Microtheories for Spatial Data Infrastructures - Accounting for Diversity of Local Conceptualizations at a Global Level

Stephanie Duce[1] and Krzysztof Janowicz[2]

[1] Department of Languages and Systems, University Juame I, Spain
[2] Department of Geography, The Pennsylvania State University, USA

Abstract. The categorization of our environment into feature types is an essential prerequisite for cartography, geographic information retrieval, routing applications, spatial decision support systems, and data sharing in general. However, there is no a priori conceptualization of the world and the creation of features and types is an act of cognition. Humans conceptualize their environment based on multiple criteria such as their cultural background, knowledge, motivation, and particularly by space and time. Sharing and making these conceptualizations explicit in a formal, unambiguous way is at the core of semantic interoperability. One way to cope with semantic heterogeneities is by standardization, i.e., by agreeing on a shared conceptualization. This bears the danger of losing local diversity. In contrast, this work proposes the use of microtheories for Spatial Data Infrastructures, such as INSPIRE, to account for the diversity of local conceptualizations while maintaining their semantic interoperability at a global level. We introduce a novel methodology to structure ontologies by spatial and temporal aspects, in our case administrative boundaries, which reflect variations in feature conceptualization. A local, bottom-up approach, based on non-standard inference, is used to compute global feature definitions which are neither too broad nor too specific. Using different conceptualizations of rivers and other geographic feature types, we demonstrate how the present approach can improve the INSPIRE data model and ease its adoption by European member states.

1 Introduction and Motivation

In 2007 the European Union launched the Infrastructure for Spatial Information in the European Community (INSPIRE) which aims at creating a Spatial Data Infrastructure (SDI) supporting cross-scale, cross-language, and cross-border interoperability and access to geodata[1]. This involves the development of spatial data themes, web services, agreements on data and service sharing, coordination and monitoring mechanisms, and especially also common metadata standards and geographic feature (object) type catalogs. The European Union, however, is very heterogeneous in terms of ecosystems, climatic and physical conditions,

[1] INSPIRE Directive http://inspire.jrc.ec.europa.eu/index.cfm

cultures, languages, and administrative systems. This makes the definition of a shared conceptualization of geographic features a difficult task. If the guidelines set up by INSPIRE are too generic, i.e., do not sufficiently restrict possible interpretations [1], they will fail to establish interoperability or at least require manual, application specific, and error-prone adjustments. Overly specific guidelines could hinder implementation and reduce the usability of the data. In general, creating such a broad and multipurpose infrastructure to ensure overarching interoperability carries the danger that important nuances in the local and contextual terminology will be lost. For the INSPIRE initiative to be effective, efficient and successful, all parties should be free to define geographic feature types in a manner most suited to their unique environment and culture though still consistent at an all-encompassing upper level. This need introduces a struggle to create, integrate, and maintain conceptualizations at a local and European level.

The importance of local conceptualizations of geographic features has been widely acknowledged and discussed in the literature. Geographic features are susceptible to sorites vagueness and are characterized by vague boundaries [2, 3], vague adjective-based definitions [4, 5], meso-scale [6], and temporal dynamics [7]. This means that human perception, language, and social agreement play a strong role in our conceptualization of geographic features and can lead to semantic heterogeneities [8–11].

For instance, a forest can be a protected area, plantation, recreational area, agricultural area, habitat, and so forth. These different perspectives give rise to potential socio-economic conflicts but also hinder classification and retrieval. Lund [12], for instance, lists over 900 (often contradictory) definitions of forest. As forests do not stop at borders, a forest in Spain may be regarded as meadowland in France. Whether an area is categorized as forest or not may have legal and economic consequences as in the case of deforestation.

Given the indeterminacy of geographic features used for land cover classification and their increasing availability to the public, Comber and Fisher [10] argue that there is an urgent need for the semantics of data to be made explicit to users. An ontology for the geographic domain should reflect and capture multiple conceptualizations of geographic features [13]

The challenge of handling local [14], i.e., domain specific, conceptualizations at a global level is not new and has been a core topic in Artificial Intelligence (AI) research for 30 years [15, 16]. The key idea is to be consistent at the local level but allow contradicting conceptualizations within the global overall knowledge base. One promising approach to handle semantic heterogeneity is to structure knowledge in domain specific microtheories (also called contexts). This approach has been first implemented in the OpenCyC ontology which contains hundreds of thousands of terms and assertions. Each microtheory is designed as a coherent set of statements and can be thought of as a single ontology. Separate microtheories can hold information about the same concept but contain incompatible facts. For instance, one microtheory may be strict about physical properties and laws of nature, while other microtheories may have weaker constraints to support *naïve physics* [17].

Usually microtheories are organized in subsumption hierarchies, i.e., facts specified in the super-microtheory must also hold in each of its sub-theories. Sibling-theories, however, may contain contradicting conceptualizations. Note that microtheories are not the only approach to ontology modularization [9, 18–20]. Kokla and Kavouras [21, 22] discussed the use of concept lattices to identify overlapping relationships and manage different geographic domain ontologies. While Guha et al. [23] revitalized the notion of context for the Semantic Web. Batemann et al. [19] discuss how to develop multi-perspectival ontologies of space using algebraic specifications and DOLCE as foundational ontology. The microtheories approach proposed here calculates a the Least Common Subsumer rather than using a concept lattice to identify commonalities and overlaps between different microtheories.

