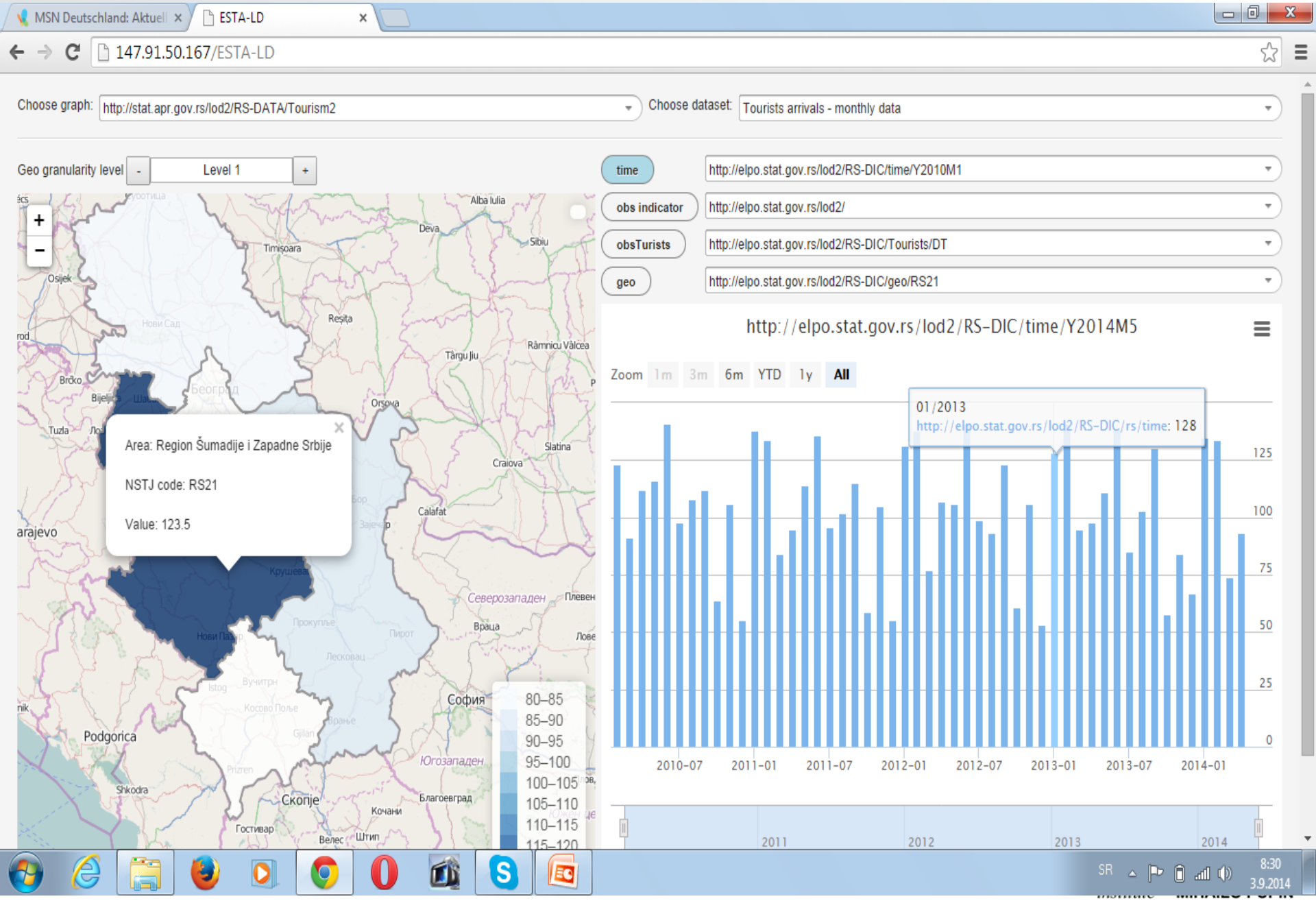
Exploration and Visualization



Exploration and Visualization

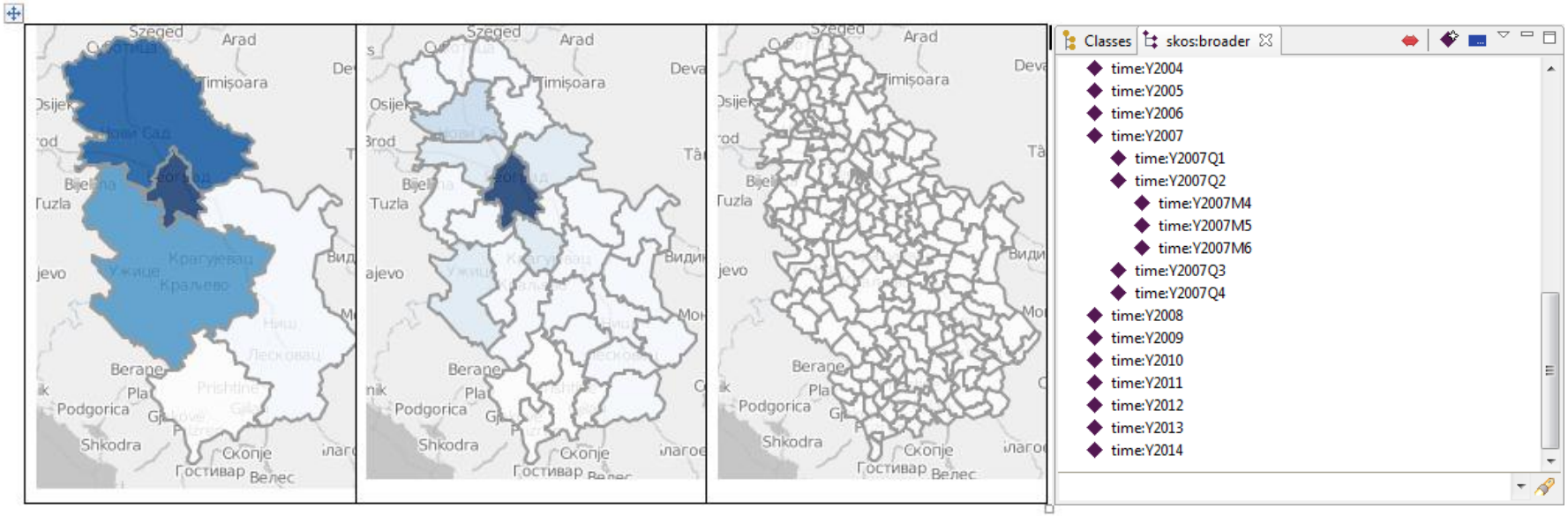


Tourism data



ESTA-LD Implementation

- Transformation of shape files into GeoJSON format
- Preparing the code lists (space, time dimensions)
- Querying statistical indicators from local Virtuoso RDF Data store using
- Visualization using *Highcharts*



Benefits for early adopters

- improve the accessibility and transparency of data by extending the existing public services with new features
- standardize the data publishing / consumption process
- ensure interoperability (e.g. integration of data from the Register with data from the Dissemination database of the Statistical Office of the Republic of Serbia, SORS)
- allow advanced analysis and visualization of available indicators
  (i.e. spatio-temporal data on a geographical map)

Conclusions

- need for guided analysis of the Linked Data retrieved from arbitrary SPARQL endpoint
- ESTA-LD contributes to further development and standardization of the Linked Data technologies
- Serbian e-Government Use Case showcased the potential of Linked Data tools and technologies for integrating data and building interactive dashboards

