No. 20 (2020): JOSIS' 10th Anniversary

Ontologies for geospatial information: progress and challenges ahead

Revisited… 30/03/2022

Christophe Claramunt

https://josis.org/index.php/josis/article/view/111

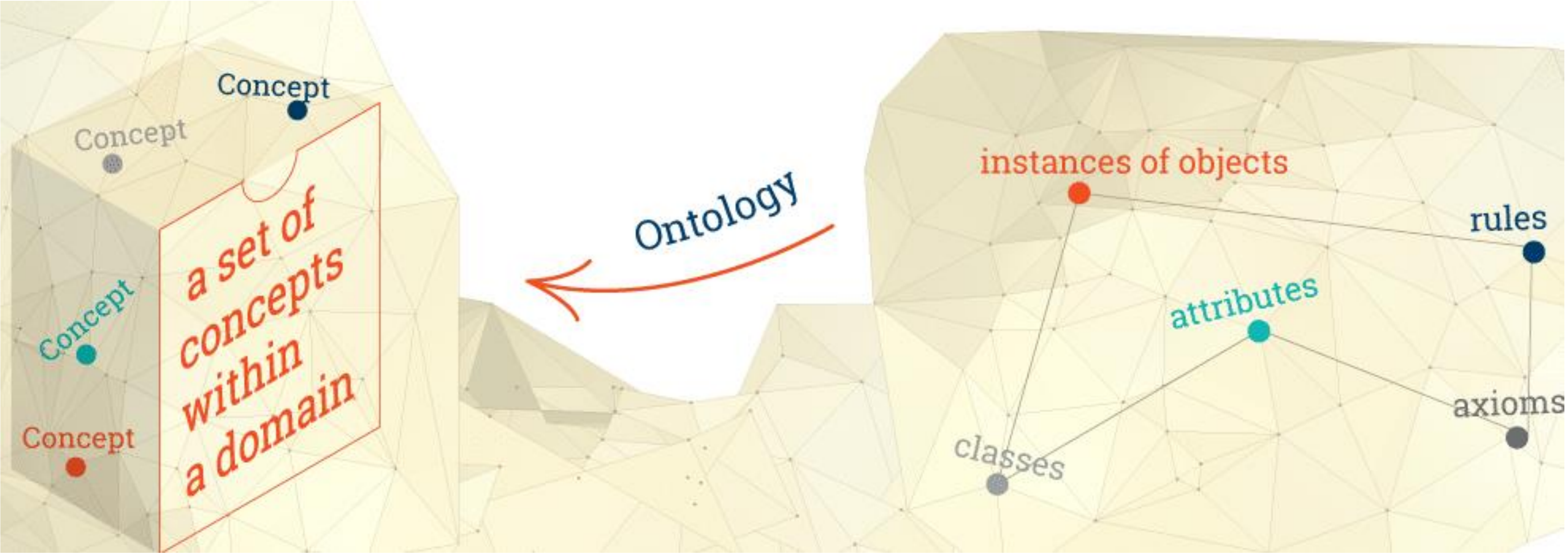
Short introduction to the principles behind semantic ontologies and how they can be applied to complex geospatial information, by evaluating their potential and limitations

The big issue: how can we design a conceptual bridge between current GIS technologies, models, and human beings on the one hand, and the necessary theoretical GIS foundations on the other hand, and how to do so?