The main difference between our approach and previous work on microtheories is the use of alternative ordering principles. In previous work we proposed to introduce time and space as additional first class ordering principles for microtheories [24]. For instance, the definition of rivers differs markedly between southern European and northern European countries and hence microtheories specifying local conceptualizations may contradict. However, these microtheories have to be consistent with an EU-wide theory. As semantic hererogeneity is not a problem but a challenge, such an approach supports the diversity of different feature type conceptualizations across Europe, while creating and maintaining a consistent global ontology at a European scale to support interoperability. In this work, we discuss how microtheories can be used to define local conceptualizations and demonstrate how non-standard inference [25] and similarity reasoning can be employed to automatically infer an appropriate top-level as a common compromise. While our work is not restricted to INSPIRE or SDI, they will serve as running examples throughout the paper.

2 Structuring Microtheories by Administrative Containment

This section introduces the role of microthoeries and the methodology used to compute a top-level conceptualization from local knowledge.

The use of microtheories for knowledge representation and reasoning has numerous advantages [26] – the ability to support multiple conceptualizations for the same terminology and to provide structural relationships between these theories are the two most relevant benefits for the presented work. As each microtheory is considered an object in its own right and is only evaluated in a given context, two microtheories can hold conflicting facts without undermining the reasoning capacity of the entire knowledge base [27, 26]. In addition, microtheories provide modularity for ontologies [18–20]. This makes reasoning and querying more efficient as only relevant parts are used to answer a query [26]. Modularization also eases the updating of ontologies and allows their evolution without having to make widespread changes to the overall system. This is highly desirable as concepts in geospatial domains are regularly evolving as better understanding is achieved [28].

From the INSPIRE perspective, different conceptualizations of the same geographic feature may conflict with each other. Germany's conceptualization of river may state that it contains flowing water. However, in Spain, where rivers may by dry for most of the year, the definition of river cannot rely on the presence of flowing water. Most ontologies developed for the semantic (geospatial) Web [29], are strongly bound by the rules of logic and cannot cope with such conflicts. Therefore, in order to merge the definitions of rivers in Spain and Germany to create a Europe-wide conceptualization of rivers one of them would need to be changed to a definition that does not reflect the nature of the features in that country. This is undesirable and undermines the success of the INSPIRE initiative. Rivers and forests are by no means the only examples - in fact most terminologies require a spatially bounded context for their interpretation.

2.1 Structuring Microtheories

So far, microtheories have only been structured by establishing hierarchical relationships between them, i.e, by generalization. Other potential ordering principles such as space, time, or cultural background have received nearly no attention in the Semantic Web community. While their importance has been recognized recently, existing work reduces space and time to simple latitude-longitude pairs and time stamps. Tobler's First Law of Geography states that 'Everything is related to everything else, but near things are more related than distant things'. Climatic, geographic and geological factors, all of which adhere to the above law, govern the character of geographic features and hence influence their categorization. Besides their role in the gradual change of the environment, space and time are the most fundamental ordering relations used in human cognition and language – spatial metaphors are just one prominent example [30].

More formally, the hierarchy of microtheories is created using a generalization relationship between microtheories called *genlMt* in OpenCyC and *specializes* by McCarthy and Buvac [31]. If $ist(mt, p)$ is the *is true in* relation between a microtheory mt and a predicate p, then $genlMt$ is the anti-symmetric, reflexive, and transitive, binary predicate by which the theory hierarchy is constructed by adding axioms of the form

$$mt_0 : \forall p \; ist(mt_g, p) \wedge genlMt(mt_g, mt_s) \longrightarrow ist(mt_s, p)$$

to the topmost theory mt_0; where mt_g is the more general and mt_s the more specific theory[2]. Figure 1a depicts the relation between an overall geographic microtheory and a more specific version for volunteered geographic information. The second may introduce new vocabulary for navigation, such as landmarks, instead of relying on latitude and longitude only and may redefine the notion of distance.

[2] Note that we do not allow cycles; see [23] for details.

