NATIONAL TECHNICAL UNIVERSITY OF ATHENS
SCHOOL OF RURAL, SURVEYING AND GEOINFORMATICS ENGINEERING
MSC GEOINFORMATICS
SPATIAL DATABASES

Interoperability - Integration of Spatial Databases

Margarita Kokla
Associate Professor
Cartography Laboratory

Co-funded by the European Union

ΕΣΠΑ 2021-2027

Human Resources and Social Cohesion Programme

The project "Support for Internationalization Actions of Foreign Language Speaker Postgraduate Studies of the National Technical University of Athens" (MIS 6004804) is co-financed by the European Union through the Operational Programme "Human Resources and Social Cohesion"

Example

- Aim: creation of an application for the detection and control of pollution along the Danube River.

http://www.ssqq.com/travel/images/danuberiver2016x007.jpg

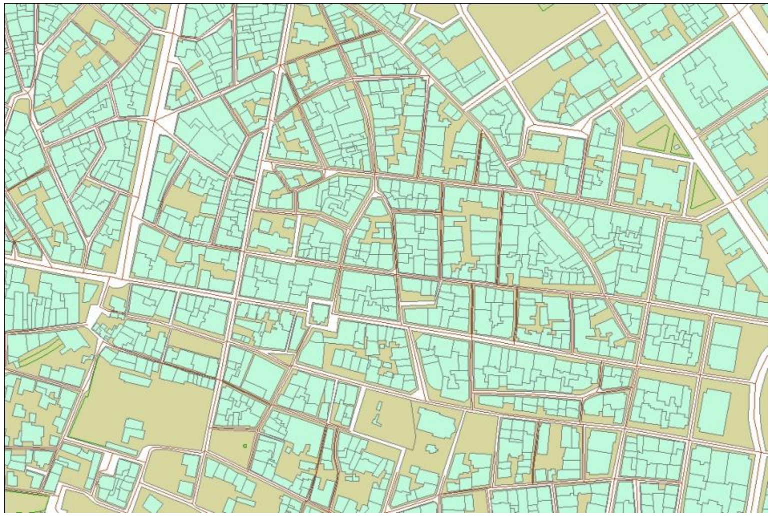
- The river passes through or touches the borders of ten countries: Germany, Austria, Slovakia, Hungary, Croatia, Serbia, Bulgaria, Romania, Moldova, and Ukraine. Its drainage basin extends into nine more countries.
- Problem: different databases from different countries — differences in language, formats, CRS, schemas, update cycles, terminologies, meaning

Types of heterogeneities

- Syntactic heterogeneities
- Schematic heterogeneities
- Semantic heterogeneities

Syntactic heterogeneity

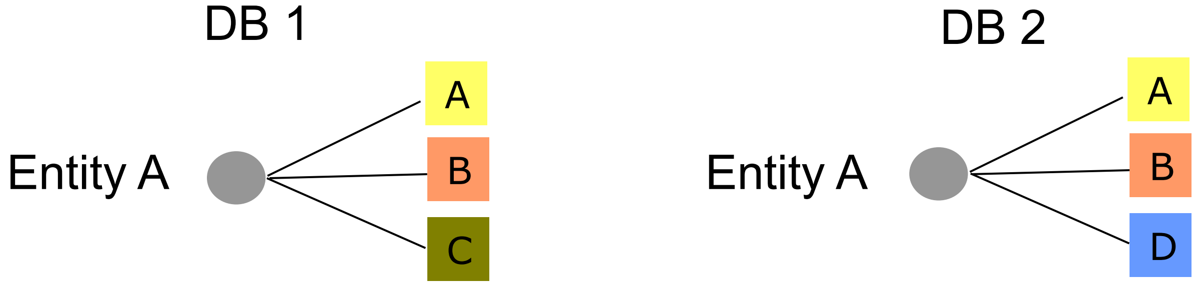
- DBMSs that use different logical data models (e.g., relational and object-oriented).
- Different geometric representations of geographic objects, e.g., vector and raster.



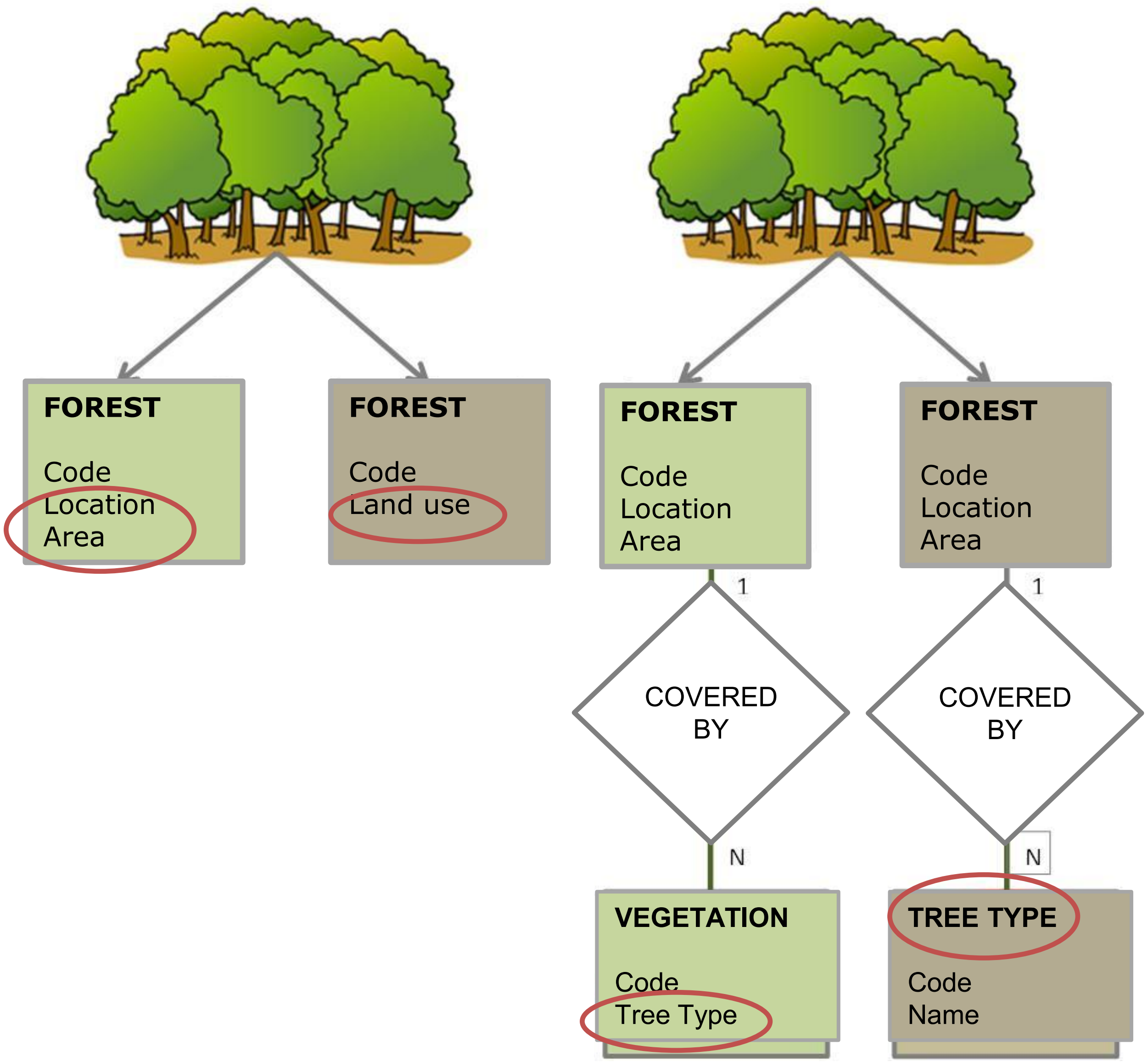
Schematic heterogeneity

Different conceptual data models due to different ways of classifying real-world entities:

- Entity - Entity (1:M and N:M relationships, different attributes or constraints).
- Attribute - Attribute (e.g., different field values).