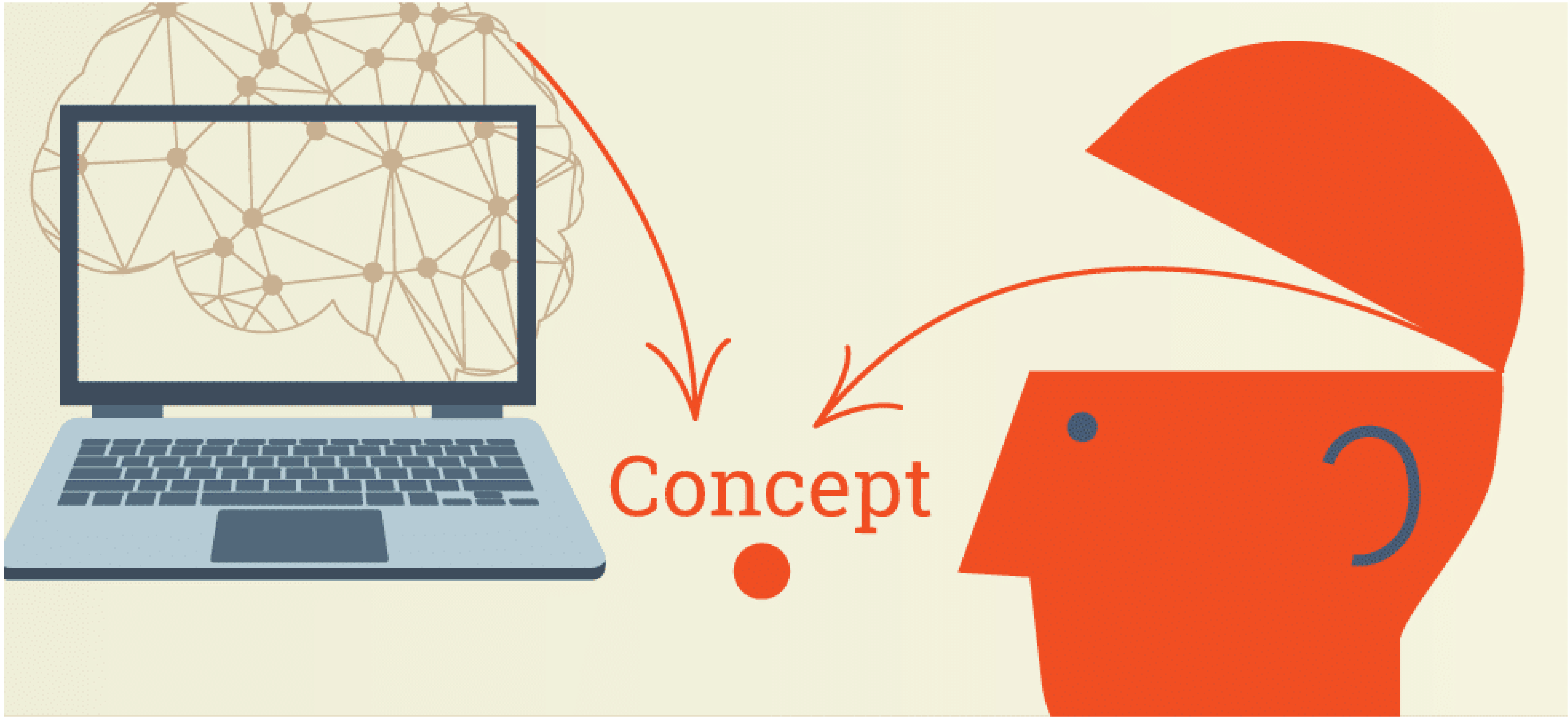
How do humans conceptualize space and time? What are the roles of language and cognition when doing so? This also stresses the close link between what reality is and how interpretations should materialise it as much as possible in modelling and computerised frameworks.

How people perceive the world, cognitive conceptualizations of geographic features and appropriate abstraction paradigms should be developed to support computerised representations

Ontologies do not only introduce a shareable and reusable knowledge representation but can also add new knowledge to a given domain



Ontologies function like a 'brain', they should work and reason with concepts and relationships in ways that are close to the way humans perceive interlinked concepts...



An ontology can be formalized by description logics through the definition of classes, relations, functions, and axioms. In description logic data is represented using a hierarchy of classes, relations and instances.

Ontologies can be implemented according to semantic web formalisms such as the Web Ontology Language (OWL). OWL offers a formal logic-based semantics and is complemented by the Resource Description Framework (RDF) and query standards such as SPARQL

→ schema and query mechanisms and reasoning rules to manipulate the represented data.

