0.1 Semantic GIS

0.1.1 GIS Background

GIS Background

- Current GIS systems work mostly with 2-dimensional coordinates, but can be 2.5-D, or fully 3-D.
- Traditionally, the most widely used storages are relational databases (Oracle Spatial, ESRI ArcSDE, MS SQL, PostgreSQL PostGIS).
  - Upon which complex server and desktop systems operate (ESRI ArcGIS, Integraph Geomedia), performing processing, analyses, and map outputs, and providing services.
- The data stored there are usually not publicly available (only the map outputs), and have custom structure.
- Because of the different structures and different technologies, it is not easy to integrate and search the data.

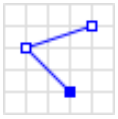
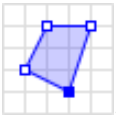
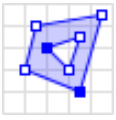
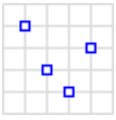
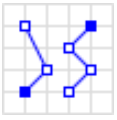
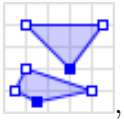
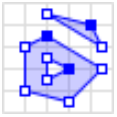
Catalogues

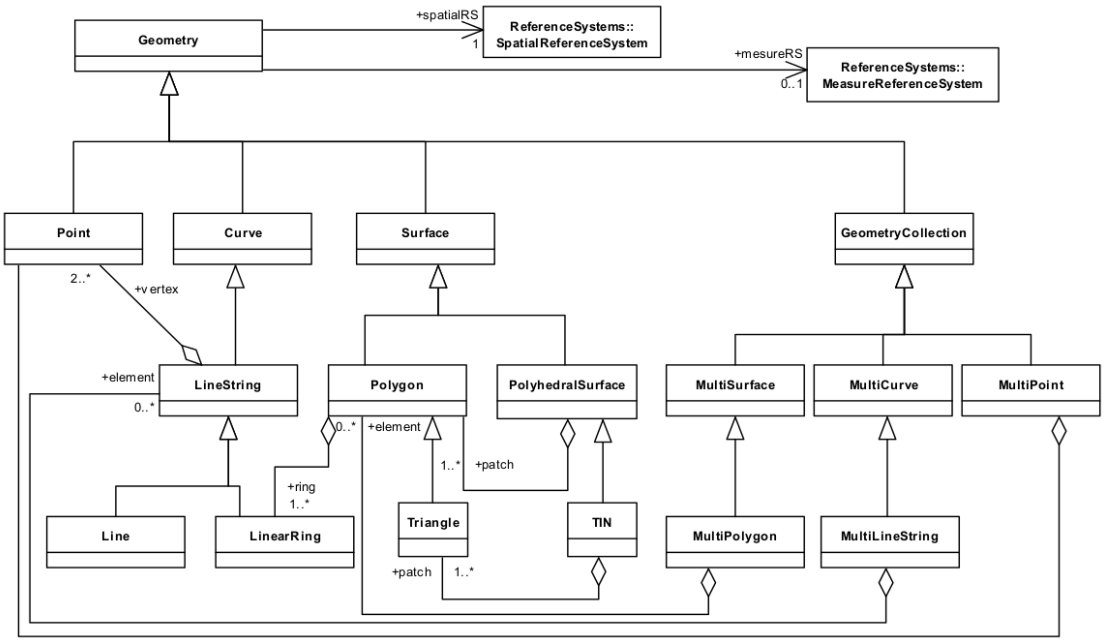
- Catalogue Service (CSW), standard by OGC (Open Geospatial Consortium, `http://www.opengeospatial.org/`).
- Interface to discover, browse, and query **metadata** about GIS data and services.
- Uses Dublin Core (small vocabulary to describe web resources, see lecture 9 – Linked Data).
- Searching by metadata – keywords, author, date, ...
- E.g. at ČÚZK (Czech Office for Surveying, Mapping and Cadastre): `http://geoportal.cuzk.cz/SDIProCSW/service.svc/get?REQUEST=GetCapabilities&SERVICE=CSW`.

