

Getting across information communities

Embedding semantics in the SDI for the Amazon

Patrick Maué · Jens Ortmann

Received: 1 June 2009 / Accepted: 23 September 2009 / Published online: 20 November 2009
© Springer-Verlag 2009

Abstract Amazon's deforestation affects citizens from varying information communities. Experts like scientists or journalists acquire relevant data on-site and publish valuable information on the Web. But different and sometimes conflicting views on reality impede sharing of information across communities, relevant content remains far too often undiscovered. We introduce rule-based semantic annotations as solution to facilitate the discovery and evaluation of geographical information. With the distinction between shared domain and local information source ontologies, the proposed architecture of a semantically supported SDI for the Amazon also takes the plethora of GI formats into account. Creating semantic annotations is challenging, a recommender system for semantic annotations enables even the non-IT experts to participate. The benefits of the proposed techniques are further illustrated by a scenario which spans across information communities of economics, ecology, and ethnology.

Keywords Information communities · Information networks · Rule-based mappings · Semantic annotations · Semantic web · Spatial data infrastructures

Introduction

By avoiding contact with western culture, the Ayoreo-Totobiegosode Indians managed to keep their traditional way of live on their land in northern Paraguay. They co-exist with the surrounding sub-tropical rainforest, which belongs to one of the last remaining parts of the Chaco. The Great Chaco is the second biggest ecosystem on the American continent after the Amazon. And like the Amazon, the Chaco is endangered by ecological destruction (Zak 2004). Due to the growing demand for cheap beef, livestock companies clear the forests to obtain pasture land. Rising popularity of biofuels makes sugarcane plantations profitable. And, although banned by public authority, illegal logging still occurs. The steadily growing population of countries like Brazil is pushing the urban areas into the forests, resulting in the construction of new roads, which again leads to new pastures, plantations, and finally large-scale deforestation. Deforestation is accelerating at an alarming rate: late summer 2008 satellite images were published which depicted how vast areas of the tribe's land had been cleared by a Brazilian cattle ranching company. A plea of the Indian organizations resulted in a withdrawal of the company's working license (Survival International 2009a), but only a few months later the same company is seeking working permission again (Survival International 2009b). Cattle ranching require pasture land, the forests and accordingly the tribe's natural environment would have to be destroyed. Globally acting NGOs like Greenpeace have to lobby policy makers in the local governments to protect the remaining forest, and consequently the existence of Indians.

P. Maué (✉) · J. Ortmann
Institute for Geoinformatics (IFGI), University of Münster,
Weseler Str. 253, 48151 Münster, Germany
e-mail: patrick.maué@uni-muenster.de

J. Ortmann
e-mail: jens.ortmann@uni-muenster.de

We only managed to draw a very rough picture of the current situation of ecosystems like the Amazon and the Great Chaco. Many different processes and external forces have to be considered to understand the reasons for deforestation. The global demand of resources for cheap prices is one factor. But even originally well-intended decisions on a local level, such as the construction of a new road to foster local economy, may have a severe negative impact on broader scales (Davis et al. 2009; Asner et al. 2005). Local settler only burn down small patches of forest to obtain arable land. Selective logging is not considered to be harmful for the forest. But the events and processes initiated by actions, such as the creation of trails by woodcutters to sawmills, have long-term consequences like intrusion of new settlers who use these trails to explore new farmland. Policy makers from the local authorities have to anticipate such consequences, but far too often they are blinded by the promises of the short-term benefits.

The public, supported by globally operating NGOs, has to be involved. Citizens, often experts in their domain, observe and monitor environmental issues and contribute the facts in form of geographical information (GI) on the Web. The ongoing trend of volunteered information, and accordingly the availability of free tools to create such content, results in an abundance of potentially relevant information. The nature of the geographical information source (GI source) usually depends on the intended target community. Intuitive Mashups visualizing complex information on interactive maps address a completely different user than Web Services delivering raw sensor data. In both cases, the data may bear information about temperature. Format doesn't change relevance, it only restricts its applications. In "[Getting across information sources](#)" we further elaborate on the distinction between format and content. The whole spectrum including volunteered GI (Goodchild 2007), georeferenced documents, and traditional spatial data should be considered. Here, we don't want to focus so much on the citizen's role as content producer. Following the idea of public participation GIS (PPGIS) (Schlossberg and Shuford 2005), this paper discusses how the access to GI, regardless of its origin or original structure, can be facilitated. Getting across communities means to address the second-level (Hargittai 2002) *Digital Divide* between the information expert and the computer illiterate. The content—coming in varying forms from varying communities—exists, but remains hidden to users lacking IT expertise.

Spatial data like satellite images or VGI, as well as a news article about the Indian tribe's land, yield information about processes and entities in geograph-

ical space. Content is defined as GI if a spatial and temporal coverage is either explicitly or implicitly (e.g. place names in the text) specified. Geographic Information Retrieval (GIR) (Larson 1996) refers to all tasks needed to find GI on the Web by using spatially-aware search engines. Traditional web search engines are restricted to keywords and indexing algorithms focusing on the text of websites. But for efficient discovery of geospatial data or georeferenced content, the spatial and temporal aspects have to be considered as well. Access to raw data is often managed by Web services with well-defined interfaces standardized by the Open Geospatial Consortium (OGC). Depending on the type of information, standards for requesting and visualizing feature or raster data are available as implementation guides from the OGC website. Embedded in Spatial Data Infrastructures (SDIs) (Nebert 2009), all these Web services are supposed to connect to a network which enables access of GI to all potential stakeholders. A commonly used definition by the US Federal Government explains an SDI as *the technology, policies, standards, human resources, and related activities necessary to acquire, process, distribute, use, maintain, and preserve spatial data* (Office of Management and Budget 2009). Hence, an SDI is more than just an infrastructure for the delivery of professionally created spatial data. It should integrate the interests of stakeholders from many different information communities. And solutions are required which not only simplify the discovery and assessment of data across information communities. We believe the envisioned SDI for a sustainable development of Amazon (Davis et al. 2009) should not only serve, but disseminate relevant geospatial data and georeferenced content to all potential stakeholders.

This paper will not provide a generic approach to enable collaboration across information communities, as deemed necessary (Davis et al. 2009) for an SDI for the Amazon. We are rather going to discuss in detail how semantics help to overcome some of the mentioned problems. We explain how semantic annotations building on top of Semantic Web (Berners-Lee et al. 2001) technologies can help to create an SDI which facilitates the involvement of different information communities, and accordingly the different GI sources. We additionally illustrate how rules in formalized models of community-specific domain knowledge approximate the chains of cause and effect leading to deforestation. Understanding and applying these technologies are quite challenging, the creation of the rules for mappings and semantic annotations requires expert knowledge. Recommender system suggesting appropriate terms from the domain support computer

illiterates to participate in and benefit from an SDI for the Amazon.

The following “[Semantic web for SDIs](#)” details how ontologies and semantic annotations can help to better describe relevant GI. How to get across the different information communities with the help of such technologies is subject of “[Getting across information communities](#)”, why the same approach also supports the integration of all kinds of GI formats is discussed in “[Getting across information sources](#)”. In “[Supporting users to create the rules](#)”, Recommender Systems for semantic annotations are explained. Before coming to the conclusion in “[Conclusion](#)”, we show how the presented approaches can be embedded within an SDI in “[Reasoning across information communities: walking through the scenario](#)”.

Semantic web for SDIs

Usefulness of non-textual information in the Web, such as photos or raw data, relies on textual descriptions. Traditional keyword-based search engines use these annotations for the indexing process. Annotations also enable the content’s evaluation regarding the searching user’s intended use. The descriptions should be flexible enough to take varying needs of different information communities into account. But they also have to be explicit enough to ensure correct and unambiguous interpretation. Using the example of data served by an OGC WFS, this section illustrates how Semantic Web techniques prove to be a suitable approach to meet both these requirements.

OGC standards¹ specify how to access and process spatial data served by OGC Web services. Functional parameters such as the operations and parameters are well-defined. With just the URL of the service as information, desktop clients automatically request service metadata to retrieve an interface description. The format of the provided data, like the Geography Markup Language (GML) for Web Feature Services, is predetermined. Once retrieved, the data can be directly processed and visualized without the need for understanding the encodings or the data models. Standards enable syntactic interoperability between the various components within an SDI. They do not, however, explain what it represents. A process like an interpolation may be invoked automatically on data. It may accept any point-based spatial data, even if attributes used

to compute the interpolated values are missing. The computation will either simply fail. Or even worse, it will produce visually correct, but content wise useless results. Interoperability is not restricted to the syntax, the semantics must be considered as well. This section introduces shared models of domain knowledge as tool to capture the meaning of any GI source, for example raw sensor data from satellites. Spatial data is explicitly grounded in space through coordinates. Free text including implicit place names is considered here as GI source as well, since such georeferenced content may be as relevant as the sensor data. Semantic Annotations establish a link between the GI source and common domain vocabularies well accepted within the particular GI community. Finding ways to mediate between similar or closely related concepts from different domain vocabularies may help to realize an SDI for an Amazon spanning across the many different GI communities.

