

# Designing SDI4Apps POI Base

---

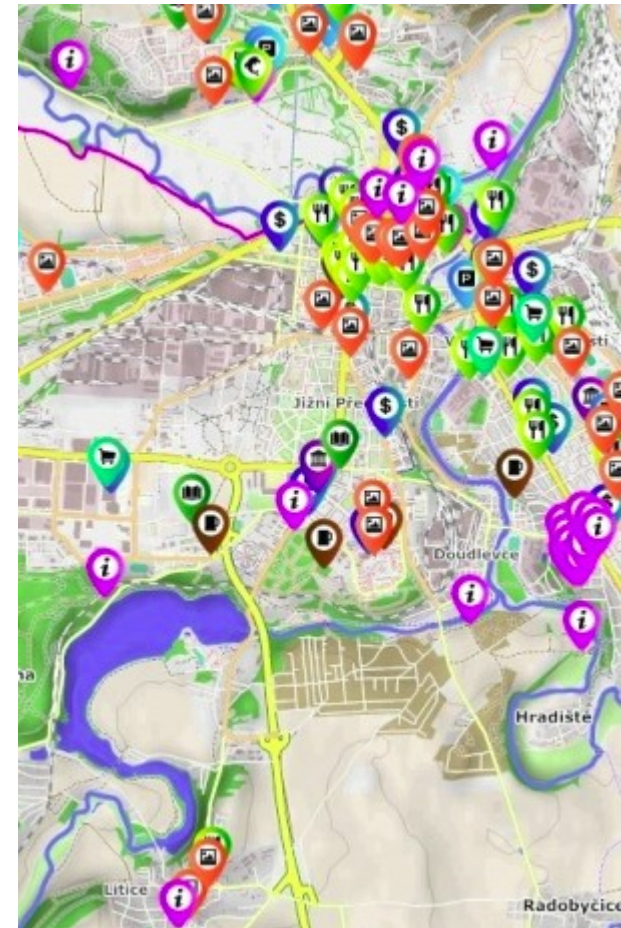
Otakar Čerba\*, Tomáš Mildorf\*, Raitis Bērziņš\*\*

\*University of West Bohemia, Czech Republic

\*Baltic Open Solutions Centre, Latvia

# Objectives

- SPOI introduction
- Data model
- SPOI vs. OpenPOIs
- SPOI publication
- Future steps & Conclusions



# SPOI data set

- Open data for tourism and travelling
- Many heterogeneous input data → data harmonization process
- Based on standards, semantic description and Linked data
- Data stored as RDF triples in Virtuoso
- Published on map portal and SPARQL endpoint
- Open Database License (ODbL)



# SPOI data set – Why?

- Request of project partners and subjects participating in tourism
- Needs of
  - Enrichment of existing data and applications
  - Promotion of particular places and territories
- Support of all types of participants of tourism