67/ESTA-LD

rs/lod2/id/Register/RegionalDevelopmentMeasuresandIncentives

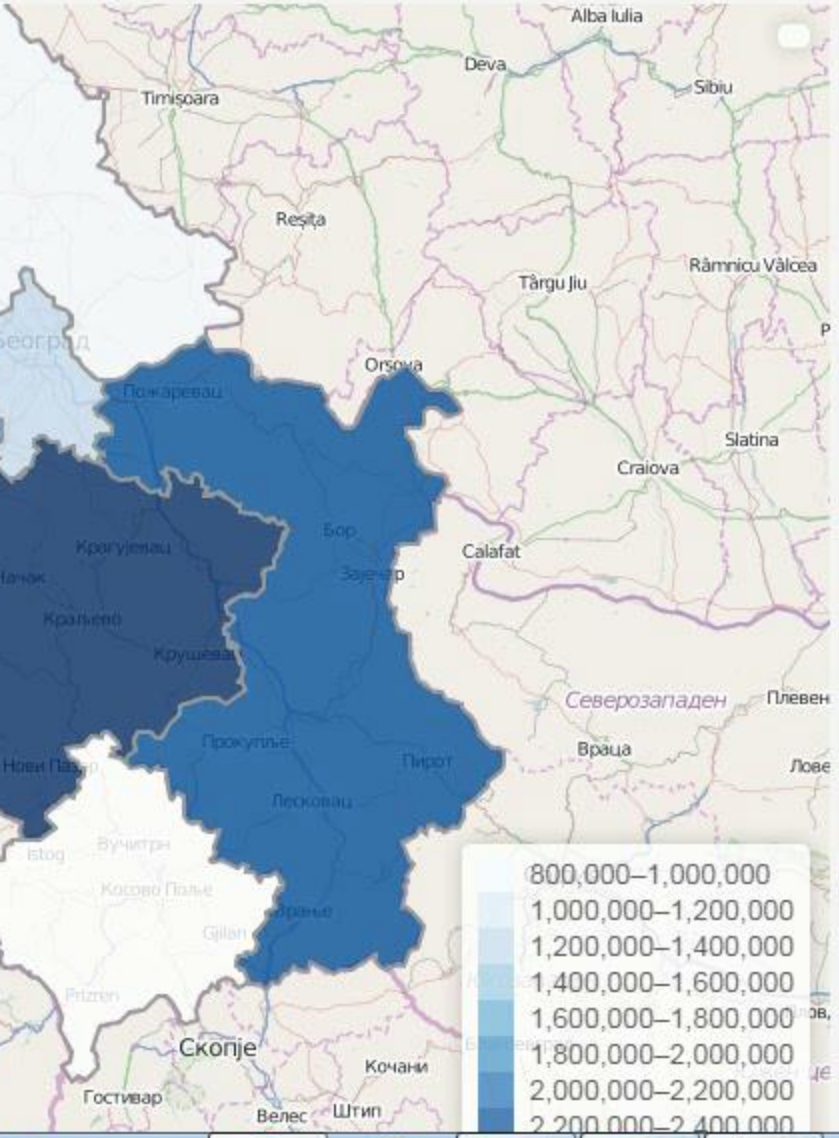
Choose dataset: Total regional development incentives by purpose

Level 1 +

time http://elpo.stat.gov.rs/lod2/RS-DIC/time/Y2009

incentiveAim http://stat.apr.gov.rs/lod2/RS-DIC/IncentivePurpose/PT01

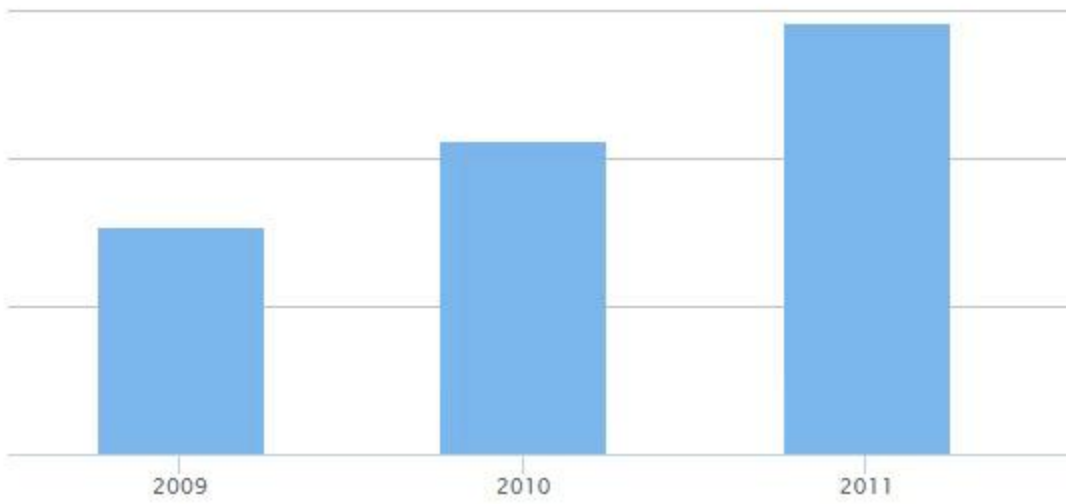
geo http://elpo.stat.gov.rs/lod2/RS-DIC/geo/RS



http://elpo.stat.gov.rs/lod2/RS-DIC/time/Y201

Zoom 1m 3m 6m YTD 1y All

2012
http://elpo.stat.gov.rs/lod2/RS



Thank you for your attention!

Thank you for your attention!

Thank you for your attention!