- Entity - Attribute, when entities in one database are considered as attributes in another database.
- Different representations of the same data, e.g., different units, different spatial resolutions


FOREST
Code
Location
Area
FOREST
Code
Land use
FOREST
Code
Location
Area
FOREST
Code
Location
Area
1
COVERED BY
N
1
COVERED BY
N
VEGETATION
Code
Tree Type
TREE TYPE
Code
Name

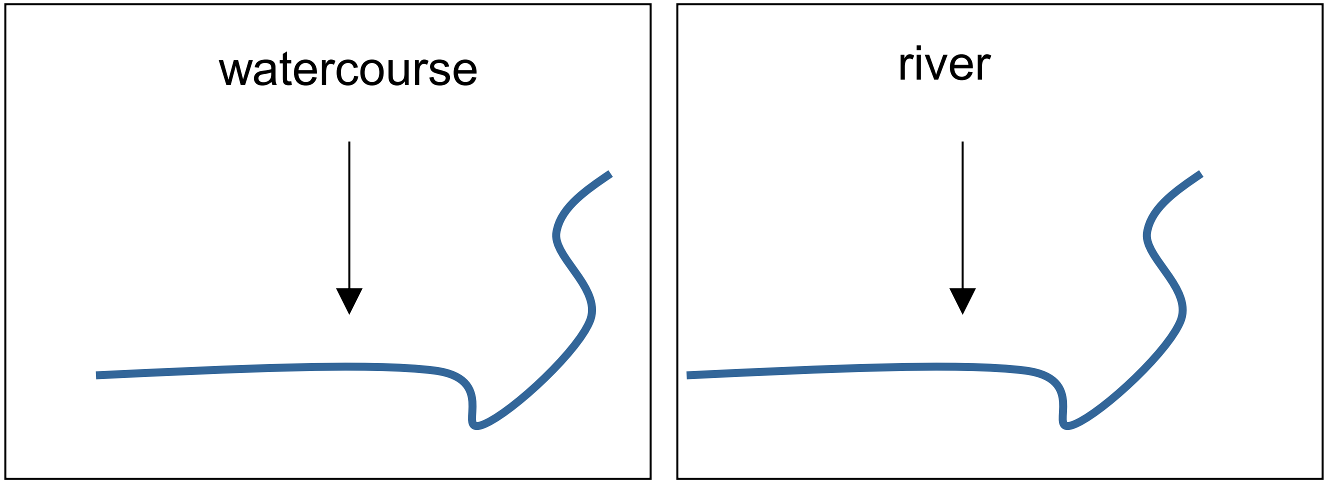
Semantic heterogeneity

Differences in meaning, interpretation, or suggested use of the same or related data:

- Naming heterogeneity
- Cognitive heterogeneity

Naming heterogeneity

synonyms: the same concept is described using different terms



Naming heterogeneity

Homonyms: the same terms are assigned to different concepts.

Bank (WordNet 3.0)**:**

1. Sloping land (especially the slope beside a body of water)

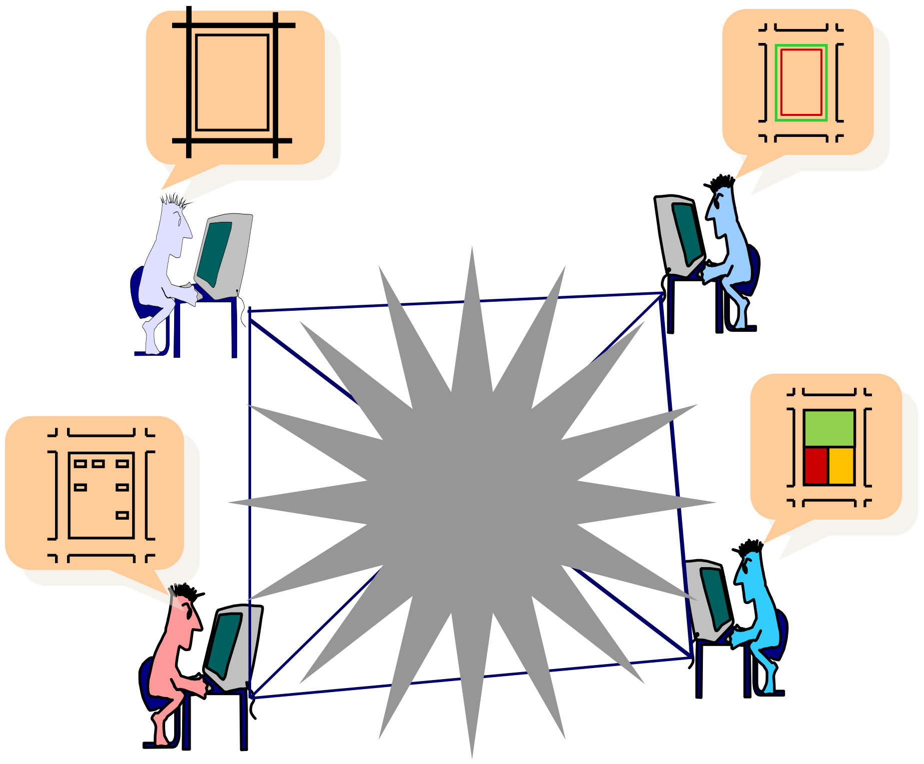
2. *A financial institution that accepts deposits and channels the money into lending activities*

3. A huge bank of earth

4. An arrangement of similar objects in a row or in tiers

Cognitive heterogeneity

Different perceptions of space, e.g., different definitions of categories, different geometric descriptions



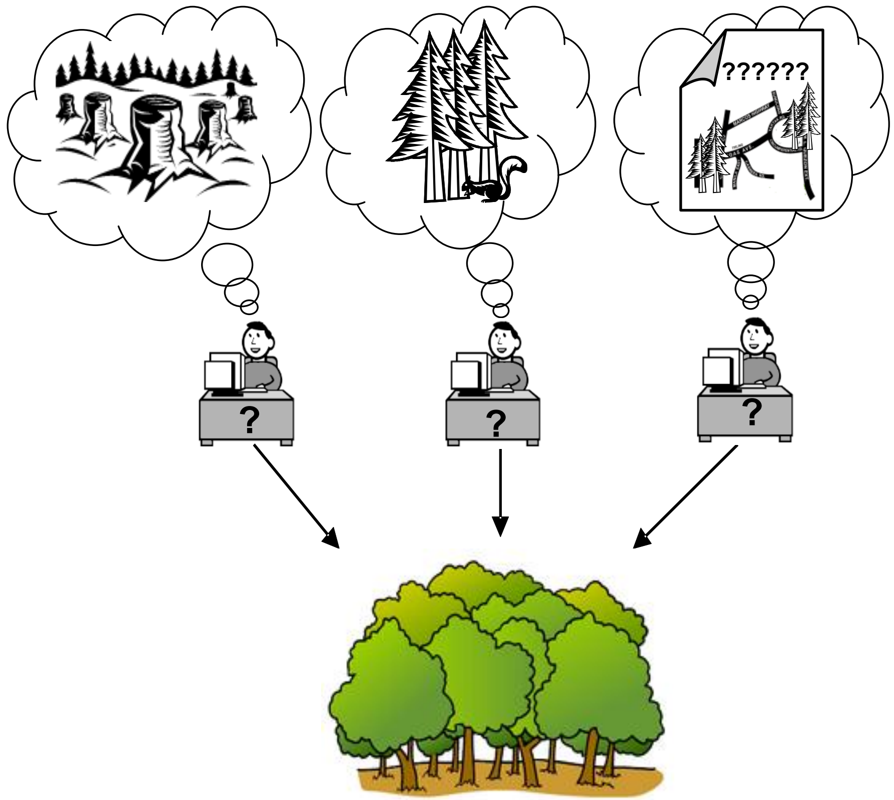
Cognitive heterogeneity

Forest: an area within the administrative boundaries of a forestry service, with or without trees (legal or administrative definition).

