International Journal of Applied Earth Observation and Geoinformation 37 (2015) 133–141

Contents lists available at ScienceDirect

International Journal of Applied Earth Observation and Geoinformation

journal homepage: www.elsevier.com/locate/jag

Using ontological inference and hierarchical matchmaking to overcome semantic heterogeneity in remote sensing-based biodiversity monitoring

Simon Nieland*, Birgit Kleinschmit, Michael Förster

Geoinformation in Environmental Planning Lab, Technische Universität Berlin, Straße des 17. Juni 145, 10623 Berlin, Germany

ARTICLE INFO

Article history:
Available online 16 October 2014

Keywords:
Earth observation
Semantic reasoning
Interoperability
Nature conservation
Natura 2000

ABSTRACT

Ontology-based applications hold promise in improving spatial data interoperability. In this work we use remote sensing-based biodiversity information and apply semantic formalisation and ontological inference to show improvements in data interoperability/comparability. The proposed methodology includes an observation-based, "bottom-up" engineering approach for remote sensing applications and gives a practical example of semantic mediation of geospatial products. We apply the methodology to three different nomenclatures used for remote sensing-based classification of two heathland nature conservation areas in Belgium and Germany. We analysed sensor nomenclatures with respect to their semantic formalisation and their bio-geographical differences. The results indicate that a hierarchical and transparent nomenclature is far more important for transferability than the sensor or study area. The inclusion of additional information, not necessarily belonging to a vegetation class description, is a key factor for the future success of using semantics for interoperability in remote sensing.

Introduction

Semantic interoperability's importance for data harmonisation has often been discussed in geographic information integration (Janowicz, 2012; Hess et al., 2007; Kavouras et al., 2005; Rodriguez and Egenhofer, 2003; Visser et al., 2002; Lutz et al., 2009). In addition to syntactic interoperability, thoroughly defined by Open Geospatial Consortium (OGC) standards, heterogeneities of underlying semantics represent an unsolved barrier for data integration, data discovery and knowledge sharing – especially in a variety of remote sensing-based applications (Arvor et al., 2013; Blaschke, 2010).

Although remote sensing products and classification procedures often implicitly use semantics for developing rule-sets or indicators there is a lack of structured, computer-readable formalisation within the given classification approaches. Remote sensing-based classification conceptualises a real world object or phenomenon (entity) and produces its mapping (symbol). When trying to compare classification results, naming conflicts (different descriptions for the same conceptualisation or one ambiguous description for different conceptualisations) and conceptual conflicts (different conceptualisations for the same mapping) occur, leading to semantic heterogeneity (Kuhn, 2005).

The described heterogeneities hamper the examination of remote sensing output information which is especially problematic when it is required for multi-national legal processes as is the case for the EU Habitats Directive (Council Directive92/43/EEC, 1992) (HabDir) and the Water Framework Directive (Council Directive2000/60/EEC, 2000). A comparable thematic depth is needed for the subsequent decision making process. Due to the semantic diversity of remote sensing results, such products are either not considered useful or in the early stages of development (Manakos and Hellas, 2013). Several approaches (Lutz et al., 2009; Visser et al., 2002; Rodriguez and Egenhofer, 2003; Durbha et al., 2009; Mena and Illarramendi, 2000; Kavouras et al., 2005; Schwering and Raubal, 2005) were applied to achieve semantic interoperability of spatial data by using ontologies based on the Resource Description Framework (RDF) or Web Ontology Language (OWL). What these approaches all have in common is the matchmaking process; a technique used to find equivalent information that fit to the particular subject of interest. Matchmaking between spatial datasets can be generated by using similarity values. Often based on dictionaries, thesauri, other RDF/OWL-based data structures (Hess et al., 2007; Rodriguez and Egenhofer, 2003; Kavouras et al., 2005; Fonseca et al., 2006) or geospatial concepts and their geometrical models (Schwering and Raubal, 2005), similarity values indicate the degree of

* Corresponding author. Tel.: +49 30 314 72601.
E-mail address: simon.nieland@tu-berlin.de (S. Nieland).

http://dx.doi.org/10.1016/j.jag.2014.09.018
0303-2434/

correspondence between entities. Additionally, matchmaking can be achieved by reasoning about the formalised concepts of one specific domain ontology (Decker et al., 1998) and its aligned upper level ontology (Cruz and Sunna, 2008). Multi-ontology systems in combination with query rewriting techniques have also been used to generate matchmaking (Mena and Illarramendi, 2000). Furthermore, a "hybrid ontology" approach where a shared vocabulary is applied for the formalisation of concepts and inter-ontological reasoning was used in several systems (Lutz et al., 2009; Visser et al., 2002; Durbha et al., 2009).

Several studies have proposed using observations for geo-ontologies (Janowicz, 2012; Couclelis, 2010; Frank, 2003). In so-called observation-driven engineering approaches ontological primitives (classes in the ontology that are not conceptualised) represent elementary concepts that can be derived from observations. Therefore, included primitives are restricted to observations or derived by aggregation of observed phenomena. Conceptualised classes in the developed ontology can be assigned to upper level ontologies to foster a broader interoperability. Starting with the semantic descriptions of observations in a bottom-up ontology engineering approach preserves the benefit of semantic diversity and local conceptualisations without giving up interoperability (Janowicz, 2012).

