

Inspiring data?

Cross-domain interoperability for EU spatial data

Andrea Perego

“Using Open Data” Workshop, 19-20 June 2012, Brussels



www.jrc.ec.europa.eu

*Serving society
Stimulating innovation
Supporting legislation*



The Joint Research Centre (JRC)

JRC is the European Commission's in-house science service. It provides the science for policy decisions, with a view to ensuring that the EU achieves its Europe 2020 goals for a productive economy as well as a safe, secure and sustainable future.

The JRC plays a key role in the European Research Area and reinforces its multi-disciplinarity by networking extensively with leading scientific organisations in the Member States, Associated Countries and worldwide.

The *Digital Earth and Reference Data Unit* is part of the Institute for Environment and Sustainability, and it is the largest group in Europe devoted to research on Spatial Data Infrastructures and to the implementation in support of relevant EU policies (e.g., the INSPIRE EC Directive)

Outline

- Interoperability in INSPIRE
- INSPIRE, eGov and Linked Open Data
- Open issues
- The INSPIRE LOD prototype
- Conclusions and future work

The INSPIRE Directive

- General rules to establish an infrastructure for spatial information in Europe, as a support to
 - Community environmental policies
 - Policies or activities which impact on the environment
- To be based on spatial data infrastructures (SDIs) established and operated by EU Member States
 - INSPIRE is a distributed infrastructure
- Does not require collection of new spatial data
- Does not affect existing Intellectual Property Rights (IPRs)
- Entry into force: 15 May 2007



Some figures

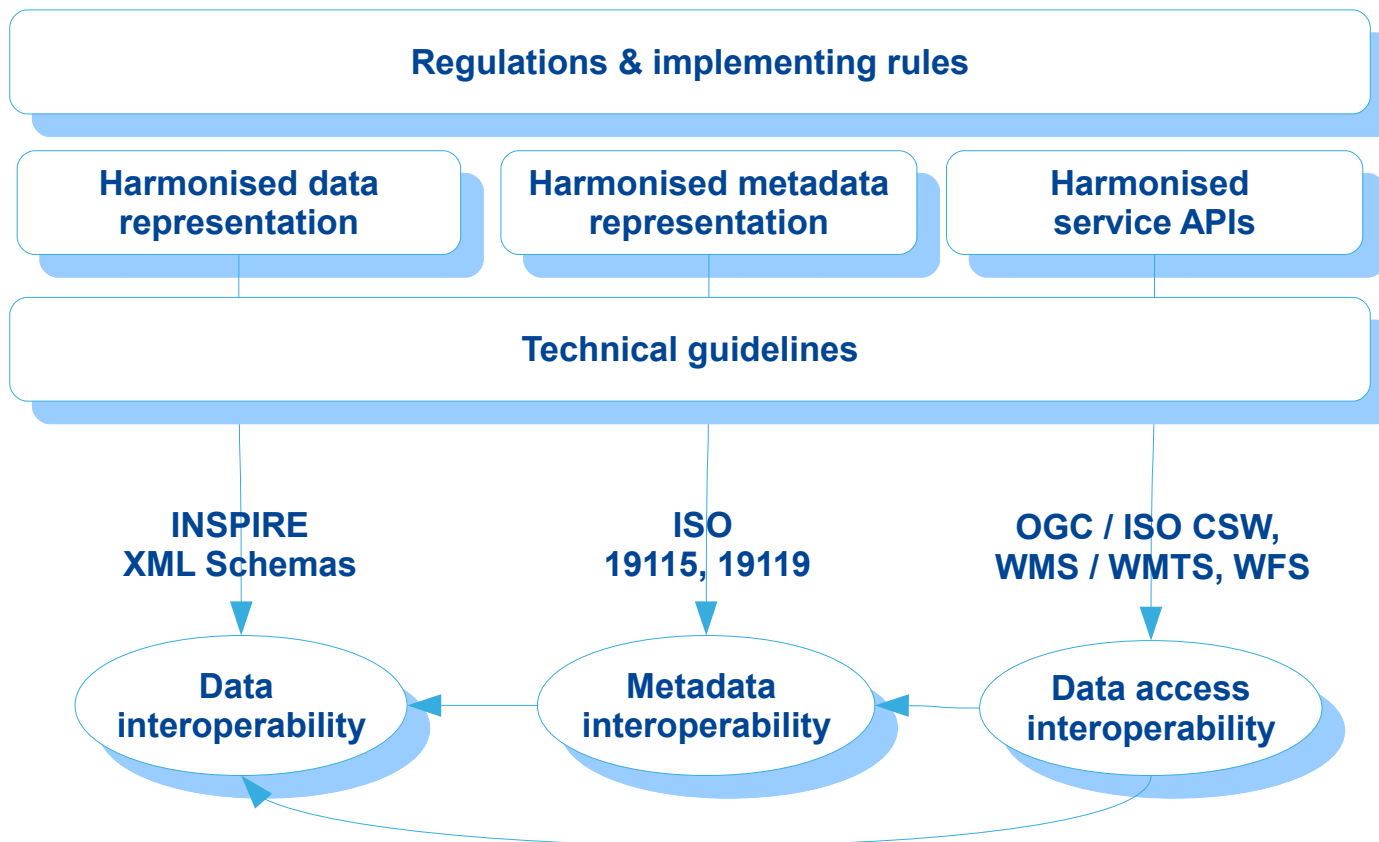
November 2011:

