

Towards Semantic Interoperability in Biodiversity Monitoring

Simon NIELAND, Michael FÖRSTER and Birgit KLEINSCHMIT

Technische Universität Berlin/Germany · simon.nieland@tu-berlin.de

This contribution was double-blind reviewed as extended abstract.

Abstract

Ontology based applications offer big opportunities in data management in regard to interoperability of spatial datasets. This work demonstrates a data management application, which mainly focuses on the semantics of biodiversity monitoring data. It suggests a possibility of interoperable data management of remote sensing based biodiversity data with the help of semantic formalization and ontological inference.

1 Introduction

Data interoperability is a big challenge for the GI community and plays an important role in European spatial harmonization efforts like INSPIRE. Especially for the Natura 2000 framework with its strict reporting obligations, harmonization and comparability of spatial biodiversity data is a very important issue. Remote sensing based assessment of Natura 2000 habitats provides automatized algorithms that give the possibility to generate comparable biodiversity information across Europe. This work suggests a possibility of interoperable data management of remote sensing based biodiversity data with the help of semantic formalization and ontological inference.

2 Input Data and Test Cases

Data input for this study are nomenclatures and results of remote sensing based classification of Natura 2000 heathland areas in the regions Kalmthoutse Heide (Belgium/Flanders) and Döberitzer Heide (Germany/Brandenburg). Main focuses of this work are heathland and grassland habitats. Classification concepts (indicators and other object primitives) could be collected by analysing remote sensing classification procedures. A detailed description of derived indicators and classification techniques can be found in (FRICK et al. 2000; THOONEN et al. 2013). In the heathland Area of Kalmthout a 4-level hierarchical classification system has been developed by analysing collected field data in combination with spectral signatures (THOONEN et al. 2013). It consists of 6 classes at level 1, representing broad habitat land cover classes: heathland, grassland, forest, sand dunes, water and arable land. These 6 classes 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. For the heathland area Döberitzer Heide different methodologies were applied to identify habitat patches. In this work we are focusing on a knowledge-based classification approach of very high resolution imagery (FRICK et al. 2000). The class hierarchy of this technique were created in regard to the federal habitat descriptions and field measurements in the Brandenburg area.

3 Methods of Data Integration

3.1 Ontology architecture

Basis of this work is a hybrid ontology model, which combines two different ontology approaches. (WACHE & VOEGELE 2001) Basic concept of this methodology is to describe classification results of different regions in its own ontology. But in contrast to multiple ontology approaches the concepts of each local ontology are based on primitives and properties of a shared vocabulary, stored in an upper level domain ontology (see Fig. 1). Main advances of this methodology are to keep the flexibility of a multi-ontological approach and preserve comparability by using shared vocabulary.

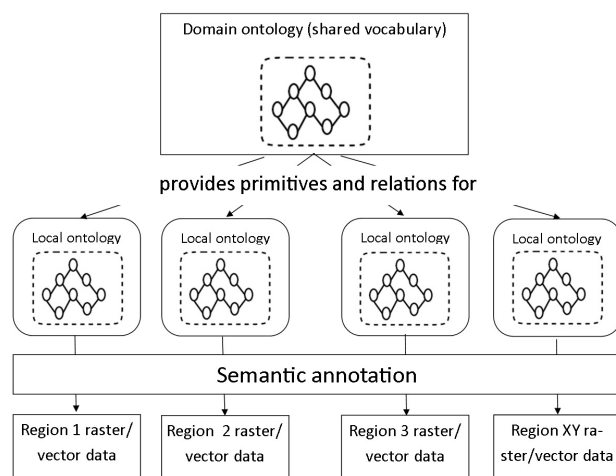


Fig. 1:
Hybrid ontology model.
Adapted from (WACHE &
VOEGELE 2001)

3.3 Equality reasoning

In order to achieve data interoperability, an equality-reasoning algorithm has been developed. The reasoning is based on the Hermit reasoner (<http://www.hermit-reasoner.com/>) which performs subsumption between objects of different regions and infers class equality. That means, the reasoner produces one hierarchy of class descriptions of all imported datasets and related regions. For example, class Hgmd (heathland, grass encroached, molinia, dry) from Flanders will be assigned to the super classes “Hd” (heathland dry/Flanders) and “HZSO” (heathland, dry, sandy, shrub-encroached/Brandenburg), which have the common super classes “H” (heathland/ Flanders) and “Heathland” (Brandenburg). Because of the fact that Hgmd is described as grass dominated it is also a subclass of

“Species of dry grassland” (Brandenburg), which is assigned to the super classes “Grassland” (Brandenburg) and “G” (Flanders). Furthermore, the developed algorithm checks if classes of the origin region have equivalents in the destination region. If not, the algorithm performs “semantic up-scaling” processes until it finds an equivalent super class.

