

# SPOI

## SDI4apps: Points of Interest



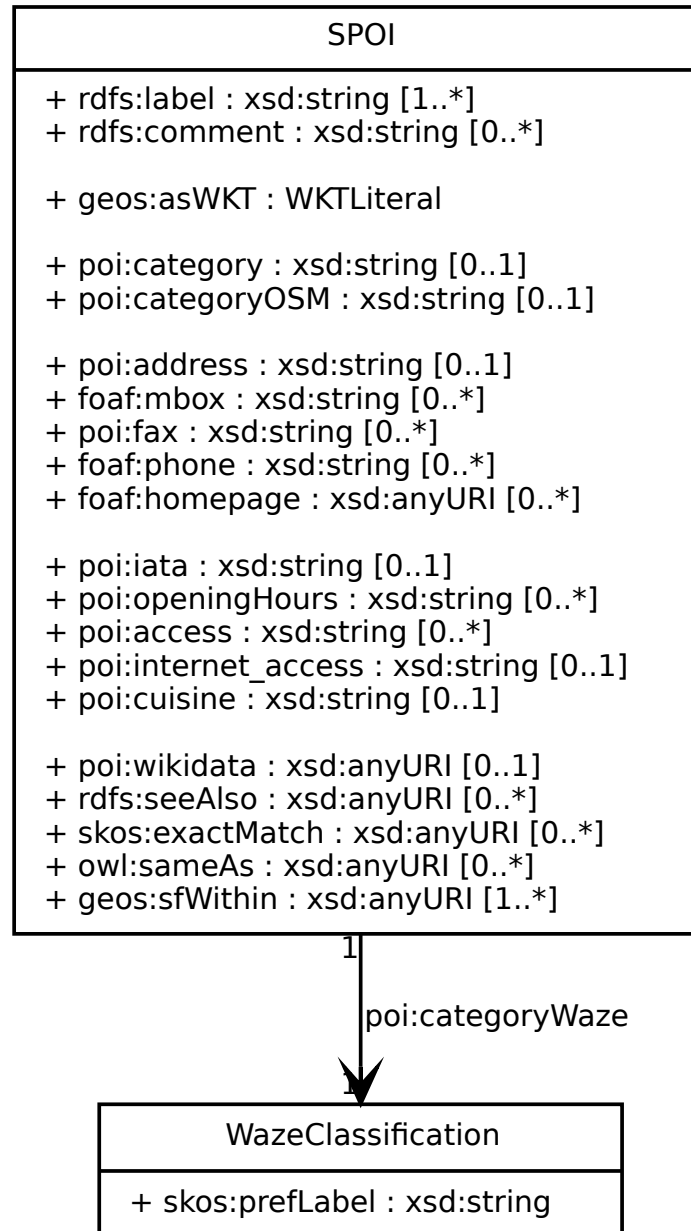
# SPOI data set

- Open data for tourism and travelling
- Many heterogeneous input data
- Data harmonization process
- Based on standards, semantic description and Linked data
- Published on map portal and SPARQL endpoint

# Source data

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

# Data model



# Data model

- **Identifier** – persistent URI
- Description – **label(s)** & description / comment (rdfs:) – multilinguality
- **Geometry** (WGS 84 W3C Geo Positioning → GeoSPARQL WKT as a link)
- Classification (**Waze**, OSM)
- Contact information (foaf:) – address, email...
- Tourism information – opening hours, cuisine...
- Links – **country**...