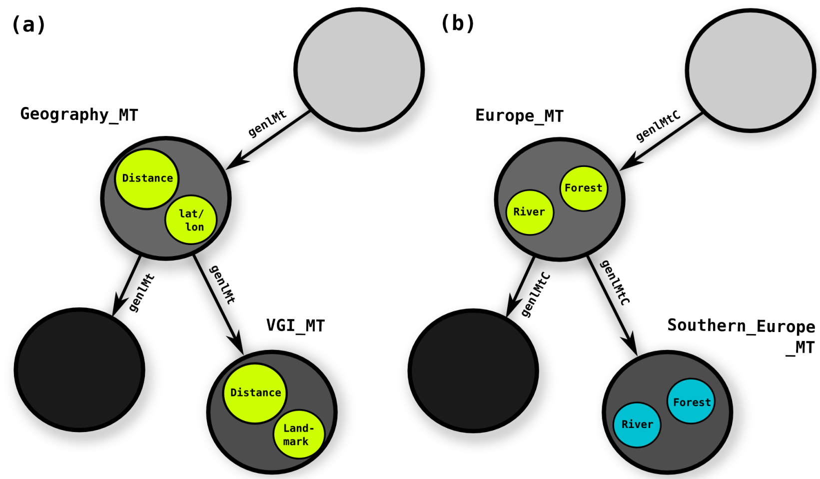

Fig. 1. Structuring microtheories by (a) generalization (*genlMt*) and by (b) generalization and (spatial) containment (*genlMtC*); see also [24].

To structure microtheories by spatial (or administrative) containment, we introduce the *genlMtC* relation which extends *genlMt* as follows:

$$mt_0 : \forall p \; ist(mt_g, p) \wedge genlMtC(mt_g, mt_s) \longrightarrow genlMt(mt_g, mt_s) \wedge \circledcirc(mt_g, mt_s)$$

Consequently, $genlMtC(mt_g, mt_s)$ holds if mt_s is a sub-theory of mt_g and all footprints of individuals of geographic feature types specified in mt_s are (spatially or administratively) contained in mt_g[3]. Examples, depicted in figure 1, include the river and forest case discussed previously.

In the following, we will use these relationships as meta-theory for local ontologies represented using description logics. Hence, a more detailed specification of the containment relation is left for further work.

2.2 Spatial versus Administrative Containment

Dividing Europe into appropriately structured microtheories using spatial containment is difficult. A division by geographic factors such as climate and geology may lead to scale problems and especially, to administrative challenges as one country could fall into more than one theory and multiple countries could belong to the same microtheory. A more fine grained solution would be to decide to which theory each feature type belongs. However, this would again be impractical from an administrative perspective.

[3] The second part is denoted by the $\circledcirc$-predicate and requires a spatial footprint for the individuals as well as for the spatial scope of the theory. A formal semantics for $\circledcirc$ including RCC is left for further work.

This paper describes the possible structuring of microtheories based on administrative boundaries. This method takes geographic and climatic factors into account (to some extent) and offers intuitive divides from a political perspective. Using this method each EU member state would define its own microtheories, best reflecting the conceptualization of geographic features in its country. This would overcome some of the administrative difficulties and align well with present data models which are usually created on the national level.

Nevertheless, administrative structuring is not ideal as the territories of countries are large and diverse themselves. Also, a country may possess outside territory where geographic features may be very different (e.g., the UK and Gibralta). To overcome these issues, autonomous or independent regions could make their own microtheory where necessary. A nation-wide microtheory could then be generalized from the internal regions. We cannot offer a definitive solution here as multiple situations may require different choices. As INSPIRE acts as a running example to illustrate our theoretical approach, we use administrative containment in the following.

2.3 Methodology

Features such as rivers, forests, and estuaries demonstrate the benefits of the microtheory-approach as their conceptualization is strongly based on factors that vary in space (rainfall, geology, topography...). They are of great importance from economic, social and environmental perspectives and involve various stakeholders. An effective SDI, based on well defined and semantically interoperable feature definitions, is imperative to understand, study and successfully manage these features.

Several steps are required to demonstrate the use of a bottom-up approach to compute an appropriate[4] global definition as a compromise between local conceptualizations.

1. Natural language definitions of geographic feature types have to be selected from the literature. These definitions should reflect the local (i.e., country specific) viewpoints. In our case, we present definitions for the feature *River*. Spain and Germany were chosen for treatment as their rivers represent different ends of the spectrum of contrasting river conditions across Europe. These natural language definitions are expanded into concept maps and related to other features in the domain.
2. To support non-standard inference and similarity reasoning these definitions are formalized using the Web Ontology Language (OWL) and the Protégé editor. Note that some parts of the informal concept map definitions cannot be adequately represented in OWL.
3. To generate an appropriate top-level for the global ontology, the Least Common Subsumer (LCS) [25] will be computed as it fulfils the requirements

[4] Appropriate is defined here as a conceptualization that is neither too broad nor too specific in the number and type of geographic features that are covered.

of *appropriateness* described above. The computation of the LCS between DL-based concepts requires a trade off between the expressivity of the conceptualizations and the reasoning capabilities of the methodologies and tools used. Similarity reasoning and computing the LCS can only be performed on a subset of OWL. Hence, further reductions to the concept maps are required. In many cases, these restrictions are caused by the tools selected and can be resolved in the near future with new implementations. One typical example, namely the problem of handling logical disjunction in case of the LCS will be discussed in the formalization section.

4. Finally, after computing the LCS – which in our running example serves as the EU wide definition of *River* – we use subsumption and similarity reasoning to evaluate our results, i.e., to check whether the LCS provides a more appropriate top-level conceptualization than the existing INSPIRE definition(s).