Forest: a large area with dense coverage of trees or other plants (land cover definition).

Forest: an area with forested vegetation managed for the production of timber and other products or for indirect benefits such as recreation or protection of watershed areas (land use definition).

Forest: an ecological system consisting of biotic and abiotic elements that interact, where trees dominate with coverage greater than 20% of the area (ecological definition).



Typical conflicts among spatial databases

Naming conflicts

- parcel_id vs lot_no vs property_unit

Unit Conflicts

- meters vs feet

Classification Conflicts

- residential_low_density vs R2

Scale Conflicts

- road network at 1:5,000 vs 1:100,000

Contextual Conflicts

- The meaning of “Vacant land” in planning differs from taxation.

Challenges with Spatial Databases

- Spatial databases are more challenging during integration because they combine:
  - geometry
  - topology
  - time
  - scale
  - uncertainty

INTEROPERABILITY

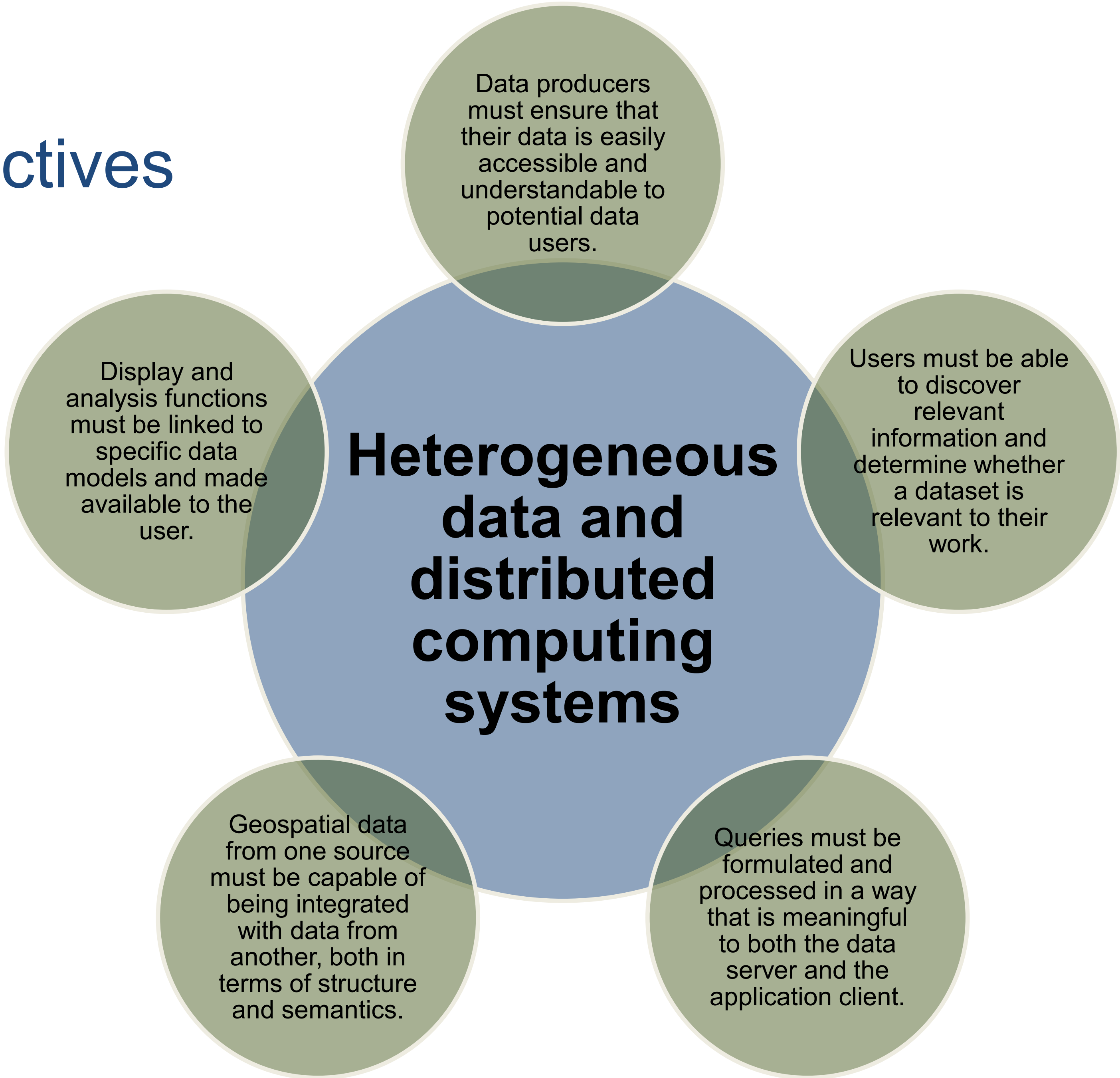
Interoperability

- *“the ability of two or more systems or components to exchange information and to use the information that has been exchanged”* *(IEEE Standard Computer Dictionary)*
- *“the ability of diverse systems and organizations to work together seamlessly. That means exchanging information and using the information that has been exchanged”.* *(International Organization for Standardization)*

Interoperability

- Interoperability usually refers to bottom-up efforts that are neither imposed by a central authority nor driven by a single application.
- Users' systems and data models are heterogeneous and have been developed independently of one another.

Objectives


Data producers must ensure that their data is easily accessible and understandable to potential data users.
Users must be able to discover relevant information and determine whether a dataset is relevant to their work.
Queries must be formulated and processed in a way that is meaningful to both the data server and the application client.
Geospatial data from one source must be capable of being integrated with data from another, both in terms of structure and semantics.
Display and analysis functions must be linked to specific data models and made available to the user.
Heterogeneous data and distributed computing systems

Interoperability means:

- "Openness“ in the software industry, meaning the free publication of internal data structures.
- Free exchange of data: each system knows the format of the data from other systems — exchange standards.
- Commonality in user interaction: meaning the degree of familiarity users have with the environments they interact with.

Interoperability means:

- Simplification in data formats and standards, in user-system interaction, and in the knowledge required from the user.
- Transparency: the detachment of the user from implementation details.
- Similarity: the degree to which data sets, software systems, or organizations use the same vocabulary, follow the same conventions, and therefore can interoperate.

Interoperability levels
Organizational interoperability
Semantic interoperability
Syntactic interoperability
Technical interoperability

Technical interoperability

- It addresses the technical issues of connecting systems and services.
- It includes key aspects such as open interfaces, interoperability services, data exchange, integration and presentation, accessibility, and security of services.
- It is usually associated with the hardware/software, systems, and platforms that enable communication between systems.
- It focuses on communication protocols and the necessary infrastructure for their operation.

Syntactic interoperability

- It is usually associated with data formats.
- The messages transferred by communication protocols must have a well-defined syntax and encoding.
- Many protocols transfer data or content, and this can be represented using markup languages such as HTML and XML.

Semantic interoperability

- It is concerned with ensuring that the exact meaning of the exchanged information is understood by any other application that was not originally developed for this purpose.
- It provides the ability for systems to combine received information with other data and process them in a meaningful way.

Organizational interoperability

- It deals with defining the goals of organizations or businesses, modeling processes, and facilitating collaboration between organizations or businesses that wish to exchange information but have different internal structures and processes.
- It aims to address the needs of the user community by making services available, easily recognizable, accessible, and user-oriented.