→ RDF triples made of subject-predicate-object are easily understandable by machines (though perhaps not so well by humans…).

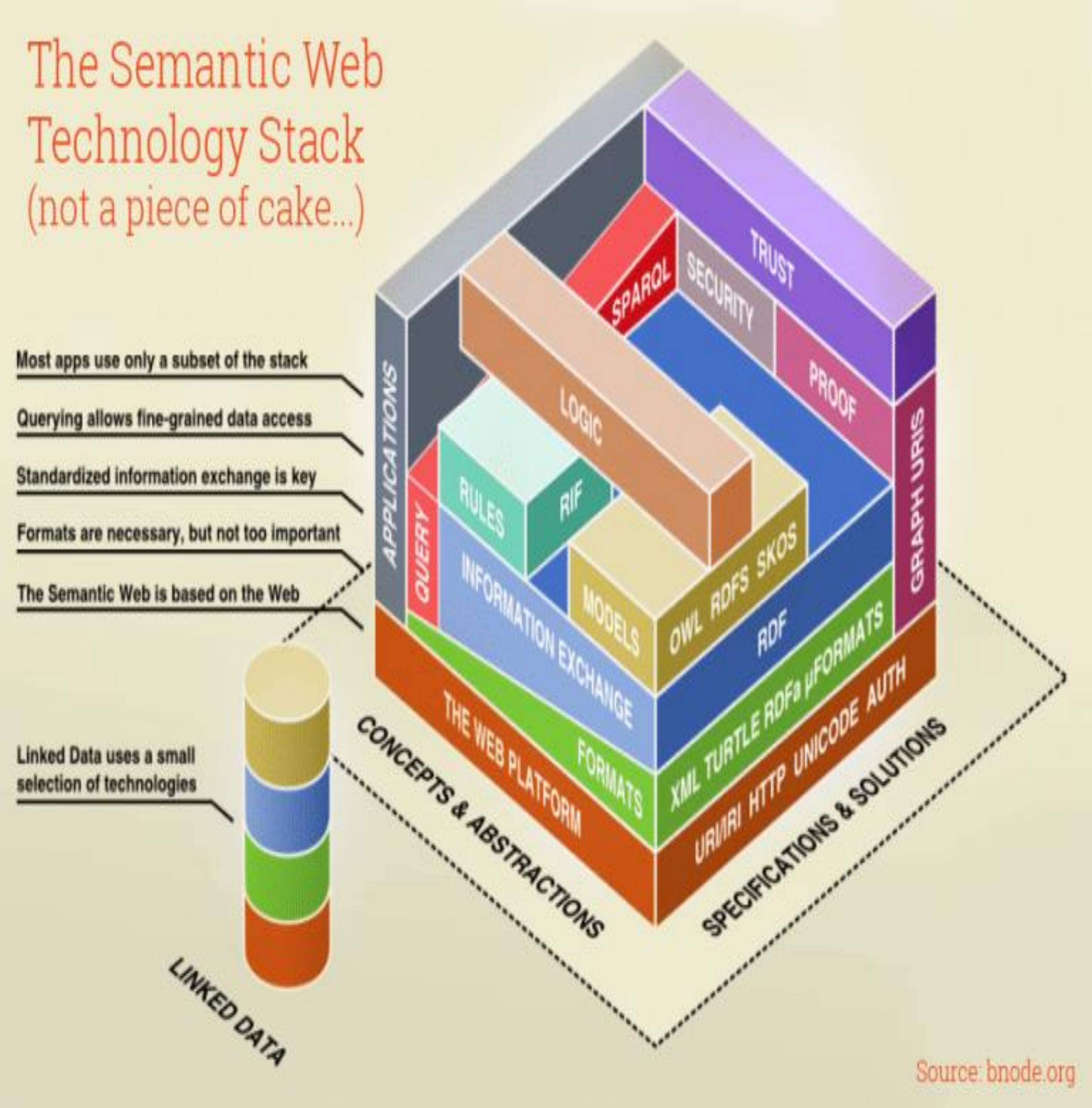
→ Several formats nowadays support RDF implementations including RDF/XML,N-Triples, JSON-LD, Turtle, and Notation.