Recent approaches of remote sensing classification are bound to semantic web standards proposed by the World Wide Web Consortium (W3C) and therefore allow the utilisation of semantic reasoning in the classification process (Andres et al., 2012; di Sciascio et al., 2013; Belgiu et al., 2014). These approaches are not broadly used in the remote sensing community, despite recent discussions touting their benefits (Arvor et al., 2013).

Heterogeneities in classes based on remote sensing analysis are resulting from the fact that classification procedures are specified through electromagnetic signals, whereas indicators of field surveys and nomenclatures reflect the composition of parameters defined in the particular area of research or cognitive interest. In many cases remote sensing classification techniques are adapted to classes which are optimised for manual interpretation of aerial imagery or fieldwork through the aggregation of the primitive classification results to the target classes. Consequently, these primitive classifications conceal information because users or customers only have the generalised mapping without its underlying conceptualisation.

A better conceptualisation of remote sensing outputs in RDF/OWL-based structured metadata would not only lead to a better re-usability and exchangeability, but would additionally improve spatial information retrieval (Arvor et al., 2013). Inferring relations between data requirements or products and existing data or nomenclatures is a benefit that is already broadly used in other research areas (Bard and Rhee, 2004).

The main objectives of this work are to

- propose an observation-based, bottom-up ontology engineering approach for remote sensing applications, which will be used for solving semantic heterogeneity problems in remote sensing classification results by taking into account ontology-based automatic reasoning in combination with matchmaking processes based on generalisation,
- give a practical example of semantic mediation of geospatial data in the field of remote sensing-based biodiversity monitoring
- and analyse certain criteria of selected areas (similarity of sensors, number of classes, similarity of geographical region) in regard to their influence on used indicators and subsequently their effect on data interoperability.

A big future challenge of remote sensing research is to transform local or regional classification outputs into interoperable, comparable information. Since there are existing interoperability approaches in other research domains, we contribute to the existing research by addressing the need for interoperability with a novel semantic approach that is based on ontological subsumption.

Methodology

This section proposes a bottom-up, observation-based ontology engineering approach and shows how it can be used for data interoperability in a prototype application.

Study sites and existing habitat data

We analyse remote sensing classification results of Natura 2000 heathland areas and corresponding nomenclatures for this study. The Natura 2000 sites are heathland and grassland habitats in Flanders (Belgium) and Brandenburg (Germany).

Kalmthoutse Heide[1] (abbr. FL), located in northern Belgium is mainly covered with dry and wet heathland, inland dunes, water bodies and forests (Chan et al., 2012). The 6 broader habitat classes at level 1 are gradually arranged into subcategories that reflect the definitions of the habitat structure as well as the structures and functions that are crucial for the assessment of habitat quality (Thoonen et al., 2013).

The second study site, Döberitzer Heide,[2] located in eastern Germany is characterised by heathland and grassland vegetation, humid meadows and woods on predominantly dry and sandy soils.

For the Döberitzer Heide, two classification hierarchies are available (see Table 1). The nomenclature for multi-temporal, high resolution (HR) and hyper-spectral analysis (abbr. BB-HyMap) extends the federal nomenclature towards specific plant communities.

These plant communities are used as indicators for the evaluation of habitat conservation status (Schuster et al., 2015). Additional class attributes were already included in the federal nomenclature. Since only parts of the developed and well-formalised BB-HyMaP nomenclature have been classified within this study, a synthetic dataset was created to acquire more significant information about the quality of the semantic transformation process. It uses the extent and pixel-size of BB-VHR and includes one band with corresponding class values. The multi-temporal classification was performed by Schuster et al. (2015) with 21 RapidEye scenes covering dates from March to October (Schuster et al., 2015). The classes were created using federal habitat descriptions and field-based mapping in Brandenburg.

Habitat classification with VHR imagery was realised using a knowledge-based classification approach. The development of the classification procedure can be divided in the following steps. Initially, suitable indicators are selected, which are limited to those that can be derived from very high spatial resolution (VHR) remote sensing imagery. In the next step, these indicators can now be used to develop a hierarchical classification schema. To validate the separability of the determined classes by using a discriminant analysis, each class has to be associated with representative ground truth areas. The developed hierarchical schema is the basis for the multi-level, pixel-based classification procedure. To be able to analyse the variability in the imagery statistically, the classification procedure uses a hybrid system of supervised and unsupervised classification modules (Frick, 2006). Therefore, image data has to be structured and relative values (NDVI, texture, spatial arrangement) have to be derived.

[1] http://natura2000.eea.europa.eu/Natura2000/SDF.aspx?site=BE2100015
[2] http://natura2000.eea.europa.eu/Natura2000/SDF.aspx?site=DE3444303