

Observation-Driven Geo-Ontology Engineering

KRZYSZTOF JANOWICZ
University of California, Santa Barbara, USA
jano@geog.ucsb.edu

Abstract

Big Data, Linked Data, Smart Dust, Digital Earth, and e-Science are just some of the names for research trends that surfaced over the last years. While all of them address different visions and needs, they share a common theme: How do we manage massive amounts of heterogeneous data, derive knowledge out of them instead of drowning in information, and how do we make our findings reproducible and reusable by others? In a network of knowledge, topics span across scientific disciplines and the idea of domain ontologies as common agreements seems like an illusion. In this work, we argue that these trends require a radical paradigm shift in ontology engineering away from a small number of authoritative, global ontologies developed top-down, to a high number of local ontologies that are driven by application needs and developed bottom-up out of observation data. Similarly as the early Web was replaced by a social Web in which volunteers produce data instead of purely consuming it, the next generation of knowledge infrastructures has to enable users to become knowledge engineers themselves. Surprisingly, existing ontology engineering frameworks are not well suited for this new perspective. Hence, we propose an observation-driven ontology engineering framework, show how its layers can be realized using specific methodologies, and relate the framework to existing work on geo-ontologies.

1 Introduction and Motivation

The ever increasing data universe and studies showing that the amount of new data created surpasses our capacities to store it, call for improved filtering, knowledge mining, and retrieval capabilities. While we hope that more data with a higher spatial, temporal, and thematic resolution enables us to better address complex scientific and social questions, understanding, sharing, and reusing data becomes more challenging. Methods to integrate data from heterogeneous sources gain importance, e.g., for achieving the vision of a multi-thematic and multi-perspective Digital Earth. The same trend can be observed in other scientific domains. For instance, the integration of massive data has been named a new frontiers for Bioinformatics ([Jones et al., 2006](#)).

The Semantic Web has been proposed to address the integration problem

and improve information retrieval beyond simple keyword matching via knowledge representation languages and reasoning. However, at its very core, the Semantic Web is a technology stack developed with questions of decidability and inferential power rather than conceptual modeling in mind. Consequently, the formal semantics defined for languages such as the Web Ontology Language (OWL) does not always integrate well with modeling paradigms in the geosciences, i.e., with numerical, statistical, and geometric methods. In the past, this has frequently led to misconceptions on the suitability of OWL for checking the logical consistency of geo-data against schema knowledge. Moreover, the Semantic Web is mostly feature-driven, a field-based view (Goodchild et al., 2007) is difficult to incorporate. Generally speaking, the majority of Semantic Web research and technologies does not answer the question of *how* to engineer geontologies and integrate them with GIS or Spatial Data Infrastructures (SDI). Most importantly, a layer specifying the transition from observation data to classes and relations is missing.

Among other factors, the lack of progress in realizing a Geospatial Semantic Web is due to a misunderstanding that ontologies are a replacement for numerical and statistical modeling. However, ontologies are best understood as a communication and exchange layer. While geo-scientists rely on sensor observations, models, and simulations, results are often shared using categories, i.e., as nominal data. For example, we publish scientific results on the role of the forest industry for rural economics by referring to the categories *Forest* or *Town*. However, the meaning of these terms differs between and within domains to a degree where they may become incompatible. Consequently, data combined on the category level is often incommensurable due to the hidden ambiguities in the definition of these categories. Nevertheless, this semantic heterogeneity should not be misunderstood as a burden but is a challenge imposed by the situated nature of conceptualization; geographic features and their types are a product of cognition and social convention (Brodaric and Gahegan, 2007; Mark, 1993; Barsalou, 2003; Lehar, 2003; Janowicz, 2010).

For instance, under California law, *Town* and *City* are explicitly defined as interchangeable by Government Code Section 34502. In contrast, Bloomsburg is the only incorporated town in Pennsylvania, while all other municipalities are either boroughs or townships. Following the Utah Municipal Code 10-2-301, five types of cities can be distinguished based on their population, while those municipalities with a population under 1000 are towns. Thus, a comparison of studies on rural economics which refer to the category *Town* without giving unambiguous definitions is on shaky grounds. The same argument can be made for the *Forest* category. Lund (2011), for instance, lists hundreds of different definitions for forest, deforestation, afforestation, and reforestation – while none of them is right or wrong, they all have a direct impact on whether a certain area on the Earth’s surface is categorized as *Forest* or not. There is nothing like a *real forest* in the physical world and, moreover, it turns out that this assumption is rather hindering for ontology engineering (Kuhn, 2009). As these terms appear frequently in plain text, readers are often not aware of their different meanings and potential incompatibilities. To address some of these problems,

publishers and e-Science researchers envision *executable papers* through which users can directly reproduce the presented results, e.g., deconstruct nominal data to observations (Nowakowski et al., 2011).

The resulting problems in sharing and reusing nominal data become even more pressing taking the current Big Data trend into account (Weinberger, 2011). By running a global query for the type *Town* (e.g., using the DBpedia class *dbpedia-owl:Town*), scholars and especially software agents may easily miss the fact that among thousands of low population municipalities their dataset also lists Los Angeles, CA or Stuttgart, Germany.¹ Analysis performed on such a collection may, therefore, yield misleading results. Experience and recurring failures suggest that we are combining data which was created under different assumptions on a regular basis (Bishr, 2006). This is especially true for the re-usage of observation data collected from sensors. The advent of Linked Data (Bizer et al., 2009), where raw data is shared, interlinked, and combined on-the-fly, will most likely act as catalyst, especially given the fact that the ontologies used so far are rather lightweight and may fail to restrict the interpretation of terms towards their intended meaning (Jain et al., 2010).

Finally, the lack of geo-ontologies is accompanied by a lack of statistical methods to evaluate nominal geo-data. The prevalent GIS textbook view on data quality will introduce positional accuracy and completeness in great detail and only provide marginal and abstract information on logical consistency (often reduced to topological relations between features) and attribute accuracy. While measuring and reporting positional accuracy is well standardized, measures for *semantic accuracy* are largely missing. The argument that this is due to missing higher-order measures or ground-truth only holds partially; as will be argued in this work, Volunteered Geographic Information (VGI) may provide a baseline.

Surprisingly, existing ontology engineering methodologies are not well suited to bridge the gap between observation data and knowledge engineering. In this work, we therefore propose a novel, observation-driven framework to the engineering of geo-ontologies (ODOE) that can be integrated with existing methodologies. Starting with the semantic description of sensors and their observations, we show how to combine semantics with methodologies such as geostatistics, data mining, and machine learning to construct ontological primitives using reification. Next, these primitives are integrated with so-called geo-ontology design patterns acting as strategies to assist domain experts in becoming knowledge engineers. These patterns can, in turn, be combined to crisp application-level ontologies, so called microtheories, that are organized in a lattice. Mappings between such theories are inferred, e.g., via the least common subsumer, similarity reasoning, ontology alignment, reasoning in the presence of inconsistencies, or out of provenance data. Our methodology preserves semantic diversity and local, application-centric conceptualizations without giving up interoperability. Provenance information is maintained while going up the layers from observations to ontologies. Therefore, scientists exchanging data on the category level can track down how classes were constructed and, thus, uncover hidden incom-

¹via `dbpedia:Stuttgart rdf:type dbpedia-owl:Town`.

patibilities before combining datasets. The proposed framework should enable users to become active knowledge engineers. First steps into this directions are already visible in form of Google’s Freebase in which users can not only contribute information but also the schema knowledge to structure this information (Bollacker et al., 2008). By fostering semantic diversity, the ODOE framework addresses some of the concerns voiced against a purely engineering view on geographic space (Schuurman, 2006).

Observations have been proposed as the foundation of geo-ontologies before. The presented work is especially influenced by Frank’s ontology tiers (Frank, 2003), Couclelis’ ontology of geographic information (Couclelis, 2010), and Kuhn’s notion of semantic engineering (Kuhn, 2009). However, the ODOE framework differs in coverage, scope, and applicability. Following Kuhn’s approach, and in contrast to the work by Couclelis, our work takes a strictly engineering perspective on ontologies and defines them as networks of constraints. In terms of applicability, i.e., proposing concrete technical and methodological solutions, the ODOE framework is most similar to Frank’s proposal. However, the presented framework is based on existing Semantic Web technologies, implementations, and ontologies, and, hence, can be realized within existing knowledge infrastructures. While provenance is a key component of the ODOE framework, it is not covered in Frank’s and Couclelis’ work. Similarly, while both fundamentally rely on observation data, they do not account for sensor semantics.² Finally, Coulelis’s (but also Frank’s) work has a strong focus on the stepwise construction of objects, while the presented frameworks focuses on the engineering of geo-ontologies. Hence, tier 0 and level 1 of those approaches are not covered here. For the same reason, Frank and Couclelis situate social and cognitive aspects toward the top of their frameworks while ODOE assumes that all stages, including observations, are driven by social and cognitive aspects. A similar argument has been made by Scheider (2011) based on the notion of inter-subjectivity.

Finally, a note on terminology is deemed appropriate. Throughout the paper, we will use Big (Geo-)Data for the research vision and Linked Data for the knowledge infrastructure to realize this vision.³ The same is true for e-Science and Linked Science. We will use the term framework for a generic, template-like approach, while methodology will be used for specific realizations of layers within such a framework. To stay in line with the related Semantic Web work, we will use the terms class and concept, as well as role and relation as synonyms. Category, however, will refer to the extension of classes.

