Designing SDI4Apps POI Base

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

*University of West Bohemia, Czech Republic
*Baltic Open Solutions Centre, Latvia

Joint International Workshop on Semantic Sensor Networks and Terra Cognita
Bethlehem PA, October 11, 2015
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: 78%
- Africa: 22%

- OSM: 97%
- GeoNames: 3%
- Other: 0%
Data model

Labels & descriptions
- SPOI
  - rdfs:label: xsd:string [1..*]
  - rdfs:comment: xsd:string [0..*]
  - geos:asWKT: WKTLiteral

Classification
- poi:category: xsd:string [0..1]
- poi:categoryOSM: xsd:string [0..1]
- poi:address: xsd:string [0..1]
- foaf:mbox: xsd:string [0..1]
- poi:fax: xsd:string [0..1]
- foaf:phone: xsd:string [0..1]
- foaf:homepage: xsd:anyURI [0..1]
- poi:iata: xsd:string [0..1]
- poi:openingHours: xsd:string [0..1]
- poi:access: xsd:string [0..1]
- poi:internetAccess: xsd:string [0..1]
- poi:cuisine: xsd:string [0..1]
- poi:photo: xsd:anyURI [0..1]
- poi:wikidata: xsd:anyURI [0..1]
- rdfs:seeAlso: xsd:anyURI [0..1]
- skos:exactMatch: xsd:anyURI [0..1]
- owl:sameAs: xsd:anyURI [0..1]
- geos:sfWithin: xsd:anyURI [1..1]

Geometry

Contact information

Tourist information

Links

WazeClassification
- skos:prefLabel: xsd:string
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

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

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

Virtuoso SPARQL Query Editor

Default Data Set Name (Graph IRI)

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

Results Format: HTML
Execution timeout: 0 milliseconds (values less than 1000 are ignored)
Options: Strict checking of void variables

(Security restrictions of this server do not allow you to retrieve remote RDF data, see details.)
(The result can only be sent back to browser, not saved on the server, see details)

Run Query  Reset

Map client

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

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

The SDI4Apps POI data set is available for other data set containing road and other geographic information. Its principal target is to harmonize data set with other data set containing road and other geographic information.

The added value of the approach as comparison to other similar solutions consists in implementation of linked data, using well established 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

linkedin.com/in/otakarcerba