INSPIRE

- Infrastructure for Spatial Information in the European Community.
- European directive, in the Czech Republic is the responsible authority CENIA (Česká informační agentura životního prostředí).
- To make geographical information more accessible and interoperable.
- Topic categories (19), spatial data service types (6), classification of spatial data services (70).
- Keywords from GEMET (general environmental multilingual thesaurus, see lecture 5 – KOS).

Basic GIS objects

- a point
- a linestring 
- a polygon  , 
- a multi-point 
- a multi-linestring 
- a multi-polygon  , 



Images from http://en.wikipedia.org/wiki/Well-known_text and http://www.opengeospatial.org/standards/sfa, 2014

Spatial serializations

Many serializations exist

- WKT, WKB, GML, KML
- GeoJSON, GeoJSON-LD, GeoRSS, GPX, ...

We will focus on:

WKT

```
POINT (30 10)
LINESTRING (30 10, 10 30, 40 40)
POLYGON ((30 10, 40 40, 20 40, 10 20, 30 10))
```

GML

```
<gml:LineString gml:id="p21"
   srsName="http://www.opengis.net/def/crs/EPSG/0/4326">
  <gml:coordinates>45.67,88.56 55.56,89.44</gml:coordinates>
</gml:LineString>
```

Spatial Reference Systems

- Defines how to describe point locations on Earth by 2-D coordinates.
- Complete list available at http://spatialreference.org/.
- SRID identifier.

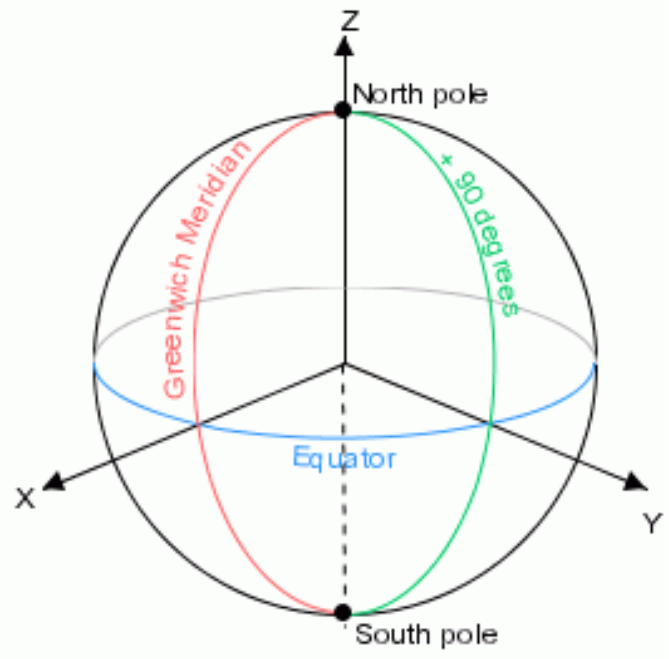


Image from `http://pic.dhe.ibm.com/`, 2014

Some SRSs

- WGS 84 (EPSG:4326), ex. `14.41982, 50.07591`, or `50°04′33.3″N 14°25′11.3″E`
- Google Mercator projection (EPSG:900913), ex. `1605205.41178, 6459433.66549`
- in CZ: Křovákovo zobrazení (S-JTSK) (ESRI:102027), ex. `-743087.47789, -1044271.17375`

0.1.2 Vocabularies

Basic Geo Vocabulary

- Standardized by W3C, widely used.
- Available at `http://www.w3.org/2003/01/geo/`.
- Namespace (typically prefixed `geo:`) `http://www.w3.org/2003/01/geo/wgs84_pos#`.
- Uses only WGS 84 SRS.
- Very simple, for points only.
- Class `SpatialThing`, and its subclass `Point`.
- Properties `lat` (latitude), `long` (longitude), `alt` (altitude); (domains: `SpatialThing`).