Semantic conflicts within GI source descriptions

Figure 1 presents an exemplary schema of a GML feature type.² The served data coming from a satellite yields information about deforested areas in Brazil.

The cryptic attribute names make the evaluation of the usefulness of the content a tedious and challenging task. The following list presents some of the more typical examples of semantic conflicts part of this example which can be addressed with semantic annotations:

Application-specific knowledge: Different types of deforestation exist: one might, for example, distinguish between legal or illegal logging and fire clearing or selective logging. The range of values for the `tipo` (eng: type) attribute is application-specific. Without appropriate documentation, the user has no other choice but to request this information directly from the providing organization to understand its meaning.

Community-specific knowledge: The `satellite` (eng: satellite) attribute is in this case an identifier of the satellite which took the source image. The implications, e.g. possible high uncertainty due to low resolution, are unknown to users not aware of the technical specifications of this satellite.

Semantic heterogeneities: Spoken language is ambiguous by nature; contextual information is needed to infer the meaning of the terms. The meaning of `data`

¹More information about the OGC standards mentioned here is available from their website at: <http://www.opengeospatial.org/standards>.

²Such data is, for example, provided by the Brazilian National Institute for Space Research (INPE) by its Real Time Deforestation Monitoring System (DETER, <http://www.obt.inpe.br/deter/>) or Global Forest Watch (<http://www.globalforestwatch.org/>).

Fig. 1 GML schema for deforestation data

```
<complexType name="desmatamentoType">
  <extension base="gml:AbstractFeatureType">
    <sequence>
      <element name="geometry" type="gml:Geometry..." />
      <element name="satelite" type="xs:string"/>
      <element name="data" type="xs:dateTime"/>
      <element name="tipo" type="xs:string"/>
    </sequence>
  </extension>
</complexType>
```



(eng: date) in the example remains unclear: it could refer to the time of the incident, the time of the observation, or even to last modification. In this case it tells when the satellite has made the picture.

Multilingualism: Even though terms in a foreign language may be confusing, the semantics remain the same. In this case, *desmatamento* is the Portuguese translation of deforestation. Dictionaries offer solutions for simple case. But in combination with other semantic conflicts like ambiguities, other approaches are more suitable.

The attributes in the example obviously share several of the described conflicts. *tipo* is also an ambiguous term, and the interpretation of *data* requires application-specific knowledge. But it illustrates that the data provider has to be able to explicitly describe the intended meaning of these elements to make this data useful for people from other information communities. After describing the concept of ontologies as tools to make the interpretation of these terms explicit, we explain how a data provider can make use of semantic annotations to perform this task.


Domain ontologies for shared vocabularies

The community-specific domain vocabulary comprises well-defined terms controlled by a commonly accepted authority. Specific examples of thesauri are introduced in “[Getting across information communities](#)”. Although some thesauri support basic relationships between terms as described in the ISO 2788 (International Organization for Standardization (ISO) 2007), they alone are not expressive enough to address the mentioned semantic conflicts. Being able to map between community vocabularies, including reasoning across domains, requires more sophisticated techniques.

Some years ago Berners-Lee introduced the notion of the Semantic Web (Berners-Lee et al. 2001). He envisioned decentralized extensions to the traditional

Web which give well-defined meaning to its content. Embracing this view, Egenhofer (2002) extended it to the Geospatial Semantic Web, which builds on top of Semantic Web technologies to ease the discovery, integration, and inference on GI. In Kuhn (2005), Kuhn further elaborates on the peculiarity of Geospatial Semantics. On top of unstructured web sites or raw spatial data served by Web services, machine-processable ontologies describe in a well-defined way the concepts used within the data or web site. Ontologies are formal specifications of conceptualizations (Gruber 1995). Guarino (1998) further details this definition by defining the ontology as “an engineering artefact, constituted by a specific vocabulary used to describe a certain reality, plus a set of explicit assumptions regarding the intended meaning of the vocabulary words”. The thesaurus is such a specific vocabulary: it captures a certain view on reality by one particular information community. Sustainable infrastructures depend on user commitment and therefore requires vocabularies users are familiar with. But although they refer to the same entities in reality, these communities have sometimes conflicting views. Deforestation is either regarded as opportunity for industrial development, or as destruction of an ecosystem. Both views are valid and have to be taken into account for the SDI for the Amazon. The following Fig. 2 depicts a domain ontology excerpt which models the concept *DEFORESTATION* and its relation to the concept *ECOSYSTEM*. Note that we refer to common domain vocabularies like GEMET to reflect where the vocabulary originally comes from. Concepts in examples are aligned to the foundational ontology DOLCE (Lehmann et al. 2004), which serves a generic, domain-independent vocabulary. All examples presented here have been created and tested, and are available for download.³

³The ontology files can be downloaded from <http://purl.org/ifgi/~pajoma/amazonSDI/>.

Fig. 2 Domain ontology with the concept deforestation


```

owl:Class: ecology:Deforestation
Annotations:
  rdfs:seeAlso "http://www.eionet.europa.eu/gemet/concept/3406"
SubClassOf:
  ecology:EcosystemDestruction
  dolce:participant only ecology:Forest

ecology:affectedEcosystem
  rdf:type owl:ObjectProperty ;
  rdfs:domain ecology:EcosystemDestruction ;
  rdfs:range ecology:Ecosystem
  
```

In Fig. 2 we use the Manchester OWL Syntax (Horridge et al. 2006). OWL, the Web Ontology Language (Bechhofer et al. 2004), enables the specification of concepts, relations, and constraints on these relations. OWL and its encoding, the Resource Description Framework RDF, are official W3C Standards. Contrary to XML with its hierarchical model, RDF's flexible graph model allows for extending the data without raising compatibility issues. Individual elements in the graph can be uniquely identified using a URI. Triple statements model the relations between these elements. The required seamless integration of remote (loosely coupled) ontologies restricts the use for identifiers to URLs.⁴ Inference algorithms can then move along the relations and jump across different domain vocabularies, making it possible to find relevant GI without even knowing which vocabulary of which community has been used for its semantic annotation. Note that the presented pattern for semantic annotations is independent from the chosen technology. The pattern has been originally developed by Klien (2007) with the Web Service Modeling Language WSMML, an application is discussed in Maué and Schade (2009).

Semantic annotations

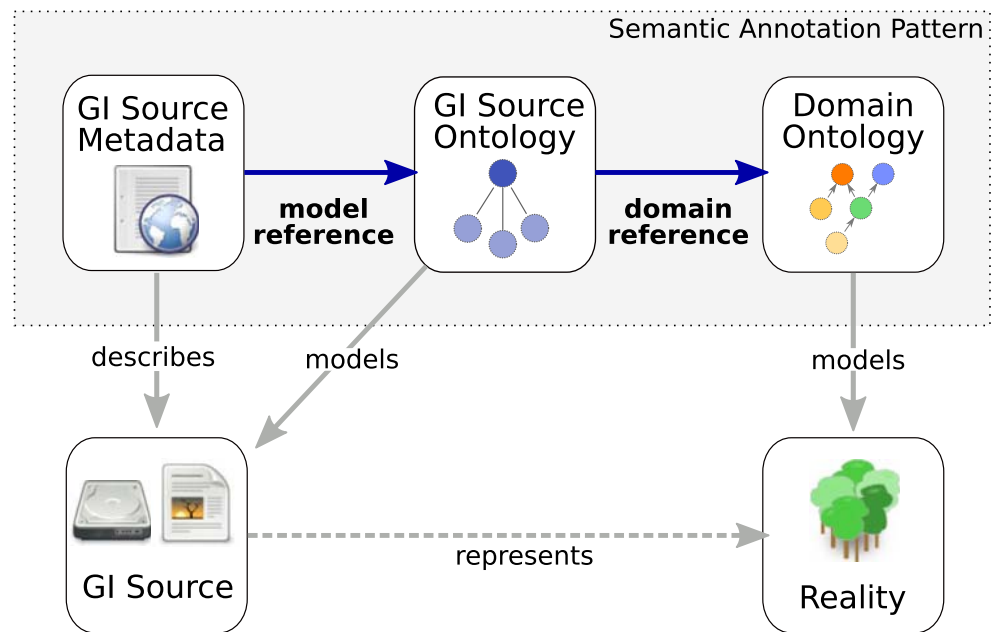
A domain ontology has to be globally accessible on the Web to allow local applications for downloading and integrating the definitions. Local ontologies capture the semantics of one particular application. Local ontologies are aligned to global, shared ontologies; we can then re-use the domain vocabulary to find and evaluate the content referring to it. The local *GI Source Ontology* describes the semantics of one particular GI, such as a website or a Web service's data. Figure 3 illustrates

the *Semantic Annotations* pattern. The semantic annotation is a two-step process, comprising the SAWSDL model reference (another W3C standard (Farrell and Lausen 2007)) and the domain reference. The discovery of information sources depends on structured metadata which describes application-specific details of the GI source. This metadata can be automatically translated into its counterpart, the GI source ontology. The translation result does not capture any real world semantics, its purpose is to (re-)model the information source in form of an ontology. In this form it is processable by inference algorithms. The original metadata is coupled to the according GI source ontology via the model reference. Domain ontologies represent a certain view of reality. They are not explicitly associated with any kind of content; their purpose is to provide a vocabulary which is re-used within the GI source ontology. In the second step of the semantic annotation pattern, the concepts representing the data elements are linked to domain concepts via rules including domain references (more details about this are discussed later). Once a GI source has been semantically annotated, a person is able to understand what the content represents. Reasoning algorithms can support this person to identify appropriate vocabularies and to infer implicit (not explicitly defined) information about the data.