Semantic data and rules can be used and reused without rewriting code, thus reducing maintenance and evolution costs.

RDF can be thought of as a grammar in which facts about the world are expressed in RDF as triplets of <subject, predicate, object>.

RDF, usually in the form of XML, can be embedded in HTML so that browsers, search engines, and other programs can manipulate the represented data and infer additional knowledge.

Alternative models to RDF and XML have been suggested, for example, the JSON JavaScript Object Notation which is both more compact and" easier for humans to read and interpret".



Formal limitations

- Usability ….

- Limited set of properties (e.g., OWL)

- Constraints (rigidity regarding data updates)

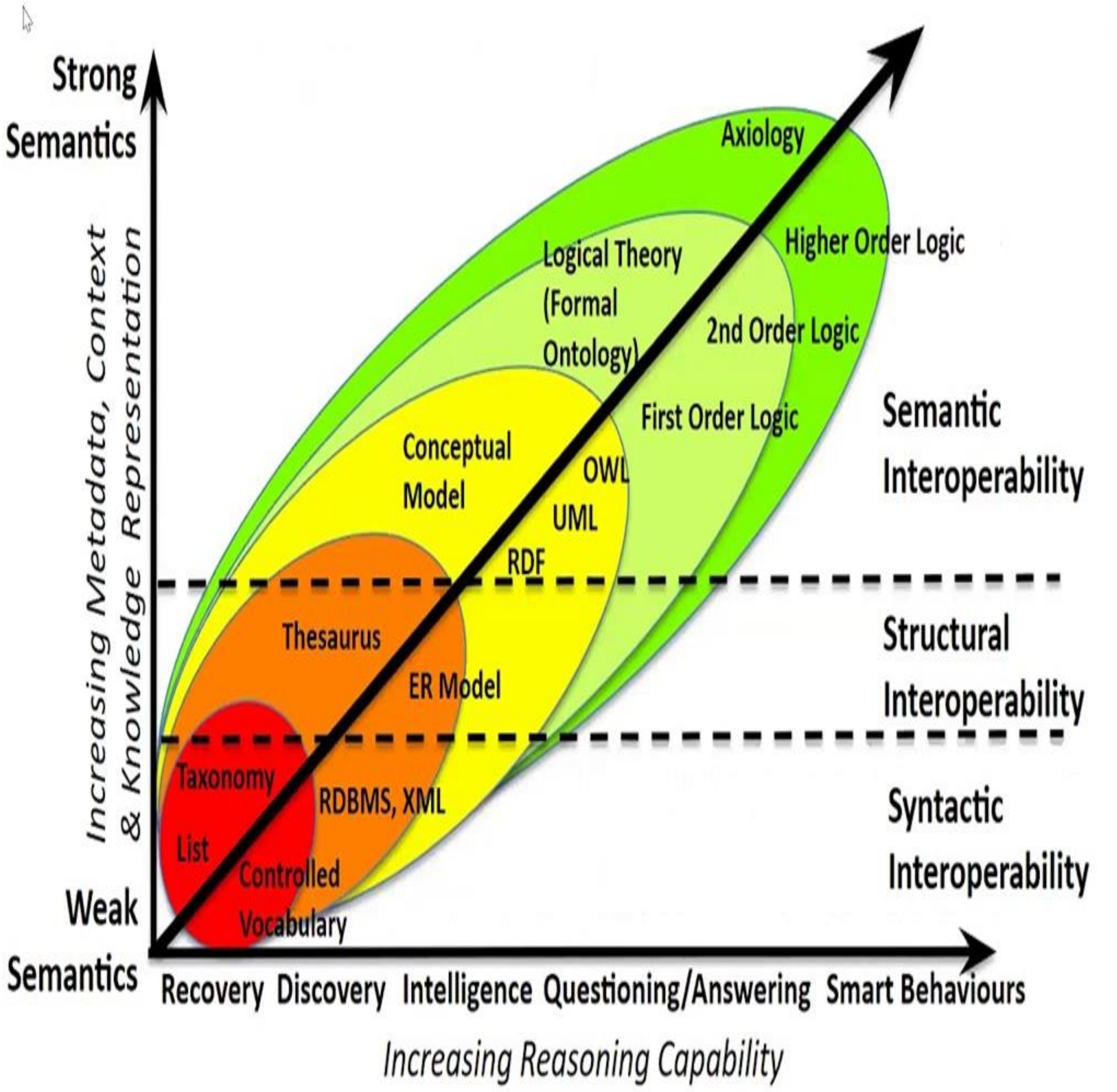
- Data schema updates too

- Lack of versatility

- Independent of the computational level

  - Interoperability with conceptual/manipulation level, e.g., graph databases, e.g. Neo4J

  - **Not a piece of cake, really !!!**


Strong
Semantics
Weak
Semantics
Increasing Metadata, Context & Knowledge Representation
Axiology
Higher Order Logic
Logical Theory
(Formal
Ontology)
2nd Order Logic
First Order Logic
Conceptual
Model
OWL
UML
RDF
Thesaurus
ER Model
Taxonomy
RDBMS, XML
List
Controlled
Vocabulary
Semantic
Interoperability
Structural
Interoperability
Syntactic
Interoperability
Recovery
Discovery
Intelligence
Questioning/Answering
Smart Behaviours
Increasing Reasoning Capability

GIScience still searching for novel theoretical pathways to re-engineer GIS data models… and establishing a close link between reality and data representations

Ontology re-appeared and offered canonical descriptions of knowledge domains as defined as "a neutral and computationally tractable description or theory of a given domain which can be accepted and reused by all information gatherers in that domain".

A geospatial ontology should be defined as a formal, explicit specification of a shared conceptualization providing a non-ambiguous and formal representation of a geospatial domain (or space as a whole).

Ontologies should abstract the world as it is, using formal and primitive geographical entities and properties at large (events, processes etc.)