The remainder of this article is structured as follows. Section 2 introduces the observation-driven ontology engineering framework in detail. To give an intuition of how the framework can be realized, a specific methodology is highlighted for each level and related to alternative approaches. Examples are given to evaluate each level. To improve readability, instead of a dedicated related

²However, Frank’s work does include a detailed discussion about measurement and observation procedures.

³In doing so, we explicitly focus on the challenge of heterogeneous data integration and leave issues of scalability and storage, such as addressed by MapReduce or Hadoop, aside.

work section, all relevant materials will be discussed at the beginning of each subsection. Next, section 3 revisits the framework layers in reverse order to discuss how provenance is integrated as an essential component of the framework. Finally, conclusions and an outlook to further work are given in section 4.

2 Observation-Driven Ontology Engineering

This section introduces a layered framework for the engineering of geo-ontologies out of observation data. To remain adjustable, the framework does not stipulate specific realizations for its layers and multiple methods can be combined. However, to demonstrate our approach and to make it directly applicable for researchers, we highlight a realization for each layer based on our previous work, show examples, point out how other approaches can be integrated, and relate the layers to existing work on geo-ontology engineering. The presented approach should not be misunderstood as a replacement of these methodologies. For instance, we do not define all of the stages and actors that are provided by Methontology (Fernandez-Lopez et al., 1997) but show how to make geo-ontology engineering ready for the challenges of a data-intensive science.

2.1 Beyond Technology Stacks

The Semantic Web Layer Cake is an evolving and regularly updated, hierarchical technology stack of standards that address different aspects of the Semantic Web; see 1(a). For instance it defines RDF as data-interchange language, SPARQL as one of the query languages, and OWL and its profiles as a knowledge representation language for developing ontologies. *How* to engineer ontologies is out of scope for the layer cake. This question is addressed by ontology engineering methodologies such as Methontology. However, these methodologies have been neither developed with a data-driven semantics nor geo-ontologies in mind.

In this work, we propose a stack that closes this gap; see figure 1(b). The observation-driven ontology engineering layer cake starts with observations acquired from sensors, which also includes derived and aggregated observations. In a next step statistical and numerical methods are used to learn ontological primitives out of these observations. On the third layer, these primitives are used within ontology design patterns that act as strategies for the creation of application ontologies on the fourth level. While the patterns and application ontologies can be aligned with top-level ontologies, the proposed framework focuses on mappings between local ontologies. This follows an established paradigm in Artificial Intelligence research, namely developing a lattice of theories that is consistent at the local level while allowing for contradicting conceptualizations within the global knowledge (McCarthy, 1987; Uschold, 2000a; Wachsmuth, 2000). While researchers and ontology engineers move up the stack layer by layer to create and maintain ontologies, the stack can be queried downwards to access provenance information.

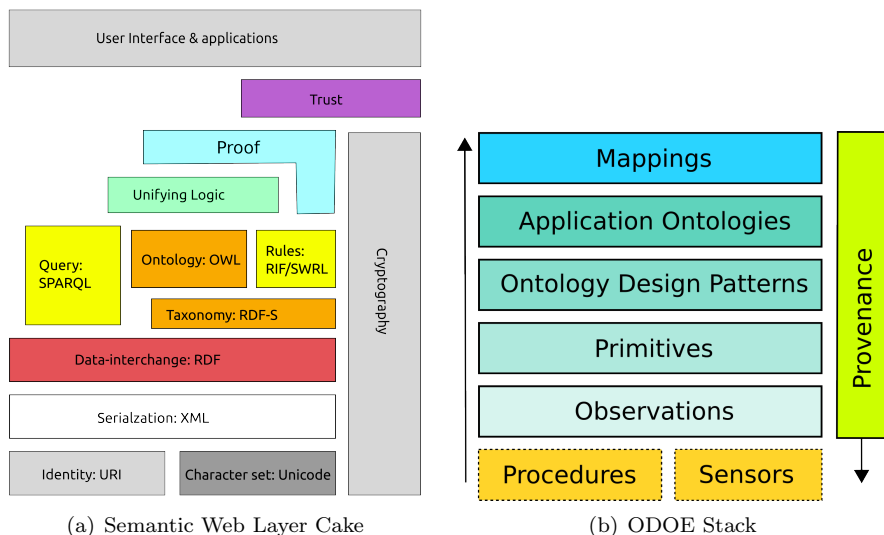


Figure 1: The layers of the Semantic Web and ODOE stacks.

The following subsections discuss the layers and their relationships in detail.

2.2 Sensors, Observation Procedures, and Observations

As physical devices, sensors are out of scope for the ODOE framework. However, they play two important roles. First, they produce the observations that form the first level of the framework. Second, we consider sensors to be the physical realizations of semantic datums (Scheider et al., 2009), and, therefore, sensors and observation procedures provide crucial provenance information; see section 3. As pointed out by Probst and Espeter (2006), observation procedures are required to assign values to observed properties. For example understanding water-level measurements, e.g., as numerical values, requires an understanding of how these measures were performed, e.g., at the deepest point of a waterbody.

The observation data, i.e., the observation result, sampling time and location, targeted feature of interest, sensor used, observation procedure, and so forth, can be stored in any format and infrastructure as long as the raw data is available and directly accessible online. Hence, we propose to either store the data using a Sensor Observation Service (SOS) as part of the Open Geospatial Consortium’s (OGC) Sensor Web Enablement (SWE) initiative, or as Linked Data. OGC services cannot be directly connected to the Linked Data cloud and the Semantic Web. Therefore, we have introduced a transparent Semantic Enablement Layer (SEL) that translates between both infrastructures (Janowicz et al., 2010). A RESTful (Richardson and Ruby, 2007) and transparent proxy that serves Linked Data from a SOS has been implemented as part of the SEL and is available as free and open source software (Janowicz et al., 2011a). Al-

ternative solutions to make observation data available on the Linked Data cloud have been presented by [Page et al. \(2009\)](#) and [Henson et al. \(2009\)](#). Finally, to accommodate geometric approaches, e.g., those based on conceptual spaces ([Gärdenfors, 2000](#); [Raubal, 2004](#)), and to ease the integration of a field-based view ([Fonseca et al., 2002](#)), we have proposed a mapping from regions in conceptual spaces to OWL ontologies that directly follows the ODOE framework; see ([Adams and Janowicz, 2011](#)) for more details.

The Semantic Sensor Network (SSN) ontology and the Stimulus-Sensor-Observation (SSO) ontology design pattern have been proposed by the World Wide Web Consortium’s (W3C) SSN-XG incubator group ([Lefort et al., 2011](#); [Janowicz and Compton, 2010](#)) to model, annotate, and store observation data. The SSO pattern will be discussed in more detail in subsection 2.4, while the SSN ontology was developed on top of it to specify a broad range of sensor characteristics such as the survival range, accuracy, and precision. An overview of alternative ontologies was recently presented by [Compton et al. \(2009\)](#). The ODOE framework is not restricted to a specific type of sensors or observations, humans as sensors and derived observations, e.g., from aggregation, can be used as well.

To give a concrete example, the listing below shows some of the assertions for an observation taken by a smartphone at the French Press café in Santa Barbara, CA where the observed property is the user’s location. Such observations are key to Location-based Social Networks, such as Foursquare or Whrrl, which allow users to check-in at specific places. This information, together with the check-in time and optional tags, is then shared with the social network of the user. A novel methodology to exploit such observations to construct ontological primitives for Points Of Interest (POI) is discussed in subsection 2.3.

```

<SSO:Sensor rdf:ID="Droid2Global_Smartphone">[...]</SSO:Sensor>
<SSO:FeatureOfInterest rdf:ID="FrenchPress">
  <SSO:hasProperty rdf:resource="#Location"/>[...]
</SSO:FeatureOfInterest>
<SSO:Procedure rdf:ID="GPS_Procedure"/>
<SSO:ObservedProperty rdf:ID="Location"/>
<SSO:Observation rdf:ID="Observation_1">
  <DUL:includesObject rdf:resource="#Droid2G"/>
  <DUL:satisfies rdf:resource="#GPS_Procedure"/>[...]
</SSO:Observation>
[...]
```

The observation layer is best compared to Frank’s ontology tier 1 and 2. It is more difficult to map to Couclelis’ levels which have a strong focus on the construction of individuals. Her observation level 2 is clearly separated from the simple and composite objects level 4 and 5. The listing above shows the difference to the ODOE framework by referring to an optional feature of interest, i.e., in ODOE observations can refer to objects. Finally, Frank’s and Couclelis’ approaches do not account for sensors metadata; see section 3 for details.

2.3 Ontological Primitives

Primitives, also called base symbols, are those classes that are not defined within an ontology and represent the borders of formalization. The choice of primitives depends on the scope of the ontology; classes that may be defined in one ontology may be primitives in others. For instance, an ontology used in Human Geography research may declare the class *Human* as primitive, while an ontology from the medical domain will define humans via parts of their bodies, chemical processes, and so forth.

The ODOE framework proposes to semi-automatically derive ontological primitives from the observation layer using methodologies such as machine learning, data mining, or geo-statistics. This bottom-up approach may seem like an obvious step underlying all ontology engineering methodologies. However, in fact, it is a radical change in the way how ontologies are developed in practice. It restricts potential primitives to those that can either be directly observed, i.e., those for which we have sensors, or those that can be derived from observation data, e.g., by aggregation. The resulting primitives and classes defined based on them may be very different from those derived top-down. For instance, primitives defining hills and mountains may be derived from digital elevation data (Sinha and Mark, 2010b), while rivers and lakes may be distinguished by observable characteristics such as length to width ratios (Bennett et al., 2008; Santos et al., 2005). While the ODOE framework does not stipulate a specific method, we introduce a novel approach that we call *semantic signatures* to illustrate this level.

