GEOSPATIAL DATA HARMONIZATION FROM REGIONAL LEVEL TO EUROPEAN LEVEL

A Use Case in Forest Fire Data

Kaori Otsu

A thesis submitted in partial fulfillment of the requirements for the degree of Master of Science in Geospatial Technologies

February 2010

Universitat Jaume I (UJI), Dept. Lenguajes y Sistemas Informáticos (LSI), Spain
Westfälische Wilhelms-Universität Münster (WWU), Institute for Geoinformatics (IFGI), Germany
Universidade Nova de Lisboa (UNL), Instituto Superior de Estatística e Gestão de Informação (ISEGI), Portugal

Dissertation supervised by

Ms. Laura Díaz Sánchez
Research Associate
Department of Information Systems
Universitat Jaume I
Av. Vicente Sos Baynat s/n
12071 Castellón, Spain
E-mail: diazl@uji.es

Dr. Sven Schade
Postdoctoral Researcher
European Commission, Joint Research Centre
Institute for Environment and Sustainability
Spatial Data Infrastructures Unit, TP 262
Via E. Fermi, 2749, I-21027 Ispra (VA), Italy
E-mail: sven.schade@jrc.ec.europa.eu

Prof. Miguel de Castro Neto
Professor
Instituto Superior de Estatística e Gestão de Informação
Universidade Nova de Lisboa
Campus de Campolide
1070-312 Lisboa, Portugal
E-mail: mneto@isegi.unl.pt

ACKNOWLEDGEMENTS

I feel very fortunate that I was given this research opportunity funded by European Commission's Erasmus Mundus Scholarship. I thank my committee members, Ms Laura Diaz, Dr. Sven Schade, and Prof. Miguel Neto for their support with my thesis work. I owe special thanks to my supervisors Ms Diaz (regarding syntactic interoperability) and Dr. Schade (regarding schematic and semantic interoperability) for their significant contributions, insightful advices, and editorial time.

I also appreciate collaboration from European Forest Fire Information System (Mr. Cristiano Giovando at Joint Research Centre) and the Valencia government (Mr. Jorge Suarez at Conselleria de Medi Ambient, Aigua, Urbanisme i Habitatge) for providing GIS data and consulting the data content regarding burned areas. In addition, Prof. Mar Marcos at Universitat Jaume I kindly guided me in ontology engineering.

Lastly, I am deeply grateful that my family, friends, professors, and colleagues who have encouraged me in many ways throughout my master's program. I include many thanks to co-director Prof. Joaquin Huerta and administrator Ms. Dolores Apanewicz from Master of Science in Geospatial Technologies for organizing the program and making my stay in Spain comfortable during the two semesters I spent at Universitat Jaume I.

ABSTRACT

Geospatial data harmonization is becoming more and more important to increase interoperability of heterogeneous data derived from various sources in spatial data infrastructures. To address this harmonization issue we present the current status of data availability among different communities, languages, and administrative scales from regional to national and European levels. With a use case in forest data models in Europe, interoperability of burned area data derived from Europe and Valencia Community in Spain were tested and analyzed on the syntactic, schematic and semantic level. We suggest approaches for achieving a higher chance of data interoperability to guide forest domain experts in forest fire analysis. For testing syntactic interoperability, a common platform in the context of formats and web services was examined. We found that establishing OGC standard web services in a combination with GIS software applications that support various formats and web services can increase the chance of achieving syntactic interoperability between multiple geospatial data derived from different sources. For testing schematic and semantic interoperability, the ontology-based schema mapping approach was taken to transform a regional data model to a European data model on the conceptual level. The Feature Manipulation Engine enabled various types of data transformation from source to target attributes to achieve schematic interoperability. Ontological modelling in Protégé helped identify a common concept between the source and target data models, especially in cases where matching attributes were not found at the schematic level. Establishment of the domain ontology was explored to reach common ground between application ontologies and achieve a higher level of semantic interoperability.

Keywords: Geospatial Data Harmonization, Interoperability, Schema Mapping, Ontology, Forest Fire Model

ACRONYMS

ArcIMS	Arc Internet Map Server
CMA	Conselleria de Medio Ambiente, Agua, Urbanismo y Vivienda (Ministry of Environment, Water, Urban Planning and Housing)
CSW	Catalogue Service
DOLCE	Descriptive Ontology for Linguistic and Cognitive Engineering
EEA	European Environmental Agency
EFDAC	European Forest Data Centre
EFFIS	European Forest Fire Information System
EFICP	European Forest Information and Communication Platform
ETL	Extract, Transform, Load
FAO	Food and Agriculture Organization of the United Nations
FIS	Forest Information System
FME	Feature Manipulation Engine
FRA	Forest Resources Assessment
GIS	Geographic Information System
GML	Geography Mark-up Language
IDEE	Infraestructura de Datos Espaciales de España (Spatial Data Infrastructure of Spain)
IGN	Instituto Geográfico Nacional (National Geographic Institute)
INSPIRE	Infrastructure for Spatial Information in Europe
JRC	Joint Research Centre
MMA	Ministerio de Medio Ambiente, y Medio Rural y Marino (Ministry of Environment, and Rural and Marine)
NUTS	Nomenclature of Territorial Units for Statistics
OGC	Open Geospatial Consortium
OWL	Web Ontology Language
RDF	Resource Description Framework
SDI	Spatial Data Infrastructure
SIOSE	Sistema de Información sobre la Ocupación del Suelo en España (Land Cover Information System in Spain)
WCS	Web Coverage Service
WCTS	Web Coordinates Transforming Services
WFS	Web Feature Service
WMS	Web Map Service
WPS	Web Processing Service
WSML	Web Service Modeling Language
XML	Extensible Mark-up Language

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1. INTRODUCTION

Over the past two decades the distribution of geospatial data has significantly increased as information technologies advanced [Masser 2005]. As the data often derive from different sources, it is necessary to establish a common framework for sharing and exchanging data [INSPIRE 2004]. The common framework can be designed in a spatial data infrastructure (SDI) [Nebert 2004] where geospatial data can be readily accessible in cooperation with various stakeholders including governments, organizations and private sectors through agreed policies and common standards [Phillips *et al.* 1998]. Today, SDI plays a key role to support users and providers for decision making where they can discover, visualize, and evaluate geospatial data at regional, national and global levels [Nebert 2004; Masser 2005]. Geospatial data in an SDI include various products and services ranging from security, census, environment, health, emergency response, transportation, agriculture to forestry [Masser 2005]. In this context data harmonization is becoming more and more important to increase interoperability of heterogeneous data in a SDI [INSPIRE 2007].

In case of forestry, geospatial data are essential for monitoring and managing forests to be sustainable. At the European level, European Forest Data Centre[1] (EFDAC) is under development to improve a Forest Information System (FIS). EFDAC is being implemented in compliance with the guidelines of the Infrastructure for Spatial Information in Europe[2] (INSPIRE), which attempts to establish common standards in an SDI to make different SDI nodes interoperable with each other in Europe [INSPIRE 2003]. Upon implementation, the European FIS will enhance data harmonization and improve the efficiency of data collection. At national level, Spain is a Member State of European Union which is nowadays adopting the INSPIRE Directive to be compliant with it at different administrative

[1] European Forest Data Centre. European Commission Joint Research Centre. URL: http://efdac.jrc.ec.europa.eu (last accessed on December 1st 2009).
[2] Infrastructure for Spatial Information in Europe. European Commissions. URL: http://www.inspire-geoportal.eu (last accessed on December 1st 2009).

level, being the Spanish national SDI (IDEE[3]) as the main node connected to the Directive. While IDEE contains basic forest cover data, the Ministry of the Environment and Rural and Marine[4] (MMA) are currently attempting to allow other forest data managed by the national forestry program accessible in the IDEE [MMA 2009a]. At regional level, forest data collected from autonomous regions in Spain are accessible through the national forestry program [MMA 2009a].