- EU Member States expose INSPIRE-compliant discovery and view services
- A pilot version of the INSPIRE Geoportal is made publicly available. This is meant to provide a single access point to all data available through the INSPIRE infrastructure

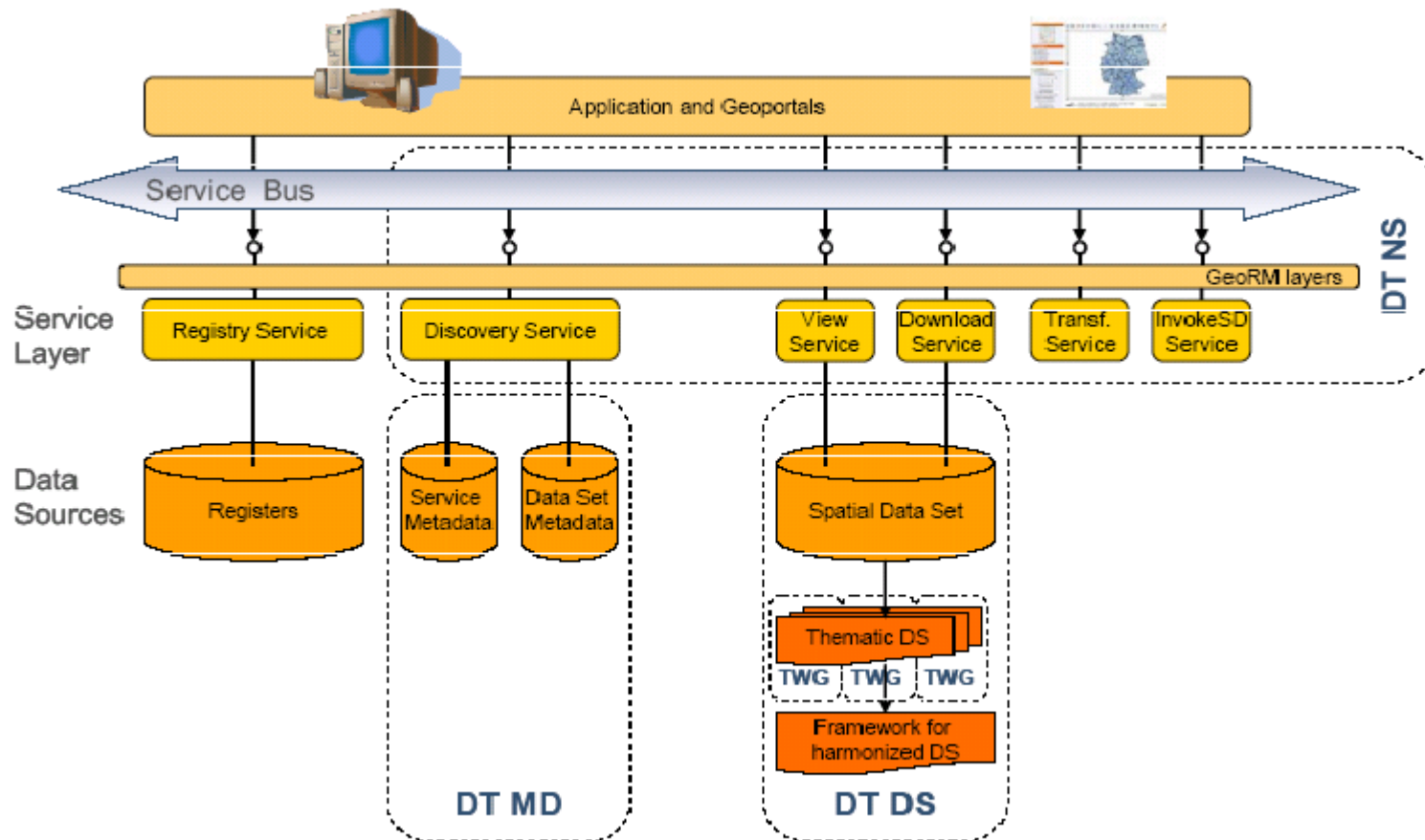
Current situation:

- ~200,000 geospatial resources from 15 Member States
- We are still in the testing phase

Interoperability in INSPIRE




The INSPIRE architecture



Data and INSPIRE thematic scope (not only geospatial...)

Annex I

- 
1. Coordinate reference systems
 2. Geographical grid systems
 3. Geographical names
 4. Administrative units
 5. Addresses
 6. Cadastral parcels
 7. Transport networks
 8. Hydrography
 9. Protected sites

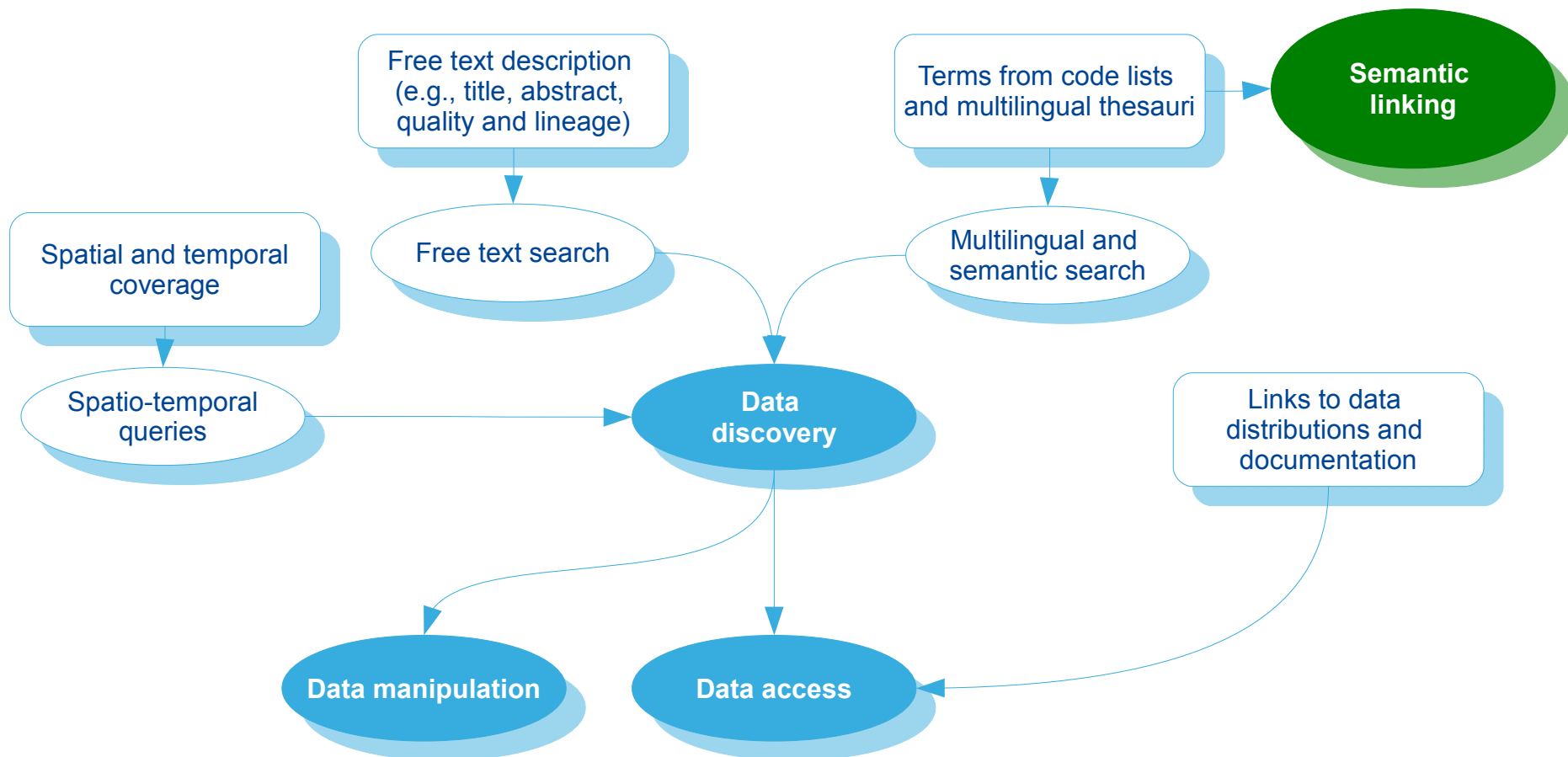
Annex II

1. Elevation
2. Land cover
3. Ortho-imagery
4. Geology

Annex III

1. Statistical units
2. Buildings
3. Soil
4. Land use
5. Human health and safety
6. Utility and governmental services
7. Environmental monitoring facilities
8. Production and industrial facilities
9. Agricultural and aquacultural facilities
10. Population distribution – demography
11. Area management / restriction / regulation zones & reporting units
12. Natural risk zones
13. Atmospheric conditions
14. Meteorological geographical features
15. Sea regions
16. Bio-geographical regions
17. Habitats and biotopes
18. Species distribution
19. Energy resources
20. Mineral resources

Resource metadata in INSPIRE



INSPIRE in the eGov framework

- Geospatial data are a relevant part of eGov data
- The integration of INSPIRE in the broader eGov framework would allow **cross-domain search and data aggregation** – and this would also be a benefit for policy making
- Such integration should be based on a **domain-independent data model**, and might lead to cross-domain interoperability
- RDF and Linked Open Data (LOD) satisfy such requirements, and in addition provide enhanced data access by linking resources
- But what is needed to make INSPIRE LOD-compliant?

INSPIRE and Linked Open Data

- ★ Available on the Web, with an open data licence ✓
- ★★ Available as machine-readable structured data ✓
- ★★★ Use non-proprietary formats ✓
- ★★★★ Use URIs to identify things ***Partially***
- ★★★★★ Linked your data to other data to provide context ***Partially***

From Open Data to Linked Open Data