- Property `location` (range: `SpatialThing`, subproperty of `http://xmlns.com/foaf/0.1/based_ne`

```
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
     xmlns:geo="http://www.w3.org/2003/01/geo/wgs84_pos#">
  <geo:Point>
    <geo:lat>50.07591</geo:lat>
    <geo:long>14.41982</geo:long>
  </geo:Point>
</rdf:RDF>
```

RCC8 topology relations

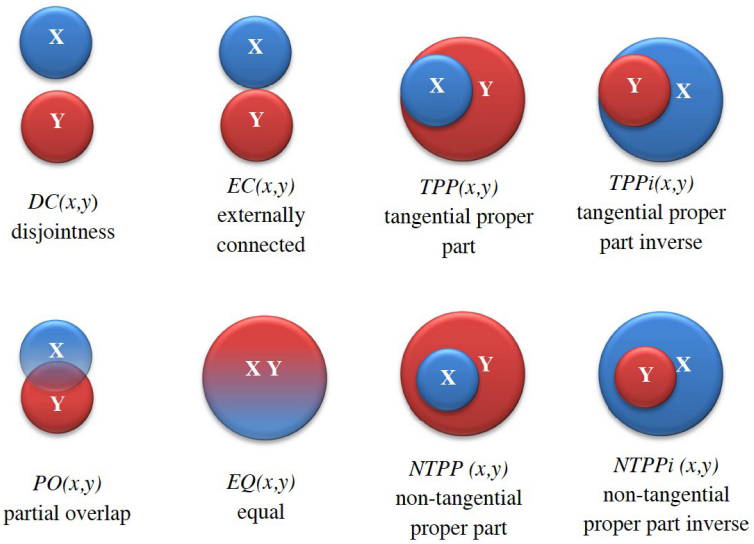


Image from `http://www.strabon.di.uoa.gr/terracognita/papers/moeller-terra-cognita.pdf`, Ralf Moeller, 2014

GeoVocab (NeoGeo) Vocabulary

- Available at `http://geovocab.org/`.
- Namespaces `http://geovocab.org/geometry#` (usually prefixed `geom:`), and `http://geovocab.org/spatial#` (usually prefixed `spatial:`).
- Classes `spatial:Feature` and `geom:Geometry`, property connecting them `geom:geometry`.
- Spatial topology relations, as defined in RCC8: `spatial:C`, `spatial:DC`, `spatial:EQ`, ...
- No stable support for representing geometries.
- Not widespread.
- Result of a community effort.