The INSPIRE Directive aims to regulate various spatial data themes needed for environmental applications [INSPIRE 2003]. To date guidelines for INSPIRE Data Specifications on some data themes have been established such as protected sites [INSPIRE 2009a] and transport networks [INSPIRE 2009b]. To interoperate FISs in Europe, data specifications on forestry need to be defined in the same manner, which is addressed as one of ongoing projects by INSPIRE [INSPIRE 2003]. Thus, we intent to investigate the status quo of forest data at different administrative levels and how they can be harmonized within Europe for forest domain experts to enhance analysis.

1.1 Motivation

Increasing interoperability can help create a seamless global FIS where regional, national, and global systems are better interconnected. For example, under the initiative of Food and Agriculture Organization of the United Nations[5](FAO), Global Forest Resource Assessment requires reports from each country on forest health and productivity [FAO 2009]. The global FIS is expected to enhance the efficiency of data collection and contribute to monitoring and managing forests efficiently from regional to global scales.

Using the common thematic forest data, our research aims to test data harmonization on different levels of interoperability in different scales within Europe. As the most fundamental forest data in a FIS, forest cover is available in

[3] Infraestructura de Datos Espaciales de España. Gobierno de España. URL: http://www.idee.es (last accessed on November 17th 2009).
[4] Banco de Datos de la Biodiversidad. Ministerio de Medio Ambiente, y Medio Rural y Marino. URL: http://www.mma.es/portal/secciones/biodiversidad/banco_datos (last accessed on November 17th 2009).
[5] Food and Agriculture Organization of the United Nations. URL: http://www.fao.org (last accessed on November 18th 2009).

different information systems from regional, national and European levels. Forest cover information can be used as input for the further analysis in forest fire monitoring, wildlife habitat protection, and watershed management [European Commission 2003].

One of EFDAC components, European Forest Fire Information System[6] (EFFIS), mainly provides the information about fire danger forecast, hotspots and burned areas across Europe. For example, burned areas are delineated by overlay of satellite images and land cover map, which include forest cover affected by fires [JRC 2009]. At national level in Spain, a member state of EFFIS, the MMA provides an access point to regional forest fire data where each autonomous region is in charge with data collection [MMA 2009c]. Using a forest fire scenario, our intent is to guide how regional forest domain experts can perform analysis by utilizing geospatial data and applications from regional to European level. This scenario illustrates where geospatial data can be obtained and how interoperability of such data can be achieved.

1.2 Research Problems and Questions

When we refer to interoperability, there are many aspects to consider. Interoperability does not only address one type of integration but can be categorized into mainly four types: *system, syntactic, schematic and semantic* [Bishr 1998; Goodchild *et al.* 1999]. In our research, we use a concept of geospatial data harmonization as a factor to increase interoperability in the context of syntaxes, schemas and semantics [Lehto 2007; Schade 2009]. Syntactic interoperability refers to integrating the elements in various systems such as data formats and standards. Schematic interoperability is explained by the common classification and hierarchical structure [Bishr 1998]. Semantic interoperability harmonizes meanings of terms and expression according to how the terms are named and described [Bishr 1998].

[6] European Forest Fire Information System (EFFIS). European Commission. URL: http://effis.jrc.ec.europa.eu (last accessed on December 1st 2009).

A. Syntactic Interoperability

Finding common means to access heterogeneous data can increase syntactic interoperability. At service level, SDI web services are implemented by using standard interfaces defined by Open Geospatial Consortium (OGC) [Percival 2003]. The most common OGC interfaces include mapping images (OGC Web Map Service [OGC 2009], WMS), manipulating geographic features (OGC Web Feature Service [OGC 2009], WFS), and manipulating grid coverages (OGC Web Coverage Service [OGC 2009], WCS). When multiple data layers are not accessible together via web services, Geographic Information System (GIS) software applications at client level, which support different formats, are other tools to achieve syntactic interoperability.

As the central point at European level, EFFIS is playing an important role to provide web services to the public with the updated forest fire information across Europe. EFFIS manages extensive datasets produced by the JRC[7] and individual national forestry programs of member states [EuroGEOSS 2009]. In theory, the geospatial data mapped by JRC and national forest fire programs should be consistent for describing the same area. In reality, when the data come from different sources in different standards and formats, they do not always match [EuroGEOSS 2009; HUMBOLDT 2009]. To examine the status of current syntactic interoperability between these geospatial data in different administrative scales, we aim to answer the following questions:

- *Are forest fire data from EFFIS and member states syntactically interoperable at service level and/or client level?*
- *Are there any scale issues of forest fire data from different sources? Are there any discrepancies in the total burned areas between EFFIS and member states?*

Approaching these questions, we define the following hypothesis:

[7] Joint Research Centre, IES, European Commission. URL: http://www.jrc.ec.europa.eu/ (last accessed on December 1st 2009).

Hypothesis A

By establishing standard web services and common tools, we can increase the chance of achieving syntactic interoperability between multiple geospatial data derived from different sources.

B. Schematic and Semantic Interoperability

EFFIS applies burned areas as one of input data in its forest fire model to estimate CO2 emission [JRC 2009]. In case a regional forest domain expert attempts to apply their regional data in the CO2 emission model developed by EFFIS, they cannot simply apply the burned areas derived from their region as input data. This is due to the difference in semantics and schemas used in burned area data between different communities, languages, and administrative scales. Thus, the regional data need to be harmonized into common semantics with the EFFIS data first to reclassify schemas used in regional data for establishing common schemas with EFFIS [Bishr 1998]. In cases where different languages are harmonized into one common language, the semantics and schemas used in original languages may get lost in translation as well. In terms of semantic and schematic interoperability, our research aims to answer following question:

- *How can forest fire data be transformed and mapped into common schemas and semantics across different administrative scales, where the forest fire data are heterogeneous among different communities and languages?*

This question is examined against the following hypothesis:

Hypothesis B

By identifying common schemas and concepts, we can transform the regional data model to the European data model on the semantic level.

1.3 Testing Approach and Expected Results

To test interoperability of forest fire data between EFFIS and Spain, burned areas are collected from both sources. In our case study, burned areas in Valencia Community (one of autonomous regions) are selected to represent Spanish data and further compared with European data which include burned areas in Valencia

Community. Figure 1 shows the workflow of interoperability testing on the syntactic, schematic and semantic level.

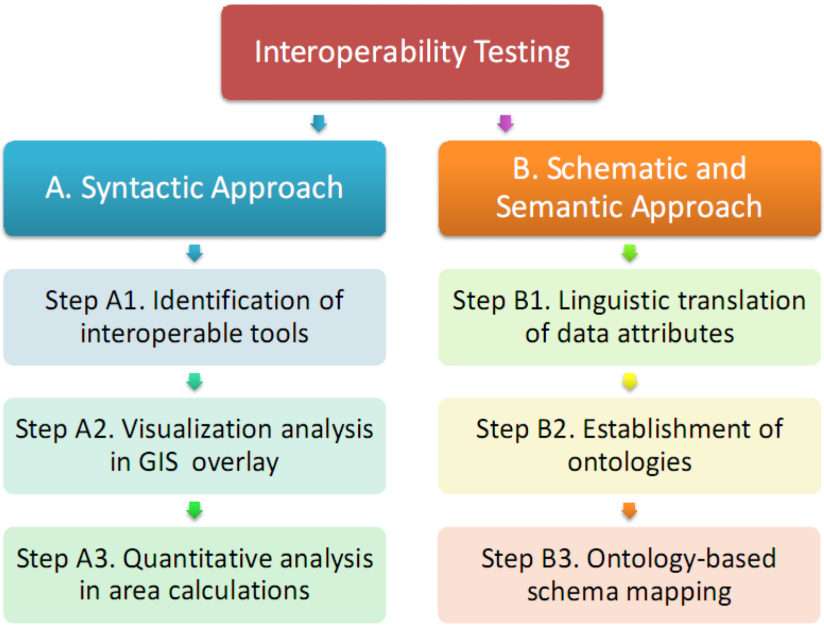


Figure 1. Testing approach to syntactic interoperability, and schematic and semantic interoperability.