Similarity reasoning, using the SIM-DL reasoner, is employed to test how well the definitions capture the domain and reflect human conceptualizations. SIM-DL is an asymmetric, context-aware similarity measurement theory used for information retrieval. It compares a DL search concept with one or more target concepts, by measuring the degree of overlap between their definitions. See [32, 33] for details on SIM-DL and similarity estimations.

The least common subsumer (computed in step 3) is defined as follows:

Definition 1. *Given a description logic $\mathcal{L}$, and a set of concepts $C_1, ..., C_n$, a particular concept D is the least common subsumer with respect to $C_1, ..., C_n$ iff it satisfies the following conditions:*

(a) $C_i \sqsubseteq D$ for all $C_{1,...,n}$
(b) All concepts D' satisfying $C_i \sqsubseteq D'$ (for all $C_{1,...,n}$) also satisfy $D \sqsubseteq D'$, i.e., D is the least $\mathcal{L}$ concept satisfying (a) and unique.

3 Application

This section demonstrates the application of the above methodology. We introduce local microtheories for rivers in Spain and Germany. These are related via *genlMtC* to the EU microtheory which was computed using the LCS i.e., the least common subsumer of the conceptualizations provided by the local microtheories. While the full study also includes forests and estuaries in the theories we focus on the river example due to lack of space.

3.1 Natural Language Definitions

The traditional northern European perspective of a river is a continuously flowing body of water which may also be navigable [34]. This view is reflected in the INSPIRE context. We especially refer to the *INSPIRE Feature Concept*

Dictionary[5], the *INSPIRE Consolidated UML Model*[6], the *EuroRegionalMap Specification and Data Catalogues*, as well as the *Water Framework Directive (WFD)*. In these classifications, *Watercourse* is defined as '[a] natural or man–made flowing watercourse or stream'[7], while *WFD River* is defined as '[a] body of inland water flowing for the most part on the surface of the land but which may flow underground for part of its course'[8]. This definition seems broad, however, its requirement of flowing water may be too specific to encompass rivers in the Mediterranean climes of southern Europe – especially taking the effects of global warming into account. For example, rivers in southern Spain are highly ephemeral and may only contain water during flood events. In these regions, the conceptualization of rivers may include channels or depressions through which water flows, even if they are dry [35].

Rivers are highly complex ecosystems and commercialized anthropogenic entities. Hence, the definitions presented in this work do not claim to encompass all their elements. They show how diverse elements can be used to better define local conceptualizations without undermining global interoperability. The ecosystem functions and anthropogenic services performed by rivers are considered in this work to be (thematic) *roles*, similar to the notion of affordances used for modeling by Kuhn [36] and others in GIScience.

These properties of rivers may to some extent transcend the spatio-temporal vagueness and variability which hamper the use of mereotopology in defining rivers and are likely to represent commonalities and distinctions in different local conceptualizations. For example, rivers, wherever they are, play the role of transporting water. In Germany a river can also play the role of providing transport to humans and goods (as their constant flow of water makes them navigable). However, in Spain the frequent lack of water means rivers are not perceived as navigable. In the following, rivers in Spain and Germany are defined in natural language terms. These definitions were derived from multiple sources to ensure they are not biased by a particular point of view.

A Spanish river is a channel, with a bed and more or less defined banks, which transects a river basin at a low point in the topography. It drains water which falls as precipitation on the river basin. It has a flow regime which refers to the average presence or absence of water within the channel throughout a year. It may participate in flood events and provides the ecological service of protecting against these events. Spanish rivers also participate in droughts and can provide terrestrial or aquatic habitat, terrestrial or aquatic recreational areas and play the role of supplying water.

A German river is a channel, with a river bed and river banks which contains flowing water and transects a river basin with another waterbody as its desti-

[5] https://inspire-registry.jrc.ec.europa.eu/registers/FCD
[6] http://inspire-twg.jrc.ec.europa.eu/inspire-model/; Generatied 24 August 2009 v3, Revision 873
[7] http://inspire-registry.jrc.ec.europa.eu/registers/FCD/items/105 as of 05-Dec-08
[8] http://inspire-registry.jrc.ec.europa.eu/registers/FCD/items/421 as of 19-Jan-10

nation. It represents the above ground expression of the groundwater table and also drains water, from precipitation or snow melt, in the river basin. It may participate in flood events and provides the ecological service of protecting against these events. German rivers provide aquatic habitat and aquatic recreational areas and play the role of supplying water and transportation.

3.2 Conceptual Modeling

According to the presented methodology, the natural language definitions are encoded as semi-formal concept maps and aligned to the top-level classes, *Physical Endurant*, *Perdurant*, *Role* and *Quality*, proposed by the DOLCE (Descriptive Ontology for Linguistic and Cognitive Engineering) top-level ontology [37]. The concept maps are depicted in the figures 2 and 3 with the main differences marked red[9]. They show that German rivers contain flowing water and have river banks as their proper parts. However, for Spanish rivers these relationships are optional (indicated by dotted lines) to reflect the temporal variability and vagueness of these properties. Furthermore, German rivers are defined as having waterbodies as their destinations which is not required for Spanish rivers as they may simply peter out. German and Spanish rivers were defined as having precipitation and ground water as sources of water with Germany having snow melt as an additional source. Flood events (and drought events in Spain) as well as erosion are of particular management importance and thus were included in the conceptual models.