GeoVocab spatial topology relations

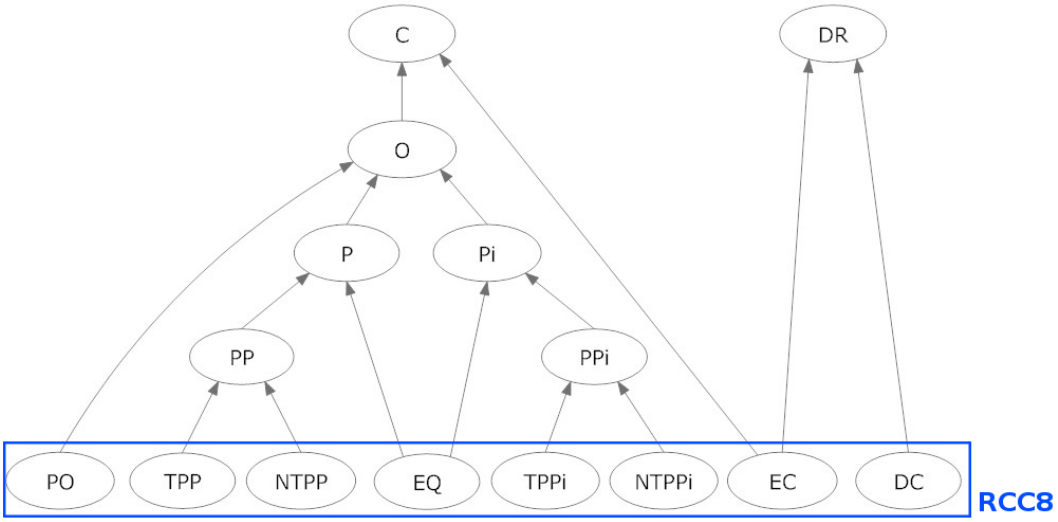


Image from `http://geovocab.org/doc/neogeo/`, 2014

0.1.3 Existing Data

GeoNames

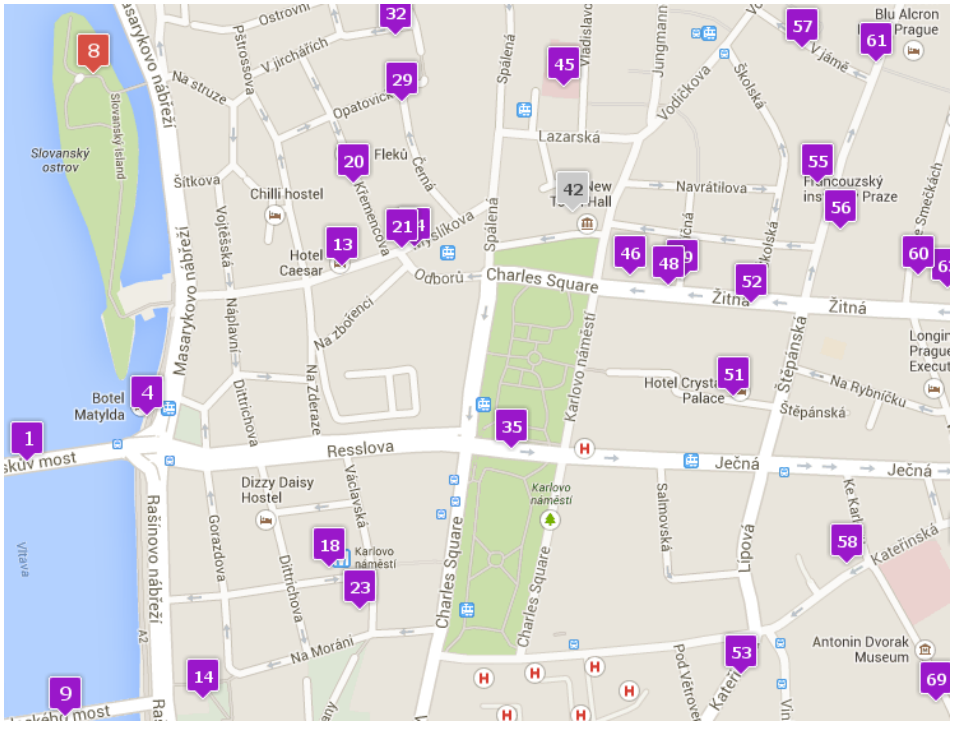
GeoNames

- Important points all over the world with precise categorization.
- Not many (today cca 8.5 M), but high quality.
- Available at `http://www.geonames.org/`.
- Prefix `http://www.geonames.org/ontology#`.
- Localized to *many* languages.
- Categorization into 9 feature classes and 645 feature codes.
- Uses only Basic Geo Vocabulary.
- Dereferencable, many REST webservices, download in CSV.
- Possibility to pay for checked releases.

GeoNames Feature Classes

1. A – country, state, region, ...
2. H – stream, lake, ...
3. L – parks,area, ...
4. P – city, village, ...
5. R – road, railroad
6. S – spot, building, farm
7. T – mountain,hill,rock, ...
8. U – undersea
9. V – forest,heath, ...

GeoNames map of Karlovo náměstí (over Google Maps)



GeoNames example of London

```
<http://sws.geonames.org/2643743/> a gn:Feature ;
   gn:name "Londres" ;
   gn:alternateName "Londen"@af,
      "London"@als,
      "Londres"@an,
      "Lunden"@ang ; [...  total 161 languages]
   gn:countryCode "GB" ;
   gn:featureClass gn:P ;
   gn:featureCode gn:P.PPLC ;
   gn:locationMap <http://www.geonames.org/2643743/london.html> ;
   gn:nearbyFeatures <http://sws.geonames.org/2643743/nearby.rdf> ;
   gn:parentCountry <http://sws.geonames.org/2635167/> ;
   gn:population "7556900" ;
   gn:wikipediaArticle <http://en.wikipedia.org/wiki/London> ;
   rdfs:isDefinedBy <http://sws.geonames.org/2643743/about.rdf> ;
   rdfs:seeAlso <http://dbpedia.org/resource/London> ;
   geo:lat "51.50853" ;
   geo:long "-0.12574" .
```