One common way to understand the distribution of geographic features on the surface of the Earth is remote sensing. Different types of land use/cover produce characteristic reflection and absorption patterns in different wavelengths, called bands, of electromagnetic energy. In many cases, studying the reflection values within a particular band, such as the visible light, is sufficient to distinguish between land use types such as hydrographic features versus soil. In other cases, the reflection patterns become only distinguishable in the near infrared or thermal band and, hence, multiple bands have to be sensed. A typical example are conifers and deciduous trees which cannot be distinguished by comparing the visible band alone; see figure 2. These unique patterns are referred to as *spectral signatures*. The idea of signatures has also been successfully applied to other domains such as Internet security, chemistry, and astrophysics.

This does not mean that ontologies are not necessary to describe land use or other geographic feature types but that ontologies are about communication and retrieval and not about replacing numerical models (Sheth et al., 2005). The spectral signatures have to be named, e.g., *Developed, Low Intensity* (Anderson et al., 1976; Homer et al., 2004), and the potential interpretations of their plain text definitions have to be restricted and made machine-readable using ontologies to support data retrieval and integration. For example, the class *Developed, Low Intensity* is vaguely defined as 'Includes areas with a mixture of constructed materials and vegetation. Impervious surfaces account for 20-49 percent of total cover. These areas most commonly include single-family hous-

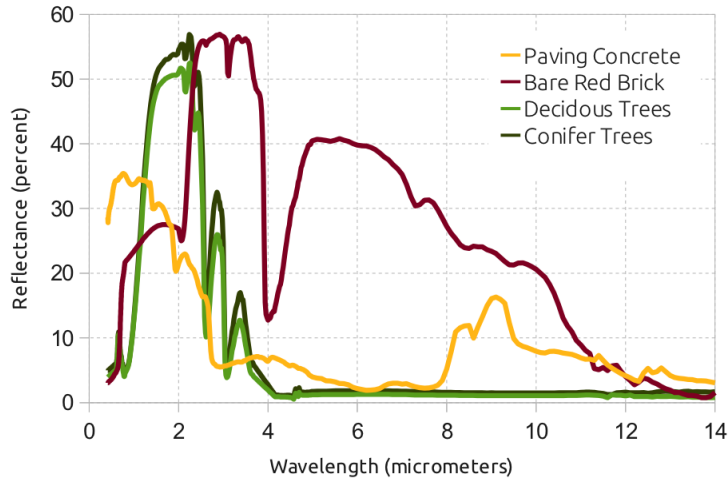


Figure 2: Selected spectral signatures from the NASA ASTER Spectral Library.

ing units'.⁴ Similarly, the classification method used and the number of classes are arbitrary to a certain degree and have to be documented and justified. Recently, Gahegan et al. (2011) discussed the political and economic dimension of choosing specific land use/cover classes in context of the Kyoto protocol.

Moreover, raw spectral signatures cannot be combined in a way to answer queries about *man-made land use* types as this does not correspond to single reflection patterns but to a set of land use/cover types connected by a common creation history and function; see also (Barsalou et al., 2005). Previous work, however, shows that a combination of ontologies, observation data, and rule engines can be used to handle such queries (Henson et al., 2009; Keßler et al., 2009; Devaraju and Kauppinen, 2012; in press). Another relevant example are rivers which cannot be distinguished based on remote sensing data only. The reflection values do not contain information about shape, water flow, and topology, nor can they unveil thematic roles (Sowa, 2000) such as *navigability*.

To bridge the gap between numerical and statistical models and ontologies, we propose to introduce the notion of *semantic signatures* as an analogy to spectral signatures. For instance, POI type, such as *Restaurant* and *University*, can be discriminated from others using signatures. Similarly, as a single band may not be sufficient to establish a unique spectral signature, several semantic characteristics may be required to create a semantic signature. While there are multiple methods to compute and learn such signatures, we have shown how to mine them from massive observation data created by smartphones for Location-based Social Networks. We demonstrated how semantics-enabled geo-statistics (Ahlqvist and Shortridge, 2006) can be used to compute a spatial (Mülligann et al., 2011) and several temporal bands (Ye et al., 2011b,a). Ontological primi-

⁴http://www.mrlc.gov/nlcd_definitions.php.

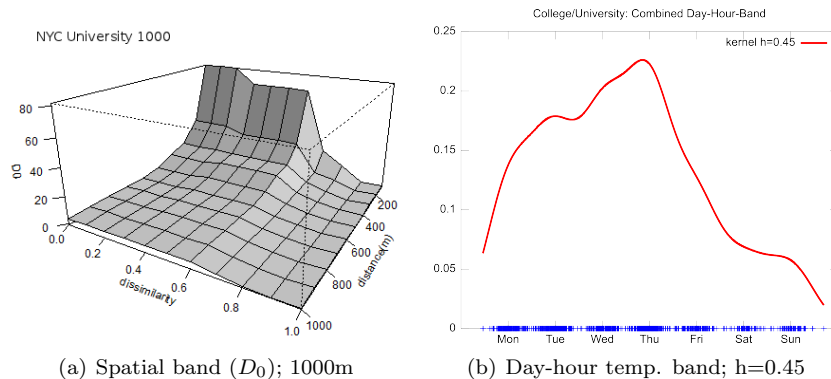


Figure 3: Spatial and temporal bands for the POI type *University*.

tives can be reified out of these bands. For instance, following figure 3, universities are spatially clustered with buildings of similar types (e.g., other university buildings or libraries) and mostly visited during weekdays and working hours. In contrast, bars show different clustering patterns and are mostly visited during evenings and on weekends. Hence, while the primitives *clustered*, *weekday*, and *working_hours* can be used to describe universities, other POI types show different characteristics. While most of the bands that we have studied so far are robust, e.g., the spatial bands for university buildings from Münster, Germany, London, UK, and New York City, US are very similar, bands are local. For instance, they may differ between cultures. This should not be misunderstood as drawback but makes semantic heterogeneities explicit. Two datasets that refer to the same class names, but have different signatures should only be combined with care. So far, we have demonstrated how these bands can be used for data cleaning, recommendation systems, and automatic semantic annotation (Mülligann et al., 2011; Ye et al., 2011b).

Summing up, the ontological primitives extracted out of these bands allow to define geographic feature types in a novel way. Instead of describing POI types by walls, tables, or menus, they can be specified as clustered or uniformly distributed, weekend or evening locations, and so forth. The classical definition of geographic information consisting of a spatial, temporal, and thematic component should not lead to the conclusion that those components are separated. The thematic component may as well be described by spatial and thematic characteristics.

Similarly, and to demonstrate that the proposed methodology is not restricted to POI types, we are studying (Adams and Janowicz, 2012; forthcoming) how a thematic band for geographic feature types can be learned out of massive amounts of unstructured text, e.g., from Wikipedia and travelblogs, using Latent Dirichlet Allocation (Blei et al., 2003). Figure 4, for instance, shows first results for the *Mountain* type. Different characteristics of the studied observation data can be used to create signatures and, therefore, this choice

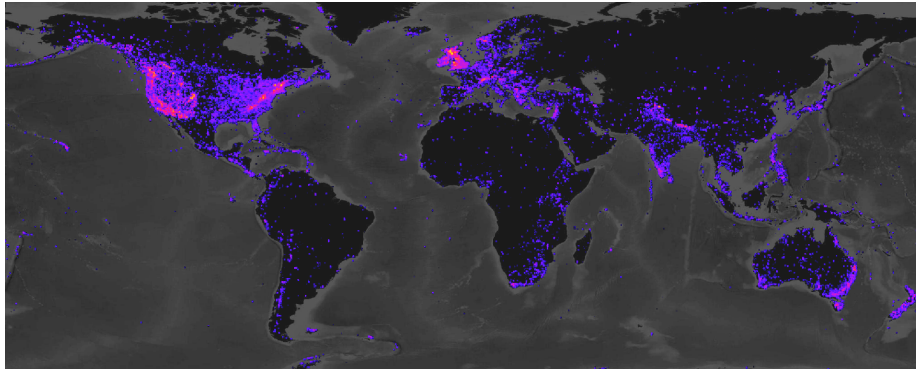


Figure 4: A probability surface visualizing a thematic band for the *Mountain* type learned from unstructured text by LDA and Kernel Density Estimations.

has to be documented as provenance information, see section 3. Combining the different bands to even more discriminative semantic signatures is left for future work.

Semantic signatures are not the only approach to construct primitives out of observation data. Alternatively, one could also exploit the structure underlying existing databases, transform their data to a machine-understandable language such as RDF, and discover potential links to external sources. Several techniques have been developed to support this so-called triplification process, examples include Triplify (Auer et al., 2009) for publishing and the SILK framework (Volz et al., 2009) for link discovery. Finally, scientific workflows and procedures (Gil et al., 2011) offer an additional source that could be studied to derive ontological primitives.