The definitions deliberately avoid reference to rivers being artificial or natural as these terms are vague and can cause confusion. For example, natural rivers can have artificial components (e.g. bank stabilization measures) or an artificial flow regime (e.g., due to the presence of a dam). While these characteristics may help distinguish between some feature types (e.g., canal and river), they do not provide identity to rivers. To support grounding by observations, we also include properties such as water depth to be linked to the currently developed measurement ontologies [38].

3.3 Formalization

To use the Semantic Web infrastructure and reasoners the conceptual models are represented in OWL. The Protégé versions 4 and 3.3.1 were used as ontology editors as they provides plug-ins to the SIM-DL reasoner which supports, subsumption and similarity reasoning as well as the computation of the LCS [32, 33][10]. Several simplifications and ontological commitments are necessary to represent the conceptual models of rivers. For instance, due to the open world assumption, optional relations are not specified in the river definitions.

[9] The presented conceptual models are simplified for reasons of readability; the original and more detailed versions can be downloaded at http://www.personal.psu.edu/kuj13/GIScience2010MT.zip.

[10] Note that the current version only supports a subset of OWL-DL and the computation of the LCS is even further restricted; download at http://sim-dl.sf.net/.

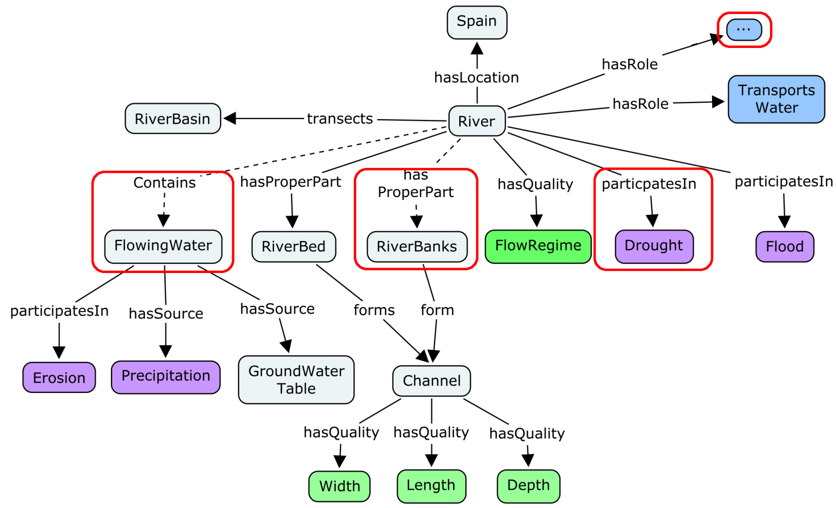

Fig. 2. Simplified conceptual model showing the relations between entities defining a *SpanishRiver*. The entities are divided roughly into the DOLCE top-level classes: physical endurant (white), perdurant (purple), role (blue) and quality (green). Elements of difference to the German river definition are marked in red.

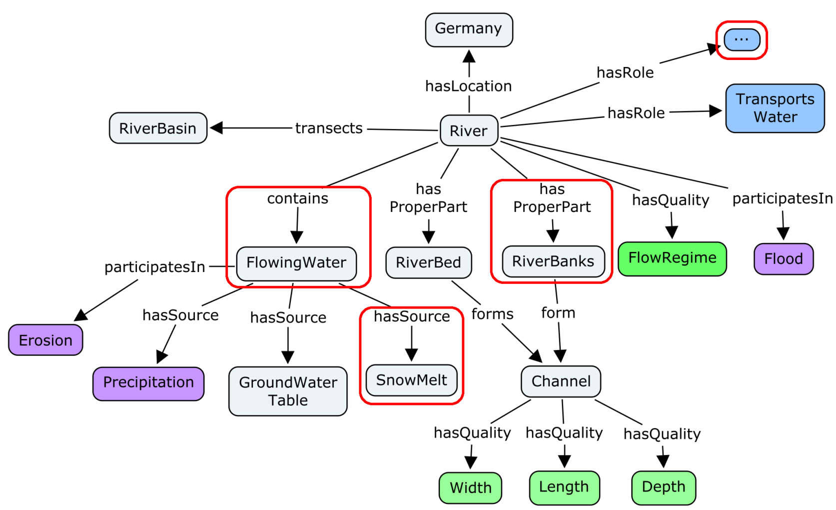

Fig. 3. Simplified conceptual model showing the relations between entities defining a *GermanRiver*. Elements of difference to the Spanish river definition are marked in red. This model has been simplified for presentation.

Most importantly, the LCS is only meaningful for description logics without disjunction as the LCS would simply be the disjunction of compared concepts. There are two solutions to this problem. First, to reduce the expressivity of the language used and hence approximate the conceptualizations, e.g., by vivification [39]. Second, to compute a *good* instead of the *least* common subsumer; see [40] for details.