The model reference in the pattern does not directly link to the domain vocabulary. There are various reasons to it, which will become clear in the remainder. GI source ontologies represent the data and its properties (such as quality, encoding, size, access rights, licenses, creators, and more). Domain ontologies capture shared views on a certain, community-dependent reality. Information represents (but is not part of) reality, the domain reference reflects this distinction. In addition, information is usually application-specific (adding it to domain ontologies would unnecessarily clutter the domain vocabulary) or may even be sensitive. We believe using a GI source ontology results in a clear separation between the local application-specific information

⁴This is unfortunately rarely the case. In the GDI-Grid (<http://www.gdi-grid.de>) project we developed a concept repository which assigns URLs to all ontology elements.

Fig. 3 The semantic annotations pattern



focusing on the data, and the global domain-specific information focusing on reality. The following example in Fig. 4 presents an extract of an application-specific GI Source Ontology which re-models the GML schema from Fig. 1. We focus on the attribute with the identifier of the satellite's remote sensor only.

In “[Getting across information sources](#)” we explain the meaning of types like `gml:FeatureAttribute` and how these relations help to support different file formats. Figure 5 shows how the GML schema from Fig. 1 can be extended by a model reference to explicitly link the data to its GI source ontology.

The abstract syntax of the Semantic Web Rule Language (SWRL) (Horrocks et al. 2004) is used in Fig. 6 to define the rules including the domain references. The Horn-like SWRL rules enable direct use of OWL's conceptual model. Since an RDF encoding is supported, the rules can be seamlessly integrated into OWL. A SWRL rule consists of an antecedent and its consequent. The antecedent makes a claim about our ontology. If there is a model that suffices the antecedence—individuals fulfill the requirements of the claim—the rule engine applies the consequent. The consequent contains

Fig. 4 GI source ontology example

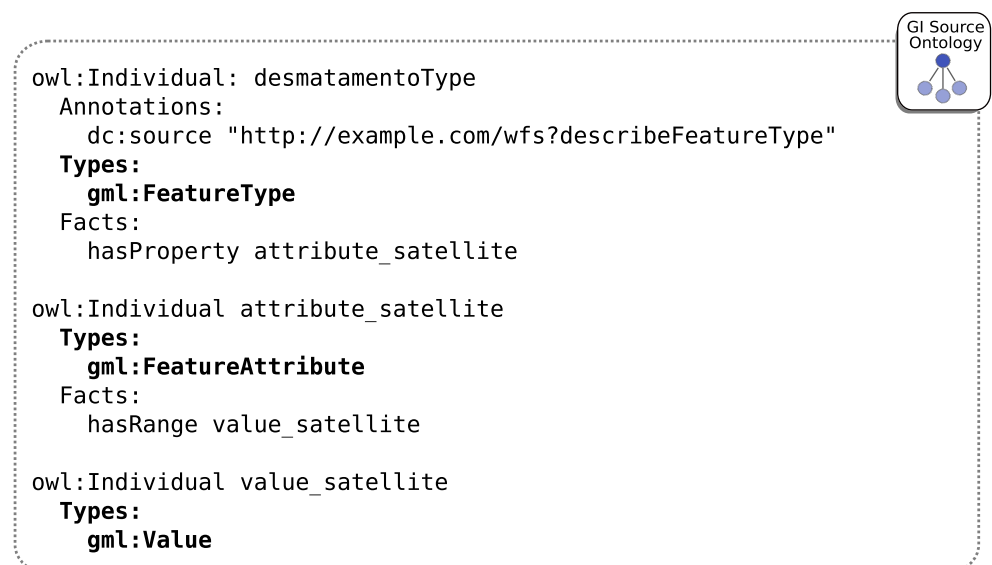
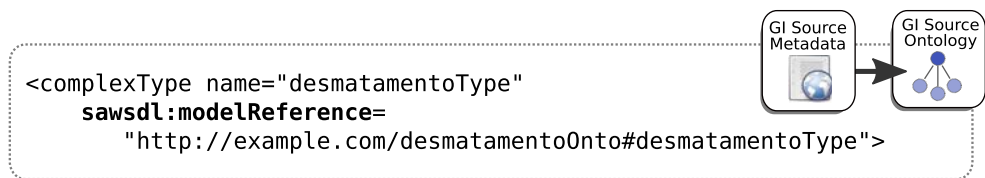


Fig. 5 Extract of GML schema extended to include the W3C SA-WSDL model reference



new facts or axioms that we want to be explicitly stated if the antecedent claim is true. Logically the rule defines an implication between two conjunctions which can be also read as an "if...and...and...then ...and..."-sentence. A rule `Deforestation(?x) ∧ hasParticipant(?x, forest) → Endangered(forest)` can be read as follows: If an individual forest participates (`hasParticipant`) in an instance (`?x`) of Deforestation, then (\rightarrow) this forest has to be classified as Endangered. Note that SWRL comes with the expense of decidability: OWL itself is based on decidable Description Logics (Bechhofer et al. 2004). Including SWRL rules may result in ontologies which cannot be efficiently reasoned on. DL-safe rules restrict the expressiveness of SWRL, and are for example supported by open source reasoners like Pellet (Sirin et al. 2007).

The rules only relate the feature type `desmatamentoType` and its attribute `satellite` to the appropriate terms from the domain vocabulary. The first rule is a direct (1:1) relation. There is one concept in the domain ontology which bears the required meaning, we can therefore directly relate to it. The mentioned distinction between information and reality forbids to simply model the `desmatamentoType` as a special kind of `DEFORESTATION`. The first is a data element encoded in GML, the latter is real world process. Only the *domainReference*-Relation should be used to jump from information to reality. The second rule is slightly more complex. The value of the attribute `satellite` includes the name of the satellite's remote sensor. This example illustrates the big advantage of the rule-based approach. The simple solution of the first rule (which doesn't rely on rules,

OWL alone would allow for such expressions) is not sufficient. Linking the value to a domain concept, e.g. `IDENTIFIER`, does not resolve the semantic conflicts mentioned earlier. A user cannot infer what the identifier refers to, it could be the satellite itself. Rules allow for rebuilding the *inner relationships* of the data model. In this example, we explicitly state that the value of the attribute is the identifier of a remote sensor, which itself is part of a satellite. In the end, this rule-based approach allows for a far more precise discovery and evaluation of data. They are obviously also far more complex to create. In "Supporting users to create the rules" we discuss how a recommender system can support authors to create semantic annotations.