A. Testing Syntactic Interoperability

We test syntactic interoperability in the context of data access and format. Firstly, availability of tools that allow both data from EFFIS and Spain to be interoperable is identified in Step A1. These tools include OGC standard web services (WMS, WFS and WCS) and GIS software applications. In Step A2, the two data layers are visualized and overlaid using interoperable tools to detect any qualitative discrepancy. Step A3 then presents the quantitative difference between burned areas mapped by EFFIS and Valencia Community by area calculations. Expected outputs include a summary table of syntactic interoperability, screen shots of overlay visualization, and burned area calculations.

B. Testing Schematic and Semantic Interoperability

The research question on schematic and semantic interoperability can be answered by identifying common schemas and concepts between the regional data to EFFIS data. Firstly, Spanish terms used in data attributes and values are translated into English in Step B1. For reasoning the semantics translated from Spanish terms, ontologies are established in Step B2 to identify matching attributes with a shared concept between the two data models. Finally, in ontology-based schema mapping Step B3, source data attributes from Valencia Community are transformed to the target attributes of EFFIS data according to the type of mapping operations required. Expected outputs include a table of data attributes translated from Spanish to English; ontology-based schema mapping tables from source to target data attributes; a summary of mapping operations corresponding to matching attributes; and an example of mapping rules saved as output file.

1.4 Thesis Structure

This chapter introduces key issues of interoperability of geospatial data within Europe, followed by the background of standard web services in SDIs, geospatial data harmonization, data transformation tools, ontology languages, and ontology-based schema mapping in Chapter 2. Chapter 3 describes the current status of forest data in study areas on different scales from Europe to Spain and Valencia Community. In Chapter 4, testing methods on interoperability are described using syntactic, schematic, and semantic approaches. Then the testing results are presented with a comparison of data availability and interoperability, burned area calculations, and schema mapping tables and rules in Chapter 5. In Chapter 6 we analyze the results and discuss with recommendations on syntactic, schematic and semantic interoperability. The final chapter (Chapter 7) concludes our research on geospatial data harmonization from regional level to European level and suggests future work to be done.

2. BACKGROUND

In this chapter, the background information is introduced in relation to methods for testing interoperability. First, we describe how standard web services are implemented in SDIs. Second, we emphasize the importance of geospatial data harmonization for web transformation services. Then, as tools for data transformation, schema mapping software and ontology languages are introduced. Finally, we present recent researches related to ontology-based schema mapping.

2.1 Standard Web Services in Spatial Data Infrastructures

The shift in technological terms from national SDIs to multinational SDIs has been emphasized over the last decade along with the development of the World Wide Web (WWW) [Masser 2005]. Current trends in SDI development include a shift from a product to a process model where the WWW enables end users to share spatial data in decentralized structures [Masser 2005]. Another trend emphasizes implementation of multilevel SDIs in the context of hierarchy. Hierarchy can be a bottom-up as well as a top-down structure in SDI implementation. The top-down structure aims to achieve harmonization while the bottom-up structure aims to preserve heterogeneity. The challenge for implementing multilevel SDIs is to agree with a common standard in consideration with heterogeneity of various stakeholders [Masser 2005].

In Europe, the European Commission took an initiative to study the development of multinational SDIs, INSPIRE. INSPIRE Directive addresses the need for web services to discover, view, transform, invoke, and download geospatial data, which enable various stakeholders to share data in the multilevel hierarchy on the multinational scale [INSPIRE 2007]. Such web services require technical specifications commonly agreed by the Member States for the interoperability and harmonization of their SDIs [INSPIRE 2007]. Currently INSPIRE adopts the OGC specifications, existing OGC Web Services (OWS) standards, as a technical guidance for implementing those web services [INSPIRE 2008].

OGC specifications refer to standard web services such as Catalogue Service (CSW) for discovering, Web Mapping Service (WMS) for viewing, Web Coordinates

Transforming Services (WCTS) for transforming, Web Processing Service (WPS) for invoking, and Web Coverage Service (WCS) and Web Feature Service (WFS) for downloading [INSPIRE 2008]:

- CSW supports discovery, evaluation and use of spatial data and services through their metadata properties.
- WMS allows requests over geo-referenced data belonging to the themes and provides a visual representation of these data, rendered in an image format such as PNG, GIF or JPEG.
- WCTS performs schema transformation and coordinate transformation.
- WPS provides client access across a network to pre-programmed calculations and/or computation models.
- WCS provides client access to potentially detailed and rich sets of geospatial information and returns coverages.
- WFS allows clients to retrieve and update geospatial data encoded in Geography Mark-up Language (GML) [OGC 2007] from multiple WFSs.

GML represents geographical features expressed in Extensible Mark-up Language (XML) that enables geospatial data to be stored, transferred, exchanged, processed and transformed though web services such as WFS and WPS in a standard format [Percival 2003; Diaz et al. 2009]. As illustrated in Figure 2, WFS plays a role of wrapper in SDIs for accessing and editing heterogeneous geospatial data in a standard way while WPS acts as a mediator for processing such data by linking WFS and data sources [Diaz et al. 2009].

Beside OGC web services, ArcIMS[8] (Arc Internet Map Server) developed by ESRI offers map service and feature service using ArcXML, which follows ESRI's XML specification [ESRI 2009]. ArcIMS also provides web links of map service and feature service in the same manner as OGC WMS and WFS, however, it does not follow OGC standards.

---

[8] ArcIMS. ESRI. URL: http://www.esri.com/software/arcgis/arcims/index.html (last accessed on December 12th 2009).

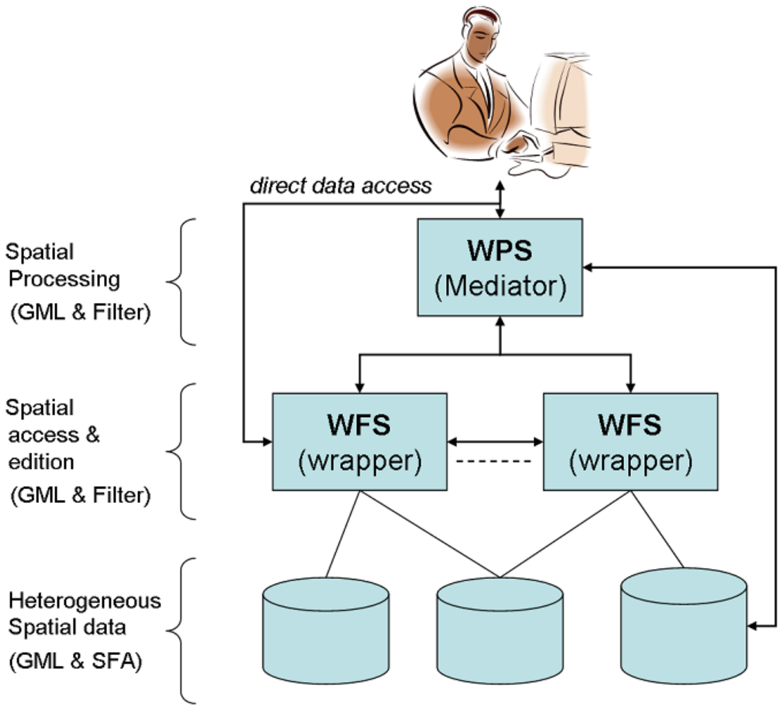


Figure 2. Components for spatial data integration over the Web [Diaz et al. 2009].