LinkedGeoData

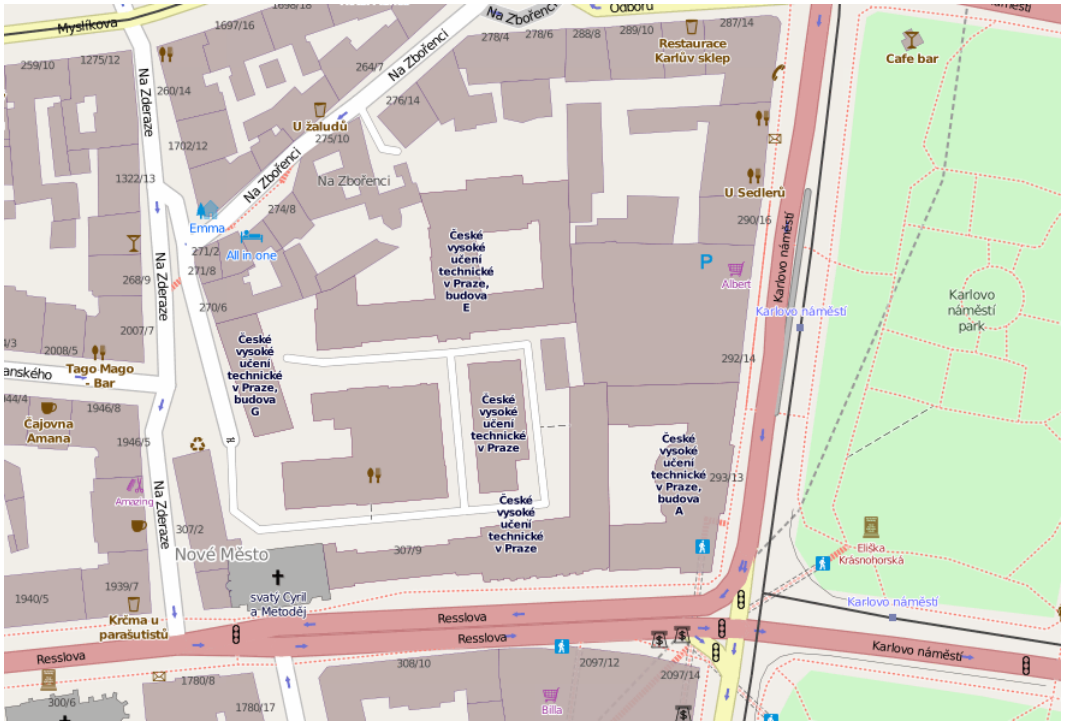
LinkedGeoData

- Publishing data from OpenStreetMap (OSM) as RDF.
- Available at `http://linkedgeodata.org/`.
- 20 billion triples.
- Dereferencable, REST webservice, SPARQL endpoint, download.
- In WGS 84 SRS, geometries serialized in WKT.
- Classes and properties defined by OSM tags.
- Uses partially GeoVocab and partially GeoSPARQL vocabularies.

OpenStreetMap

- Publicly available (as vectors!) geographic data set of the World.
- Available at `http://www.openstreetmap.org/`.
- 2348 M nodes, 233 M ways, 2.5 M relations – very detailed.
- Geometries annotated with tags (key-value pairs).
- Many errors, anybody can add own data, categories and attributes (in form of tags) not enforced.