Getting across information communities

We have to consider three different types of actors (expert, citizen, and policy-makers) to understand the complex interactions between the different GI communities (Davis et al. 2009), and how Semantic Web technologies can help to facilitate information exchange across them. The expert, maybe scientist or journalist, is responsible for the observation, analysis, and representation of information. Scientists are responsible for observing and analyzing phenomena, e.g. using satellite images to detect recently cleared patches of forests, and to publish this information in form of reports or scientific articles. Journalist can be described in a similar fashion: they research and write articles about current events and publish them, for example, as articles on the Web. Though not always true, the expert has the

Fig. 6 SWRL rules with domain references

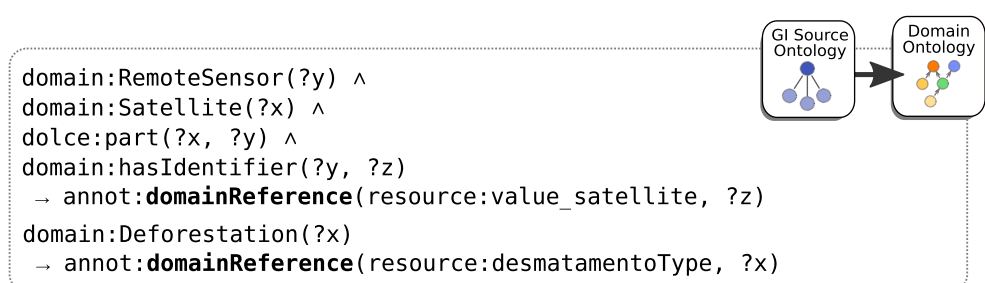
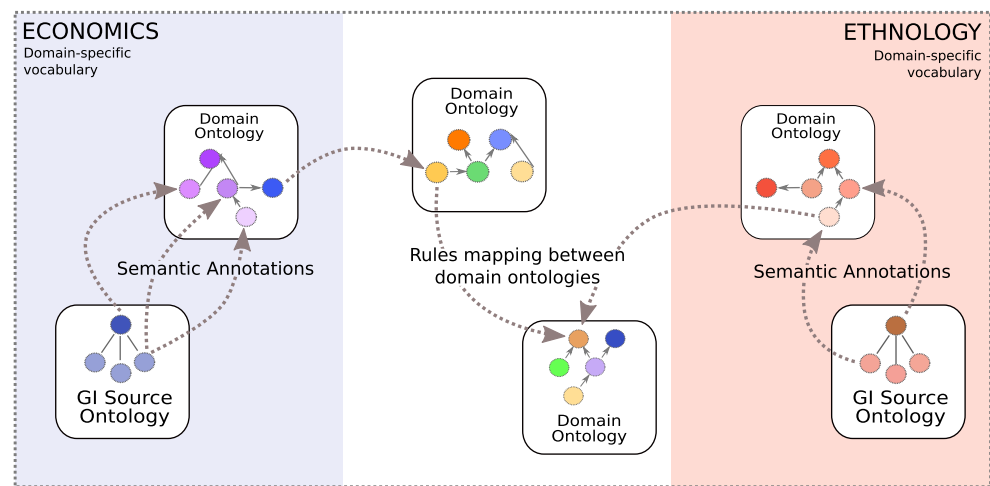


Fig. 7 Connecting information community vocabularies with rules



responsibility to create such information in an objective manner. Target group for published information are the citizens and the policy makers. On the basis of facts within the information, the policy makers are entrusted to make decisions which respect the interests of the majority of the citizens. In the end it is this last group which is affected by decisions, and it mostly comprises the local population with their interest in matters like growing economy or a steady supply of food.

GI communities integrate the various kinds of actors. A journalist specialized in business affairs may write about the current plans of a livestock company to explore new farmland. Scientists study the consequences of economic development on the environment. Their discoveries may then again be used by journalists to illustrate the deforestation trends. In the following examples we focus on three GI communities, which we simply call the economics, ecology and ethnology community. In the case of deforestation, actors of these communities have their own (often conflicting) view on the matter. An economics researcher reporting on the economic development of a region most probably will not use the term “deforestation” as concept to describe his data. But since it is relevant, the environmental policy maker studying destruction of the ecosystems should be able to find it. Getting across GI communities requires techniques like ontologies and rules, which are able to mediate between the different vocabularies.

The usefulness of domain ontologies, and consequently their use for semantic annotations, depends in particular on the used vocabulary. Such vocabularies are controlled by trusted authorities, i.e. international organizations like the United Nations, governmental institutions, or established companies. Examples for controlled vocabularies used by the economics community are the *International Standard Industrial Classifica-*

tion of All Economic Activities United Nations Statistics Division (UNSD 2009) or the *STW Thesaurus for Economy* (German National Library of Economics (ZBW) 2009). On the other side, commonly used thesauri for the environmental domain are the *GEMET General Multilingual Environmental Thesaurus* or the *AGROVOC Thesaurus* (Food and Agricultural Organization of the United Nations (FAO) 2009). The last three are of particular interest for us: they already provide Web Service interfaces to retrieve the vocabulary in RDF. STW and GEMET make use of the W3C Standard SKOS (Simple Knowledge Organization) (Miles and Bechhofer 2009; Miles and Pérez-Agüera 2007). AGROVOC is currently implementing an interface to retrieve its content as OWL ontology with a certain number of well-defined relations (Liang et al. 2006).

It is the GI community’s responsibility to create and serve its specific domain ontology. But one GI community alone cannot create the mediation rules mapping towards other domains. Sophisticated knowledge is needed in both, the domain of discourse and logics. Global community-independent knowledge bases like Wikipedia or Freebase⁵ prove that, as long as its usefulness is apparent, experts from varying fields are motivated to collaborate to create high-quality content. Here we have to assume that these mappings exist, and that repositories on the web serve these rules. Figure 7 illustrates how rules can establish a link between the economics and the ethnology community.

A GI source, e.g. a WFS embedded in an SDI, has been semantically annotated with concepts coming from a domain ontology build on top of the STW

⁵Freely available from <http://www.wikipedia.org/> and <http://www.freebase.com/>.

thesaurus. On the right side, another source has been semantically annotated, this time with concepts from an ethnology thesaurus. The GI source ontology in this case may not only represent spatial data. For semantic data integration, it may also refer to the user's request.

The Semantic Web is seamlessly embedded in the traditional Web. The same applies for the Geospatial Semantic Web or an SDI integrating varying communities. The proposed SDI is not explained by its Web Services or a catalog which serves as entry point. All spatial content on the Web which yields information about one region or theme of interest is implicitly part of the SDI. The semantic annotation linking between one SDI's relevant domain vocabulary and the content can establish an explicit relation. After all, it is not one user's task to identify where to publish data. Once it is somewhere on the Web, crawling algorithms, e.g. deployed by Google, index the data and make them accessible to a broader public. Hence, the semantic-enabled SDI for the Amazon should be defined though its purpose. Its various components including Web Services, RDF repositories for the rules and ontologies, portals for searching content, and more, are simply linked with each other. Figure 7 illustrates the open aspect of the architecture, in this case for the mappings between various domain ontologies.

In “[Semantic conflicts within GI source descriptions](#)”, hierarchical knowledge (e.g. too application- or too community-specific knowledge) has been identified as one potential source for semantic conflicts. In the given example, we argued that a given name of the remote sensor of a satellite yields information about the data resolution. This expert knowledge unnecessarily restricts the potential audience of this particular information. Users simply searching for high-resolution data without explicitly specifying the remote sensor will either fail to discover it, or will not be able to evaluate its fitness for use. He might use the following rule to specify a query: `Coverage(?x) ∧ hasResolution(?x,high) → sqwrl:select(?x)`. The built-in predicate `sqwrl:select` in the head (consequent) returns all `?x` which match the criteria in the body (antecedent). The user only requests coverages `?x` with a high resolution, the content semantics are not defined here. If the data has been semantically annotated using the rules from Fig. 6, a reasoner supporting SWRL can re-classify the content. The rule `RemoteSensor(?a) ∧ hasIdentifier(?a,?b) ∧ swrlb:matches("CCD") → hasResolution(?a,high)` defines, if the identifier of a remote sensor has the value “CCD”, then its resolution can be classified as high.

The feature type named `desmatamentoType` has been semantically annotated with the domain concept `DEFORESTATION`. The concept serves information about recently detected patches of cleared forests in the Amazon region. In a second example we detail how rules can help to span across the individual community vocabularies. The rule `gemet:Deforestation(?x) → annot:domainReference(desmatamentoType, ?x)` creates an explicit link to the GEMET (European Environment Agency (EEA) 2009) vocabulary from the environmental community. But researchers studying influencing factors for the region's economic development would rather use terms like “logging” or “forestry” for describing (and searching) such content. The following rule models an equality relation: `stw:Logging(?x) ∧ stw:Forest(?y) ∧ stw:operatesIn(?x,?y) → gemet:Deforestation(?x)`. It basically states that a forest participates in the process `Deforestation` if it (the same individual) also participates in the industrial activity `Logging`. If a user requests GI semantically annotated with `gemet:Deforestation`, the reasoner-supported query processing algorithm will then also consider GI described with `stw:Logging` to be potentially relevant. The rule is obviously over-simplifying the relation between the two terms. `Logging` as industrial activity is performed in all parts of the world, and in many case in a sustainable manner. `Deforestation` as process is a long-term degradation of woodlands driven by illegal logging, forest fires, or forest clearing to acquire farmland. But a perfect representation of the explicit and implicit relations between the terms of different vocabularies cannot be the goal of this approach. An automatic discovery, evaluation, and integration of information, is not the intent of this approach. It should rather enlarge the number of potentially relevant GI (the recall); it remains in the responsibility of the user to evaluate its suitability. To reach semantically correct rules, if possible at all, requires experts from the various communities to collaborate. This is a challenging task, which may exclude non IT experts, including the majority of citizens and policy makers, from the process. But to realize our vision to deliver information to all interest groups, regardless their cultural and educational background, we have to acknowledge that inconsistencies and imprecise definitions will be part of the domain ontologies. Hence, semantic components in the Amazon SDI have to deal with uncertain and inconsistent knowledge.