Challenges for geospatial data integration

- Different formats, e.g., shapefiles, geoJSON, KML, raster files, relational databases, etc. – need for conversion tools (complex and time-consuming)
- Different schemas
- Different semantics
- Different spatial reference systems
- Varying timeframes for temporal data
- To address these challenges we need appropriate frameworks, standards, and technologies

STANDARDS

Standards for data sharing and consistency

- Standards reduce ambiguity by formalizing concepts.
- Important standards:
  - ISO/TC standards
  - Open Geospatial Consortium (OGC) standards
  - INSPIRE (Infrastructure for Spatial Information in Europe) Directive
  - CityGML / IndoorGML
  - Land Administration Domain Model (LADM)

ISO/TC 211 geographic information standards https://committee.iso.org/home/tc211

- Reference model for Open distributed processing (RM-ODP)
- Geometry
- Topology
- Temporal concepts
- Coverage geometry
- Tracking and navigation
- Multimodal routing and navigation
- Moving features
- Spatial referencing by coordinates
- Data quality
- Metadata
- Rules for application schema
- Methodology for feature cataloguing

Open GeoSpatial Consortium (OGC)

- An international, non-profit organization focused on the development and implementation of open standards for geospatial information and services to support interoperability between different sources, services, and geographic data applications.
- These standards are developed through a consensus process and are freely available for use.

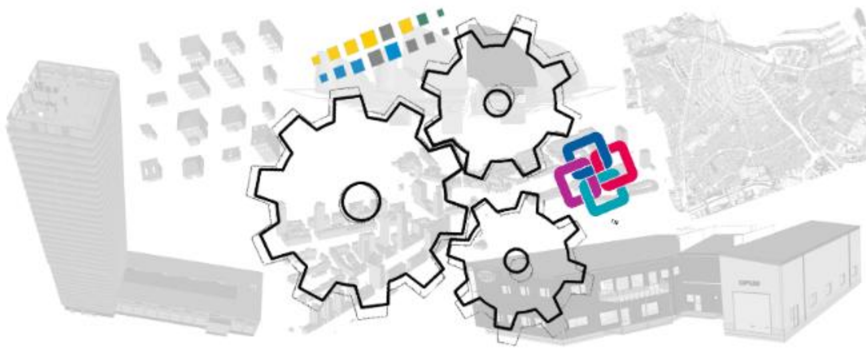
OGC standards

http://www.ogc.org/docs/is/

- Geography Markup Language Encoding Standard (GML)
- Observations and Measurements conceptual model
- OGC GeoSPARQL - A Geographic Query Language for RDF Data
- Web Feature Service (WFS) Implementation Specification
- Web Coverage Processing Service (WCPS)
- CityGML
- …..

GeoBIM benchmark   background ▾ tasks ▾ data software participants events faq presentations & publications

GeoBIM benchmark 2019



Recent news

Jan 11 The three final journal papers are published and open access online. See them in the 'presentations&publications' page.

Sep 03 Final results of the benchmark are now available in the 'presentations&publications' page. Thank you to all the people being involved in the initiative.

Dec 22 The benchmark is finished, thank you to all the involved people for the participation! The website is up-to-date with final participants and (almost) final list of software. In the 'event' page of this website, you can also find the materials and recordings of the GeoBIM benchmark workshop. Have a nice Christmas break everyone!!

All news

Important dates and next steps

March 2019

- Complete materials available
- Start of declaration of interest from participants

July 8th, 2019

- GeoBIM benchmark meeting with participants and proponents

October 31st, 2019

- Deadline for data processing and benchmark answer submission

December 2nd-3rd, 2019

- EuroSDR-AMS GeoBIM benchmark workshop

Team

Francesca Noardo
Ken Arroyo Ohori
Jantien Stoter
Filip Biljecki
Claire Ellul
Lars Harrie
Thomas Krijnen
Giorgio Agugiaro
Margarita Kokla

isprs information from imagery
EuroSDR
TU Delft 3Dgeoinfo
Lund University
NUS National University of Singapore
UCL
National Technical University of Athens

https://3d.bk.tudelft.nl/projects/geobim-benchmark/

Semantic web technologies

- technologies to establish semantically enriched data models, reduce ambiguities, and enhance platform interoperability.
- offer advanced querying capabilities
- facilitate extraction of meaningful insights from complex spatial and non-spatial datasets
- improve data consistency and accessibility

What is the Semantic Web?
https://www.w3.org/2001/sw/

- provides a common framework that allows data to be shared and reused across application, enterprise, and community boundaries.
- It is a collaborative effort led by the World Wide Web Consortium (W3C) with participation from a large number of researchers and industrial partners.
- It is based on the Resource Description Framework (RDF).

Resource Description Framework (RDF)

- RDF is a standard model for data interchange on the Web.
- facilitates data merging even if the underlying schemas differ and supports the evolution of schemas over time without requiring all the data consumers to be changed.
- allows structured and semi-structured data to be mixed and shared across different applications.

RDF statements

- RDF allows us to make statements about resources.
- A statement always has the following structure:

  <subject> <predicate> <object>

- An RDF statement expresses a relationship between two resources.
  - the subject and the object represent the two resources being related;
  - the predicate represents the nature of their relationship.
  - the relationship is phrased in a directional way (from subject to object).
- Because RDF statements consist of three elements they are called triples.

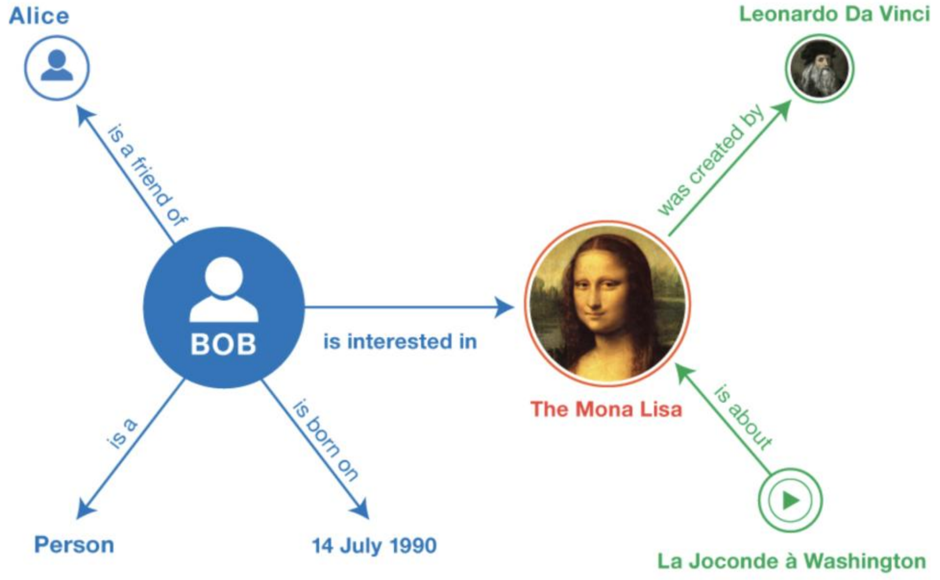
What is a triple?

- Triples (also known as *facts*) are a way to express graph data.
- A triple consists of three components: A subject, a predicate, and an object.
- The predicate describes the relationship between two nodes (or a node and a literal—a string, number, date, etc.), one the subject and the other the object.

Examples of RDF triples

EXAMPLE 1: Sample triples (informal)

```
<Bob> <is a> <person>.
<Bob> <is a friend of> <Alice>.
<Bob> <is born on> <the 4th of July 1990>.
<Bob> <is interested in> <the Mona Lisa>.
<the Mona Lisa> <was created by> <Leonardo da Vinci>.
<the video 'La Joconde à Washington'> <is about> <the Mona Lisa>
```



International Resource Identifiers (IRIs)