2.2 Geospatial Data Harmonization

Geospatial data harmonization plays an important role for on-the-fly data transformations [Lehto 2007; Schade 2009]. To achieve interoperability via web transformation services, geospatial data need to be harmonized between different sources [INSPIRE 2007; HUMBOLDT 2008]. In Europe, HUMBOLDT projects have focused on the improvement of harmonization issues such as data format, type of web service, spatial reference system, data model, classification schemes, terms and concepts, and metadata profile, which can be harmonized on different levels of interoperability [HUMBOLDT 2008]. In our research, data harmonization refers to harmonizing heterogeneous data to achieve interoperability on the syntactic, schematic and semantic levels [Visser 2001; Stuckenschmidt 2003; Friis-Christensen et al. 2005; Lehto 2007; Vaccari et al. 2009].

Syntactic interoperability can be achieved by transforming data format, type of web service, and spatial reference system [Visser 2001; Vaccari et al. 2009]. Currently syntactic harmonization issues in the context of geospatial web service are considered as minor [Lehto 2007]. As an example of forest fire models, retrieving burned areas and land cover via OGC WFS and transforming one of their coordinate systems via OGC WTCS (if necessary) can render the overlay image and provide statistics such as area calculations via OGC WPS for the selected area [Friis-Christensen et al. 2007].

On the schematic level, data heterogeneity is due to differences in data models and classification schemas [Friis-Christensen et al. 2005]. Geographical features such as polygons, lines, and points are often represented by different geometrical and data schemas [Vaccari et al. 2009]. Schematic interoperability can be achieved by transforming the structure of the source data model to the target source model, where the schemas refer to the respective XML Schema documents [Schade 2009]. To perform schema transformation, INSPIRE follows the ORCHESTRA Schema Mapping Service (SMS) to provide functionalities that are related to the mapping of features from a source into a target schema [INSPIRE 2008]. Specifically, schematic transformation involves filtering, renaming, reclassifying, merging/splitting, reordering, converting, morphing, and augmenting geographic features and their properties used in the schema [Lehto 2007]. This transformation is called schema mapping. In forest fire data models between source and target schemas, we may find heterogeneity in structures and attributes based on individual forest fire data standards and land cover classifications.

Semantic transformation is required in cases where the exact match from source to target schema cannot be found [Lehto 2007]. Semantic heterogeneity arises from the use of different terms in specific contexts established by different communities [Friis-Christensen et al. 2005]. For example, entities with the same term can have different semantics while entities termed differently can be semantically the same [Friis-Christensen et al. 2005; Vaccari et al. 2009; Abadie 2009]. In many cases the exact match does not exist, therefore, semantic transformation can be supported by exploiting ontologies and metadata [Goodchild 1999; Lehto 2007; Schade 2009]. Ontology in the context of computer science refers to “an explicit specification of a conceptualization” to express a common understanding of entities [Guarino 1998]. Metadata summarizes the information about data and it can contain the semantic content as well as syntactic and schematic details [Goodchild 1999]. Currently, many projects related to web services focus on semantic interoperability issues due to the complexity of semantic matching [Diaz et al. 2009]. In case of forest fire models, it is possible that the terms such as 'forests' and 'burned areas' may be defined differently between source and target data models.

2.3 Data Transformation Tools for Schema Mapping

Today spatial ETL (Extract, Transform, Load) platform is widely deployed for schema transformation in SDIs. The spatial ETL efficiently enables us to extract spatial data from source suppliers to SDI, transform the source data model to a new output in any format or application to be loaded as the target data model requested by end users [Safe Software 2009]. Data transformation for schema mapping restructures geometry and attributes, such as manipulating geometry, feature type, attribute name, attribute value, and attribute type [Safe Software 2009].

Feature Manipulation Engine (FME[9]) established by Safe Software is leading spatial ETL software that now is implemented in the ArcGIS Data Interoperability extension [ESRI 2009]. FME provides various data transformation functionalities including schema mapper within the geoprocessing environment. While FME is a most widely used tool, other tools such as GoPublisher[10] (Snowflake Software), Spatial Data Integrator[11] (Camp to Camp), and GeoXSLT[12] [Klausen 2006) are currently available for schema mapping. These four tools were compared and analyzed previously [Beckman et al. 2009; Schade 2009; Chunyuan et al. 2010, forthcoming]. Key criteria include GML support, web service, GUI for mapping rule generation, support for mapping rules, and type of software [Chunyuan et al. 2010, forthcoming]. While all of them support GML as output format, only commercial

---

[9] Feature Manipulation Engine. Safe Software. URL: www.safesoftware.com (last accessed on January 13th 2010).
[10] GoPublisher. Snowflake Software. URL: http://www.snowflakesoftware.co.uk (last accessed on January 13th 2010).
[11] Spatial Data Integrator. Camp to Camp. URL: http://www.spatialdataintegrator.com (last accessed on January 13th 2010).
[12] GeoXSLT. http://www.svisj.no/fredrik/geoxslt (last accessed on January 13th 2010).

tools FME and GoPublisher provide a functionality of standard web service interface for OGC WFS. SDI is open source software and offers a graphical user interface (GUI) with many built-in transformation operators as offered by FME.

2.4 Ontology Languages and Reasoning

The Semantic Web evolved from the existing WWW with the advance of knowledge representation, using machine-understandable Web content that can be processed by computer [Berners-Lee et al. 2006]. In the context of the Semantic Web, ontology plays a key role of providing a shared understanding of a domain model to harmonize the heterogeneity in terminology [Berners-Lee et al. 2006]. Domain ontology refers a shared conceptualization between different application-specific models while application ontology is only based on a local knowledge model established by the application service provider [Klien and Probst 2005; Duchesne 2008]. At the top level of ontology architecture, a foundational ontology is formalized by philosophers and cognitive engineers to link different domain ontologies [Klien and Probst 2005; Gruber et al. 2006; Duchesne 2008; Schade 2009].

Currently, commonly used ontology languages are Resource Description Framework (RDF), RDF Schema, and Web Ontology Language (OWL) [W3C 2006]. Ontology languages mainly require a well-defined syntax for machine processing of information, reasoning support for checking consistency of the ontological knowledge, and a formal semantics for describing the meaning of knowledge in a domain [Antoniou and van Harmelen 2008]. While the expressivity of RDF and RDF Schema is limited to representing information at the structural level, OWL expands the expressivity to reasoning based on description logic [Breitman et al. 2007]. In the context of the Semantic Web, reasoning refers to machine-supported inference expressed in a language that can be processed by the algorithm [Duchesne 2008]. Description logic is a formalization of knowledge representation that defines the domain concepts using classes, relations, attributes, and properties in the domain [Breitman et al. 2007]. Description logic can be modeled by reasoning algorithms such as FaCT, RACER, and Pellet [Breitman et al. 2007; Antoniou and van Harmelen

2008;]. Such reasoners may be known as classifiers that compute the inferred class hierarchy [Horridge 2009]. These reasoners can be installed as a plug-in to a software application such as Protégé[13], which is a platform for ontology modeling and knowledge acquisition [Knublauch et al. 2004].

In case of defining forests, Figure 3 illustrates a simple example of conceptualizing of the class hierachy of forests. ‘Land’ is a generic class of ‘Forests’ while ‘Coniferous’, ‘Mixed’, and ‘Broad-leaved’ are more specific classes of ‘Forests’, where a taxonomic structure is indicated in the direction from left to right [Horridge 2009].