**SPOI**

**4 247 371**

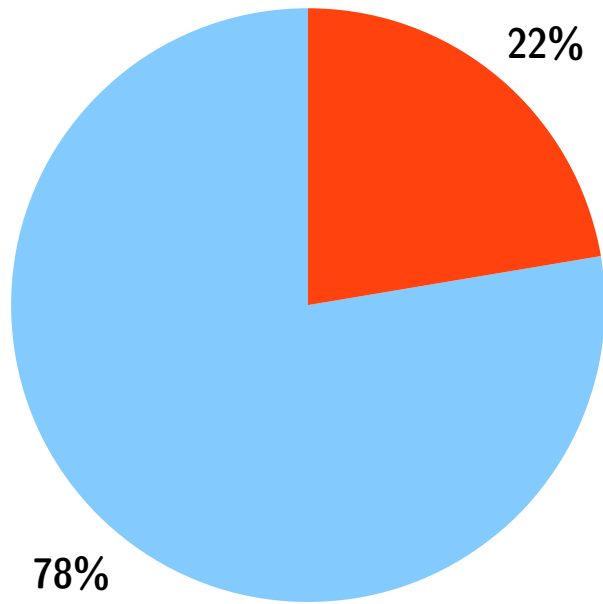
**POIs**



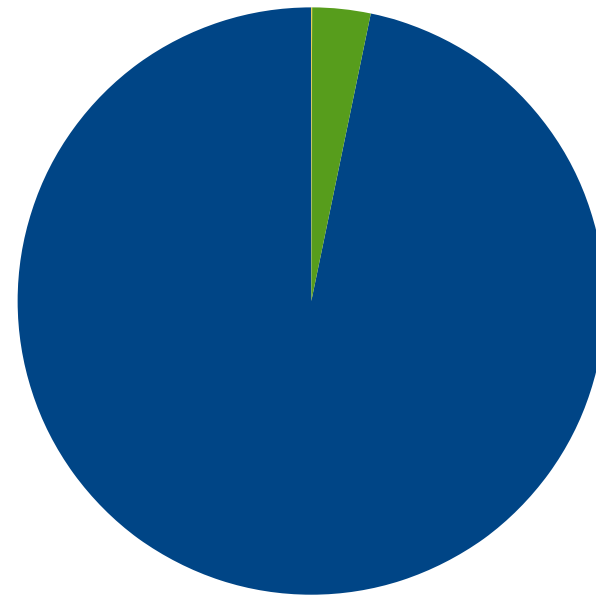
# Source data

- **OpenStreetMap**
- **GeoNames.org (dumps)**
- **Local data – documents from Posumavi region, Sicily and travel agency**
- **Semantic data – experimental ontologies (OWL) of UWB (ski resort, sight in Rome)**