2.4 Geo-Ontology Design Patterns

Ontology Design Patterns (ODP) have been first proposed as an analogy to the successful software engineering design patterns (Gamma et al., 1995) by Gangemi (2005). Several types of patterns have been discussed in the literature, e.g., logical patterns, content patterns, or alignment patterns.⁵ Logical patterns can be thought of as workarounds to solve problems that are a consequence of the formal semantics of the knowledge representation language used. Therefore, logical patterns are independent of specific domains or application areas. In contrast, content patterns are building blocks or strategies that model repeating domain facts, e.g., how to assign places to entities, or how to model the movement of objects along a path. ODPs can vary in size and complexity and can also be composed out of other patterns. However, in general, ontology design patterns should be self-contained and directly applicable to increase their

⁵A collection of ontology design patterns can be found at <http://ontologydesignpatterns.org>.

reusability. As a consequence, ontology design patterns are developed by minimizing the amount of ontological commitments. For instance, the *Place* pattern⁶ does not specify types of places nor does it restrict the introduced *isLocationOf* relation to geographic space. Surprisingly, this reduction of ontological commitments and not the coverage is the key difference between ontology design patterns and top-level ontologies. Over the last years, the NEON project has transformed many parts of the DOLCE Ultra Light foundational ontology into a collection of patterns. Hence, while different content patterns can be combined to a common alignment level to better integrating local ontologies, using a single pattern does not force ontology users and engineers to understand and agree with the overwhelming set of abstract ontological commitments introduced by top-level ontologies.

Ontology design patterns can be developed following classical ontology engineering methodologies such as Methontology; but what are potential candidates for *geo*-ontology design patterns? While the pattern level cannot be matched to any of Couclelis' levels, Frank proposes to use so-called image schemata (Johnson, 1987) to ground semantics in physical operations. Over the years, some of those image schemata have been formalized, e.g., by Kuhn (2007). While the ODOE framework uses a different grounding approach, we believe that image schemata can be thought of as strategies for reoccurring problems as well. For instance, the *Path* image schema may form the basis of a transportation pattern. The idea of *geo*-ontology design patterns is new and, therefore, only a few patterns have been proposed so far, e.g., Ortmann and Daniel (2011) implemented patterns for referential qualities, while Janowicz and Compton (2010) have developed the Stimulus-Sensor-Observation pattern (SSO).⁷ To give a concrete example, and because the ODOE framework makes use of the SSO pattern to establish data provenance, the last-mentioned will be discussed in more detail.

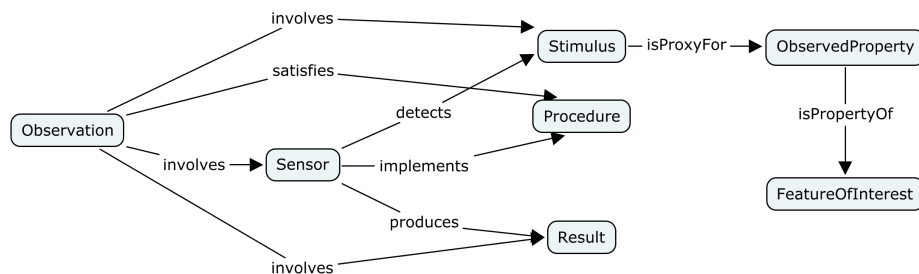


Figure 5: Concept map depicting the Stimulus-Sensor-Observation ontology design pattern without alignment to a top-level ontology (Janowicz and Compton, 2010).

The SSO pattern emerged from the need to reach an agreement between 24

⁶<http://ontologydesignpatterns.org/wiki/Submissions:Place>

⁷Geo-patterns for Points Of Interest, path & movement, as well as for events and change have been proposed recently. The Simple Event Model (SEM) could also be considered as a more complex pattern (van Hage et al., 2011).

domain experts and ontology engineers from 15 different organizations in how to develop a Semantic Sensor Network ontology (SSN) within the W3C SSN-XG incubator group. The SSO pattern was proposed as a minimal and reusable OWL ontology to form the core of the more complex SSN ontology and at the same time serve the needs for light-weight ontologies for Linked Data. As depicted in figure 5, the SSO pattern introduces 7 classes and relations between them. It defines observations as a nexus that connects sensors, the stimuli they receive and the results they produce in transforming the stimuli. It also states that observations and sensors follow procedures that define how and what is measured. For instance, a thermometer can be used to measure surface air temperature or soil temperature. Therefore, an observation procedure has to specify how the measurement should be performed (e.g., 2m above the ground) to make the results reproducible and comparable. Additionally, sensors for the same observable property can be constructed in different ways, e.g., based on expansion versus resistance. Therefore, sensors implement certain procedures which are required to make their results comparable. The pattern does not include any ontological commitments on the nature of the proposed classes. For instance, it does not define stimuli as events or restrict sensors to physical objects. While this makes the SSO pattern flexible, reusable, and easy to extend, it puts more burden on semantic interoperability. Two datasets can refer to the same classes and still have radically different conceptualizations in mind. To address these needs, the pattern was published together with an optional alignment to the foundational ontology DOLCE (not shown in figure 5; details are given in (Janowicz and Compton, 2010)). This alignment, for instance, defines the class *Stimulus* to be a subclass of DOLCE’s *Event* class. While this alignment allows to further restrict the potential interpretations of the classes towards their intended meaning, it comes at the cost of accepting a set of abstract ontological commitments (Guarino, 1998). For example, the pattern alignment proposes to restrict sensors to physical objects and defines observations as situations and, hence, social objects. While accepted for the SSN ontology, other researchers have objected these alignments and proposed to define observations as events and allow for virtual sensors as well.

For this reason, the alignment of ontology design patterns is optional in the ODOE framework. Patterns are rather viewed as strategies which ease the creation of application ontologies by non-experts. Consequently, geo-patterns should build on the discussed observations and primitives levels.

2.5 Application Ontologies

The classification of ontologies based on their granularity and thematic scope into top-level, domain, task, and application ontologies was first introduced by Guarino (1998). Top-level ontologies, such as DOLCE (Masolo et al., 2002), formalize generic knowledge about the world in an application and domain independent way. For instance, they introduce the distinction between endurants, e.g., physical objects, and perdurants, e.g., events. Ontologies from lower levels should align their conceptualization to those provided by the top-level to im-

prove interoperability. A transportation domain ontology may define the class *TrafficJam* as a subconcept of *Perdurant*. Domain ontologies specify the concepts related to generic domains such as geography. In contrast, task ontologies focus on generic and reoccurring activities such as travel and transportation. Finally, Guarino defines application ontologies as specializations of domain and task ontologies. For instance, an ontology-driven trip planning system may refine the vocabularies coming from both the travel and the geography ontologies. Nevertheless, the boundaries between the types of ontologies are rather fluid.

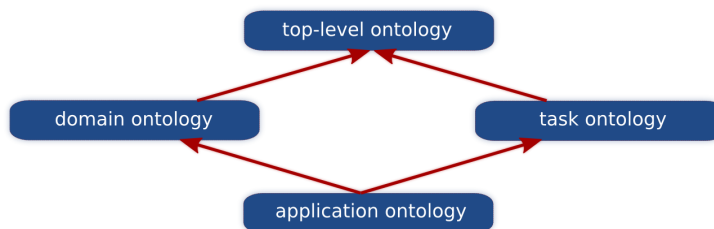


Figure 6: Types types of ontologies and their specialization relationships (based on (Guarino, 1998)).

Top-down distinctions, e.g., the dichotomy between endurants and perdurants, play a significant role in research on geospatial semantics but at the same time are a major source of limitations. For example, lenticular clouds can be categorized as events or physical objects at the same time (Davidson, 1985; Galton, 2004). More recently, Sinha and Mark (2010a) demonstrated that geographic feature types such as *Hill* can be described as a *Physical Object*, a *Feature*, or an *Amount of Matter*. These classes are among the core distinctions proposed by DOLCE for *Physical Endurants*. In other words, some geographic feature types are multi-aspect phenomena (Galton, 2004) to a degree where even top-level distinctions cannot be applied without reference to context. In practice, trying to make a group of users, domain experts, or ontology engineers agree on abstract concepts turns out to be the key obstacle in reaching agreement.

In reaction, Uschold (2000b) proposed a more flexible and pragmatic classification into local and global ontologies. The assumption that global ontologies are, in fact, top-level ontologies while local ontologies are application ontologies is misleading as both approaches focus on different aspects. A global ontology can be on the domain level, but also on the level of a single company. All that matters for characterizing global ontologies is that they are the common and shared agreement of multiple units or working groups. If all departments of an international company agree on a vocabulary this may be regarded as a global ontology for this company. In contrast, local ontologies have to be aligned to other local ontologies or to a global ontology (while global ontologies do not require connections to further global ontologies as they are the primary source of reference). Similar to Guarino’s classification, the distinction between global and local ontologies is rather fluid.

Infrastructures such as the envisioned Digital Earth have to query over multiple, highly heterogeneous ontologies. Hence, the ODOE framework proposes to develop local, application-centric ontologies which arise from clear needs and are a means to an end rather than ends in themselves. While these ontologies can be developed following established methodologies such as Methontology, a strict distinction between the roles of a user, domain expert, and ontology engineer is no longer necessary. In fact, scientists should be able to develop ontologies that document their datasets on their own.