LinkedGeoData (OpenStreetMap map) of Karlovo náměstí

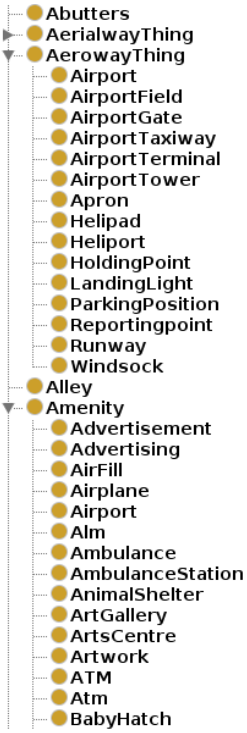


LinkedGeoData example of CTU, FEE building E

```
@prefix lgd: <http://linkedgeodata.org/triplify/> .
@prefix lgdo: <http://linkedgeodata.org/ontology/> .
@prefix lgdm: <http://linkedgeodata.org/meta/> .
@prefix lgd-geom: <http://linkedgeodata.org/geometry/> .
@prefix spatial: <http://geovocab.org/spatial#> .
@prefix geom: <http://geovocab.org/geometry#> .
@prefix ogc: <http://www.opengis.net/ont/geosparql#> .
lgd:way27753990
   a lgdm:Way, spatial:Feature, lgdo:University, lgdo:Amenity ;
   rdfs:label "Tschechische␣Technische␣Universitat␣Prag,␣Gebaude␣E"@de,
      "Ceske␣vysoke␣uceni␣technicke␣v␣Praze,␣budova␣E" ,
      "Czech␣Technical␣University␣in␣Prague,␣Building␣E"@en ;
   geom:geometry lgd-geom:way27753990 ;
   lgdo:building "true"^^xsd:boolean ;
   <http://linkedgeodata.org/ontology/id%3A%C4%8Dvut> "KN:E" ;
   lgdo:source "cuzk:km" ; lgdo:version "6"^^xsd:int ;
   dcterms:contributor lgdo:user152074 ;
   dcterms:modified "2013-03-29T17:29:31"^^xsd:dateTime .

lgd-geom:way27753990
   a geom:Geometry ;
   lgdo:posSeq lgd-geom:posSeq27753990 ;
   ogc:asWKT "LINESTRING(14.4172133␣50.0766042,␣...)"^^ogc:wktLiteral .
```

LinkedGeoData Classes



DBpedia

DBpedia

- Extract of structured information on Wikipedia.
- Available at `http://dbpedia.org/`.
- Lot of information, some mistakes.
- Dereferencable, SPARQL endpoint, download.
- Contains only point data, mostly for populated places, but large amount of non-spatial data.
- Always good idea to link to this dataset, if possible – many other link to it.

DBpedia example of Cambridge

```
@prefix dbpedia: <http://dbpedia.org/resource/> .
@prefix dbpedia-owl: <http://dbpedia.org/ontology/> .

dbpedia:Cambridge a dbpedia-owl:Settlement ,
      dbpedia-owl:Place ,
      dbpedia-owl:City ,
      dbpedia-owl:PopulatedPlace ;
   rdfs:label "Cambridge"@fr ,
      "Cambridge"@en ,
```

```
    "\u041A\u0435\u043C\u0431\u0440\u0438\u0434\u0436"@ru ;
  geo:geometry "POINT(0.119000␣52.205002)"^^virtrdf:Geometry ;
  geo:long "0.119"^^xsd:float ;
  geo:lat "52.205"^^xsd:float ;
```

GeoWordNet

- GeoWordNet is an extended thesaurus/semantic network based on WordNet (see lecture 5 – KOS) and GeoNames.

- Available at `http://geowordnet.semanticmatching.org/` in various formats incl. RDF.

- Hyponyms, hypernyms, holonyms, ...