The German and Spanish river classes are not modeled as disjoint because, given their broad scale, a single river can have multiple conceptualizations, multiple links and may fall within one or more member states. Instances may be attributed to one, or more than one, of the microtheories. Figure 4 shows a fragment of the ontology. Note that we have combined the definitions of German, Spanish, and EU rivers in a single ontology to perform reasoning in Protégé. In fact, they are in separate microtheories and hence are all named *River*. Difficulties arising from semantic heterogeneities are captured by our approach as all definitions have a spatial (or in our case administrative) context.

3.4 Computing the Top-Level

The spirit of modeling on the Semantic Web has often been confused. In contrast to specifying multiple taxonomies by hand, the driving idea is to let the reasoner do the *untangling*, i.e., reclassify a developed ontology, discover, and add implicit subsumption relations. In this spirit, but using the reverse direction, we propose to specify the local and member state specific conceptualizations and let the reasoner compute the common top-level. Consequently, we do not use subsumption reasoning but compute the least common subsumer as most specific top-level concept for each feature type. A similar approach was also proposed and implemented into SIM-DL in previous work [41]. Computing *EuropeanRiver* as the last common subsumer of the German and Spanish river definitions yields:

$$EuropeanRiver \equiv \exists.transects(RiverBasin) \sqcap \exists.hasPart(RiverBed) \sqcap ... \sqcap \exists.hasLocation(MemberState) \sqcap \exists.hasQuality(FlowRegime) \sqcap [\texttt{vivification}]$$ [11]

As the Spanish definition lacks the *contains FlowingWater* and *hasPart RiverBanks* restrictions, these were excluded from the *EuropeanRiver* definition. The common filler between the two definitions for the *hasLocation* property, *MemberState*, was used. Thus, based on the above methodology, a European river transects a river basin and has a river bed, flow regime, is located in an EU member state, and performs a suite of roles. Reclassification of the ontology showed that the *SpanishRiver* and *GermanRiver* are subsumed by *EuropeanRiver*, while the INSPIRE definition of river excludes the *SpanishRiver* (see figure 4). Consequently, the INSPIRE definition is too specific even when just two local definitions are compared. Adding more definitions from other member states is likely to further broaden the EU definition. While restricting rivers to flowing watercourses is too exclusive, in other respects the INSPIRE definition is too generic and could be more specific (and hence improving semantic interoperability). For instance, the river definition could include a relation to river basins;

[11] A richer approximation of the LCS could be determined by vivifivation or by computing the *good common subsumer*; both require manual interaction.

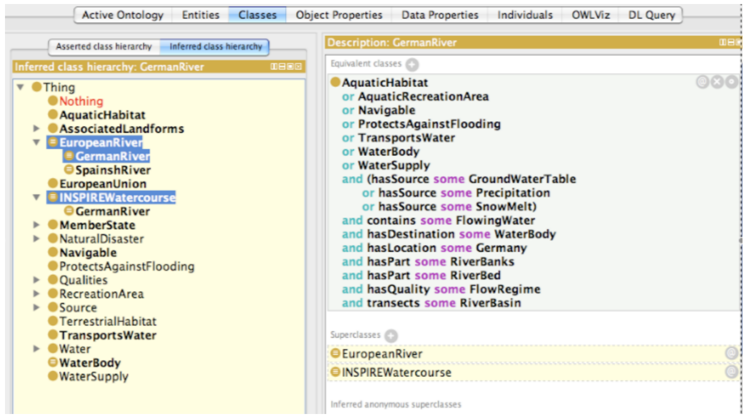

Fig. 4. Screenshot showing the inferred class hierarchy and the restrictions used to define *GermanRiver*. The reasoner inferred that *GermanRiver* and *SpanishRiver* are both kinds of *EuropeanRiver* (as this is the LCS). The INSPIRE *Watercourse* (and *River*) can only act as superclass for the *GermanRiver* and excludes the *SpanishRiver*.

especially as *RiverBasin*[12] is already listed in the INSPIRE Feature Concept Dictionary.

3.5 Similarity Results

Finally, *EuropeanRiver* was compared to the other definitions using the SIM-DL similarity server running the *maximum* & *asymmetry* modes [32]. As expected, from the perspective of the European river definition, the similarity to the German and Spanish definitions is 1.0. This means that users searching for *EuropeanRiver* will be satisfied retrieving both kinds of river (e.g., using a semantics-enabled interface for Web gazetteers [32]). This is not surprising as the EU wide definition is the super concept of both. The comparison to the INSPIRE definition results in a low similarity (0.11). This is to be expected as SIM-DL measures the conceptual overlap and the INSPIRE definition does not contain several of the statements made for the EU wide definition. Note, however, that adding more member states would broaden the European river definition and make it more similar to the INSPIRE version.