# Statistics



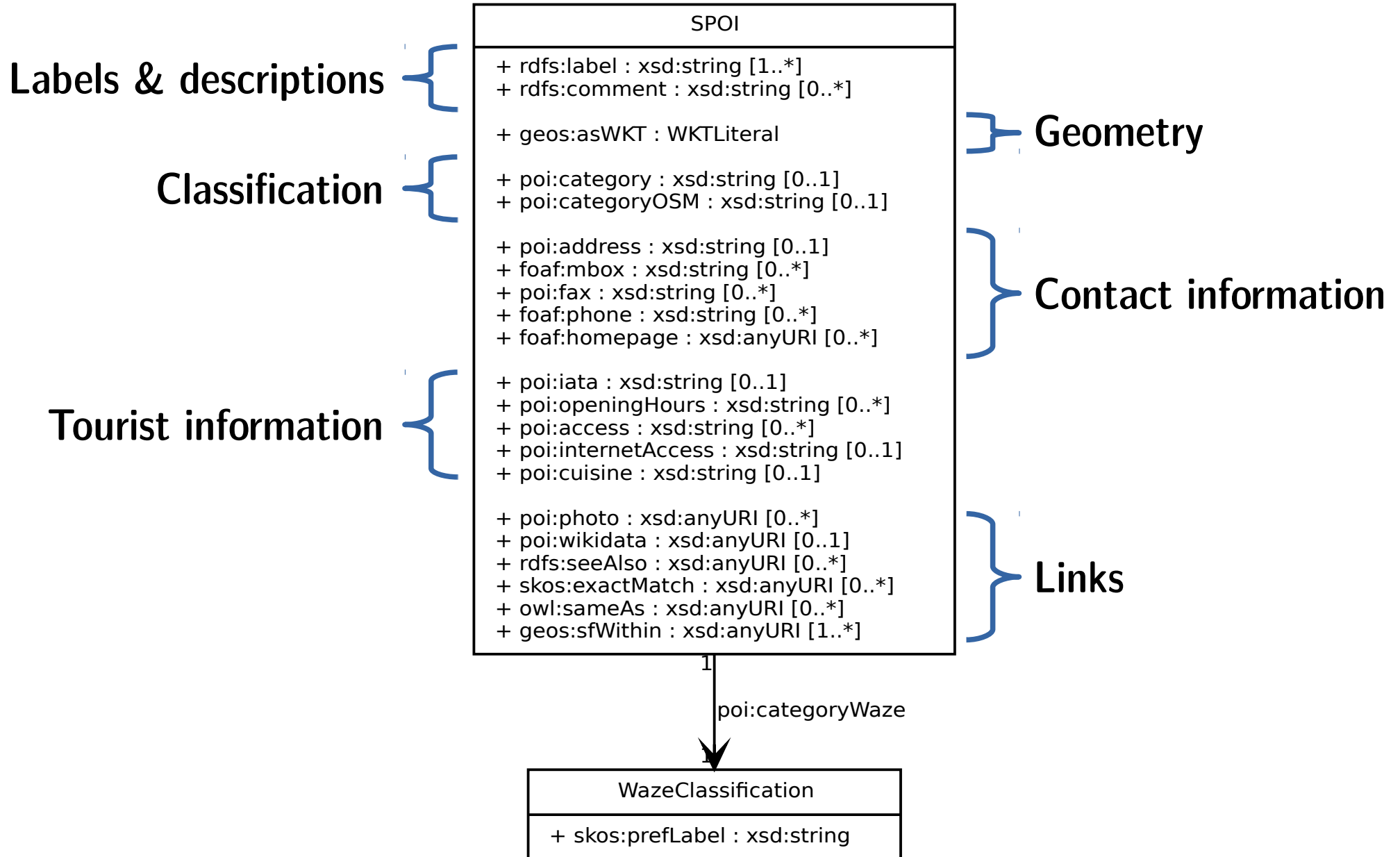
■ Europe  
■ Africa



■ OSM  
■ GeoNames  
■ Other

■ 4109425  
97%

# Data model





# Identifier

- URI
- Combination
  - ISO 3166-1 alpha-2 country code
  - Acronym of category of POI according Waze navigation data
  - Coordinates (long\_lat)

```
<rdf:Description rdf:about="http://www.sdi4apps.eu/poi/ML_NAT_0.8712_14.9746">
```

# Links

- **Classification – vocabulary to re-use**
- **Photos and pictures**
- **Same features**
  - **Web pages (Wikipedia, Wolfram|Alpha)**
  - **Linked data resources (DBpedia, GeoNames.org)**
- **Topological relation – to countries (GeoSPARQL)**

# SPOI & 5-star rating system

- \* **Accessibility: SPARQL endpoint & Open Database License (OdbL)**
- \*\* **Structured data: JSON, XML, CSV or RDF**
- \*\*\* **Non-proprietary format: RDF and other exports**
- \*\*\*\* **URIs: Identifiers of each POI**
- \*\*\*\*\* **Links: see the previous slide**

# Data harmonization

- Transcription to structured data (table, text → XML)
- Preparation of common vocabularies (Waze) and mappings (OSM, GeoNames.org → Waze)
- Filtering
- Adding information (links to countries)
- Transformation to common data model
- Export to common data format (RDF)

# Data harmonization – technology

- XSLT 2.0 templates
  - Transformation language based on XML
  - Process XML based files and non-structured files
- Saxon processor
  - Java-based
  - XSLT + input data

# SPOI vs. OpenPOIs

Property	SPOI	<u>OpenPOIs</u>
Number of <u>POIs</u>	> 3.2 millions	> 9.5 millions
Coverage	Europe	World
Main sources of data	<u>OpenStreetMap</u>	<u>GeoNames</u> , <u>DBpedia</u> (these resources are mentioned in <u>Singh, 2012</u> , a short survey of data demonstrated that many objects originated from <u>OpenStreetMap</u> )
Ways of data providing	SPARQL endpoint	Custom API, WFS
Output data formats	Formats provided by Virtuoso tool (RDF, JSON, CSV, <u>Javascript...</u> )	XML, JSON, <u>microdata</u> , RDF