dependent political entity

Gloss

a kind of political entity

Words

1. dependent political entity

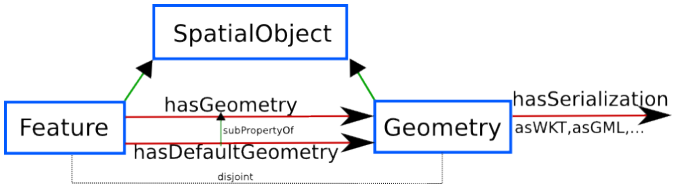
Relations

1. *hypernym*: political unit
2. *instances hyponym*: Nouvelle-Caledonie
3. *instances hyponym*: Bouvet Island
4. *instances hyponym*: Territory of Christmas Island
5. *instances hyponym*: Saint-Pierre et Miquelon
6. *instances hyponym*: Saint Helena
7. *instances hyponym*: Tristan da Cunha
8. *instances hyponym*: South Georgia and the South Sandwich Islands
9. *instances hyponym*: British Virgin Islands
10. *instances hyponym*: Montserrat
11. *instances hyponym*: Cayman Islands
12. *instances hyponym*: Gibraltar
13. *instances hyponym*: Territory of Heard Island and McDonald Islands
14. *instances hyponym*: Territory of Norfolk Island
15. *instances hyponym*: British Indian Ocean Territory

0.1.4 GeoSPARQL

GeoSPARQL

- Standard by OGC [4].
- Extension of SPARQL query language with geospatial vocabulary.
- Fairly recent, standardized on 10.9.2012.
- Available at `http://www.opengeospatial.org/standards/geosparql`.
- Class `SpatialObject`, its two disjoint sub-classes: `Feature` and `Geometry`.
- A `Feature` can be connected to its `Geometry` by the property `hasGeometry` and `hasDefaultGeometry`.



GeoSPARQL data properties

- `hasSerialization` – base property for linking a `Geometry` to its serialization. It has two standard subproperties (can be extended):
  - `asWKT` for WKT, range: custom datatype `wktLiteral`,
  - `asGML` for GML, range: custom datatype `gmlLiteral`.
- Other auxiliary properties:
  - `dimension` (topological dimension),
  - `coordinateDimension` (dimension of direct positions),
  - `spatialDimension` (dimension of the spatial portion of the direct positions),
  - `isEmpty` (no points),
  - `isSimple` (contains no self-intesections except of its boundary).

GeoSPARQL topology relations

- Defined or inferred topology relations of spatial objects in a plane.
- For a plane, it is possible to divide relation of any two objects into 8 groups.
- The relations are divided in three standard ways – families of topology relations:
  - *Simple Feature* relations: `sfEquals`, `sfDisjoint`, `sfIntersects`, `sfTouches`, `sfWithin`, `sfContains`, `sfOverlaps`, `sfCrosses`.

  - *Egenhofer* relations: `ehEquals`, `ehDisjoint`, `ehMeet`, `ehOverlap`, `ehCovers`, `ehCoveredBy`, `ehInside`, `ehContains`.
  - *RCC8* relations: `rcc8eq`, `rcc8dc`, `rcc8ec`, `rcc8po`, `rcc8tppi`, `rcc8tpp`, `rcc8ntpp`, `rcc8ntppi`.
- All relations modeled as properties with domain and range `SpatialObject`.
- Precise definitions of topological relations by the DE-9IM model – uses 3×3 matrices; for details see [2].

GeoSPARQL functions

Functions for more complex operations, used in SPARQL FILTER section.

- `distance` (gets distance of two geometry literals measured in given units),
- `buffer` (get geometry literal as an input literal with a buffer added, given the radius and units of the buffer),
- `convexHull` (get convex hull of a geometry literal),
- `intersection` (get intersection of two geometry literals),
- `union` (get union of two geometry literals),
- `difference` (get difference of two geometry literals),
- `symDifference` (get set symmetric difference of two geometry literals),
- `envelope` (get bounding box of a geometry literal),
- `boundary` (get boundary of a geometry literal),
- `getsrid` (get SRS URI of a geometry literal).

GeoSPARQL notes

- Has also OWL ontology defined with its vocabulary and definitions, by OGC at `http://schemas.opengis.net/geosparql/`. The ontology also includes ontologies for Simple Feature and GML geometries.
- RIF rules.
- WKT serialization can contain SRS as:

```
"<http://www.opengis.net/def/crs/OGC/1.3/CRS84>POINT(-122.4192 37.7793)"^^ogc:
    wktLiteral
```

- All topology relation properties have also version as a boolean function for use in filters.