4 Conclusions and Further Work

In this work we have discussed the importance of local conceptualizations of geographic space. Different communities have developed their own understanding and terminology for good reasons [14]. By introducing microtheories and structuring them by spatial or administrative containment, we have shown how local

[12] http://inspire-registry.jrc.ec.europa.eu/registers/FCD/items/409 as of 19-Jan-10

and potentially contradictory conceptualizations can be reconciled in a common knowledge base. Next, we have presented a methodology to compute the top-level of such knowledge bases instead of standardizing common feature types manually – which may exclude local conceptualizations. We have tested our approach by specifying local river definitions and computing the least common subsumer as a common EU wide definition. Our results show that the definitions proposed by INSPIRE and the Water Framework Directive are too specific in some respects while lacking other relations, e.g., to river basins. The presented EU wide definition could contain more details and hence be a better approximation of a common compromise. As argued above, this could be done semi-automatically by computing the good common subsumer or using vivification. Both approaches are promising and will be investigated in future work. Our full study also takes forests and estuaries from multiple member states into account; the concept maps and ontologies are available online. While incorporating ontologies and Semantic Web reasoners into SDIs has been difficult so far, recent work on Semantic Enablement for Spatial Data Infrastructures may ease their integration [42]. A first reference implementation of a Web Reasoning Service (WRS) for similarity reasoning is available online at the 52°North semantics community[13].

The structuring of microtheories by spatial and temporal relations presented here gives initial insights into the role of space and time for ontology modularization [24]. However, the approach is still at an early stage and requires an improved and rigid, formal underpinning. The work also points to limitations in the ability of existing Semantic Web representation languages and reasoners to adequately deal with the expressive conceptualizations necessary to effectively define vague, dynamic and highly interlinked features in geographic space. Future work should especially also focus on machine learning approaches to derive the top-level conceptualizations and should be compared to the results of the deductive approach taken in this work.

References

1. Kuhn, W.: Semantic engineering. In Navratil, G., ed.: Research Trends in Geographic Information Science, Springer (2009) 63–74
2. Fisher, P.: Sorites paradox and vague geographies. Fuzzy Sets and Systems **113** (2000) 7–18
3. Smith, B., Mark, D.M.: Do mountains exist? towards an ontology of landforms. Environment and Planning B: Planning and Design **20**(2) (2003) 411–427
4. Mark, D.M.: Toward a theoretical framework for geographic entity types. In Campari, I., Frank, A.U., eds.: Spatial Information Theory: A Theoretical Basis for GIS. Volume 716 of Lecture Notes in Computer Science., Springer, Heidelberg (1993) 270–283
5. Bennett, B., Mallenby, D., Third, A.: An ontology for grounding vague geographic terms. In: Formal Ontology in Information Systems - Proceedings of the Fifth International Conference (FOIS 2008). Volume 183., IOS Press (2008) 280–293

[13] http://www.52north.org/semantics

6. Smith, B., Mark, D.M.: Ontology and geographic kinds. In: International Symposium on Spatial Data Handling, Vancouver, Canada (July 1998) 308 – 320
7. Frank, A.: A linguistically justified proposal for a spatiotemporal ontology. In Kuhn, W., Warboys, M., Timpf, S., eds.: Proceedings of the COSIT03 International Conference. Position Paper in COSIT-03 Ontology Workshop. Number 2205, Springer Berlin (2003)
8. Egenhofer, M., Mark, D.M.: Naive geography. In Frank, A.U., Kuhn, W., eds.: Spatial Information Theory: A Theoretical Basis for GIS. Volume LNCS 988., Springer-Verlag, Berlin (1995) 1–15
9. Bishr, Y.: Overcoming the semantic and other barriers to gis interoperability. International Journal of Geographical Information Science **12**(4) (1998) 299–314
10. Comber, A., Fisher, P.: What is land cover? Environment and Planning B: Planning and Design **32** (2005) 199–209
11. Kuhn, W.: Geospatial semantics: Why, of what and how? Journal of Data Semantics **III** (2005) 1–24
12. Lund, H.G.: Definitions of forest, deforestation, afforestation, and reforestation. Technical report, Forest Information Services, http://home.comcast.net/ gyde/DEFpaper.htm (2009)
13. Janowicz, K., Maue, P., Wilkes, M., Schade, S., Scherer, F., Braun, M., Dupke, S., Kuhn, W.: Similarity as a quality indicator in ontology engineering. In Eschenbach, C., Grueninger, M., eds.: 5th International Conference on Formal Ontology in Information Systems (FOIS 2008), IOS Press (2008) 92–105
14. Uschold, M.: Creating, integrating and maintaining local and global ontologies. In Horn, W., ed.: Proceedings of 14th European Conference on Artificial Intelligence (ECAI'00), Berlin, Amsterdam, IOS Press (2000)
15. McCarthy, J.: Generality in artificial intelligence. Communications of the ACM **30**(12) (1987) 1030–1035
16. Wachsmuth, I.: The concept of intelligence in ai. Prerational Intelligence – Adaptive Behavior and Intelligent Systems without Symbols and Logic **1** (2000) 43–55
17. Smith, B., Casati, R.: Naive physics An essay in ontology. Philosophical Psychology **7/2** (1994) 225–244
18. Grau, B., Kazakov, Y., Sattler, U.: A logical framework for modularity of ontologies. In: 20th International Joint Conference on Artificial Intelligence. (2007) 183–196
19. Bateman, J., Borgo, S., Luettich, K., Masolo, C., Mossakowski, T.: Ontological modularity and spatial diversity. Spatial Cognition and Computation **7**(1) (May 2007) 97–128
20. Hois, J., Bhatt, M., Kutz, O.: Modular ontologies for architectural design. In Ferrario, R., Oltramari, A., eds.: Formal Ontologies Meet Industry, IOS Press (September 2009) 66–78
21. Kokla, M., Kavouras, M.: Fusion of top-level and geographic domain ontologies based on context formation and complementarity. International Journal of Geographical Information Science **15** (2001) 679–687
22. Kavouras, M., Kokla, M.: A method for the formalization and integration of geographic categorizations. International Journal of Geographical Information Science **16** (2002) 439–453
23. Guha, R., Mccool, R., Fikes, R.: Contexts for the semantic web. In: International Semantic Web Conference (ISWC 2004). Number 3298 in Lecture Notes in Computer Science, Springer (2004) 32–46