- IRIs identify resources such as documents, people, physical objects, and abstract concepts
- The URLs (Uniform Resource Locators) that people use as Web addresses are one form of IRI.
- The notion of IRI is a generalization of URI (Uniform Resource Identifier),
- IRIs can appear in **all three positions** of a triple.
- For example, the IRI for Leonardo da Vinci in DBpedia is:

http://dbpedia.org/resource/Leonardo_da_Vinci

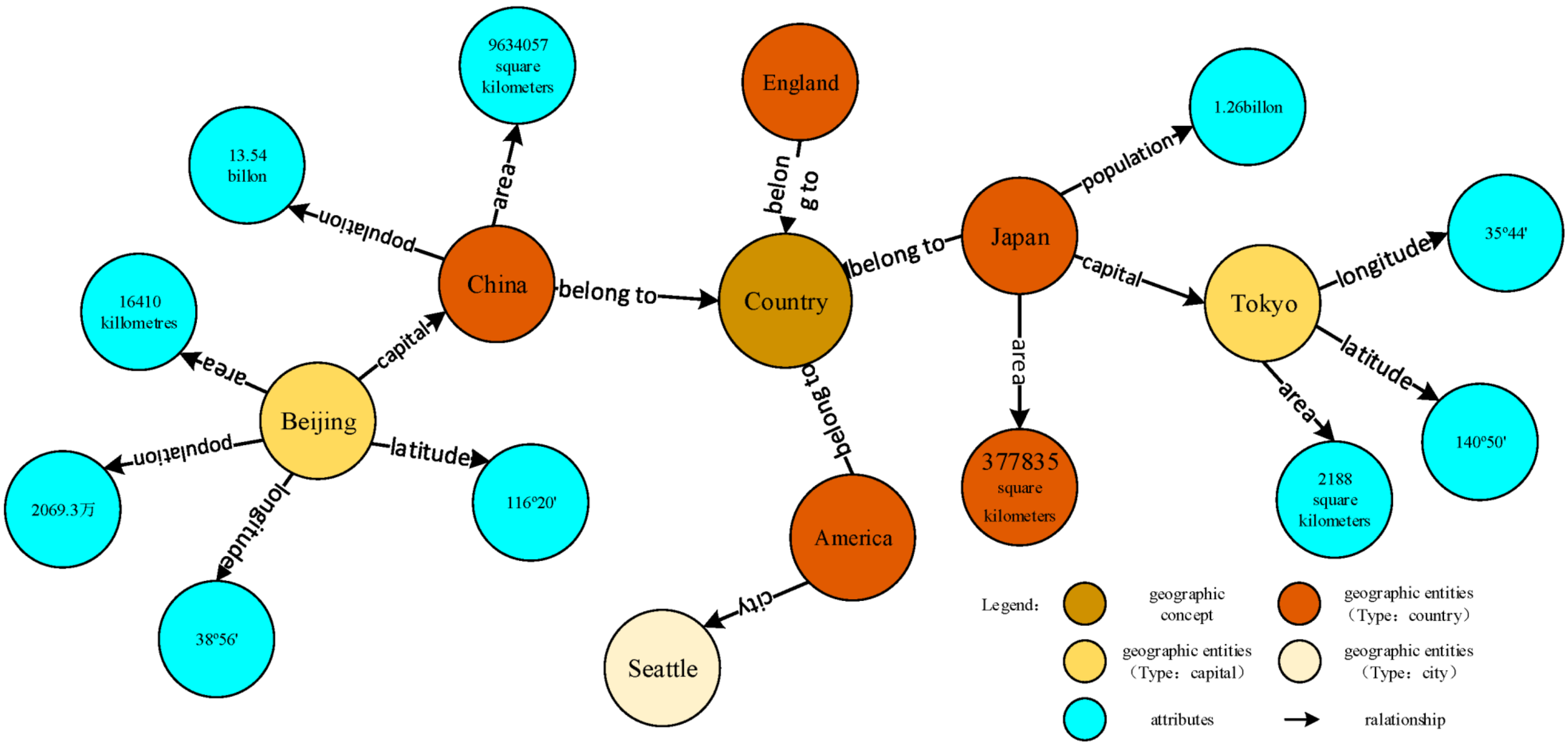
The IRI for an INA video about the Mona Lisa entitled 'La Joconde à Washington' in Europeana is:

http://data.europeana.eu/item/04802/243FA8618938F4117025F17A8B813C5F9AA4D619

- IRIs are global identifiers, so other people can re-use this IRI to identify the same thing

Knowledge graphs

- Networks of linked information about a domain or across domains
- Consist of nodes that represent concepts or entities/instances and edges that represent the connections between concepts/entities.
- Attributes and attribute values describe information about entities and concepts
- Knowledge graphs maybe queried using a suitable query language like SPARQL
- Since KGs represent information in an interconnected way:
  - they are suitable for storage, retrieval and analysis of geospatial information from diverse sources
  - They can be used for reasoning and inferring new knowledge (which is not explicitly defined)



Jiang B, Tan L, Ren Y, Li F. Intelligent Interaction with Virtual Geographical Environments Based on Geographic Knowledge Graph. *ISPRS International Journal of Geo-Information*. 2019; 8(10):428. https://doi.org/10.3390/ijgi8100428

SPARQL query language for RDF

https://docs.ogc.org/is/22-047r1/22-047r1.html

- a set of specifications that provide languages and protocols to query and manipulate RDF graph content on the Web or in an RDF store.
- can be used to express queries across diverse data sources, whether the data is stored natively as RDF or viewed as RDF via middleware.
- A SPARQL query contains a set of triple patterns called a *basic graph pattern*.
- SPARQL queries work on RDF representations of data by finding patterns that match templates in the query

Wikidata Query Service

```
#defaultView:Map
SELECT ?item ?itemLabel ?coords ?image
WHERE {
  # 1. The item has the heritage designation "UNESCO World Heritage Site"
  ?item wdt:P1435 wd:Q9259 .

  # 2. The item is located in the country "Greece"
  ?item wdt:P17 wd:Q41 .

  # 3. Retrieve the geographic coordinates of the site
  ?item wdt:P625 ?coords .

  # 4. Optionally retrieve an image of the site if it exists
  OPTIONAL { ?item wdt:P18 ?image . }

  # Automatic label retrieval (Priority: Greek, fallback: English)
  SERVICE wikibase:label { bd:serviceParam wikibase:language "el,en". }
}
```

ONTOLOGIES AND SEMANTIC INTEROPERABILITY

Semantic interoperability

- One of the major problems facing Geographic Information systems and applications today is lack of semantic interoperability .
- Ontologies are proposed as a means to solve geospatial data heterogeneity problems and support semantic interoperability and information retrieval over the Web.
- An ontology provides a formal explicit description of concepts or meanings of words in a well-defined and unambiguous manner.