A geospatial ontology should define in formal terms (axioms and reasoning capabilities) the constituents of reality *within a given domain or generic enough* and should be soundly defined and logically possible, extensible and implementable. The objective is to minimize the distance between reality and a final representation.

A geospatial ontology should encompass all the categories and modelling abstractions: from fields to objects, from events to processes as well as causal to qualitative spatial and temporal relations. Therefore, a taxonomy, a formal vocabulary that can be computerized at the software engineering level

Ontologies should also favour interoperability and knowledge sharing between different applications and users (bottom-up vs. top-down).

Towards geospatial ontologies

The benefits of a sound ontology for geographic information include not only a conceptual, logical and computational bridge between reality and machines but also a basis for an exchange of information and cross-disciplinary collaboration between different domains of science. -> domain or generic ontologies

Geospatial ontologies should share many structural similarities, regardless of the language in which they are expressed. Most ontologies describe individuals, categories, attributes, relations, rules, actions and events.

The search for a rich geospatial ontology generalizable across many fields and applications is still a major challenge.

Geospatial objects are complex abstractions, they have parts and can be constituents of others, they have bona fide or fiat boundaries, they are either well or vaguely defined and encompass a large range of spatial relations and are associated with categories and additional semantics.

While being potentially defined at different levels of abstraction and granularity they evolve through events and processes and generate multiple relational networks in space and time.

Geospatial ontologies

to provide complete and appropriate representations of real-world phenomena that integrate the four spatio-temporal dimensions and the whole complexity of real-world phenomena;

to create a formal and computational data model that could provide a sound representation of all the concepts identified at the ontological level.

whether a general ontological-based might support a formal umbrella that includes all four spatio-temporal dimensions within a unified framework?