# Links

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

# Data harmonization – steps

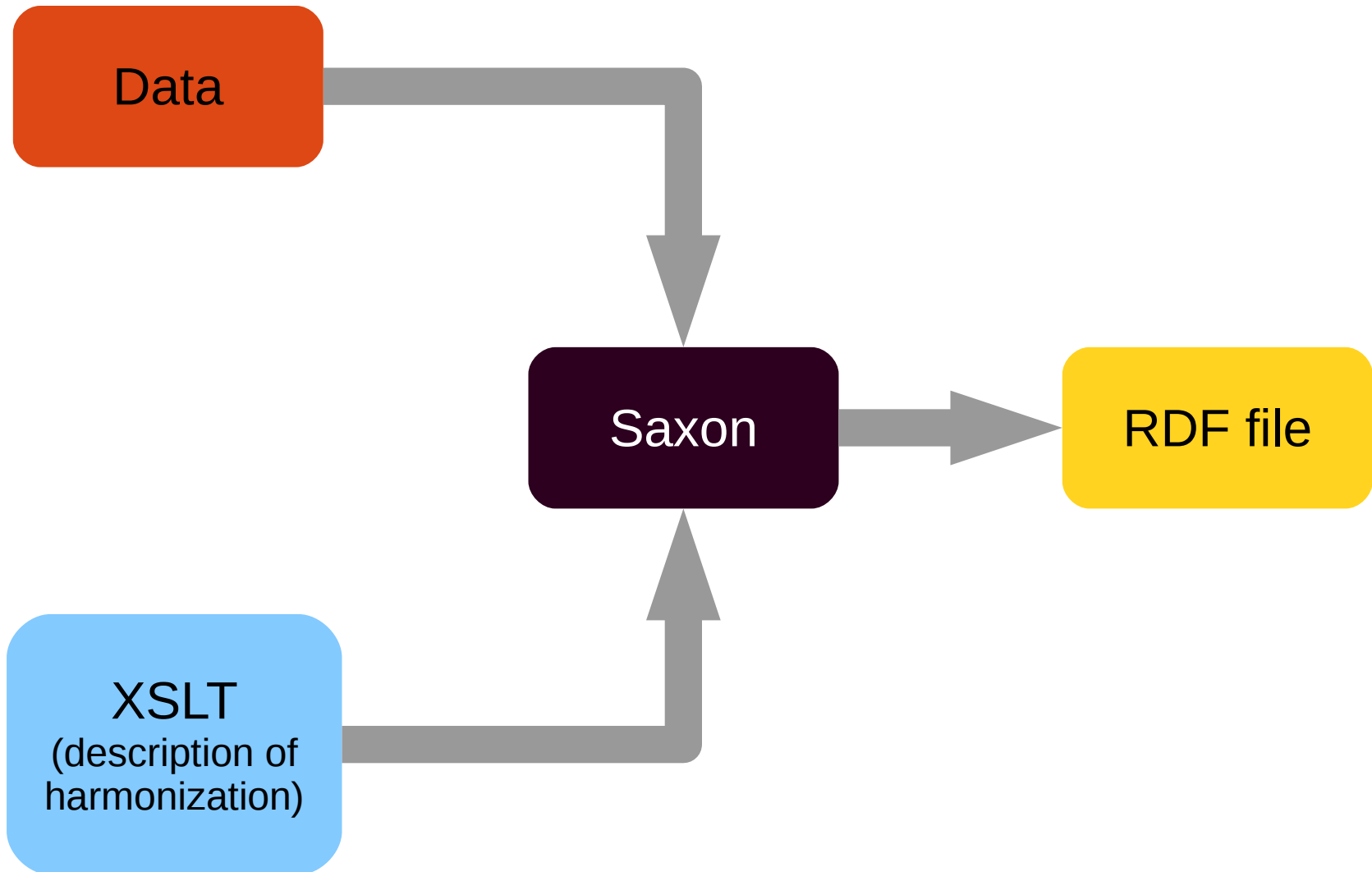
- Transcription to structured data (table → XML)
- Transformation to common data model
- Preparation of common vocabularies (Waze) and mappings (OSM, GeoNames.org → Waze, countries → Linked data resources)
- Re-classification (Waze categories; OSM)
- Filtering
- Adding information (links to countries)
- 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



# Data harmonization - scheme



# RDF example

```
<rdf:Description rdf:about="http://www.sdi4apps.eu/poi/IT\_CAE\_15.023497\_37.613992">  
<rdfs:label xml:lang="it">Casa Museo della civiltà contadina</rdfs:label>  
<rdfs:label xml:lang="en">Museum of civilization</rdfs:label>  
<geos:asWKT POINT(15.023497 37.613992)</geos:asWKT>  
<poi:categoryWaze rdf:resource="http://www.openvoc.eu/waze\_classification#Culture\_&\_entertainment"/>  
<poi:category>tourism</poi:category>  
<poi:categoryOSM>tourism.museum</poi:categoryOSM>  
<poi:address>Via Garibaldi, 58/58a – 95030 Nicolosi (CT)</poi:address>  
<foaf:phone>(+39) 095 – 910 980</foaf:phone>  
<geos:sfWithin rdf:resource="http://www.geonames.org/3175395"/>  
<geos:sfWithin rdf:resource="http://dbpedia.org/resource/Italy"/>  
</rdf:Description>
```

**4 206 573** POIs

**133 157**  
new POIs  
during Open  
Data  
Hackathon in  
Dresden, 2015

**Sept.  
2015**

# Coverage



# Comparison

Area	SPOI	<u>OpenPOIs</u>
Seaside resort (Croatia)	7	4
<u>Submontane area (Czech republic)</u>	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



July  
2015

# 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.

### What is SPOI data set?

---

The SDI4Apps Points of Interest data set is the **seamless and open resource of POIs** that is **available for other users to download, search or reuse** in applications and services.

Its principal target is to provide information for cycling as Linked data together with other data set containing road network.

The added value of the SDI4Apps approach in comparison to other similar solutions consists in **implementation of linked data**, using of standardized and respected datatype properties and development of the **completely harmonized data set** with uniform data model and common classification.

# Next steps

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

# Innovations

- Common data model
- Re-using existing standards
- Linked data
- SPARQL endpoint  
(<http://ha.isaf2014.info:8890/sparql>)