Ontology

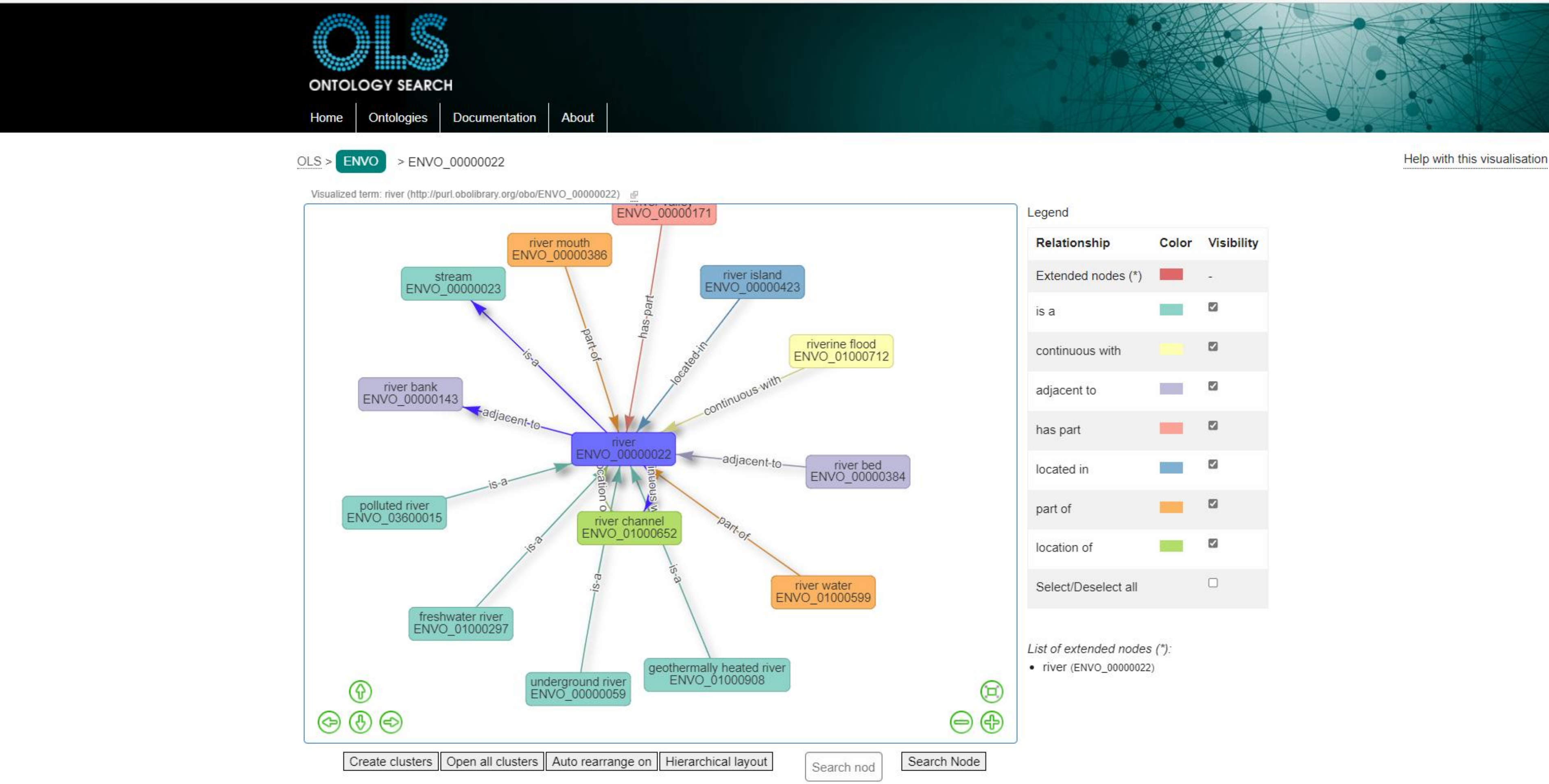
- A clear description of the key entities, as well as the potential relationships between them, referring to a specific application domain.
- A neutral, but easily implementable description or theory of a specific domain, which can be accepted and reused by everyone.
- An ontology practically provides a common vocabulary so that people or computational systems can share, reuse, integrate information about a domain.

Uses of ontologies

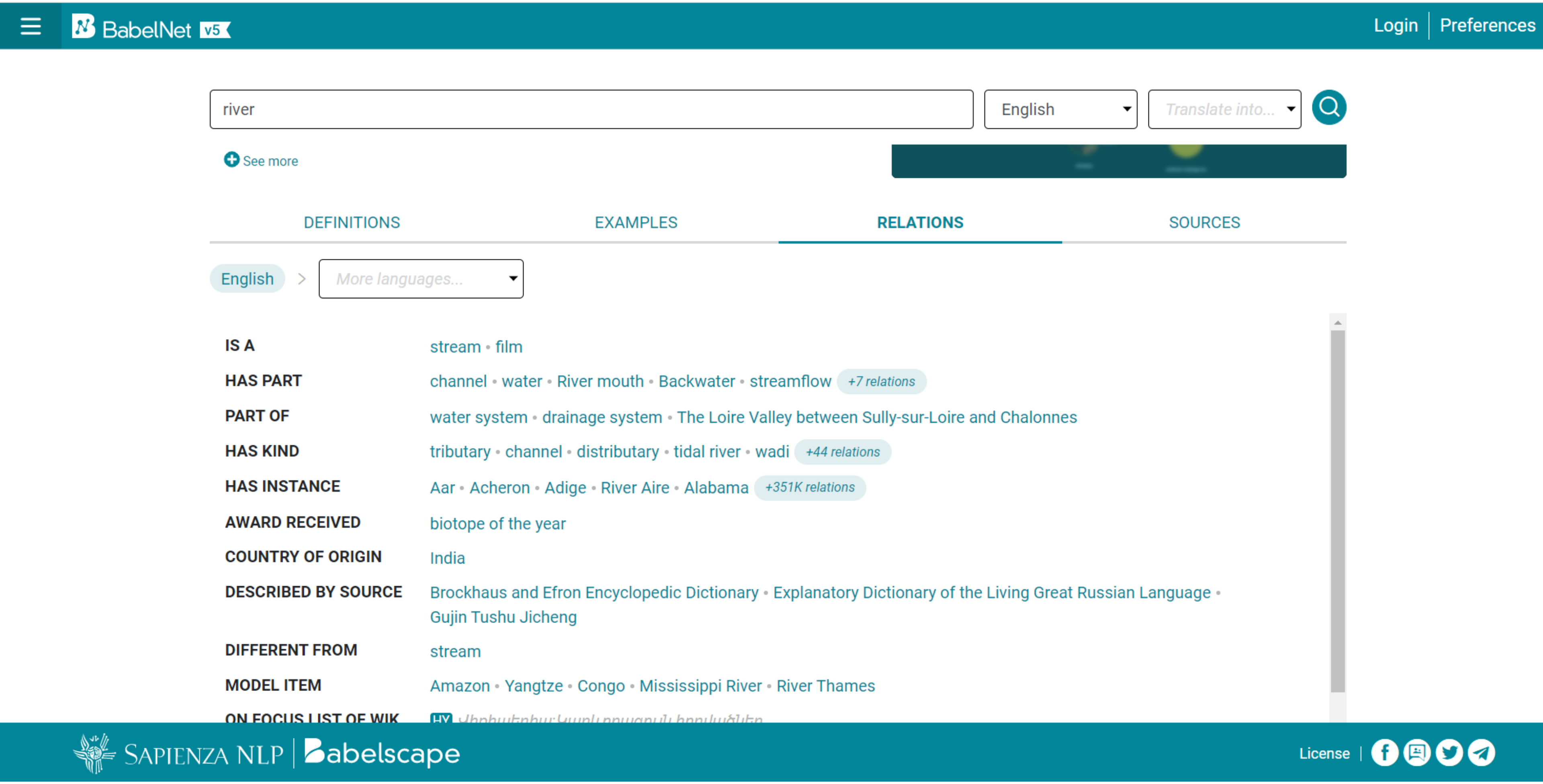
1. Conceptual database design
2. Information retrieval and extraction
3. Natural language translation
4. Representation, management, exchange, and integration of information
5. Semantic web

Ontology components

- Concepts or categories or classes and the description of their meaning, usually through definitions
- Relationships between concepts, such as:
  - hierarchical relationships (IS-A): one class C is a subclass of another class C' if every object in C is also included in C': e.g., all vineyards are agricultural land
  - non-hierarchical relationships, e.g., meronymic (hasPart, isPartOf): the riverbed, the riverbank, and the estuary, are parts of a river
- Properties and allowed values
  - Buildings have a use (residential, health, commerce, office, education, etc.)
- Axioms, constraints, rules
  - Each river must flow into a larger body of water (e.g., sea, lake, or other river).
  - Logging is not allowed in protected forests
  - Buildings have a single use
- Instances, individuals:
  - The Amazon, the Nile and the Danube are instances of the class "river"
- An ontology with a set of instances constitutes a knowledge base



https://www.ebi.ac.uk/ols4/ontologies/envo/classes/http%253A%252F%252Fpurl.obolibrary.org%252Fobo%252FENVO_00000022