Figure 3. Conceptualization of forests in a taxonomic structure.

These concepts can be formalized as entities in Protégé. As shown in Figure 4, the entity ‘Forests’ has a class hierarchy, a description, object properties, and data properties. A class hierarchy may be known as taxonomy and properties may be known as roles in description logics [Horridge 2009]. Object properties describe property characteristics and allow relationships between classes while data properties describe relationships between a class and data values [Horridge 2009]. ‘Forests’ is defined in Equivalent classes that have associated properties. To be classified as equivalent to ‘Forests’, all the conditions specified by object and data properties must be sufficient. ‘Land’ is a superclass of ‘Forests’, which meets necessary conditions to become ‘Land’. In this example, provided that the necessary condition is to have surface that is some ‘LandSurface’, ‘Forests’ has surface that are some ‘Trees’ (sub-class of ‘LandSurface’). In addition, ‘Forests’ is

---

[13] Protégé. Stanford Center for Biomedical Informatics Research. URL: http://protege.stanford.edu (last accessed on November 13th 2009).

disjointed with other sub-classes of 'Land' so that a sample of land cannot belong to 'Forests' and the other sub-classes at the same time in the process of land classification [Horridge 2009]. The automated reasoner infers the class hierarchy based on these descriptions of classes.

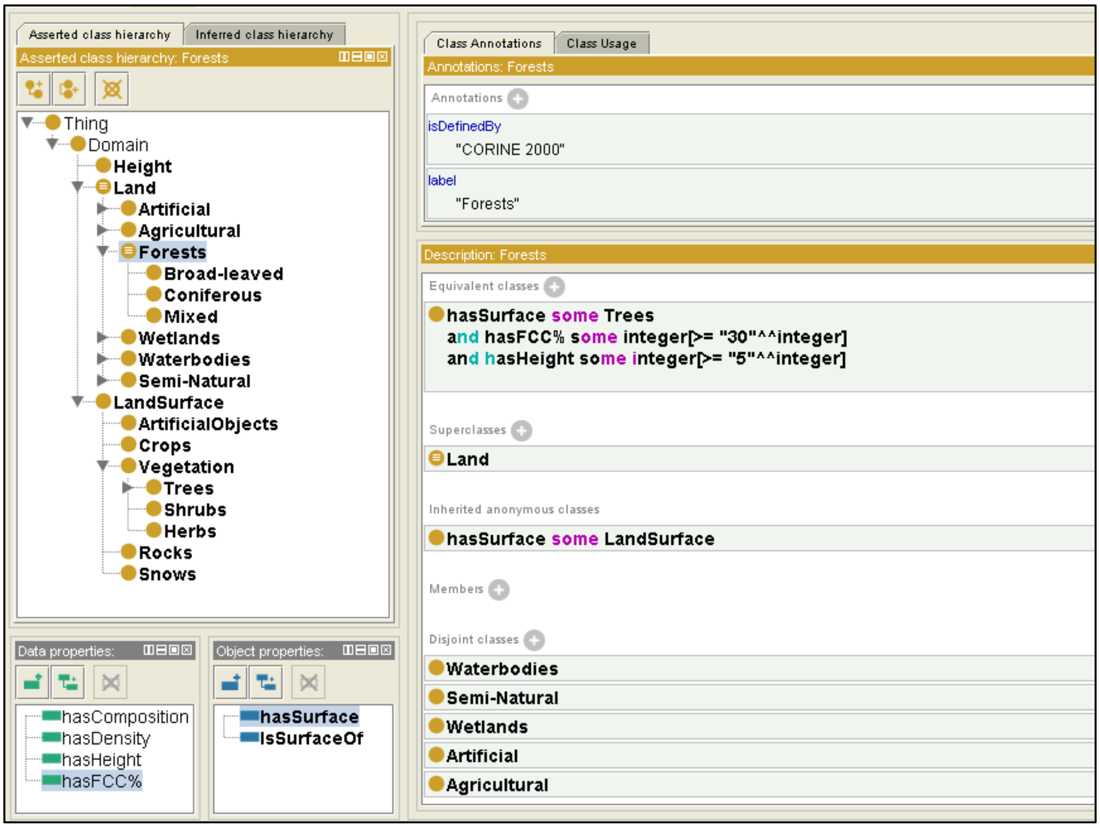


Figure 4. Protégé user interface with a description of the concept of 'Forests'.

2.5 Related Research on Ontology-Based Schema Mapping

This section focuses on recent research related to geospatial data harmonization through ontology-based schema mapping in order to achieve schematic and semantic interoperability.

Lehto and Sarjakoski (2004) presented the EU-funded project GiMoDig that performs schema transformation by using Extensive Stylesheet Language Transformations (XSLT) to encode GML geospatial data. Translation was processed from the local national schema to the jointly agreed global schema in the heterogeneous WFS environment. As a result, a prototype cross-border GML data service was developed in the common data model. In the context of a standards-based web service environment, Lehto (2007) further categorizes schema transformations into operation components which can be applied to different levels of the GML feature model. Some of those schema operations are applied in our methods for testing schematic interoperability (Chapter 4.2)

Donaubauer et al. (2006) recognized that previous projects such as GiMoDig were only executed on the schematic level, lacking transformation between different conceptual schemas. To address this issue, they proposed the project *md*WFS based on Model Driven Architecture (MDA), which supports automated schema transformations on the semantic level. Donaubauer et al. (2007) presented a use case of *md*WFS in the context of a cross-border SDI between Germany and Switzerland. They aim at implementing the prototype service in the contexts of cross-border SDI and INSPIRE, which will enable transformations guided by user-defined schema mapping. Our testing approach also addresses the issue of schema transformations on the semantic level in the context of cross-border SDI in Europe.

Using existing software applications, Friis-Christensen et al. (2005) investigated possible methods of achieving schematic and semantic interoperability of geographic data at European level. The FME was used to support transformations on the schematic level from source attributes to target attributes. This method required comprehensive knowledge of source and target schemas for all the corresponding attributes to be mapped manually. They also explored ontology-based approaches to automated schema mapping, which incorporate semantics to provide generic concepts between source and target applications. The ontology editor Protégé and the reasoner RACER were used to reclassify source and target schemas by concepts established in the domain. They found that the ontology-based classification approach by Protégé could support maximizing automation in schema mapping. In our methods for testing schematic and semantic interoperability, these existing software applications (FME and Protégé) are also used for schematic transformations and ontology-based classification.

Abadie (2009) also took an ontology approach using Protégé to test schema mapping between two national geographic databases in France. Their approach was mainly based on attribute values at the class instance level and background ontology at the domain level. They analyzed attribute values to represent semantic details described in both schemas, which are often hidden at the schematic level. They further used comprehensive domain ontology as background ontology to match two application ontologies. The domain ontology was established by discovering relationships between source and target concepts. They are currently implementing a tool to automatically compare formal specifications and detect heterogeneities. We also take class attribute values and the domain ontology into account for testing semantic interoperability.

Most recently, Schade (2009) extended one of foundational ontologies, Descriptive Ontology for Linguistic and Cognitive Engineering (DOLCE) [Masolo et al. 2003], as a semantic reference frame to define geospatial data models, mapped from a national road data model in Germany to INSPIRE Transport Networks model. They also took a logic-based approach where ontologies were implemented using the Web Service Modeling Language (WSML)-Flight reasoner IRIS (Integrated Rule Inference System), which is a variant based on logic programming that allows for inferring relations between source and target attributes. This approach supported in selecting appropriate translation rules on the attribute level. Our testing approach is based on description logic on the attribute level, however, they demonstrated that WSML based on logic programming provides an alternative to other ontology languages such as OWL based on description logic.