2.6 Mappings

The ODOE framework proposes to formalize local ontologies arising from specific needs rather than a small set of top-down engineered and authoritative ontologies. Consequently, the definitions of the same terms may differ or even become incompatible. Strictly speaking, and to demonstrate the differences between both paradigms, a question such as *'what defines a river?'* cannot be answered in a global knowledge infrastructure. Instead, one has to either ask *'What defines a river in Southern Europe?'* or *'what are the common characteristics of European rivers?'*. Therefore, ontology alignment and semantic translation become essential components of modern knowledge infrastructures (Cruz and Sunna, 2008; Shvaiko and Euzenat, 2008; Schade, 2010). This also includes work on ontology modularization and multi-perspective ontologies (Kokla and Kavouras, 2001; Bateman et al., 2007). One may argue that it will not be possible to align all classes from different local ontologies. This, however, should not be misunderstood as weakness of a mappings-based approach, to the contrary, it prevents data that was created under different assumptions to be combined. In the following, to demonstrate the feasibility of the mapping level, we briefly introduce a methodology to structure local ontologies. While the presented approach makes use of ontology alignment, it also allows to semi-automatically derive a global ontology bottom-up, e.g., out of the provenance data discussed in section 3.

Integrating local views at a global level has been a core research theme for AI since decades (McCarthy, 1987). While each of these views, also called contexts, has to be a consistent theory, a global knowledge infrastructure does not need to be contradiction free. Instead, the local views can be organized within an interconnected network. This approach has been first proposed and implemented for the CYC knowledge base (Lenat, 1995). Similar principles, called domains, are also underlying Freebase. CYC/OpenCYC structures knowledge into domain specific *microtheories*. Each of these theories consists of a coherent set of statements about classes and individuals and can store facts that are incompatible with other microtheories. The theories are interconnected using a generalization relation *genMt* (McCarthy and Buvac, 1996), i.e., facts defined within a super-microtheory must also hold for each of its sub-theories, while sibling-theories may store contradicting facts. For an *is-true-in* relation *ist(mt, p)* between a microtheory *mt* and a predicate *p*, *genMt* is defined as the anti-symmetric, reflexive, and transitive binary predicate through which the hierarchy of theories is constructed by adding axioms of the form defined in equation 1 to the root

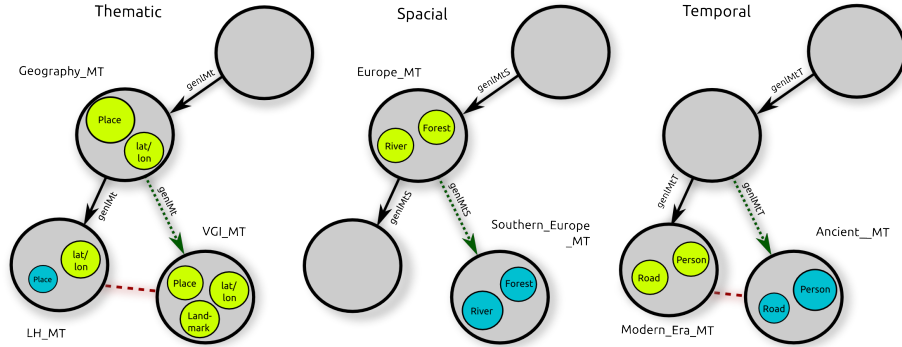


Figure 7: Lattices of microtheories structured using thematic generalization ($genMt$), generalization and spatial/administrative inclusion ($genMtS$), as well as generalization and temporal inclusion ($genMtT$).

theory mt_0 ; where mt_g is the more general and mt_s the more specific theory (Guha et al., 2004).

$$mt_0 : \forall p \text{ ist}(mt_g, p) \wedge \text{genMt}(mt_g, mt_s) \longrightarrow \text{ist}(mt_s, p) \quad (1)$$

As depicted in figure 7, a broad Geography microtheory may introduce a generic notion of *Place*, while a microtheory developed for VGI may import from this theory and refine the class *Place* by restricting it to those places for which a location on the Earth’s surface can be defined. Moreover, the VGI_MT may introduce new classes, such as a *Landmark* class for pedestrian navigation. A third microtheory developed for location hypotheses may built upon Geography_MT, but only cover places for which the location and their existence is unknown. While the VGI_MT and LH_MT are generalized by the Geography_MT (shown by the dotted lines) and all their facts are not allowed to contradict with the super-theory, the sibling theories (shown by the dashed line) contain contradicting facts.

In previous work (Duce and Janowicz, 2010), we have extended the notion of a lattice of theories by adding space and time as fundamental ordering principles. Tobler’s First Law of Geography (Tobler, 1970) is not restricted to individuals; different conceptualizations of feature types do not occur randomly but usually follow patterns that gradually change with increasing spatial and temporal distance. To give a concrete example, the meaning of *River* changes gradually from boreal climates, over maritime climates, to semiarid climates. Hence, as a consequence of local geographies, Northern European countries may have different legal river definitions compared to those in southern Europe; see figure 7. Similarly, and to give an example in which temporal aspects drive the diversification of classes, cultural heritage researchers have to handle biased, incomplete, and contradicting information. Therefore, an Ancient Times microtheory may allow for unnatural life spans of historic persons and also allow them to be at

different location at the same time. A Modern Era microtheory may exclude such cases and also differ in definitions of transportation infrastructure or settlements. Applying a modern definition of *City* to ancient settlements may be inappropriate.

We have demonstrated how this approach can account for local diversity while maintaining global interoperability for the European INSPIRE⁸ initiative (Duce and Janowicz, 2010). By creating a lattice of microtheories we ensured that feature types defined by member states that are administratively contained within the European Union must be sub-classes of an EU-wide definition. However, instead of imposing a common schema for all European member states top-down, we used local conceptualizations together with Least Common Subsumer (LCS) (Cohen et al., 1992) and similarity reasoning (Janowicz et al., 2011b) to semi-automatically infer an appropriate top-level that does not violate local definitions. While LCS reasoning is used to infer a top-level via the commonalities of the local definitions, similarity reasoning was used to measure and adjust how much information gets lost to reach this common top-level; see (Duce and Janowicz, 2010) for details.

Compared to the other frameworks, while Frank’s work explicitly acknowledges the role of context in the social reality tier (3), it does not propose how to define semi-automatic mappings.

3 Provenance

While we have discussed the steps involved in moving up the layers, the reverse way is of equal importance. In fact, data provenance (Sahoo et al., 2008) is one of the core motivations for creating the ODOE framework. Each transition between layers involves certain arbitrariness. For instance, there are multiple data mining and learning methods to classify data and, hence, to construct primitives. The geo-ontology design patterns can be combined in different ways to construct application ontologies, and the choice of an alignment methodology influences the relationship between local ontologies. Consequently, provenance information is required to make these decisions explicit and accessible (Hartig, 2009). When using scientific data, researchers should be able to deconstruct categorical information down to the sources, i.e., observations, it was derived from. While a detailed introduction to provenance research for the Semantic Web and e-Science is out of scope for this work, the following subsections give an intuition and examples of how provenance information can be collected and encoded on the different levels of the framework.

3.1 Provenance on the Mapping Level

While the formal characterization of different ontology alignment and matching operations has been an active research area, leveraging this information for the purpose of provenance has received less attention; an exception being

⁸The Infrastructure for Spatial Information in the European Community

the work of Zimmermann et al. (2006). So far, formal approaches, e.g., based on category-theory, have been proposed to understand and compare different methodologies. This line of research is especially important for benchmarking different approaches as well as for reasoning why they produce different results. For the purpose of the ODOE framework, the documentation aspect is more relevant. If scientific data is annotated by combining different ontologies, users have to be aware that another alignment methodology may have constructed slightly different mappings. While a formal characterization is the first step, the question of how to integrate the alignment provenance, e.g., matching probabilities, with provenance ontologies (Moreau et al., 2011; Hartig and Zhao, 2011) is an open research question.

To give a concrete example, a local ontology o_1 containing the classes *River* and *Creek* may have to be aligned with another ontology o_2 that only contains the class *Stream*. Even if we restrict this example to subsumption and equivalence matches and leave cases such as *skos:closeMatch* (Miles and Bechhofer, 2008) aside, several alignments are possible. Let us further assume that the classes and roles, e.g., *Watercourse* and *hasOrigin*, used for the definition of these three classes have been successfully matched in o_s . A purely TBox-driven alignment method may, based on the statements 2, 3, and 4, classify *Stream* as the superclass of *River* and *Creek*. In contrast, an alignment method that also takes ABox knowledge into account may produce a different result. For instance, if most streams in o_2 are classified as belonging to a transportation network, and this relation can be successfully matched to the navigability class in o_1 . In such a case, *River* may still be a subclass of *Stream*, while *Creek* would be a sibling class.

$$o_1 : River \sqsubseteq o_s : Watercourse \sqcap \exists o_s : hasOrigin(o_s : Spring) \sqcap o_s : Navigable \sqcap \dots \quad (2)$$

$$o_1 : Creek \sqsubseteq o_s : Watercourse \sqcap \exists o_s : hasOrigin(\top) \sqcap \neg o_s : Navigable \sqcap \dots \quad (3)$$

$$o_2 : Stream \sqsubseteq o_s : Watercourse \sqcap \exists o_s : hasOrigin(\top) \sqcap \dots \quad (4)$$

3.2 Provenance on the Ontology Level

Application ontologies reflect the local conceptualizations underlying categorical data. Hence, they can be regarded as provenance information in their own rights. The ontology design patterns used to construct these ontologies are imported and, therefore, also directly available as additional sources of provenance. However, metadata on how the application ontologies were constructed, their degree of formalization, used language, authors, change logs, and also the applied ontology engineering methodology are still relevant. The need for making this information explicit and accessible has been recognized within the last years which resulted in the development of the Ontology Metadata Vocabulary (OMV) (Hartmann et al., 2005).