This section gave an overview over the different actors and GI communities representing the users of an SDI. The different backgrounds and interests of the

various GI communities can result in different, sometimes conflicting domain vocabularies. GI community-specific organizations control vocabularies which have to be used to build the domain ontologies used for the semantic annotations. Two examples are presented which show how rules can bridge the gap between these different domain ontologies.

Getting across information sources

The approach to integrate different communities has to consider the plethora of sources for potentially relevant information as well. The selected GI format usually depends on the author's community. Scientists often distribute discoveries as raw observation data, VGI is usually published as OGC KML files, and by far the most prevalent form in the Web is implicitly referenced GI such as news articles. In this section the problems of heterogeneity among the different forms of information sources are discussed. Semantic interoperability is mainly regarded as a problem of heterogeneous content. In the following we explain why this view is not sufficient and why we additionally have to consider the semantics of content and form.

For the discovery of GI representing deforestation on the Amazon it makes a difference whether you find a report that describes the development in the last decade, a diagram that visualizes this development, or an OGC Web Mapping Service presenting the annual progress of deforestation as a map. The different representations serve different needs. A written report may highlight certain aspects or certain problematic areas. A map provides an overview of the mapped area without putting emphasis on certain details. A diagram is suitable for stressing the temporal aspect, i.e. presenting a development over time. We argue the variety of GI forms has to be taken as seriously as the already addressed variety in describing GI content.

Linguists and communication theorists have discussed the duality between form and content for a long time. Platon reflected on the duality of form and content of single signs already more than one and a half millennia ago (Nöth 2000). Modern semiotics usually refers to the studies of content and form of De Saussure (2001) and Peirce (1966). In communication science the duality of content and form is analyzed at the scale of whole messages. The message requires an expression, the content, and a medium, the form (Maletzke 1963). An idea in your mind, coming for example from observing an event in your environment, can be regarded as the content of your message. Communicating this idea requires to have it expressed according to a language.

The language is a semiotic code that is shared across a community and especially between the sender and the recipients of the message (Eco 1991). The chosen language determines the form. Expressions in a natural language usually result in a text which is read and interpreted by the recipient to gain information. Formal languages result in other forms of information sources, formulas and equations are for example expressed in a certain mathematical notation. Graphic expressions like pictures, drawings, diagrams and graphs also use a certain semiotic code to convey information (Nöth 2000).

GI sources have always been available in a wide variety of forms, simply because maps provide an additional, rich and expressive way to represent spatial information. Furthermore, maps can be easily combined with other forms like diagrams, illustrations and texts. The implications of different forms of GI sources were analyzed by Bertin (1967) more than 40 years ago. The Web gave rise to many new tools facilitating the creation of information. In the same time it also fostered rapid dissemination of ideas. Public email lists, blogs, wikis and personal websites are only some examples of the many new forms of information sources. Publishing is no longer a formal act restricted to a limited circle of people. Publishing information to communicate your ideas to a global audience is nowadays as normal as talking about it with your neighbor. New tools fostering social interaction and collaborative work resulted in a Social Web (Dickson 2001) which has the potential to enable public engagement on far larger scale as it has been anticipated in the beginnings of PPGIS research. The Brazilian environmentalist's blog describing recent developments in the Amazon may be read by people coming from all over the globe. The same accounts for the articles of a German biodiversity expert writing about the impact of these developments on the wildlife. Only ten years ago this local knowledge was nearly inaccessible, participation was limited to experts with immediate access to the information. In the Social Web, collaboration across information communities is slowly becoming reality. Merging Social Web and Semantic Web will even accelerate this phenomenon (Gruber 2008).

The form of GI is often considered to be an indicator for certain characteristics. VGI as most prevalent content in the Social Web is usually created by non-experts, which has a significant impact on potential applications for such content. Aspects like reliability, uncertainty, timeliness, or the author's credibility play an important role if the usefulness of VGI for a certain application has to be assessed. But in most cases explicit information about these quality properties is missing, either due to

the lack of support by the format, or because the author was not aware of its importance. These parameters are hard to assess and rarely communicated. Consequently, the value of VGI for location-aware applications is disputed by many. But projects like Wikipedia and OpenStreetMap or the wide-spread usage of open source software illustrates that volunteered contributions can be of high-quality and reliable enough for even professional applications. The problem of evaluating quality can actually be extended to every kind of GI source. Well-documented GI compliant to OGC standards is still useless if no information about the publisher is included. But if crucial information such as the author's background is included into the GI source ontology (which is independent from the GI's form), the assessment of the GI's usefulness regarding its fitness for use is separated from its original form.

The quality may be an aspect of the content, but information about the form is still needed to understand how to use the information. As described in “[Semantic web for SDIs](#)”, domain ontologies are community-specific and are built on the community's vocabulary. The vocabulary used to describe the special characteristics of the form, such as encoding or language, as well as the quality aspects are also described in domain ontologies. But in contrary to domain ontologies representing a certain view on reality, a domain reference as separation between information and reality is not needed here. Figure 8 illustrates how the GI source ontology extends a specific domain ontology capturing the semantics of form and content, e.g. a GML ontology to describe the encoding. Taking the example of the GI source ontology in Fig. 4, a user might take the following rule to specify desired content: `annot:`

`domainReference(?y,?x) ∧ Deforestation(?x) → sqwrl:select(?y)`. By simply adding some details, he can also constrain the form and a certain quality aspect: `hasResolution(?y, high) ∧ annot:domainReference(?y,?x) ∧ Deforestation(?x) ∧ gml:Coverage(?y) → sqwrl:select(?y)`.

In this section we introduced the need for a clear separation between form and content. The domain reference acts as bridge between these two aspects of GI. The form has to be described in the GI source ontology as well, since information such as encoding or quality are important to assess the GI's usefulness. But assessment requires a shared vocabulary again. Specific domain ontologies capturing information-specific properties have therefore to be used to create the GI source ontology. If we are able to embed this approach (including the semantic annotations) into the SDI for the Amazon, the vision of discovery and integration of heterogeneous content and heterogeneous form may become reality.

Supporting users to create the rules

Already the selection of appropriate concepts from the domain vocabularies makes semantic annotations a challenging task, which often depends on expert knowledge in the particular domain as well as in formal logics. Even more complex is the specification of the rules in SWRL. The content provider, though expert in his particular field, may find such a task too demanding. Easy creation of semantic annotations and rules depends on tools which can help to find correct concepts and accordingly to specify the rules. In this section we introduce how a Recommender System (RS) can support users.

RS render suggestions for items, in this case the concepts from a domain ontology, based on certain similarity metrics. Collaborative filtering (CF) (Manning et al. 2008) techniques analyze user habits, for example by computing similar user profiles and recommending items from a similar user's collection. The similarity of the (geospatial) data is considered for Content-based (CB) recommendation techniques (Ramezani et al. 2008). In the case of rules for semantic annotations, a RS can assume that concepts are appropriate if very similar content refers to them. For the specification of the mappings in between domain ontologies, already existing annotations and rules can provide the basis. In “[Reasoning across information communities: walking through the scenario](#)” we illustrate how the suggested methods can be embedded in the Amazon SDI. The RS

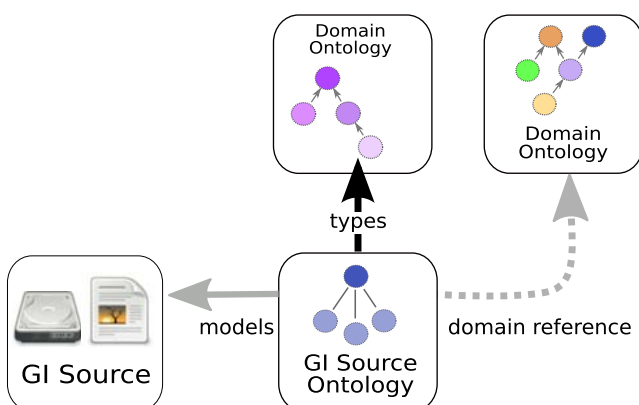


Fig. 8 GI source ontology both extend and are referenced to domain ontologies

module is part of the component which supports the GI publication, e.g. as news article on a newspaper's website. The search engine also has to include the RS to create the rules describing the required content.