3. CURRENT STATUS OF FOREST DATA AVAILABILITY

In this chapter, we describe how forest data are currently managed at the European, national and regional levels. The availability of forest data via web services is also indicated. In relation to our use case, we focus on forest cover and forest fire data from Europe, Spain, and Valencia Community respectively.

3.1 Current Situation in Europe

At European level, forest cover is identified by CORINE Land Cover Classification 2000, owned by European Environmental Agency (EEA). Land cover is classified into 44 categories where forest cover includes 3 categories: coniferous, broadleaved, and mixed [Nunes de Lima 2005].

Joint Research Centre (JRC) of the European Commission is in charge of establishing EFDAC as the central point for forest information at European level [EuroGEOSS 2009]. As a bridge of forest data flow between the world and Member States, EFDAC functions as part of a nested system for harmonized assessment to link FAO Global Forest Resources Assessment (FRA) and National Forest Inventories of Member States (Figure 5) [JRC 2009]. This way assessment can be harmonized from regional/local Forest Services to National Forest Inventories, EFDAC, and FAO Forestry.

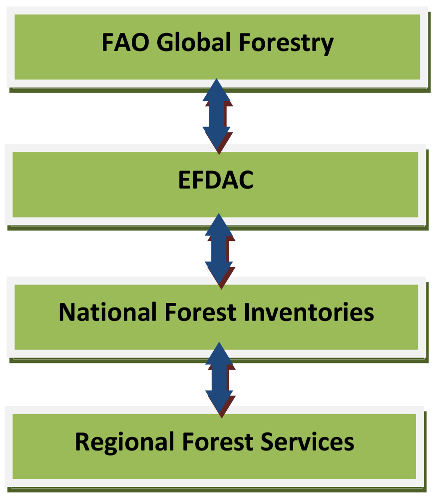


Figure 5. Harmonized assessment in a nested forest information system [JRC 2007].

The EFDAC is advancing as a European FIS, built on three existing systems, European Forest Information and Communication Platform (EFICP) for forest resources, Forest Focus Database for forest conditions, and EFFIS for forest fires [EuroGEOSS 2009]. EFICP manages forest resources data including forest area, ownership, forest type, age class distribution, and growing stock, derived from National Forest Inventories and international data sources such as FAO Forestry [EuroGEOSS 2009]. Forest Focus Database holds data related to forest conditions such as forest map, forest patterns, forest health, and air pollution at different administrative levels [JRC 2008]. Forest fires data are maintained by EFFIS as a complementary system to national and regional systems in Member States in cases where the harmonized forest fire information is required for trans-boundary collaboration [JRC 2009]. Fire related modules are accessible on the EFFIS website such as forest danger forecast, damage assessment, rapid damage assessment, EU fire database, atmospheric emissions, and potential soil erosion. The current situation of fire danger forecast, hot spots and burned area are updated everyday while fire history is maintained in a separate viewer. Table 1 summarizes availability of spatial data related to forest cover and fires in Europe. The data description includes a web address to view or download spatial data. A Map Viewer displays images directly on the addressed web page while a WMS link can be used to add the map server onto a different web portal or a software application.

Data	Description	Source	View	Download
CORINE 2000	CORINE Land Cover in 2000 at 1:100 000 scale http://www.eea.europa.eu/themes/landuse/clc-download	EEA	Map Viewer	Shape
Forest Map of Europe	Pan-European forest/non-forest map with 25 m spatial resolution derived from Landsat ETM in 2000 http://forest.jrc.ec.europa.eu/download/data/forest-map-2000-download	JRC		GeoTIFF
Forest Map of Europe	Pan-European forest/non-forest with 1 km spatial resolution derived from AVHRR and forest statistics http://efdac.jrc.ec.europa.eu/viewer/ http://efdac.jrc.ec.europa.eu/mapserv/mapserv	EFDAC	Map Viewer, WMS	
European Forest Resources	Images and graphs related to forest resources in Europe http://efdac.jrc.ec.europa.eu/index.php/efris	JRC	Map Viewer	
EFFIS Advanced Viewer	Modules include fire danger forecast, damage assessment, rapid damage assessment, EU fire database, atmospheric emissions, and potential soil erosion. http://effis-viewer.jrc.ec.europa.eu/wmi/viewer.html	JRC	Map Viewer	
EFFIS Current Situation	Modules include recent fire danger forecast, daily MODIS, hot spot, and burned areas http://effis.jrc.ec.europa.eu/current-situation	JRC	Map Viewer	
EFFIS Fire History	Annual maps of forest burned area with minimum size of 50 ha http://effis.jrc.ec.europa.eu/fire-history	JRC	Map Viewer	

Table 1. Spatial data available for forest cover and fires in Europe [EuroGEOSS 2009].

3.2 Current Situation in Spain

Forest cover at national level is identified by Land Cover Information System known as SIOSE, the latest version of national land cover classification in Spain, under development for web services by National Geographic Institute (IGN[14]) [SIOSE

[14] Instituto Geográfico Nacional. Ministerio de Fomento. URL: http://www.ign.es/ign/es/IGN/home.jsp (last accessed on December 7th 2009).

2007]. IGN also provides CORINE Land Cover map via WMS. Forest map of Spain is available based on the 3rd National Forest Inventory with 37 land cover classifications [MMA 2009b]. Table 2 shows a list of forest cover data of Spain by different sources.

Data	Description	Source	View	Download
SIOSE	National Land Cover in 2005 at 1:25 000 scale	IGN		
CORINE Land Cover 2000	CORINE Land Cover in 2000 at 1:100 000 scale http://www.idee.es/wms/IGN-Corine/IGN-Corine?	IGN	WMS	
Forest Map of Spain (1:50 000)	Detailed forest inventory map of Spain at 1:50 000 scale http://servicios2.mma.es/wmsconnector/com.esri.wms.Esrimap/BIODIV_MFE?	Biodiversity Data Bank	Map Viewer, WMS	
Forest Map of Spain (1:200 000)	Less detailed forest inventory map of Spain at 1:200 000 scale http://www.mma.es/portal/secciones/biodiversidad/banco_datos/info_disponible/mfe200.htm	Biodiversity Data Bank		Shape

Table 2. Spatial data available for forest cover and fires in Spain [EuroGEOSS 2009].

While the national FIS is not currently established in Spain, the Biodiversity Data Bank from the MMA provides geospatial data including forest inventory and management via Map Viewer and WMS [EuroGEOSS 2009]. This WMS is planned to be implemented into the National SDI, IDEE [MMA 2009a]. IDEE provides access to the main node of distribution connected to the nodes of other web services that are established by autonomous regions. Via this main node it is possible to access national, regional, and local web services (WMS, WFS, WCS, and CSW).

Forest fire data are managed by the MMA. The Coordination Center of the National Wildland Fire Information (CCINIF) was created in 2005 by MMA, which provides the daily updated information about forest fire risks through the CIRCA computer tool [MMA 2009c]. This information is also linked to the EFFIS. As spatial data of burned areas as well as fire risks are not currently accessible on the Web, they need to be requested to the MMA for a specified autonomous region by a request form.

3.3 Current Situation in Valencia Community

Forest cover at regional level is identified by SIOSE as the latest version, coordinated by IGN and CMA (Conselleria de Medio Ambiente, Agua, Urbanismo y Vivienda[15]), and the information is expected to be available in 2010 [SIOSE 2009]. Forest maps at the level of Valencia Community are accessible via Map Viewer and WMS (Table 3). These spatial data are organized in a regional SDI provided by CMA in cooperation with Valencia Cartography Institute[16]. The regional SDI contains a metadata catalogue system following ISO 19115 where the data models are described in a standardized format. While the WMS and CSW are well established, the datasets are not freely downloadable.