To give a concrete example, the following listing shows metadata created using the OMV for a POI ontology. Note that resources such as *ApplicationOntology*, *OWL-DL*, or *Vocabulary* are provided by OMV, while others, such as *VGI*, have to be defined. The metadata also shows that the ontology imports a design pattern for places.

```
<Ontology rdf:ID="OSMPOI">
  <acronym rdf:datatype="&xsd:string">OSM-POI</acronym>
  <creationDate rdf:datatype="&xsd:string">12-13-2011</creationDate>
  <description rdf:datatype="&xsd:string"
    >Ontology for OpenStreetMap Points of Interest</description>
  <hasCreator rdf:resource="#KJano"/>
  <hasDomain rdf:resource="#VGI"/>
  <useImports rdf:resource="#GODP_Place"/>
  <hasFormalityLevel rdf:resource="#Vocabulary"/>
  <hasLicense rdf:resource="#GeneralPublicLicense_GPL"/>
  <hasOntologyLanguage rdf:resource="#OWL-DL"/>
  <hasOntologySyntax rdf:resource="#OWL-XML"/>
  <isOfType rdf:resource="#ApplicationOntology"/>
  <name xml:lang="en">OpenStreetMap POI Ontology</name>
  <version rdf:datatype="&xsd:string">1</version>
</Ontology>
```

While the provided metadata can assist researcher in understanding how the ontology was developed, aspects such as the geographic bounding box specified by the ISO 19115 metadata standard are missing so far. The bounding box of an application ontology could, for instance, be defined via the locations of the observations or features of interest instantiated within the ontology or used for the construction of the primitives.

Ontology design patterns are best thought of as strategies, but are ontologies in their own rights. Hence, metadata for them can be provided using the OMV as well. From a provenance perspective, the only notable difference is that the number of patterns is small compared to the number of application ontologies and existing libraries document those patterns in detail.

3.3 Provenance on the Primitives Level

The ODOE framework proposes to use methodologies such as data mining, machine learning, and geo-statistics to construct ontological primitives out of observation data. As primitives are those classes for which no further definitions can be given within the ontology, provenance information can be gathered by deconstructing primitives back to the data they were created from. The idea of sharing the used methods and models online to make scientific results reproducible is not new. It plays a key role in e-Science and the Model Web (Pebesma et al., 2010; Brodaric and Gahegan, 2010). The Mathematical Markup Language (MathML) may provide a starting point, however, the representation of mathematical knowledge on the Semantic Web is still an open research question. An overview on the state of the art and a roadmap for future work have been recently presented by Lange (2012; in press). In terms of provenance, the core

challenge will be to represent the complex workflows used to learn the primitives. This includes more than the representation of the used algorithms, but also the used sampling strategies or parameter settings.

3.4 Provenance on the Observations Level

The ODOE framework proposes to model observations using the SSO ontology design pattern. Each observation can be either stored as a collection of assertions in a triple store or as record in a Sensor Web Enablement service. In the last-mentioned case, semantic annotations can be used to connect the observations to the ontologies. Therefore, while not always necessary, scholars can decompose ontologies down to single observations to understand nominal data and uncover hidden incompatibilities. As argued before, the interpretation of observations may require information on the used procedures. These procedures, however, cannot be defined within ontologies as this would lead to a regress loop and is at the core of the symbol grounding problem (Harnad, 1990). Therefore, observation and construction procedures form the boundary of the ODOE framework. Recently, Scheider (2011) proposed to ground observations in perceptual operations.

4 Conclusions and Outlook

In this work we have introduced a novel framework for the engineering of ontologies out of observation data. Instead of arguing for a small and authoritative set of global ontologies that have to address issues of vagueness and uncertainty due to their context-independent nature, we propose to develop small but crisp ontologies that emerge from application needs and are derived bottom-up. Starting with the formal specification of sensors and their observations, the framework proposes to employ geo-statistics, data mining, and machine learning to construct ontological primitives. These primitives are integrated with geo-ontology design patterns, which, due to their minimal amount of ontological commitments, act as flexible and reusable building blocks for local ontologies, called microtheories. The patterns take the role of classical top-level ontologies in providing a common level to ease ontology alignment. These alignments allows to match and integrate local ontologies within a lattice of theories. In addition, the framework also discusses how to deconstruct ontologies to derive the provenance information necessary to uncover hidden incompatibilities between datasets. Summing up, the presented framework accounts for local semantic diversity without giving up interoperability on the global level. To foster retrieval and reuse, scientists should avoid annotating their data with classes that are loaded in terms of their ontological commitments, and, instead, prefer those classes that can be deconstructed to observations. In this respect, *dynamic typing* as known from software engineering seems more appropriate. For example, whether a certain area with a high density of trees is a forest, should be determined when accessing the data (in the context of a local ontology).

While we argue that the ODOE framework makes ontology engineering ready for the age of data intensive science, a lot of research remains to be done. To demonstrate how the framework can be used, we introduced a specific methodologies as realizations of each of the proposed layers. However, these methodologies, while successfully applied to specific use cases, are still research topics themselves. For example, we have shown that the spatial and temporal bands for the semantic signatures can be successfully applied to tasks such as the automatic semantic annotations of Points of Interest (Ye et al., 2011b). The combination of multiple bands and the extraction of topic bands out of unstructured text using LDA is ongoing. It is also not clear how to develop scale invariant signatures. One reason for the success of spectral signatures are open libraries that allow scholars to reuse and exchange signatures. Therefore, we hope that the development of an open semantic signatures library will boost their usage in the future. As the bands are created out of massive real world data, they may also act as baseline for future semantic accuracy measures. Similarly, while ontology design patterns have been successfully applied in ontology engineering, patterns specific to the needs of geographic information are largely missing and have to be developed. It is important that such geo-patterns are closely tied to the observations and primitives levels. For instance, a Point of Interest pattern may describe places by their boundaries and define those in terms of spatial and temporal constraints using the semantic signatures. Moreover, while the ODOE framework does not stipulate specific methodologies, further work is required to understand which methods can be most successfully combined to meet specific application or domain needs. The final evaluation of an ontology engineering framework can only be performed by using it in different settings over a long period. Provenance is a key component of the proposed framework. While we have introduced the current state of the art, the discussion also shows that more research is necessary.

Finally, data intensive science was one of the key motivations for developing the framework. However, Big Data is not without problems, e.g., with respect to ethical questions or the problem of sampling (Boyd and Crawford, 2011). What are the challenges for a data-intensive Geographic Information Science?

Acknowledgments

The author would like to thank Simon Scheider, Werner Kuhn, Benjamin Adams, Pascal Hitzler, and the anonymous reviewers for their valuable comments and suggestions.

References

- B. Adams and K. Janowicz. Constructing geo-ontologies by reification of observation data. In *19th ACM SIGSPATIAL International Conference on Ad-*

- vances in *Geographic Information Systems (ACM SIGSPATIAL GIS 2011)*, pages 309–318. ACM, 2011.
- B. Adams and K. Janowicz. On the geo-indicativeness of non georeferenced text. In *International AAAI Conference on Weblogs and Social Media (ICWSM 2012)*. AAAI, 2012; forthcoming.
- O. Ahlqvist and A. Shortridge. Characterizing land cover structure with semantic variograms. In A. Riedl, W. Kainz, and G. Elmes, editors, *Progress in Spatial Data Handling -12th International Symposium on Spatial Data Handling*, pages 401–415. Springer, 2006.
- J. R. Anderson, E. E. Hardy, and R. E. Roach, J. T. and Witmer. A land use and land cover classification system for use with remote sensor data. *Geological Survey Professional Paper 964*, 2001:28, 1976.
- S. Auer, S. Dietzold, J. Lehmann, S. Hellmann, and D. Aumueller. Triplify: Light-weight linked data publication from relational databases. In Juan Quemada, Gonzalo León, Yoëlle S. Maarek, and Wolfgang Nejdl, editors, *Proceedings of the 18th International Conference on World Wide Web, WWW 2009, Madrid, Spain, April 20-24, 2009*, pages 621–630. ACM, 2009.
- L. Barsalou. Situated simulation in the human conceptual system. *Language and Cognitive Processes*, 5(6):513–562, 2003.
- L. Barsalou, S.A. Sloman, and S.E. Chaigneau. The HIPE theory of function. In L. Carlson & E. van der Zee, editor, *Representing functional features for language and space: Insights from perception, categorization and development*, pages 131–147. Oxford University Press, 2005.
- J. Bateman, S. Borgo, K. Luettich, C. Masolo, and T. Mossakowski. Ontological modularity and spatial diversity. *Spatial Cognition and Computation*, 7(1): 97–128, May 2007.
- B. Bennett, D. Mallenby, and A. Third. An ontology for grounding vague geographic terms. In *Formal Ontology in Information Systems - Proceedings of the Fifth International Conference (FOIS 2008)*, volume 183, pages 280–293. IOS Press, 2008.
- Y. Bishr. Geospatial semantic web. In Sanjay Rana and Jayant Sharma, editors, *Frontiers of Geographic Information Technology*, pages 139–154. Springer, 2006. ISBN 978-3-540-31305-2.
- C. Bizer, T. Heath, and T. Berners-Lee. Linked data – the story so far. *Journal on Semantic Web and Information Systems (IJSWIS)*, 5(3):1–22, 2009.
- D. M. Blei, A. Y. Ng, and M. I. Jordan. Latent dirichlet allocation. *Journal of Machine Learning Research*, 3:993–1022, March 2003. ISSN 1532-4435.