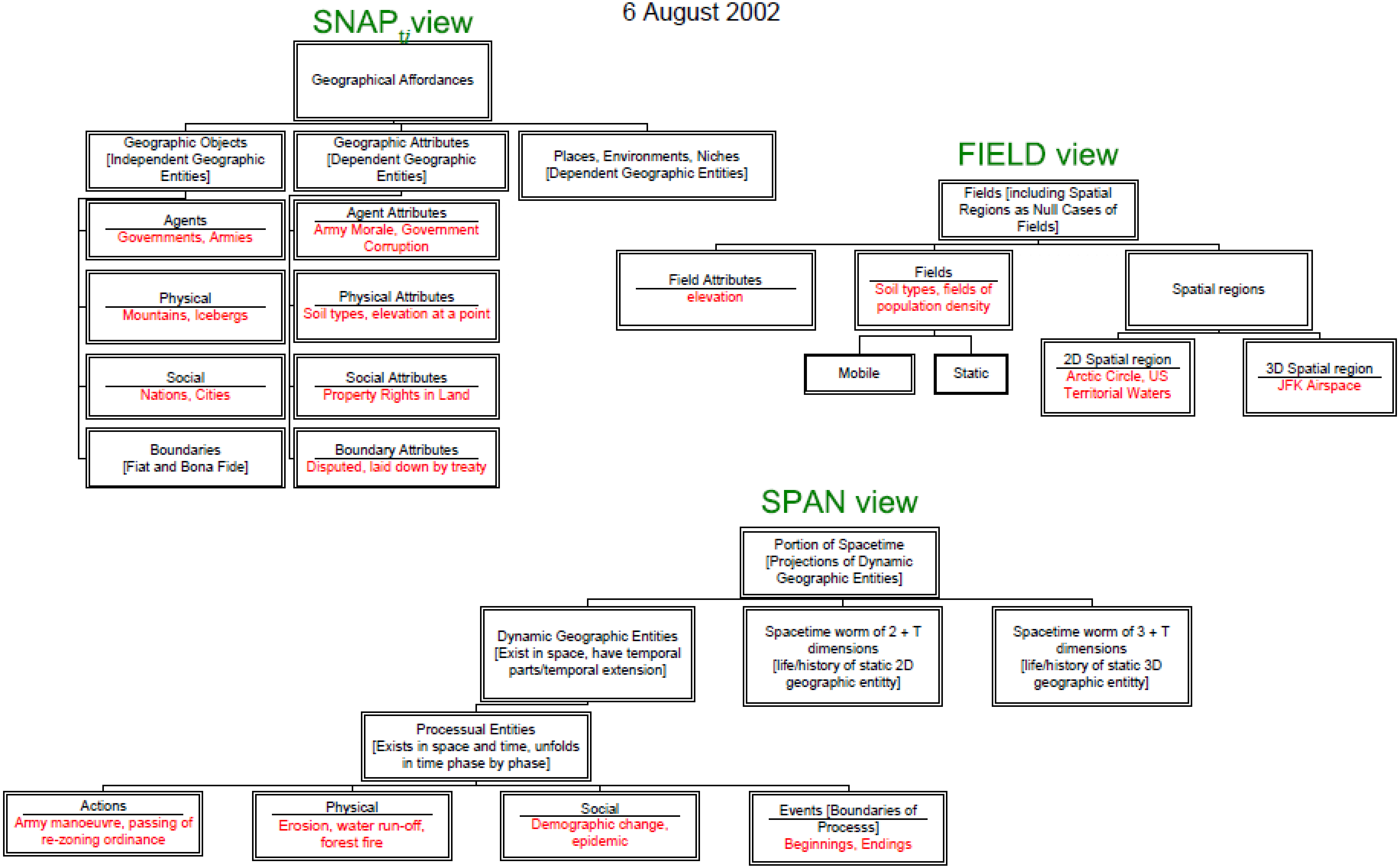
https://babelnet.org/synset?id=bn%3A00067948n&orig=river&lang=EN

Ontology Development Process

- Step 1. Identification of the domain and the objectives of ontology
- Step 2. Exploring the possibility of reusing existing ontologies
- Step 3. Identification of the concepts - classes and the relationships between them
- Step 4. Determination of the properties of classes and their values
- Step 5. Identification of instances

Geographic Ontology

6 August 2002

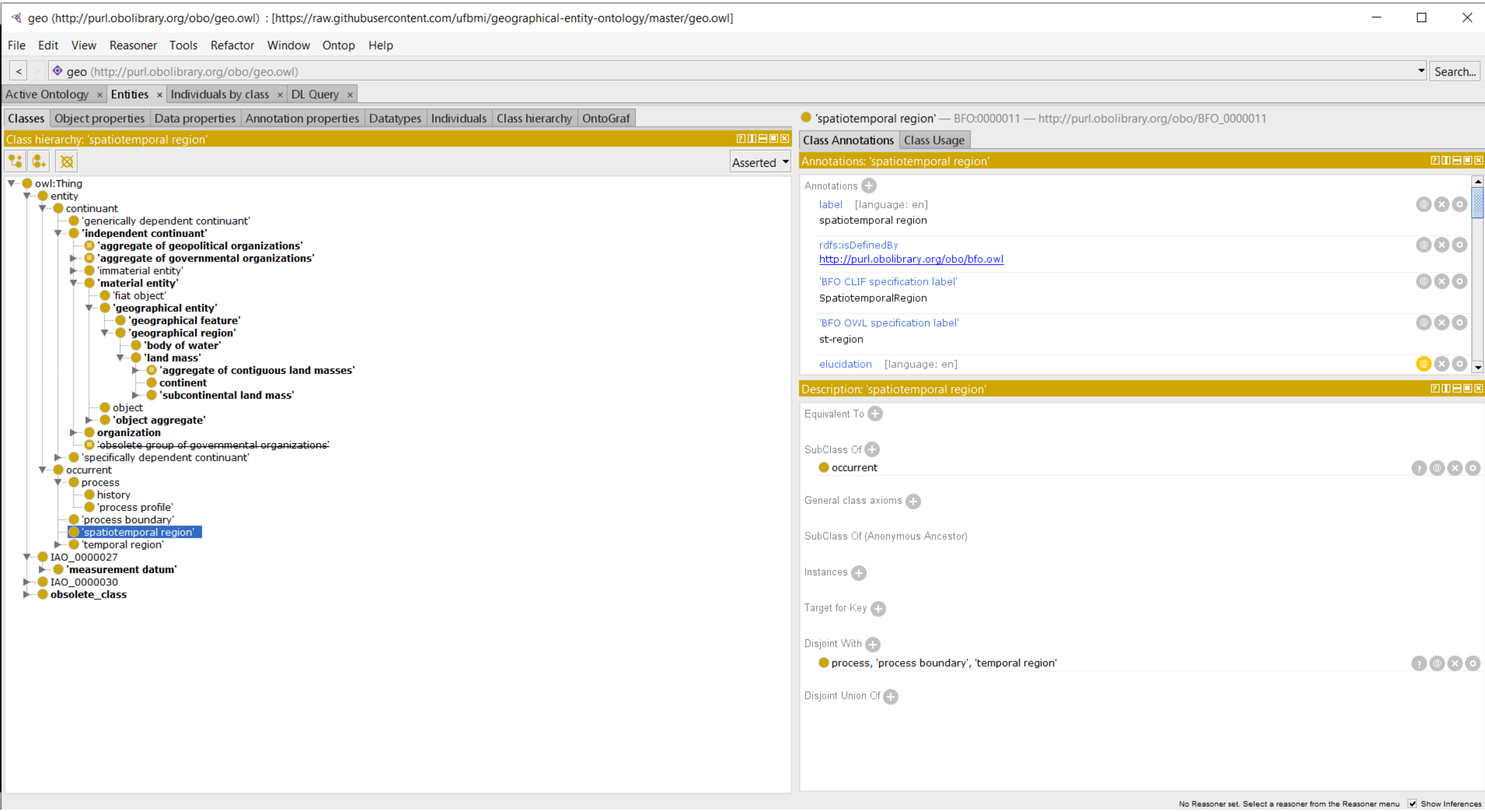


Basic Formal Ontology (https://basic-formal-ontology.org/)

OWL (Web Ontology Language)

- A semantic web language developed by the World Wide Web Consortium (W3C) (https://www.w3.org/OWL/)
- It provides standardization for describing complex knowledge about concepts and the relationships between them.
- It is based on the Resource Description Framework (RDF).
- Ontologies add more vocabulary for describing properties and classes.