According to the discussed scenario, GI sources such as raw spatial data or unstructured, but georeferenced content have to be annotated. The rules of the semantic annotations are part of the GI source ontology, which itself is automatically created from the GI source metadata. In the case of unstructured, but textual content such as a news article, the structured metadata has to be created first. Text mining (Tan et al. 2006) techniques (e.g. based on Formal Concept Analysis (Cimiano et al. 2005)) can analyze free text and extract key concepts. The result is a list of terms (often already with assigned categories such as Person, Place, Company); an example is the viewer of the Calais Web Service⁶ maintained by Thomson Reuters. The resulting collection of terms is obviously incomplete and inconsistent, e.g. only few of the extracted place names really refer to place covered by the text, and even the best mining algorithm can not extract all relevant concepts. In the end manual input by the author is required to complete and clean the list as preparation for the transformation into the according GI source ontology. In the case of spatial data such as satellite images, structured metadata is usually automatically created and extended by the providing organizations.

With the support of the author, this process can be performed without directly interacting with the ontology. Selecting, deleting, and adding key terms is not necessarily different from the tagging process which is commonly applied to content in the Web, such as bookmarks on Delicious or photos on Flickr.⁷ The extracted word lists represent an excerpt of the author's vocabulary with all its inconsistencies, ambiguities, and bias towards his personal background. In the next step, the author needs to link the key terms to the concepts from the domain vocabulary. For this step, a recommender system analyzes the existing GI source metadata, and retrieves similar content in the catalog. In the case of GI, the similarity measurement has to consider all three dimensions: space, time, and theme (Kuhn 2005). Spatial or temporal similarity, e.g. both GI sources cover the Amazonas region, is not easy to compute. We assume spatial or temporal similarity if two GI source roughly (with a scale-dependent degree of error) occupy the same region on the spatial or

temporal dimension. Thematic similarity, e.g. both GI sources show locations of ongoing deforestation, cannot be computed automatically due to the mentioned semantic conflicts. But the co-occurrence of terms in different word lists can indicate similarity, which again may indicate that both sources can be annotated with the same or closely related domain concepts. The nature of recommendations depends on the dimensions the similarity has been computed for. Computing spatial similarity, but no thematic similarity, should for example only render recommendation for place names.

In most case a direct link as in the first rule of Fig. 6 may be sufficient. For more complex rules which include relations between the concepts, a more sophisticated solution is required. User interfaces can, for example, represent rules in a controlled, but easily understandable language. In the SWING project we have studied the use of graphs⁸ in combination with intuitive interfaces to perform this task (Grčar 2008). But discussing how innovative user interfaces supported by recommendations can help users to create even complex rules (and consequently support and encourage even non-IT experts to semantically annotate their content) remains subject for future research. In this section we briefly discussed how a RS module can support even IT novices to find and publish GI in the Amazon SDI. The expressiveness of Semantic Web technologies can facilitate collaboration across information communities, but we need to find ways to abstract from the underlying complexity to make the benefits accessible to all parts of the population.

Reasoning across information communities: walking through the scenario

In the introducing scenario the cattle company's decision to clear the rain forests on the Ayoreo-Totobiegosode tribe's land was prevented by public authorities. The Paraguayan government withdrew the working permit for the company on this land (Survival International 2009a). Just recently (Survival International 2009b), the company requested again the license to clear the land to turn it into pasture land. The request's approval is endangering the ecosystem, including the wildlife as well as the Indian tribe. We mentioned in "Introduction" that an SDI for the Amazon could better support a sustainable development of the environment, if we expand our understanding on its

⁶Try it out here: <http://viewer.opencalais.com/>.

⁷Available at <http://www.delicious.com> and <http://www.flickr.com>.

⁸A video demonstrating the interface for creating such complex annotations is available at: <http://swing-project.org/showcase.html>.

original purpose. An SDI should not only include components serving GI, it should also be able to actively disseminate it to all potential interest groups. Such an active delivery has to span across the various information communities, regardless of the individual's cultural or educational background. In the remainder of this section we illustrate how we envision an early warning system embedded into the SDI build on top of the introduces rules. The presented rules approximate chains of cause and effect, and can help to understand how processes eventually endanger the tribe's existence.

Catalogs enable the registration and discovery of OGC Web Services. Most GI, including implicitly geo-referenced content like news articles as well as volunteered GI, is not included in such registries. A journalist from a local newspaper may just have written about the company's request, the information is then published on the newspaper's website. The website's target audience is restricted to the local population. But perhaps it is the newspapers policy to use semantic annotations to better categorize their own content, including explicit references to place names and concepts from an economics domain ontology. Semantic annotations are embedded into the content's metadata, a crawler (Manning et al. 2008) which is continuously searching for websites with model references does eventually index the article.

On the other side, a computer expert working for the NGO Survival International has setup a news feed which continuously publishes information about timely events which may be potentially endangering the indigenous tribes. The feed is based on GeoRSS (OGC 2009), the events are restricted to the territory of the Ayoreo-Totobiegosode tribe. The feed, used as early warning system for the organization, is created automatically. The computer expert has defined SWRL rules which states that any published GI source with content which can be considered as threat for the Indians should appear as item in the feed. Once the crawler has indexed the mentioned semantically annotated news article, the rules presented in the following section are triggered. Only seconds later, the just published article has been disseminated to all people who have subscribed to the news feed.

The traditional components of an SDI can obviously not perform these tasks. SWRL rules, domain ontologies, and GI source ontologies are encoded in RDF and can be served by traditional RDF repositories.⁹ The RDF encoding enforces URIs as identifiers for the rules

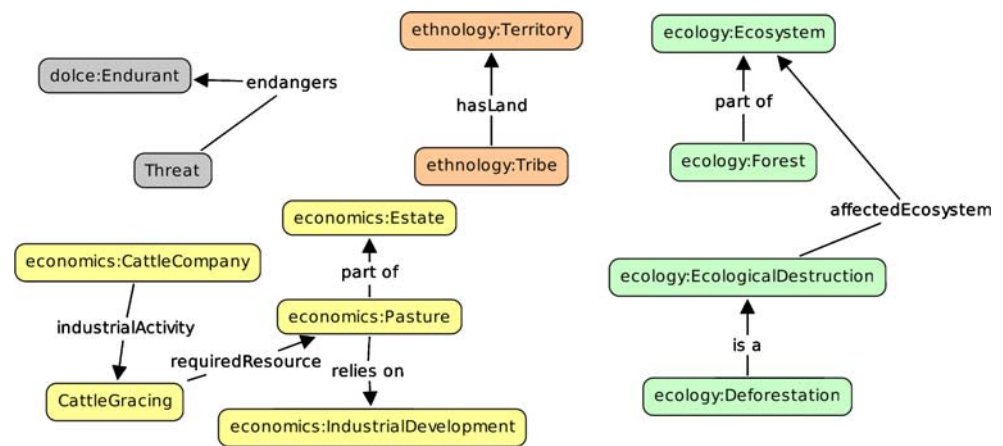
or concepts. Constraining this further to URLs enables inference algorithms working with the rules to simply move along the graph of RDF elements coming from different repositories (the URL of the next element can be resolved to the actual RDF code). The interfaces including the recommender system for semantic annotations and decidable rules can be integrated as modules into existing community websites. Created in a loosely-coupled, decentralized fashion, this network of RDF repositories and community catalogs can be seamlessly integrated into an existing SDI. The rules, which may trigger the information flow across the communities, are processed by reasoner-supported components integrated into existing applications. Hence, from a technological perspective the idea of embedding the proposed methodology into an existing SDI for the Amazon is feasible. Creating the intuitive interfaces for creating the rules and semantic annotations is far more challenging and remains subject for future research.

Still, the process of creating a rule to approximate chains of cause and effect is straightforward. The original fact is defined in the antecedent of our rule and the inferred statement is specified in the consequent. The facts in our case are the individuals modeled in the GI source ontologies. Once registered in the knowledge base, the rule engine is invoked and processes the new rules. The individuals from the local ontologies are related via rules to the domain ontologies. For the scenario we assume the existence of three community ontologies representing the domains of economics, ecology, and ethnology. The following concept map in Fig. 9 presents some of their concepts as well as the concept THREAT which can be considered domain-independent. All concepts are aligned to the foundational ontology DOLCE¹⁰ (Lehmann et al. 2004), concepts from this ontology are marked with the prefix DOLCE:. One of the CATTLE COMPANY'S INDUSTRIAL ACTIVITIES is the PRODUCTION OF CATTLE, which RELIES ON sufficient supply on food available on PASTURE LAND. The meadows have to be part of the company's ESTATE, and RELIES ON the INDUSTRIAL DEVELOPMENT of the existing land into pasture land. In the ecology ontology, the concept DEFORESTATION is modeled as A KIND OF the concept ECOLOGICAL DESTRUCTION with FOREST as AFFECTED ECOSYSTEM. For the scenario, only the relation between a TRIBE and its TERRITORY has to be specified.

The geographical region is one common denominator which can be used to map in between the domain ontologies. In particular this means that we consider

⁹Like the open source repository Sesame available at: <http://www.openrdf.org/>.

¹⁰Documentation and Ontology can be downloaded at: <http://www.loa-cnr.it/DOLCE.html>.