24. Janowicz, K.: The role of place for the spatial referencing of heritage data. In: The Cultural Heritage of Historic European Cities and Public Participatory GIS Workshop, The University of York, UK (17 -18 September 2009)
25. Küsters, R.: Non-Standard Inferences in Description Logics. Volume 2100 of Lecture Notes in Artificial Intelligence. Springer (2001)
26. Cycorp: Contexts in cyc. available online at: http://www.cyc.com/cycdoc/course/contexts-basic-module.html (2002)
27. Hovy, E.: Comparing Sets of Semantic Relations in Ontologies. In: The Semantics of Relationships: An Interdisciplinary Perspective. Kluwer Publishers (2002)
28. Brodaric, B., Gahegan, M.: Distinguishing instances and evidence of geographical concepts for geospatial database design. Second International Conference on Geographic Information Science **2478** (2002) 22–37
29. Egenhofer, M.: Toward the semantic geospatial web. In: GIS '02 Proceedings of the 10th ACM international symposium on Advances in geographic information systems, New York, NY, USA, ACM (2002) 1–4
30. Lakoff, G., Johnson, M.: Metaphors We Live By. University Of Chicago Press (1980)
31. McCarthy, J., Buvac, S.: Formalizing context (expanded notes) (1996)
32. Janowicz, K., Kessler, C., Schwarz, M., Wilkes, M., Panov, I., Espeter, M., Baeumer, B.: Algorithm, implementation and application of the sim-dl similarity server. In Fonseca, F.T., Rodriguez, A., Levashkin, S., eds.: Second International Conference on GeoSpatial Semantics (GeoS 2007). Number LNCS 4853, Springer (2007) 128–145
33. Janowicz, K., Wilkes, M.: SIM-DL_A: A novel semantic similarity measure for description logics reducing inter-concept to inter-instance similarity. In: The 6th Annual European Semantic Web Conference. Volume Lecture Notes in Computer Science 5554., Springer (2009) 353–367
34. Taylor, M., Stokes, R.: Up the creek What is wrong with the definition of a river in new south wales? Environment and Planning Law Journal **22**(3) (2005) 193–211
35. Taylor, M.P., Stokes, R.: When is a river not a river? consideration of the legal definition of a river for geomorphologists practising in new south wales, australia. Australian Geographer **36**(2) (July 2005) 183–200
36. Kuhn, W.: Ontologies in support of activities in geographical space. International Journal of Geographical Information Science **15**(7) (October 2001) 613–631
37. Masolo, C., Borgo, S., Gangemi, A., Guarino, N., Oltrami, A.: Ontology library deliverable d18. Technical report, ISTC-CNR (2003)
38. Kuhn, W.: A functional ontology of observation and measurement. In Janowicz, K., Raubal, M., Levashkin, S., eds.: GeoSpatial Semantics. Volume LNCS5892., Springer-Verlag Berlin Heidelberg (2009) 26–43
39. Cohen, W., Borgida, A., Hirsh, H.: Computing least common subsumers in description logics. In: In Proceedings of the 10th National Conference on Artificial Intelligence, MIT Press (1992) 754–760
40. Baader, F., Sertkaya, B., Turhan, A.Y.: Computing the least common subsumer w.r.t. a background terminology. Journal of Applied Logic **5**(3) (2007) 392–420
41. Janowicz, K., Wilkes, M., Lutz, M.: Similarity-based information retrieval and its role within spatial data infrastructures. In Cova, T.J., Miller, H.J., Beard, K., Frank, A.U., Goodchild, M.F., eds.: 5th International Conference on Geographic Information Science. Volume LNCS 5266., Springer (2008) 151–167
42. Janowicz, K., Schade, S., Bröring, A., Keßler, C., Maue, P., Stasch, C.: Semantic enablement for spatial data infrastructures. Transactions in GIS **14**(2) (2010) 111 – 129