- K. Bollacker, C. Evans, P. Paritosh, T. Sturge, and J. Taylor. Freebase: a collaboratively created graph database for structuring human knowledge. In *Proceedings of the 2008 ACM SIGMOD international conference on Management of data*, SIGMOD '08, pages 1247–1250. ACM, 2008. ISBN 978-1-60558-102-6.
- D. Boyd and K. Crawford. Six Provocations for Big Data. *Social Science Research Network Working Paper Series*, 2011. URL <http://ssrn.com/abstract=1926431>.
- B. Brodaric and M. Gahegan. Experiments to Examine the Situated Nature of Geoscientific Concepts. *Spatial Cognition and Computation*, 7(1):61–95, 2007.
- B. Brodaric and M. Gahegan. Ontology use for semantic e-science. *Semantic Web*, 1:149–153, April 2010. ISSN 1570-0844.
- W. Cohen, A. Borgida, and H. Hirsh. Computing least common subsumers in description logics. In *10th National Conference on Artificial Intelligence*, pages 754–760. MIT Press, 1992.
- M. Compton, C. Henson, H. Neuhaus, L. Lefort, and A. Sheth. A survey of the semantic specification of sensors. In Taylor. K., A. Ayyagari, and D. De Roure, editors, *Proceedings of the 2nd International Workshop on Semantic Sensor Networks (SSN09)*, volume Vol-522, pages 17–32. CEUR, 2009.
- H. Couclelis. Ontologies of geographic information. *International Journal of Geographical Information Science*, 24:1785–1809, 2010.
- I. Cruz and W. Sunna. Structural alignment methods with applications to geospatial ontologies. *Transactions in GIS, special issue on Semantic Similarity Measurement and Geospatial Applications*, 12(6):683–711, 2008.
- D. Davidson. Reply to quine on events. In E. LePore & B. McLaughlin. *Actions and events: Perspectives on the philosophy of Donald Davidson*, pages 172–176, 1985.
- A. Devaraju and T. Kauppinen. Sensors tell more than they sense: Modeling and reasoning about sensor observations for understanding weather events. *International Journal of Sensors, Wireless Communications and Control*, Special Issue on Semantic Sensor Networks, 2012; in press.
- S. Duce and K. Janowicz. Microtheories for spatial data infrastructures - accounting for diversity of local conceptualizations at a global level. In S. Fabrikant, T. Reichenbacher, M. van Kreveld, and C. Schlieder, editors, *Geographic Information Science, 6th International Conference GIScience 2010*, volume 6292 of *Lecture Notes in Computer Science*, pages 27–41. Springer, 2010.

- M. Fernandez-Lopez, A. Gomez-Perez, and N. Juristo. Methontology: from ontological art towards ontological engineering. In *Proceedings of the AAAI97 Spring Symposium*, pages 33–40, Stanford, USA, 1997.
- F. Fonseca, M. Egenhofer, P. Agouris, and G. Camara. Using ontologies for integrated geographic information systems. *Transactions in GIS*, 6:231–257, 2002.
- A. Frank. Ontology for spatio-temporal databases. In M. Koubarakis, T. Sellis, A. Frank, S. Grumbach, R. Güting, C. Jensen, N. Lorentzos, Y. Manolopoulos, E. Nardelli, B. Pernici, H. Schek, M Scholl, B. Theodoulidis, and N. Tryfona, editors, *Spatio-Temporal Databases: The CHOROCHRONOS Approach*, volume 2520 of *Lecture Notes in Computer Science*, pages 9–77. Springer, 2003.
- M. Gahegan, W. Smart, S. Masoud-Ansari, and B. Whitehead. A semantic web map mediation service: interactive redesign and sharing of map legends. In *Proceedings of the 1st ACM SIGSPATIAL International Workshop on Spatial Semantics and Ontologies*, SSO '11, pages 1–8. ACM, 2011. ISBN 978-1-4503-1039-0.
- A. Galton. Fields and objects in space, time, and space-time. *Spatial Cognition and Computation*, 1:39–68, 2004.
- E. Gamma, R. Helm, R. Johnson, and J. Vlissides. *Design Patterns: Elements of Resusable Object-Oriented Software*. Addison-Wesley Professional, 1995.
- A. Gangemi. Ontology design patterns for semantic web content. In Y. Gil, E. Motta, R. Benjamins, and M. Musen, editors, *The Semantic Web - ISWC 2005, 4th International Semantic Web Conference*, volume 3729 of *Lecture Notes in Computer Science*, pages 262–276. Springer, 2005.
- P. Gärdenfors. *Conceptual Spaces: The Geometry of Thought*. A Bradford Book. MIT Press, 2000.
- Y. Gil, V. Ratnakar, J. Kim, P. Gonzalez-Calero, P. Groth, J. Moody, and E. Deelman. Wings: Intelligent workflow-based design of computational experiments. *IEEE Intelligent Systems*, 26(1), 2011.
- M. Goodchild, M. Yuan, and T. Cova. Towards a general theory of geographic representation in gis. *International Journal of Geographical Information Science*, 21(3):239–260, 2007.
- N. Guarino. Formal Ontology and Information Systems. In N. Guarino, editor, *International Conference on Formal Ontology in Information Systems (FOIS1998)*, pages 3–15. IOS Press, Trento, Italy, 1998.
- R. Guha, R. Mccool, and R. Fikes. Contexts for the semantic web. In *International Semantic Web Conference (ISWC 2004)*, number 3298 in *Lecture Notes in Computer Science*, pages 32–46. Springer, 2004.

- S. Harnad. The symbol grounding problem. *Phys. D*, 42(1–3):335–346, 1990.
- O. Hartig. *Provenance Information in the Web of Data*, pages 1–9. CEUR-WS, 2009. URL <http://www.dbis.informatik.hu-berlin.de/fileadmin/research/papers/conferences/2009-ldow-hartig.pdf>.
- O. Hartig and J. Zhao. Provenance vocabulary core ontology specification). Technical Report 6, 2011. URL <http://purl.org/net/provenance/ns>.
- J. Hartmann, P. Palma, Y. Sure, M. Suárez-Figueroa, P. Haase, A. Gómez-Pérez, and R. Studer. Ontology metadata vocabulary and applications. In *OTM Workshops; On the Move to Meaningful Internet Systems 2005: OTM 2005 Workshops, OTM Confederated International Workshops and Posters, AWeSOMe, CAMS, GADA, MIOS+INTEROP, ORM, PhDS, SeBGIS, SWWS, and WOSE 2005*, volume 3762 of *Lecture Notes in Computer Science*, pages 906–915. Springer, 2005.
- C. A. Henson, J. K. Pschorr, A. P. Sheth, and K. Thirunarayan. SemSOS: Semantic sensor observation service. In *International Symposium on Collaborative Technologies and Systems (CTS 2009)*, 2009.
- C. Homer, C. Huang, L. Yang, B. Wylie, and M. Coan. Development of a 2001 national landcover database for the united states. *Photogrammetric Engineering and Remote Sensing*, 70(7):829–840, 2004.
- P. Jain, P. Hitzler, P. Yeh, K. Verma, and A. Sheth. Linked Data Is Merely More Data. In *2010 AAAI Spring Symposium Series*. Association for the Advancement of Artificial Intelligence, 2010.
- K. Janowicz. The role of space and time for knowledge organization on the semantic web. *Semantic Web*, 1(1-2):25–32, 2010.
- K. Janowicz and M. Compton. The stimulus-sensor-observation ontology design pattern and its integration into the semantic sensor network ontology. In Arun Ayyagari, David De Roure, and Kerry Taylor, editors, *Proceedings of the 3rd International workshop on Semantic Sensor Networks 2010 (SSN10) in conjunction with the 9th International Semantic Web Conference (ISWC 2010)*, volume 668. CEUR-WS, 2010.
- K. Janowicz, S. Schade, A. Bröring, C. Keßler, P. Maue, and C. Stasch. Semantic enablement for spatial data infrastructures. *Transactions in GIS*, 14(2):111–129, 2010.
- K. Janowicz, A. Bröring, C. Stasch, S. Schade, T. Everding, and A. Llaves. A restful proxy and data model for linked sensor data. *International Journal of Digital Earth*, pages 1–22, 2011a.
- K. Janowicz, M. Raubal, and W. Kuhn. The semantics of similarity in geographic information retrieval. *Journal of Spatial Information Science*, 2: 29–57, 2011b.