Fig. 9 Concept map of the domain ontologies

a particular instance representing the geometry of the economy's ESTATE to be also an instance of ECOSYSTEM from the ecology ontology. This rule is obviously oversimplifying the relation between real property and ecosystem. The rule should actually include means to check if the geometry of the estate is spatially within or overlapping with existing ecosystems like the Great Chaco. As already mentioned, SWRL supports built-in predicates which can be easily extended (O'Connor et al. 2007). Spatial extensions, which delegate the comparison of GI source geometries to spatial algorithms, is therefore feasible, but unfortunately not yet available. As depicted in Fig. 10, the mapping rules should (if appropriate) be modeled in both directions. The second rule states that an ecosystem is also the tribe's territory. Again, this rule does not model all needed aspects: the tribe's territory is probably only a small part of the ecosystem. Note that completeness (if it can be even reached) is not necessary to reach to intended goal of a warning system. The rules should only approximate causal relations, they should (and can) not model all aspects of the complex relationships between different information communities.

The first two rules in Fig. 11 model that the need for pastures may require industrial development (e.g. clearing primeval forest) of the company's estate. The intended meaning of the rules is included. In the third

rule, we jump from the economics to the ecology ontology. An early warning system for deforestation threats may already issue the first messages to potential subscribers. Here, another rule is triggered to bridge the gap between the ecology and ethnology vocabularies. A DEFORESTATION instance has been asserted to affect an ecosystem which is also home to the Indian tribe. This is modeled as direct threat to the Indians. The consequent of the last rule is just what the NGO's computer expert is interested in. Since he specified something like `Tribe(?a) ∧ endangers(?c, ?a) → sqwrl:select(?c)` for the creation of the news feed, the rule engine can add the GI source which triggered the whole chain as new item to his news feed.

Our use case has shown how we can reason across domains. If the local news article (usually restricted to a local audience) has been semantically annotated, it can be immediately disseminated to a global readership. All sorts of GI, including professional sources like raw sensor data, volunteered GI and simply georeferenced content, can benefit from the approach as long as structured metadata is made available and GI source ontologies can be generated and semantically annotated. In this section, we walked through the scenario discussed in the introduction. Altogether, we presented three examples where we believe rule-based semantic annotations can be applied to avoid some of the more

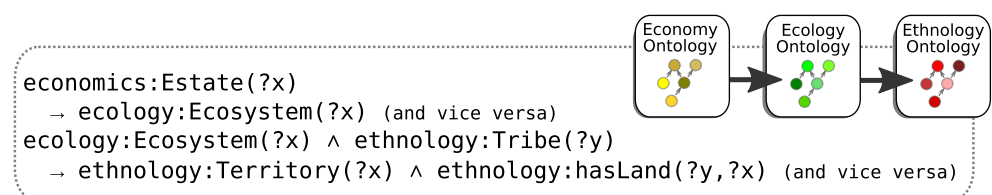
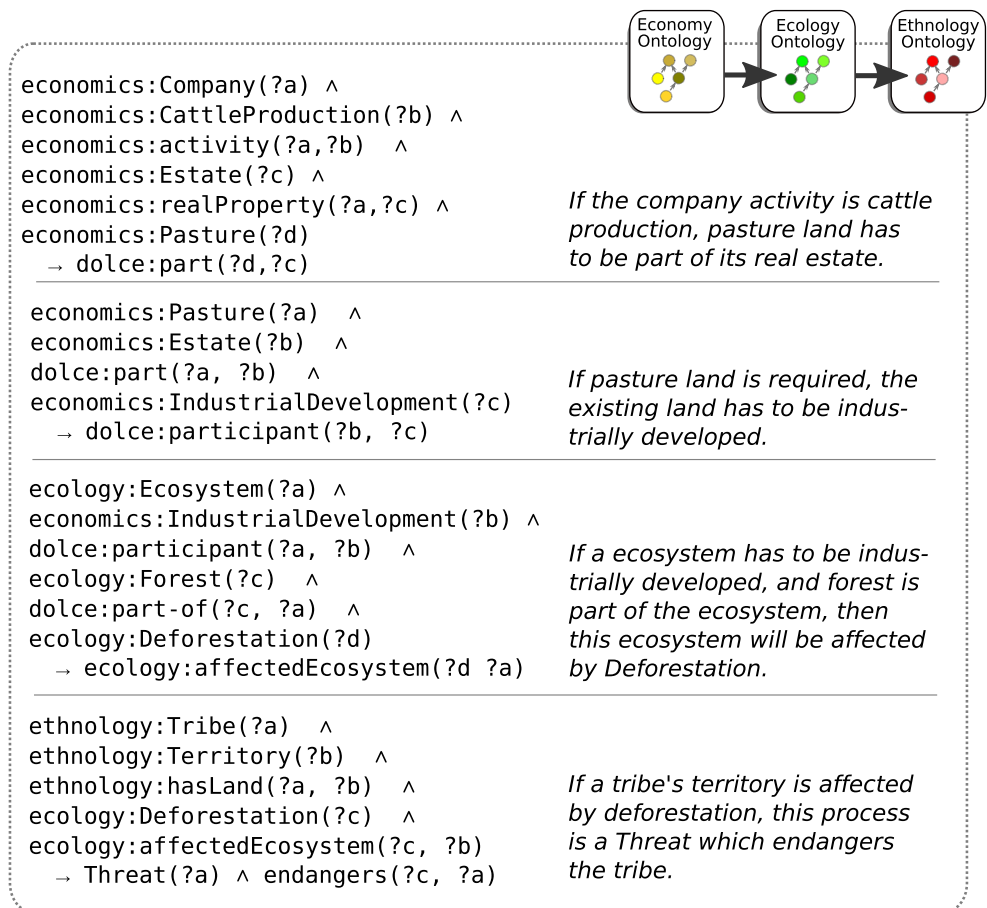
Fig. 10 Mapping spatial similarity between economy, ecology and ethnology ontology

Fig. 11 Rules to deduce new facts across domain ontologies



typical semantic conflicts. The lack of semantic “interoperability” between different information communities is one important factor for the conflicting views on reality. The presented methods help to close the vocabulary gaps between the communities, and could contribute to the SDI for the sustainable development of the Amazon.

Conclusion

We introduced an approach to embed semantics into SDIs. The methods support the integration of various GI sources from several communities provided by different actors. Moreover, the solution is not limited to semantic data integration. An early warning system has the capability of an active dissemination of content to users with potential interest in the information. The rule-based approach for mappings and semantic annotations enables deduction of new facts across domains, making it possible to users from other communities to gain access to this information as well.

We illustrated how currently available Semantic Web technologies can give well-defined meaning to content served in SDIs. Policy-makers, citizens, and experts can benefit from these extensions. Most spatial data created by official institutions already has metadata attached. These annotations provide specifications of form and sometimes even content. (V)GI contributed by other actors is often published in less structured form. Structured metadata of such GI can be extracted; semi-automatic solutions to create annotations that commit to a certain domain vocabulary exist. Domain ontologies are formal specifications of shared domain knowledge. Created and managed by domain experts within authoritative organizations, they serve as basic vocabulary for the semantic annotations. The annotation of a GI source is turned into a semantic annotation, if a link between the metadata and the concepts formalized in domain ontologies is established. We explained why SWRL rules including the domain references have to be used for this task. SWRL rules also help to bridge the vocabulary gap between different domains. Different information communities specify their knowledge in individual domain ontologies.

However, they share concepts, even though varying regarding their intended use. These shared concepts can be mapped, which enables reasoners to infer facts across different domain ontologies.

However, the presented approach based on SWRL has its deficiencies and we had to take some shortcuts. The mappings between domain vocabularies of different GI communities take the spatiotemporal dimension into account. Integrating this aspect into SWRL is rather complex and questionable. The presented methods assume that spatiotemporal processing (e.g. filtering out content with differing spatial extent) is applied beforehand. Furthermore, modeling causal relations and emerging effects poses a problem. Since decidability is crucial for the applicability of this approach, the rule's consequent must not contain individuals which are not already defined in the antecedent. Since rules embedded in the SDI are semi-automatically created by tools like the discussed recommender system, we believe we can ensure decidability even with hidden extensions which better reflect causal chains. These tools are required anyway, since the creation of rules and semantic annotations is, despite many approaches to facilitate it, still a tedious task.

Knowledge changes continuously, the same is valid for the domain ontologies. Rules linking between vocabularies deteriorate accordingly, inconsistencies due to changes in the domain ontologies have to be expected. Versioning and regular consistency checks can partially address this, but the challenge of keeping the rules up-to-date has to be addressed in future research. Additionally, a user's context influences his decisions during his interaction with the ontologies, such as the selection of the appropriate concepts for describing published content. It remains an interesting open question how the user's individual situation drives his decisions. Does a user's cultural identity predefine the language she selects for describing the content? Does a person have multiple identities (like being a conservationist, employer, and mother) simultaneously and tries to balance them against each other, or does she select the most appropriate role?