Protégé ontology editor



GeoSPARQL (Geographic Simple Protocol and RDF Query Language)

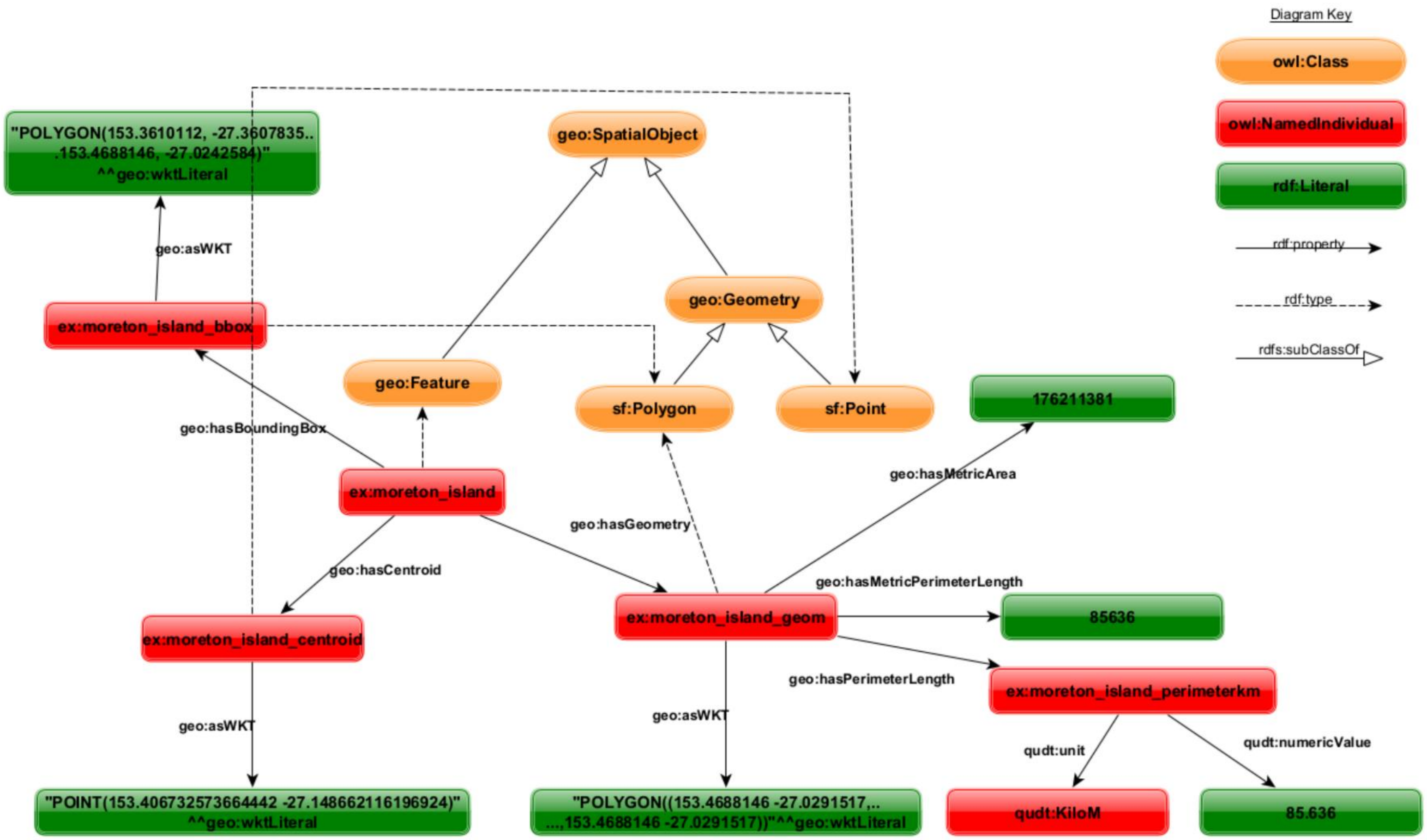
https://docs.ogc.org/is/22-047r1/22-047r1.html

- a geographic query language for RDF data.
- was proposed by OGC (Open Geospatial Consortium) as an extension of SPARQL for querying geographic RDF data.
- defines a vocabulary for representing geospatial data in RDF
  - every geospatial feature is modeled as a subclass of a top level SpatialObjectclass.
  - SpatialObject class is a subclass of the highest level class Thing in OWL.
  - spatial relationships are modeled as object properties on GeoFeatures.
- defines simple topological and non-topological relations for spatial data
- can accommodate both information systems based on qualitative spatial reasoning and quantitative spatial computations.

GeoSPARQL (Geographic Simple Protocol and RDF Query Language)

- GeoSPARQL enables semantic + spatial querying.
- Example:
- “Find schools within flood zones owned by public authority.”

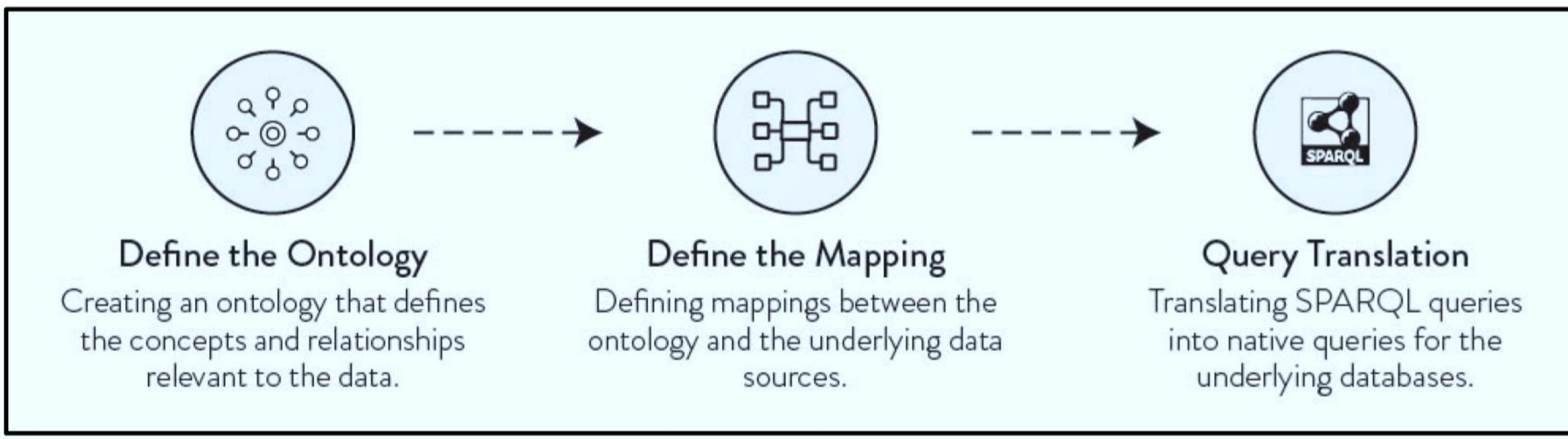
Example



Car NJ, Homburg T. GeoSPARQL 1.1: Motivations, Details and Applications of the Decadal Update to the Most Important Geospatial LOD Standard. *ISPRS International Journal of Geo-Information*. 2022; 11(2):117. https://doi.org/10.3390/ijgi11020117

Ontology-Based Data Access (OBDA)

- Also known as Virtual Knowledge Graph approach
- Addresses the problem of heterogeneous spatial data integration
- An ontology is used as a semantic layer for the data and it is linked to the data source through mappings between the relational database schemas and ontology concepts, properties, and axioms.
- data are maintained in their original relational databases
- User SPARQL queries expressed in ontology terms are automatically translated into SQL queries at runtime
- The data are queried on the fly from the relational data source



Ranatunga S, Ødegård RS, Jetlund K, Onstein E. Use of Semantic Web Technologies to Enhance the Integration and Interoperability of Environmental Geospatial Data: A Framework Based on Ontology-Based Data Access. *ISPRS International Journal of Geo-Information*. 2025; 14(2):52. https://doi.org/10.3390/ijgi14020052