A geospatial ontology nowadays offers a series of functionalities towards the geospatial semantic web where a comprehensive set of geographical properties and abstractions can be both understandable by different communities and implemented in order to better match human cognition,

Geospatial ontologies should be grounded by establishing meaningful and suitable geographical and semantic primitives and integrating time as well as different levels of abstraction and users' points of view.

A good balance should be also made between generic geospatial ontologies and domain-based ontologies, since the two views are complementary.

Major research challenges (1)

Formalization of expert knowledge is a key issue.

Transferring expert knowledge to classes, relations and rules is not always straightforward, especially as declarative languages are not user friendly.

The way triples might represent the full complexity of relational concepts is not always satisfactory, especially for some semantically complex relations and the way triples might be interpreted is another difficult issue.

Transferring specialised knowledge from texts or domain experts to abstract and effective concept representations is far from being an easy task and can often lead to misinterpretations and ambiguities.

Major research challenges (2)

Are OWL and RDF sufficient enough to represent and manipulate the whole complexity of geographical and temporal abstractions?

Although objects are relatively well represented, image data is not completely represented by RDF triples and GeoSPARQL and temporal abstractions have still to be integrated.

Similarly, 3D models and Building Information Models should be fully integrated. Last, but not least the emergence of big geospatial data is likely to bring computational issues as RDF and GOSPARQL were not designed to deal with massive geospatial datasets.

Major research challenges (3)

Are the functionalities of the current model and query languages such as GeoSPARQL rich and understandable enough to provide a high-level data manipulation level?

Is GeoSPARQL computationally effective as server loads are likely to be costly operations?

So far GeoSPARQL functions and queries are far from being intuitive and really far away from what a typical user might expect.

Major research challenges (4)

Geospatial ontologies should be extensible and reusable and possibly cross-domains and communities.

Interoperability implies leveraging existing standards and being adaptable to existing data-centric infrastructures.

Major research challenges (5)

The large range of ontology language editors, although some are well established (e.g., Protégé), does not always facilitate interoperability and results in uncoordinated software engineering efforts.

Major research challenges (6)

One of the advantages of formal and numerical representations of ontologies and geospatial ontologies lies in the visibility of the notations.

This leads to large repositories of data representation and an intermediate level, where users might manipulate such abstractions but not at a higher level of representation.

Major research challenges (7)

As scientific applications are not of high priority within the GIS industry despite the availability of many ontology standards, and as re-engineering existing applications will be extremely costly, embedding geospatial ontologies within GIS will be far from straightforward.

+ Major technological challenges

Impact of standard Web recommendations from the ISO and OGC

Dominance of the Web in the development of novel software engineering solutions, any alternative software engineering options?

Geospatial ontologies on the Web are largely based on different formats to implement RDF triples such as XML, RDFa, and JSON-LD.

GeoSPARQL language is a standard RDF SQL-based query that manipulates geospatial RDF data. It provides a GML-based representation of geometrical literals, topological relations, a SPARQL query interface and a rule interchange formal for further inferences.

Vendor-based software implementations of RDF (e.g., Oracle Spatial) and SPARQL and are associated with geometrical extensions such as (e.g. KML, GeoJSON). But SPARQL has similar limitations as SQ as far from being intuitive.

Geospatial ontologies should offer fundamental resources to remodel geospatial information.

Requires a good balance between sound and interoperable geospatial infrastructures and not re-inventing the wheel as every effort should be made to leverage existing GIS data infrastructures wherever possible.

Geospatial ontologies to be closely integrated within much broader and large contexts.

For instance, the Sustainable Development Goals Interface Ontology (SDGIO, 2020) is an example of how geospatial ontologies might act as a foundation for sustainable development.

Agenda for a sustainable environment

-> representation of environmental entities, processes, interconnections with many ecosystems and urban systems

-> geospatial ontologies in order to create open representations and standards.

-> This effort should involve many GISscience related communities, researchers and practitioners.