- M. Johnson. *The Body in the Mind: The Bodily Basis of Meaning, Imagination, and Reason*. Univ. of Chicago Press, 1987.
- M. Jones, M. Schildhauer, O. Reichman, and S Bowers. The new bioinformatics: Integrating ecological data from the gene to the biosphere. *Annual Review of Ecology, Evolution, and Systematics*, 37(1):519–544, 2006.
- C. Keßler, M. Raubal, and C. Wosniok. Semantic rules for context-aware geographical information retrieval. In P. Barnaghi, editor, *European Conference on Smart Sensing and Context, EuroSSC 2009*, volume 5741 of *LNCS*, pages 77–92. Springer, 2009.
- M. Kokla and M. Kavouras. Fusion of top-level and geographic domain ontologies based on context formation and complementarity. *International Journal of Geographical Information Science*, 15:679–687, 2001.
- W. Kuhn. An image-schematic account of spatial categories. In S. Winter, M. Duckham, L. Kulik, and B. Kuipers, editors, *Spatial Information Theory*, volume 4736 of *Lecture Notes in Computer Science*, pages 152–168. Springer, 2007.
- W. Kuhn. Semantic Engineering. In G. Navratil, editor, *Research Trends in Geographic Information Science*, pages 63–76. Springer, 2009.
- C. Lange. Ontologies and languages for representing mathematical knowledge on the semantic web. *Semantic Web*, 3, 2012; in press.
- L. Lefort, C. Henson, K. Taylor, P. Barnaghi, M. Compton, O. Corcho, R. Castro, J. Graybeal, A. Herzog, K. Janowicz, H. Neuhaus, A. Nikolov, and K. Page. Semantic Sensor Network XG Final Report. W3C Incubator Group Report 28 June 2011. Available at: <http://www.w3.org/2005/Incubator/ssn/XGR-ssn/>. Technical report, 2011.
- S. Lehar. *The World in Your Head: A Gestalt View of the Mechanism of Conscious Experience*. Lawrence Erlbaum, 2003.
- D. B. Lenat. Cyc: a large-scale investment in knowledge infrastructure. *Communications of the ACM*, 38:33–38, November 1995. ISSN 0001-0782.
- H. Gyde Lund. Definitions of forest, deforestation, afforestation, and reforestation. Technical report, Forest Information Services. Available from the World Wide Web: <http://home.comcast.net/~gyde/DEFpaper.htm>, 2011.
- D. Mark. Toward a theoretical framework for geographic entity types. In *Spatial Information Theory: A Theoretical Basis for GIS, International Conference COSIT '93*, pages 270–283, 1993.
- C. Masolo, S. Borgo, A. Gangemi, N. Guarino, A. Oltramari, and L. Schneider. WonderWeb deliverable D17. the WonderWeb library of foundational ontologies and the DOLCE ontology. Technical report, ISTC-CNR, 2002.

- J. McCarthy. Generality in artificial intelligence. *Communications of the ACM*, 30(12):1030–1035, 1987.
- J. McCarthy and S. Buvac. Formalizing context (expanded notes), 1996.
- A. Miles and S. Bechhofer. Skos simple knowledge organization system reference, 2008. URL <http://www.w3.org/TR/skos-reference/>.
- L. Moreau, B. Clifford, J. Freire, J. Futrelle, Y. Gil, P. Groth, N. Kwasnikowska, S. Miles, P. Missier, J. Myers, B. Plale, Y. Simmhan, E. Stephan, and J. Van den Bussche. The open provenance model core specification (v1.1). *Future Generation Computer Systems*, 27(6):743 – 756, 2011.
- C. Mülligann, K. Janowicz, Mao Ye, and W.-C. Lee. Analyzing the spatial-semantic interaction of points of interest in volunteered geographic information. In *Int. Conference on Spatial Information Theory*, volume LNCS 5899, pages 350–370, 2011.
- P. Nowakowski, E. Ciepiela, D. Harezlak, J. Kocot, M. Kasztelnik, T. Bartynski, J. Meizner, G. Dyk, and M. Malawski. The collage authoring environment. *Procedia Computer Science*, 4(0):608 – 617, 2011. ISSN 1877– 0509. Proceedings of the International Conference on Computational Science, ICCS 2011.
- J. Ortmann and D. Daniel. An ontology design pattern for referential qualities. In Lora Aroyo, Chris Welty, Harith Alani, Jamie Taylor, Abraham Bernstein, Lalana Kagal, Natasha Noy, and Eva Blomqvist, editors, *The Semantic Web – ISWC 2011*, volume 7031 of *Lecture Notes in Computer Science*, pages 537–552. Springer, 2011. ISBN 978-3-642-25072-9.
- K. Page, D. De Roure, K. Martinez, J. Sadler, and O. Kit. Linked Sensor Data: RESTfully Serving RDF and GML. In *2nd International Workshop on Semantic Sensor Networks (SSN09), In conjunction with the 8th International Semantic Web Conference, ISWC 2009*, volume 522, pages 49–63, Washington, DC, USA, October 2009. CEUR.
- E. Pebesma, D. Cornford, S. Nativi, and C. Stasch. The uncertainty enabled model web (uncertweb). In A. Berre, D. Roman, and P. Maué, editors, *Proceedings of the Workshop "Environmental Information Systems and Services - Infrastructures and Platforms" Workshop at EnviroInfo2010*, volume 679. CEUR-WS, 2010.
- F. Probst and M. Espeter. Spatial dimensionality as a classification criterion for qualities. In B. Bennett and C. Fellbaum, editors, *International Conference on Formal Ontology in Information Systems (FOIS 2006)*, volume 150 of *Frontiers in Artificial Intelligence and Applications*, pages 77–88. IOS Press, 2006.

- M. Raubal. Formalizing conceptual spaces. In A. C. Varzi and L. Vieu, editors, *Formal Ontology in Information Systems, Proceedings of the Third International Conference (FOIS 2004)*, pages 153–164. IOS Press, 2004.
- L. Richardson and S. Ruby. *RESTful Web Services*. O’Reilly Media, Inc., 2007. ISBN 0596529260.
- S. Sahoo, A. Sheth, and C. Henson. Semantic provenance for escience: Managing the deluge of scientific data. *IEEE Internet Computing*, pages 46–54, 2008.
- P. Santos, B. Bennett, and G. Sakellariou. Supervalue semantics for an inland water feature ontology. In *Proceedings of the 19th international joint conference on Artificial intelligence, IJCAI’05*, pages 564–569, San Francisco, CA, USA, 2005. Morgan Kaufmann Publishers Inc.
- S. Schade. *Ontology-Driven Translation of Geospatial Data*. PhD Thesis, University of Münster, Germany. IOS Press, Amsterdam, The Netherlands, 2010. ISBN 978-1-60750-071-1.
- S. Scheider. *Grounding Geographic Information in Perceptual Operations*. PhD Thesis, Institute for Geoinformatics, University of Münster, Germany. Available online at geographicknowledge.de/pdf/MyThesis.pdf, 2011.
- S. Scheider, K. Janowicz, and W. Kuhn. Grounding geographic categories in the meaningful environment. In K. Hornsby, C. Claramunt, M. Denis, and G. Ligozat, editors, *Conference on Spatial Information Theory (COSIT 2009)*, volume 5756 of *LNCS*, pages 69–87. Springer, 2009.
- N. Schuurman. Formalization matters: Critical gis and ontology research. *Annals of the Association of American Geographers*, 96(4):726–739, 2006.
- A. Sheth, C. Ramakrishnan, and C. Thomas. Semantics for the semantic web: the implicit, the formal and the powerful. *International Journal on Semantic Web and Information Systems*, 1:1–18, 2005.
- P. Shvaiko and J. Euzenat. Ten Challenges for Ontology Matching. In R. Meersman and Z. Tari, editors, *On the Move to Meaningful Internet Systems: OTM 2008*, volume 5332 of *LNCS*, pages 1164–1182. Springer, 2008.
- G. Sinha and D. Mark. Toward a foundational ontology of the landscape. In *Extended Abstracts of GIScience 2010*, 2010a.
- G. Sinha and D. Mark. Cognition-based extraction and modelling of topographic eminences. *Cartographica*, 45(2):105–112, 2010b.
- J. F. Sowa. *Knowledge Representation: Logical, Philosophical and Computational Foundations*. Brooks Cole Publishing Co., Pacific Grove, CA, USA, 2000.
- W. Tobler. A computer model simulating urban growth in the detroit region. *Economic Geography*, 46(2):234–240, 1970.

- M. Uschold. Creating, integrating and maintaining local and global ontologies. In W Horn, editor, *Proceedings of 14th European Conference on Artificial Intelligence (ECAI'00), Berlin*, Amsterdam, 2000a. IOS Press.
- M. Uschold. Creating, integrating and maintaining local and global ontologies. In *14th European Conference on Artificial Intelligence (ECAI 2000)*, Berlin, Germany, August 2000b.
- R. van Hage, V. Malaise, R. Segers, L. Hollink, and G. Schreiber. Design and use of the simple event model (sem). *Web Semantics: Science, Services and Agents on the World Wide Web*, 9(2), 2011. ISSN 1570-8268.
- J. Volz, C. Bizer, M. Gaedke, and G. Kobilarov. Silk – a link discovery framework for the web of data. In *2nd Workshop about Linked Data on the Web (LDOW2009), Madrid, Spain, April 20-24, 2009*, pages 621–630. ACM, 2009.
- I. Wachsmuth. The concept of intelligence in AI. *Prerational Intelligence – Adaptive Behavior and Intelligent Systems without Symbols and Logic*, 1:43–55, 2000.
- D. Weinberger. *Too Big To Know*. Basic Books, New York, 2011.
- Mao Ye, K. Janowicz, C. Mülligann, and W.-C. Lee. What you are is when you are: The temporal dimension of feature types in location-based social networks. In *19th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems (ACM SIGSPATIAL GIS 2011)*, pages 102–111. ACM, 2011a.
- Mao Ye, D. Shou, W.-C. Lee, P. Yin, and K. Janowicz. On the semantic annotation of places in location-based social networks. In *Conference on Knowledge Discovery and Data Minin, ACM SIGKDD*, pages 520–528, 2011b.
- A. Zimmermann, M. Krötzsch, J. Euzenat, and P. Hitzler. Formalizing ontology alignment and its operations with category theory. In B. Bennett and C. Fellbaum, editors, *International Conference on Formal Ontology in Information Systems (FOIS 2006)*, pages 277–288. IOS Press, 2006.