Forest fire data such as burned areas are only viewable via internal ArcIMS, which is not accessible to the public. Burned areas in Valencia Community are mapped annually, covering three provinces Valencia, Castellon and Alicante. Although spatial data of burned areas is not accessible, metadata is available via the web catalogue with a description of data attributes.

Data	Description	Source	View	Download
SIOSE	National Land Cover in 2005 at 1:25 000 scale	IGN, CMA		
Forest Map of Valencia (1:10 000)	Detailed forest inventory map in Valencia Community at 1:10 000 scale http://orto.cth.gva.es/wmsconnector/com.esri.wms.Esrimap/wms_invfor?	CMA	Map Viewer, WMS	
Forest Map of Valencia (1:20 000)	Less detailed forest inventory m ap in Valencia Community at 1:20 000 scale http://orto.cth.gva.es/wmsconnector/com.esri.wms.Esrimap/wms_invfor?	CMA	Map Viewer, WMS	
Burned Area Map	Burned areas in Valencia Community at 1:10 000 scale http://intranet.cma.gva.es	CMA	ArcIMS (internal)	

Table 3. Spatial data available for forest cover and fires in Valencia Community [EuroGEOSS 2009].

[15] Conselleria de Medio Ambiente, Agua, Urbanismo y Vivienda, Generalitat Valenciana. URL: http://www.cma.gva.es/intro.htm (last accessed on December 13th 2009).
[16] Instituto Cartográfico Valenciano. URL: http://www.icv.gva.es/ICV (last accessed on December 13th 2009).

4. METHODS FOR TESTING INTEROPERABILITY

Our approach to achieving geospatial data harmonization is to test interoperability on the syntactic, schematic and semantic levels. In this chapter, we introduce the methods for testing interoperability of heterogeneous geospatial data, and relate them to the example use case on forest fire data between EFFIS and CMA. Figure 6 illustrates the workflow of testing interoperability by syntactic approach, and a combination of schematic and semantic approach. National forest fire data in Spain are collected from autonomous regions, therefore, regional data from CMA are directly tested for interoperability with the EFFIS data. Specifically, we apply burned areas for 2007 in Valencia Community mapped by CMA as source data and burned areas for 2007 in Valencia Community mapped by EFFIS as target data. After we set the scene in this chapter, results are presented in Chapter 5.

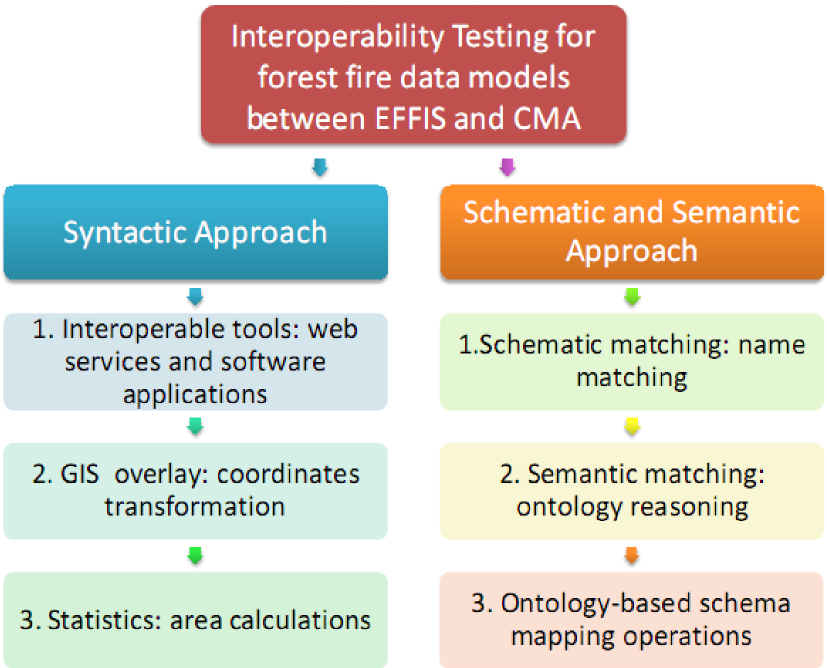


Figure 6. The workflow of interoperability testing for forest data models.

4.1 Syntactic Interoperability Testing

In this section, differences in data access and format are tested for syntactic interoperability. First, tools to interoperate the two datasets of EFFIS and CMA are identified. Then, those data layers are visualized and analyzed qualitatively in GIS overlay. Burned area calculations are also provided to analyze the datasets quantitatively.

4.1.1 Tools Available for Testing

There are various tools that enable multiple data stored in different formats (e.g., shape, Mapinfo, Oracle geo-DB, and PostGIS geoDB) to interoperate syntactically [Vaccari et al. 2009]. In our methods we categorize those tools into standard web services and software applications. Using one of those tools or a combination of them, data in different formats can be displayed in a single client view. Web services implementing standard interfaces such as OGC WMS, OGC WFS and OGC WCS are often available in SDIs, providing web links for viewing and downloading geospatial data in a standardized way. Such data can be accessible via the Web or desktop software, regardless of source formats. Currently there is a number of GIS software applications, ranging from licensed applications such as ArcGIS[17] spread globally, to open source applications such as gvSIG [18] established by the government of Valencia Community. These up-to-date software applications enable geospatial data available in original formats or available from web services to be visualized on the fly.

Thus, we will consider that two datasets can be displayed in the same client view in the following options:

- Option A - Both datasets are viewable via standard web services.
- Option B - Both datasets exist locally as created or downloaded files, which can be added onto a single software application.
- Option C - One dataset is added to a software interface from existing/downloaded files while another dataset from spatial standard web services is added to the same software interface.

In our use case, EFFIS provides a web entry point with Map Viewer to display burned area data. To view the same data outside Map Viewer on the EFFIS website,

[17] ArcGIS. ESRI. URL: http://www.esri.com/software/arcgis/index.html (last accessed on November 7th 2009).
[18] gvSIG. Conselleria de Infraestructuras y Transporte, Generalitat Valenciana. URL: http://www.gvsig.gva.es (last accessed on November 7th 2009).

it is not readily accessible in a standard way via web services, but available in Shape format upon request. On the other hand, burned area data from CMA is accessible via ArcIMS, which is only connected to an internal network[19]. Since ESRI's map and feature services are not standard web services as OGC, accessibility is limited via ArcGIS and other software applications (e.g. gvSIG) that support in displaying ArcIMS. Assuming that we are forest domain experts who work at CMA, we would be able to add CMA data via ArcIMS and EFFIS data in Shape format, using ArcGIS (Option C). If we do not have access to the internal network, we would need to request the data from CMA, for example in Shape format, to display it in ArcGIS with EFFIS data in Shape format as well (Option B).

4.1.2 Visualization Analysis

When two data layers are visualized to locate the common area in the same client view, whether via the Web or desktop software, spatial reference systems (datum, projection) need to be interoperable so that they can be overlaid. In case of rendering images via web services, WCTS transforms the coordinates of geometric elements among different spatial reference systems [INSPIRE 2008]. In case of visualization through software applications, users are responsible for transforming spatial reference systems if they are different. This overlay visualization may show some distortion of the transformed data in terms of direction, area, and shape [Iliffe and Lott 2008]. Depending on the levels of detail contained in each data, overlay analysis can also address scale issues. For example, alignment of two data layers may not be consistent along the polygon boundaries.

We use ArcGIS to do overlay analysis between two data layers from CMA and EFFIS. As the spatial reference system (GCS_European_1950) in the source data by CMA is different from the target data, it needs to be transformed to the same as the target data, GCS_ETRS_1989. This overlay analysis demonstrates visual quality and consistency of two data layers.