Yet, we argue that the benefits will, once recognized, convince information communities. The potential benefits for non-governmental organizations with interest in human rights or the environment have been illustrated with a scenario. The success of the SDI for a sustainable development of the Amazon depends largely on the participating information communities, which should not be restricted to scientists and organizations. The local population, the citizen in this paper, is largely underrepresented in traditional SDI. But community projects like Wikipedia have proven that,

if enough people are willing to share and collaborate, surprisingly high-quality (often exceeding the expert's work) content is published for free on the Web. Why shouldn't there be similar potential for the required formal specifications of knowledge?

Acknowledgements This research has been funded by the the German GDI-Grid project and through the International Research Training Group on Semantic Integration of Geospatial Information by the DFG (German Research Foundation), GRK 1498. The idea for this paper was born during the bilateral workshop between IFGI and INPE in March 2009 at Ilhabela, Brazil, we would like to thank all participants for their invaluable input. The same applies to the reviewers of this article. Some artwork is part of the public domain Tanco Icon Library (<http://tango.freedesktop.org/>).

References

- Asner GP, Knapp DE, Broadbent E, Oliveira PJC, Keller M, Silva JN (2005) Selective logging in the Brazilian Amazon. *Science* 310:480–482
- Bechhofer S, van Harmelen F, Hendler J, Horrocks I, McGuinness D, Patel-Schneider PF, Stein L (2004) OWL web ontology language reference. W3C recommendation, W3C
- Berners-Lee T, Hendler J, Lassila O (2001) The semantic web. *Sci Am* 284(4):34–43
- Bertin J (1967) *Semiologie graphique: Les Diagrammes, Les Reseaux, Les Cartes*. Mouton and Gauthiers-Villars, Paris
- Cimiano P, Hotho A, Staab S (2005) Learning concept hierarchies from text corpora using formal concept analysis
- Davis Jr CA, Fonseca FT, Câmara G (2009) Beyond SDI: integrating science and communities to create environmental policies for the sustainability of the Amazon. *International Journal of Spatial Data Infrastructure Research (IJSDIR)* (2009) Article under Review
- De Saussure F (2001) *Grundfragen der allgemeinen Sprachwissenschaft* 3rd edn. Walter de Gruyter, Berlin
- Dickson D (2001) Weaving a social web. *Nature* 414(6864):587
- Eco U (1991) *Semiotik: entwurf einer theorie der Zeichen* 2nd edn. Wilhelm Fink Verlag, München
- Egenhofer MJ (2002) Toward the semantic geospatial web. In: *GIS '02: Proceedings of the 10th ACM international symposium on Advances in geographic information systems*. ACM, New York, pp 1–4
- European Environment Agency (EEA) (2009) General multilingual environmental thesaurus. <http://www.eionet.europa.eu/gemet>. Accessed May 2009
- Farrell J, Lausen H (2007) Semantic annotations for WSDL and XML schema. W3C candidate recommendation, W3C
- Food and Agricultural Organization of the United Nations (FAO) (2009) AGROVOC thesaurus. http://www.fao.org/aims/ag_intro.htm. Accessed May 2009
- German National Library of Economics (ZBW) (2009) STW thesaurus for economics. <http://zbw.eu/stw/versions/latest/about.en.html>. Accessed May 2009
- Goodchild M (2007) Citizens as sensors: the world of volunteered geography. *GeoJournal* 69(4):211–221
- Grčar M (2008) Software module for semantic annotation of a web service. SWING Deliverable 4.5, Jožef Stefan Institute

- Gruber T (2008) Collective knowledge systems: where the social web meets the semantic web. *Web Semantics: Science, Services and Agents on the World Wide Web* 6(1): 4–13
- Gruber TR (1995) Toward principles for the design of ontologies used for knowledge sharing. *Int J Hum-Comput Stud* 43(5–6):907–928
- Guarino N (1998) Formal ontology and information systems. In: Guarino N (ed) *Proceedings of FOIS'98*, 6–8 June 1998. IOS, Trento, pp 3–15
- Hargittai E (2002) Second-level digital divide: differences in people's online skills. *First Monday* 7(4)
- Horridge M, Drummond N, Goodwin J, Rector AL, Stevens R, Wang H (2006) The manchester OWL syntax. In: Grau BC, Hitzler P, Shankey C, Wallace E, Grau BC, Hitzler P, Shankey C, Wallace E (eds) *OWLED*, vol 216. CEUR Workshop Proceedings, CEUR-WS.org
- Horrocks I, Patel-Schneider P, Boley H, Tabet S, Grosz B, Dean M (2004) SWRL: a semantic web rule language combining OWL and RuleML. W3C (submission)
- International Organization for Standardization ISO (2007) ISO 2788:1986, documentation—guidelines for the establishment and development of Monolingual Thesauri. Multiple. Distributed through American National standards Institute (ANSI)
- Klien E (2007) A rule-based strategy for the semantic annotation of geodata. *Trans GIS* 11(3):437–452
- Kuhn W (2005) *Geospatial semantics: why, of what, and how?* Lecture notes in computer science, vol 3534. Springer, New York
- Larson RR (1996) Geographic information retrieval and spatial browsing. *GIS and libraries: patrons, maps and spatial information* 81–124
- Lehmann J, Borgo S, Masolo C, Gangemi A (2004) Causality and causation in DOLCE. In: Varzi AC, Vieu L (eds) *Formal ontology in information systems, proceedings of the international conference FOIS 2004*. IOS, Amsterdam, pp 273–284
- Liang A, Lauser B, Sini M, Keizer J, Katz S (2006) From AGROVOC to the agricultural ontology service/concept server. An OWL model for managing ontologies in the agricultural domain. In: *OWL workshop*
- Maletzke G (1963) *Psychologie der Massenkommunikation: theorie und Systematik*. Hans Bredow-Institut
- Manning CD, Raghavan P, Schütze H (2009) *Introduction to information retrieval*. Cambridge University Press, Cambridge
- Maué P, Schade S (2009) Data integration in the Geospatial semantic web. *Journal of Cases on Information Technology* 11:100–122
- Miles A, Bechhofer S (2009) SKOS simple knowledge organization system reference. W3C Candidate Recommendation, W3C
- Miles A, Pérez-Agüera J (2007) SKOS: simple knowledge organisation for the web. *Cat Classif Q* 43(3):69–83
- Nebert DD (2009) *The spatial data infrastructure cookbook*. <http://www.gsdi.org/gsdicookbookindex.php>. Accessed September 2009
- Nöth W (2000) *Handbuch der Semiotik* (2., vollständig neu bearbeitet und erweiterte Auflage). 2nd edn. Metzler, Stuttgart - Weimar
- O'Connor M, Tu S, Nyulas C, Das A, Musen M (2007) Querying the semantic web with SWRL. pp 155–159
- Office of Management and Budget (2009) Circular No. A-16: coordination of geographic information and related spatial data activities. http://www.whitehouse.gov/omb/circulars/a016/a016_rev.html. Accessed September 2009
- Open Geospatial Consortium (OGC) (2009) OpenGIS catalogue service implementation specification. <http://www.opengeospatial.org/standards/cat>. Accessed May 2009
- Peirce CS (1966) *Selected writings* (Values in a universe of chance). Dover, New York
- Ramezani M, Bergman L, Thompson R, Burke R, Mobasher B (2008) Selecting and applying recommendation technology. In: *Proceedings of international workshop on recommendation and collaboration, in conjunction with 2008 international ACM conference on intelligent user interfaces (IUI 2008)*
- Schlossberg M, Shuford E (2005) Delineating 'public' and 'participation' in ppgis. *URISA J* 16(2):15–26
- Sirin E, Parsia B, Grau BC, Kalyanpur A, Katz Y (2007) Pellet: a practical OWL-DL reasoner. *Web semantics: science, services and agents on the world wide web* 5(2):51–53
- Survival International (2009a) News (November 2008): glimmer of hope for uncontacted tribes. <http://www.survival-international.org/news/3929>. Accessed September 2009
- Survival International (2009b) News (May 2009): ranchers seek license to destroy uncontacted tribes' land. <http://www.survival-international.org/news/4478>. Accessed September 2009
- Tan PN, Steinbach M, Kumar V (2006) *Introduction to data mining*. Addison Wesley, Reading
- United Nations Statistics Division (UNSD) (2009) International standard industrial classification of all economic activities. <http://unstats.un.org/unsd/cr/registry/isic-4.asp>. Accessed May 2009
- Zak M (2004) Do subtropical seasonal forests in the Gran Chaco, Argentina, have a future? *Biol Conserv* 120:589–598