Technically feasible

- Defining RDF mappings for INSPIRE meta/data specifications
- Building an abstraction layer on top of the INSPIRE infrastructure, thus making INSPIRE resources accessible and discoverable according to the LOD paradigm

A number of **issues** should be however addressed, **concerning cross-domain interoperability**, and which are **not specific to geospatial data only**

- Agreement on a **minimal common terminology** – *core vocabularies*
- Enforcement of **semantic and multilingual search and linking**

eGov Core Vocabularies

- They define a minimal, but extensible, set of recommended terms to be used to represent eGov resources
- Based as much as possible on widely used vocabularies (e.g., Dublin Core, the Data Catalog vocabulary @ W3C)

It is however not always possible re-using existing vocabularies, especially when there is no agreement on how to represent a given type of information

- E.g., often **eGov data** (and not only geospatial ones) **describe a given geographical region in a given time frame**
- There is currently **no general agreement on how to represent spatial / temporal coordinates in RDF**

Why spatial / temporal queries?

They would allow discovery and aggregation of cross-domain data based on the described geographical region, in a given time frame

Spatial coordinates in RDF:

- GeoNames / DBPedia.org are not enough flexible: we should be able to specify **arbitrary geometries**
- A number of approaches are available: W3C Geo vocabulary, the **geo** URI scheme, GeoHash.org (points only), schema.org vocabulary(also other shapes), GeoSPARQL

Temporal coordinates: the W3C Time ontology, Dublin Core (dcterms:PeriodOfTime), schema.org

A solution proposed for both (in 2003!): URI spaces for time and places (see the pioneering work by Ian Davis at PlaceTime.com)

Cross-domain interoperability @ EC

Some of the initiatives promoted in the framework of the EC ISA programme:

- Asset Description Metadata Schema (ADMS)
- eGovernment Core Vocabularies

Two new initiatives, based on INSPIRE:

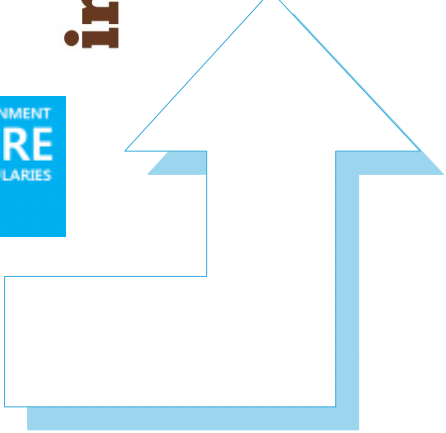
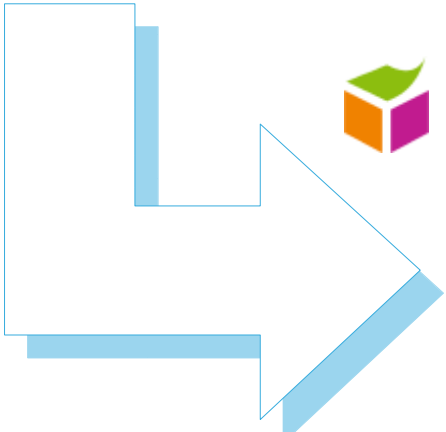
- European Union Location Framework (EULF)
- Reusable INSPIRE Reference Platform



GeoSPARQL



DCAT



VoID

Semantic and multilingual issues in eGov data

- No common terminology, even inside the same field
- EU data are available in multiple languages

A possible solution is to *use controlled vocabularies in both the annotation and discovery phase*

Preferably,

- Multilingual / language-neutral vocabularies
- Vocabularies denoting semantic relationships among the defined terms, and
- among terms defined in different vocabularies

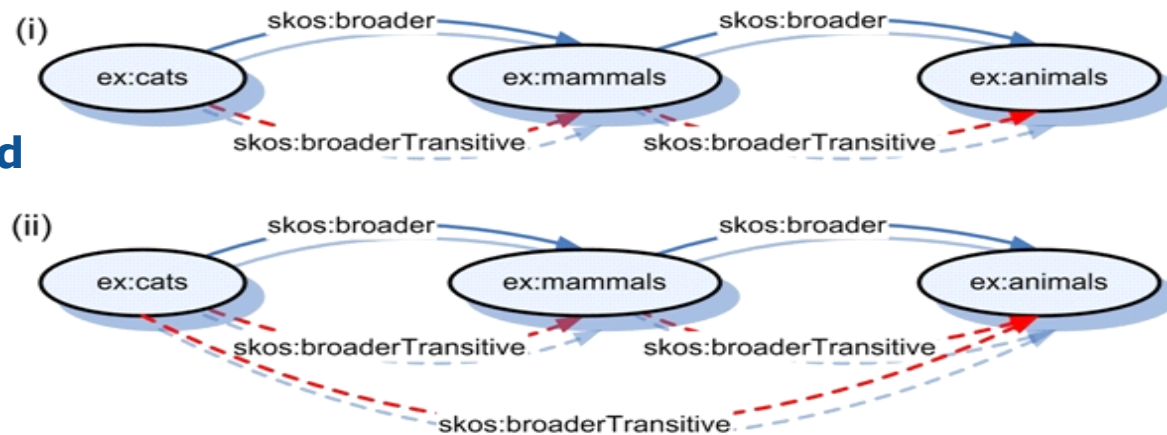
Controlled vocabularies in the SKOS format, aka thesauri, support such features