# SPOI vs. OpenPOIs

Area	SPOI	<u>OpenPOIs</u>
Seaside resort (Croatia)	7	4
<u>Submontane area</u> (Czech republic)	1	0
Mountains (France)	1	1
Rural area (Germany)	28	28
Historical site (Greece)	9	10
Large city (Italy)	57	60
Coast (Latvia)	0	0
Small towns and villages (Netherlands)	6	8
Sport center (Norway)	46	41
Industrial area (Poland)	54	57

# SPARQL endpoint

## Virtuoso SPARQL Query Editor

Default Data Set Name (Graph IRI)

Query Text

```
select distinct ?Concept where {[] a ?Concept} LIMIT 100
```

*(Security restrictions of this server do not allow you to retrieve remote RDF data, see [details](#).)*

Results Format:

HTML ▾

Execution timeout:

0

milliseconds *(values less than 1000 are ignored)*

Options:

Strict checking of void variables

*(The result can only be sent back to browser, not saved on the server, see [details](#))*

Run Query

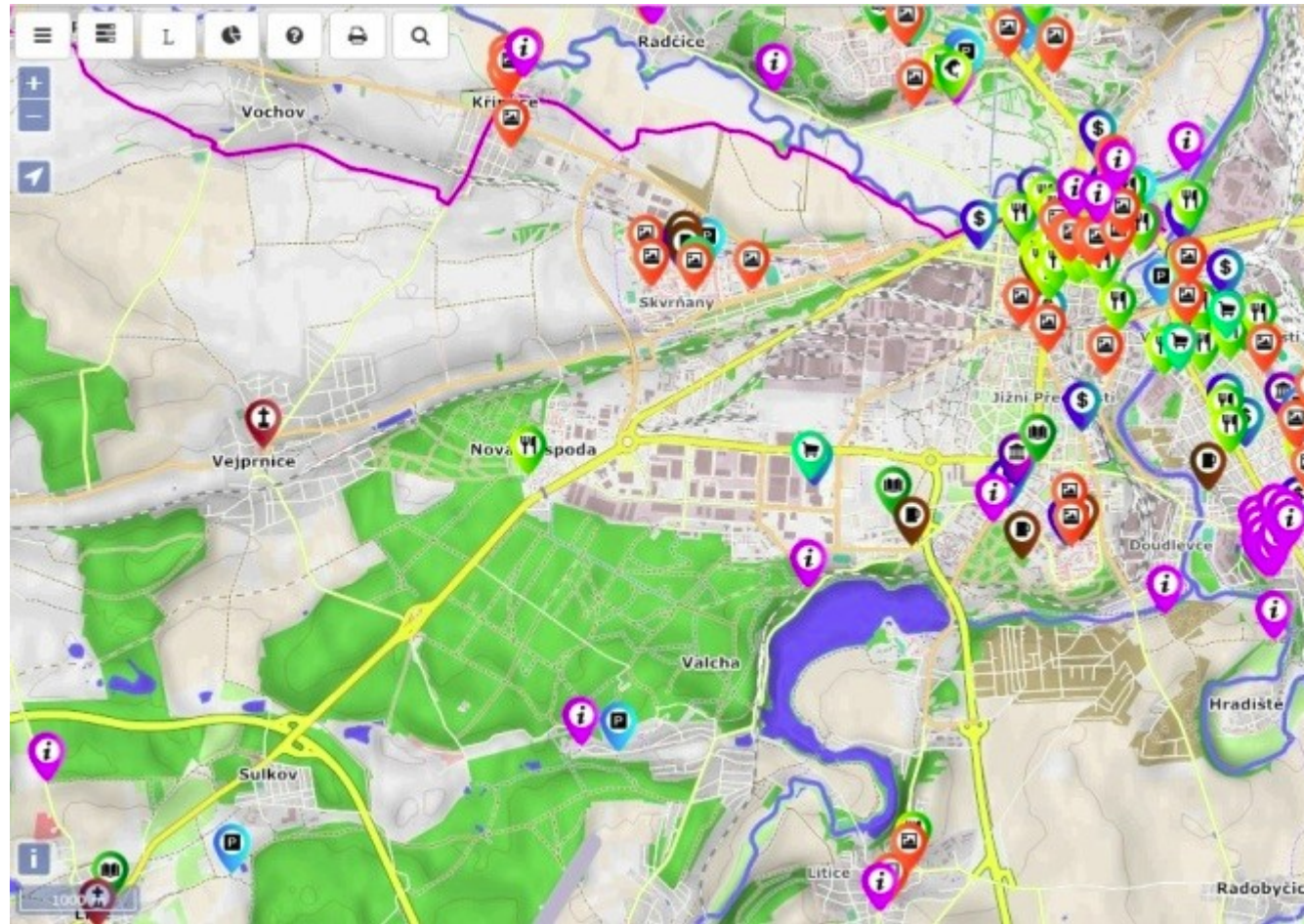
Reset

<http://ha.isaf2014.info:8890/sparql>

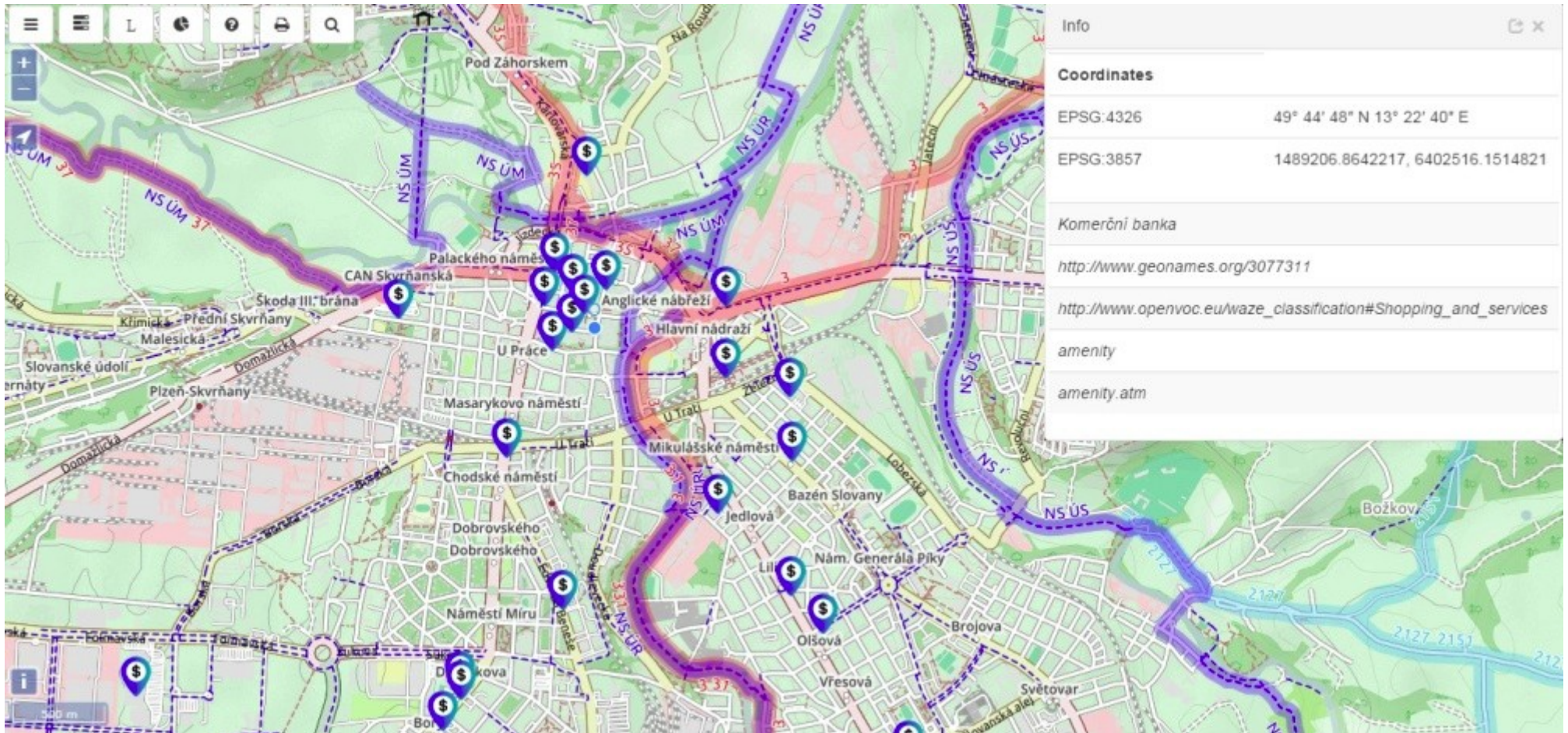


# Map client

- HS layers
- Specific tiles for cycling and biking
- OpenCycleMap, MTB map, OSM...
- Panoramio, OpenWeatherMap



# Map client



<http://ng.hslayers.org/examples/geosparql/>

# Web page

## Over 4,000,000 Points of Interest in the data set

Open and seamless SPOI data set, which is based on Linked data principles, contains over 4 million Points of Interest important for tourism from Europe and Africa.

[sdi4apps.eu/spoi](http://sdi4apps.eu/spoi)  
[gis.zcu.cz/spoi](http://gis.zcu.cz/spoi)

The SDI4Apps

available for other

Its principal target

with other data set containing road

The added value of the

comparison to other similar solutions consists in **implementation of**