- Prefixes `http://www.opengis.net/ont/geosparql#`, `http://www.opengis.net/def/function/geosparql/`, `http://www.opengis.net/def/rule/geosparql/`.

GeoSPARQL example

```
PREFIX geo: <http://www.opengis.net/ont/geosparql#>
SELECT ?m ?p WHERE {
   ?m a ex:Monument ;
      geo:hasGeometry ?mgeo .
   ?p a ex:Park ;
      geo:hasGeometry ?pgeo .
   ?mgeo geo:sfWithin ?pgeo .
}
```

```
PREFIX geo: <http://www.opengis.net/ont/geosparql#>
PREFIX geof: <http://www.opengis.net/def/function/geosparql/>
PREFIX units: <http://www.opengis.net/def/uom/OGC/1.0/>
SELECT ?p WHERE {
   ?p a ex:Park ;
      geo:hasGeometry ?pgeo .
   ?pgeo geo:asWKT ?pwkt .
   ex:WashingtonMonument geo:hasGeometry ?wgeo .
   ?wgeo geo:asWKT ?wwkt .
   FILTER(geof:distance(?pwkt, ?wwkt, units:m) < 3000) .
}
```

0.1.5 Support in Existing Systems

Parliament

Parliament

- Embedded triple store developed by BBN Technologies, open source, rule-based inference [1] [3].
- As a layer for Jena/Sesame.
- Available at `http://parliament.semwebcentral.org/`.
- Supports SPARQL and partially GeoSPARQL.
- Has spatial and temporal indexing and reasoning.
- Best support for GeoSPARQL found so far, but not widely used.
- Difficult to set up in existing environment.
- Demo during today's tutorial.

OWLIM-SE

OWLIM-SE

- Commercial triple store by Ontotext at `http://www.ontotext.com/owlim`.
- Again, coupled with Jena and Sesame.
- Semantics of RDFS, OWL 2 RL and OWL 2 QL.
- Designed for scalability, loading and query evaluation performance.
- Good support, often used.
- Supports only Basic Geo Vocabulary for spatial data, which are indexed.
- Has custom querying functions:
  - `within` – finds points within a rectangle or a polygon,
  - `nearby` – finds points within specified distance to a point,
  - `distance` – calculates distance between two points.

Virtuoso

Virtuoso

- Virtuoso Universal Server by OpenLink is multi-model data server.
- Available at `http://virtuoso.openlinksw.com/`.
- Free, commercial, and open source versions.
- Contains:
  - relational, RDF and XML data management,
  - free text content management and full text indexing,
  - document web server,
  - Linked Data server,
  - web application server,
  - web services (SOAP/REST).
- Used as SPARQL endpoint for e.g. DBpedia and LinkedGeoData.
- Uses Basic Geo Vocabulary, but converts it to custom literals, and indexes them with R-tree.
- For querying, SQL MM built-in spatial functions are used – `st_intersects`, `st_contains`, `st_within`, `st_distance`, ...

References

[1] Robert Battle and Dave Kolas. "Enabling the geospatial Semantic Web with Parliament and GeoSPARQL". In: *Semantic Web Journal, to appear* (). ISSN: 1570-0844. DOI: 10.3233/SW-2012-0065.

[2] Eliseo Clementini, Paolino Felice, and Peter Oosterom. "A small set of formal topological relationships suitable for end-user interaction". In: *Advances in Spatial Databases*. Ed. by David Abel and Beng Chin Ooi. Vol. 692. Lecture Notes in Computer Science. Springer Berlin Heidelberg, 1993, pp. 277–295. ISBN: 978-3-540-56869-8. DOI: 10.1007/3-540-56869-7_16. URL: http://dx.doi.org/10.1007/3-540-56869-7_16.

[3] Ian Emmons Mike Dean Dave Kolas. "Efficient Linked-List RDF Indexing in Parliament". In: *The Semantic Web - ISWC 2009*. 2009.

[4] Open Geospatial Consortium. *OGC GeoSPARQL — A Geographic Query Language for RDF Data*. http://www.opengeospatial.org/standards/geosparql, 2012.