3.4 Prototype development

Main focus of this work is the development of prototype software that is able to transfer classification results of a region into a classification key of another region in regard to the semantics of the nomenclatures. Thus, it generates semantic comparability between the test cases. The application can be divided in three parts. A database module, which imports/exports semantic class descriptions of raster datasets, an ontology module, which performs semantic equality reasoning of class descriptions and adds additional equality information to the database and the ontology-engineering component, that allows domain experts to add domain knowledge and classification hierarchies to the ontology (see Fig. 2).

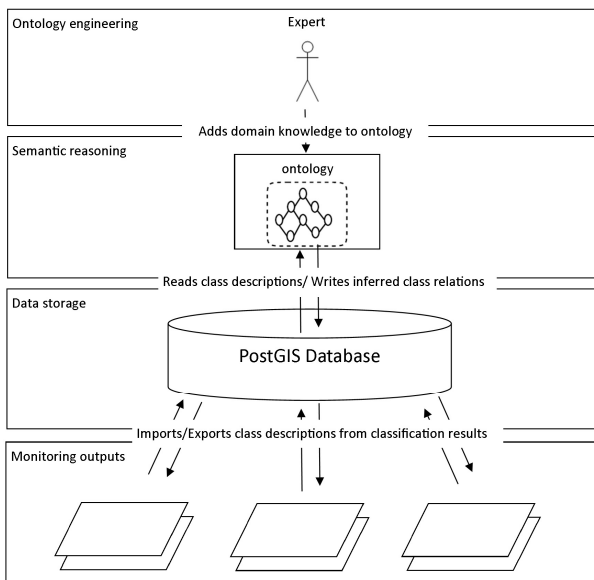


Fig. 2:
Software prototype for data interoperability

4 Results and Discussion

Fig. 3a and 3b are demonstrating the number of required up-scaling processes for achieving data equality. Areas that are displayed in white could be transferred without up-scaling. Areas in light grey needed one up-scaling process to achieve semantic equality and areas shown in dark grey were up-scaled two times. Tab. 1 illustrates the number of up-scaling processes per class in both test sites. It shows that there were less up-scaling processes in the transfer from Brandenburg to Flanders than for the transfer from Flanders to Brandenburg. A detailed view in the class descriptions shows that the Flanders nomenclature is well formalized for most of the main habitat categories (wood, grassland, heathland, water, sand), while the Brandenburg nomenclature mainly focuses on grassland habitats.

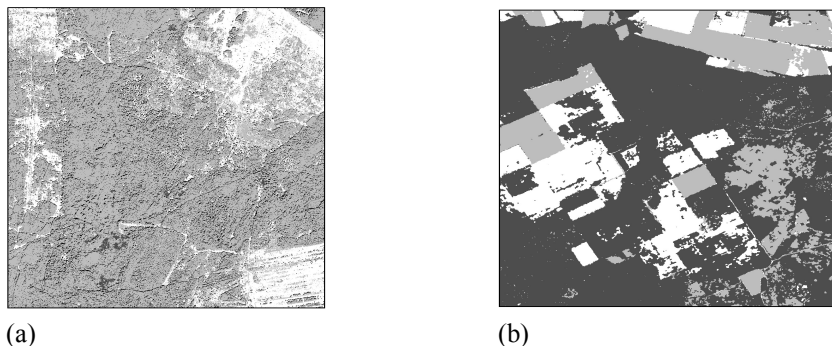


Fig. 3: Number of necessary up-scaling processes in Brandenburg (a) and Flanders (b)

Table 1: Number of necessary up-scaling processes for semantic data transfer

Number of Up-scaling processes	Number of classes (Transfer BB – FL)	Number of classes (Transfer FL – BB)
0	8	2
1	9	6
2	4	15

5 Conclusion and Outlook

That brings us to the conclusion that the level of interoperability is very much depending on the formalization of the stored data. Therefore the formalization of Natura 2000 habitats in regard to remote sensing and GIS applications would be a basis for better interoperability of the data. Furthermore the algorithm only searches for equal classes. The evaluation of classes, which only share some indicators, could be a further step towards a better comparability between biodiversity monitoring datasets.

References

- FRICK, A., WEYER, G., KENNEWEG, H., & KLEINSCHMIT, B. (2000), A Knowledge-based approach to vegetation Monitoring with Quickbird imagery. ISPRS Workshop 2005 – High-Resolution Earth Imaging for Geospatial Information.
- THOONEN, G., SPANHOVE, T., VANDEN BORRE, J. & SCHEUNDERS, P. (2013), Classification of heathland vegetation in a hierarchical contextual framework. *International Journal of Remote Sensing*, 34 (1), 96-111.
- WACHE, H., & VOEGELE, T. (2001), Ontology-based integration of information-a survey of existing approaches. *Proceedings of the IJCAI-01 Workshop: Ontologies and Information Sharing* (pp. 108-117).