SKOS: Simple Knowledge Organization System

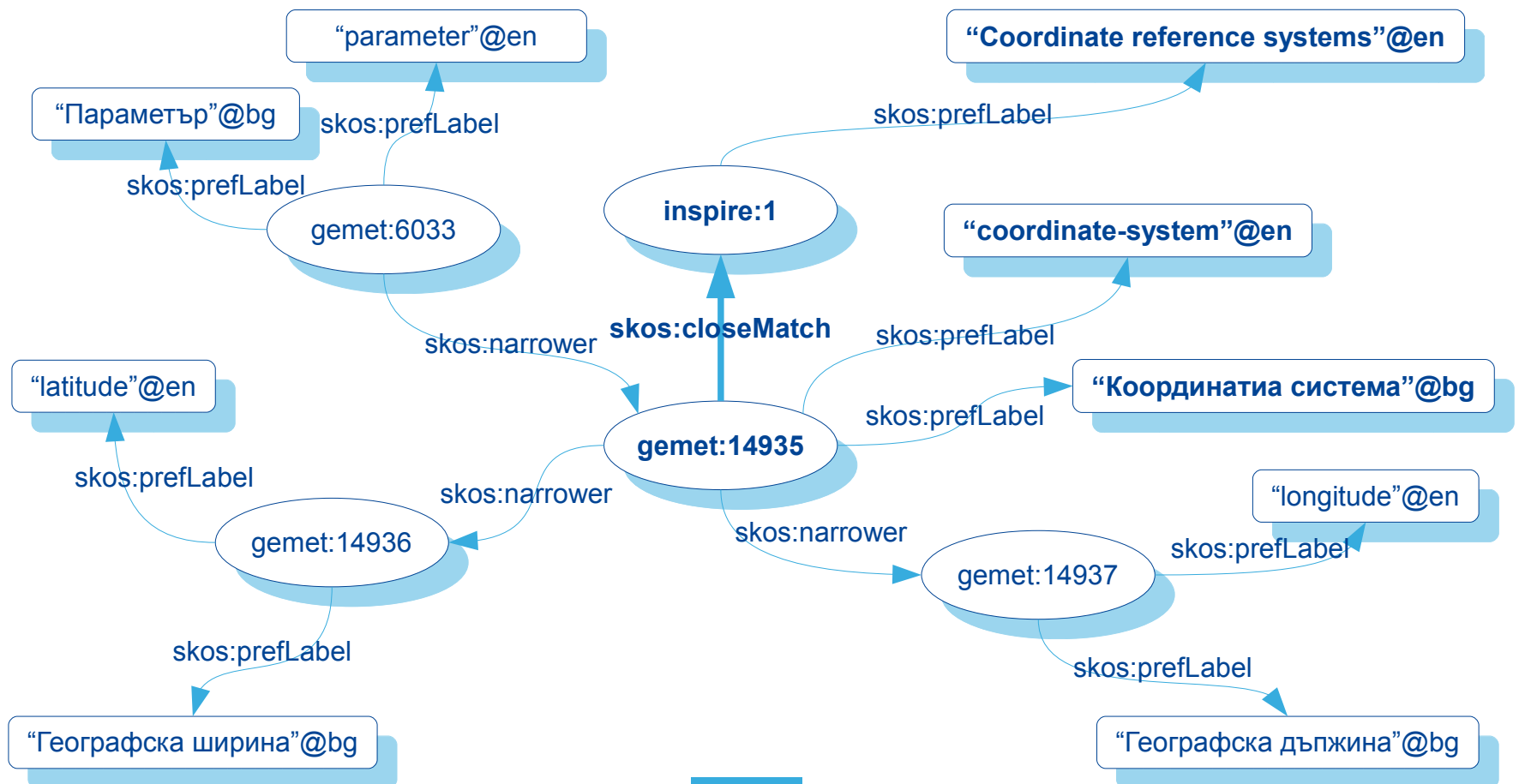
An expressive OWL ontology:

- negation of classes, transitive properties, property hierarchies, inverse and functional properties, datatypes
- Allows for the modelling of knowledge organization systems such as thesauri, vocabularies, classification schemes, etc.
- Semantic relations link terms that are **broader**, **narrower**, or just **related**

NB: *Relations may link terms from distinct thesauri*



SKOS: mapping terms from different thesauri



LOD-enabled INSPIRE prototype

- Focussing on INSPIRE metadata
 - This is what is already available in this phase – exposing INSPIRE-compliant data is not yet required
- Based on the INSPIRE Geoportal: <http://inspire-geoportal.ec.europa.eu/>
 - It collects geospatial metadata from all the EU Member States exposing INSPIRE compliant services (~200,000 resources, currently)
 - Such **metadata are served in different formats** (currently, XML, JSON, and HTML) by using HTTP content negotiation
 - It computes and stores additional information, not contained in the collected metadata, which can be used to **link resources** (e.g., a dataset and the layer(s) portraying it)

Roadmap

- The very first step (**done**):
 - Defining RDF mappings for INSPIRE metadata elements
- The second step:
 - Making RDF-encoded INSPIRE metadata available from the INSPIRE Geoportal
- The third step:
 - Making INSPIRE metadata queryable through a SPARQL endpoint

Work is underway also on the use of SKOS thesauri to enforce semantic and multilingual indexing and discovery of textual content, and to implement it in the INSPIRE Geoportal in order to support semantic search

Conclusions

- Linked Open Data provide an effective framework to enable cross-domain interoperability of eGov data, and to enhance their re-use and aggregation
- However, solutions should be devised to address interoperability open issues related to how data are represented and described
- The usage of thesauri in the whole resource life cycle, as well as the creation of mappings among the used thesauri, may play a fundamental role to enhance cross-domain data interoperability and re-use
- This, however, does not solve the issues (e.g., the not uniform representation of spatial / temporal coordinates) which limit the possible uses of LOD data, and which negatively affects the adoption of LOD

<http://ies.jrc.ec.europa.eu/DE/>