**linked data**, using **standardized and respected datatype properties** and development of the **completely**

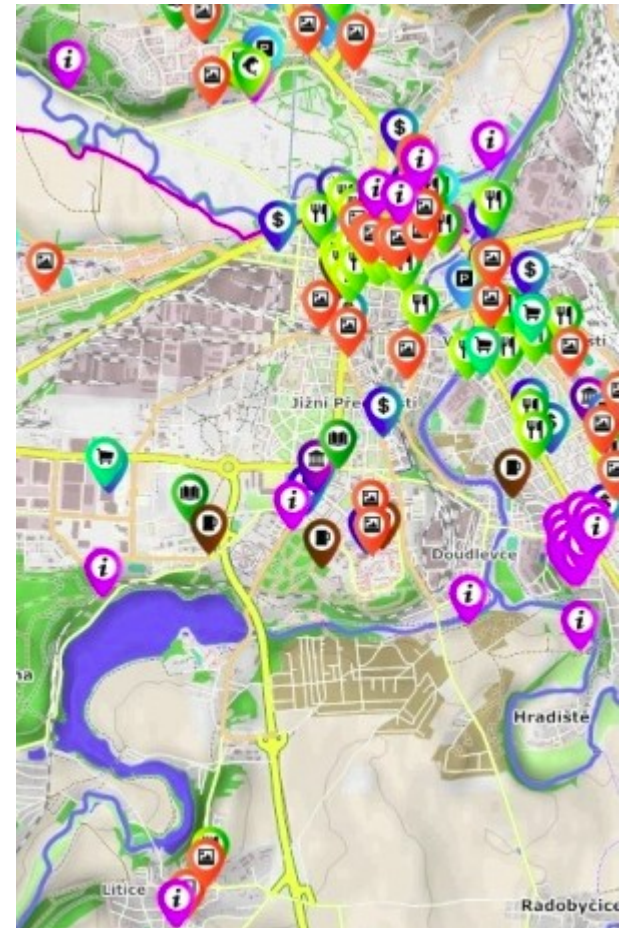
**harmonized data set** with uniform data model and common classification.

# Future steps

- Extension of information resources (imported data, links, APIs)
- Optimization of data model, data storage, data harmonization & data refining
- Cartographic challenges (clustering)
- Context-based application (user will get only information related to concrete needs)
- Analyses & itineraries & routing

# Conclusions

- Common data model
- Re-using existing standards
- Linked data
- SPARQL endpoint



# Thank you for attention

---



Otakar Čerba



[cerba@kma.zcu.cz](mailto:cerba@kma.zcu.cz)



[linkedin.com/in/otakarcerba](https://www.linkedin.com/in/otakarcerba)