[19] Visor Cartográfico Interno - Incendios y Forestal. Conselleria de Medio Ambiente, Agua, Urbanismo y Vivienda., Generalitat Valenciana. URL: http://intranet.cma.gva.es

4.1.3 Quantitative Analysis

Due to the different scales used for source data and target data, the discrepancy in area calculation may easily arise. For example, CMA may use a higher resolution to reflect more details of burned areas at the regional level while EFFIS may use a lower resolution to reflect a larger extent of burned areas across Europe. To address this scale issue, we compare the two datasets quantitatively in terms of the number of burned areas, and the minimum and maximum size of burned areas respectively. Overlay analysis further enables common burned areas mapped by CMA and EFFIS to be calculated by GIS intersection operation. As illustrated in Figure 7, the two datasets are compared for total area mapped by CMA (A), total area mapped by EFFIS (B), common area mapped by both (C), and the difference area mapped by one another (A subtracted by C and B subtracted by C) [Boschetti et al 2008].

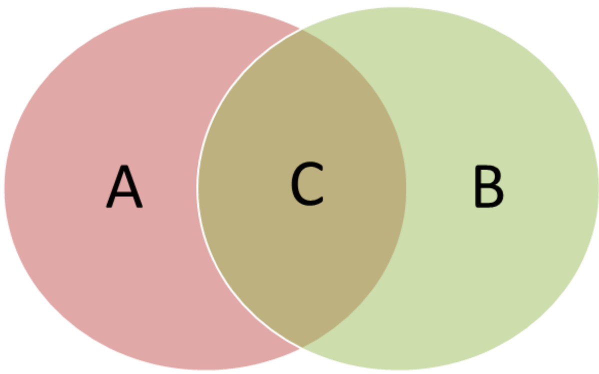


Figure 7. Venn diagram illustrating differences in burned areas mapped by CMA and EFFIS.

4.2 Schematic and Semantic Interoperability Testing

In order to apply schema mapping, interoperability of schemas and semantics between target and source data models have to be taken into account (Chapter 2.4). We explore schema mapping from direct attribute matching simply based on names, to more sophisticated semantic matching based on ontology. Our intent is to combine schematic-level and semantic-level approaches to achieve more or better matching candidates [Rahm and Bernstein 2001]. Examples of data models from CMA and EFFIS are shown in Table 4 and Table 5, respectively. The source data model represents burned areas due to forest fires in Valencia Community. CMA holds this information in its regional SDI to keep track of geographical locations of land cover affected by forest fires every year [CMA 2007]. As shown in Table 4, burned areas are categorized into non-wooded forest surface and wooded forest surface.

Attribute Name	Description in Spanish	Description translated in English
NUMPARTE	Código del parte	Code of the report
MUNICIPIO	Municipio	Municipality
PROVINCIA	Provincia	Province
COMARCA	Nombre de la comarca donde se ubica el recinto	Name of the region where the compound is located
HOJA	Hoja 1:50 000	Mapsheet 1:50 000
FECHA	Fecha del incendio	Date of the fire
TIPO_CAUSA	Causa del incendio	Cause of the fire
SUP_NARBOL	Superficie no arbolada quemada en hectáreas	Non-wooded forest surface burned in hectare
SUP_ARBOLA	Superfice arbolada quemada en hectáreas	Wooded forest surface burned in hectares
SUP_TOTAL	Superficie quemada total en hectáreas	Total forest surface burned in hectares

Table 4. Source data model for burned areas mapped by CMA [Metadata: Incendios 2007[20]].

The target data model by EFFIS represents burned areas damaged by fires in Europe. This information is publicly accessible via Map Viewer on the EFFIS website and is further used for post-fire assessments of atmospheric emissions and erosion risks [JRC 2009]. As shown in Table 5, burned areas are categorized based on land cover classification including non-forest cover types such as agricultural areas and artificial surfaces.

[20] ISO19115 Metadata: Incendios 2007. Conselleria de Medio Ambiente, Agua, Urbanismo y Vivienda, Generalitat Valenciana. URL: http://geocatalogo.cma.gva.es/geonetwork/srv/es/main.home (last accessed on November 7th 2009).

Attribute Name	Description
ID	Unique identification code
Country	Country acronyms
CountryFul	Full name of the country
Province	Province of the commune
Commune	Commune which include the largest burned area relative to the mapped fire
FireDate	Starting date of the fire
Area_HA	Total area (forest and non-forest) burned in hectares
BroadLea	% of broad leaved forest burned
Conifer	% of coniferous forest burned
Mixed	% of mixed forest burned
Scleroph	% of sclerophyllous vegetation burned
Transit	% of transitional vegetation burned
OtherNatLC	% of other natural areas burned and not related to the above mentioned classes
AgriAreas	% of agricultural areas burned
ArtifSurf	% of artificial surfaces burned
OtherLC	% of other land cover burned (not related to the above mentioned classes)
LastUpdate	Acquisition date of the most recent Modis image used to map the burned area

Table 5. Target data model for burned areas mapped by EFFIS [JRC 2009].

Based on the source and target data models presented in Tables 4 and 5, the following sections focus on the data attributes to perform schema mapping on the schematic and semantic levels. We first examine schematic matching based on names of attributes and then analyze semantic matching based on concepts of the attributes. Finally, semantically matching attributes are transformed by schema mapping operations.

4.2.1 Linguistic Matching Approach on the Schematic Level

Firstly, attribute names presented in Table 4 need to be translated from Spanish to English. When the translated source attribute can be directly matched to the target attribute, schema mapping is simple. Difference terms used in different languages can be translated by referring to a multi-language dictionaries or thesauri [Madhavan et al. 2001; Rahm and Bernstein 2001].

At the initial phase of schema mapping, we apply name matching to achieve schematic interoperability, one of linguistic matching approaches which maps attributes with equal or similar names (Figure 8) [Rahm and Bernstein 2001]. Name-based matching can be defined in exploiting short-forms (Qty for Quantity), acronyms (UoM for UnitOfMeasure), synonyms (Car and Automobile), and hypernyms (Tree and Oak) [Madhavan et al. 2001; Rahm and Bernstein 2001]. With a good knowledge of source and target schemas, name matching can be mapped and illustrated by the Spatial ETL, FME.

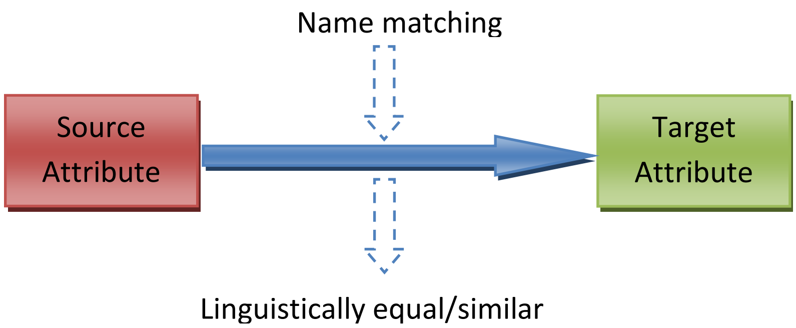


Figure 8. Linguistic matching approach to schematic interoperability (Step 1).

4.2.2 Ontology-Based Matching Approach on the Semantic Level

When the term of a source attribute is translated to match with the target attribute, accuracy of linguistic matching can be assessed by defining entities of each attribute. Ontologies (Chapter 2.2) aid to ensure whether the direct language translation is sufficient for schema mapping.

Ontology-based mapping is one possibility to determine a shared concept between source and target attributes to achieve semantic interoperability. For this purpose, we first aim to establish application ontologies from source and target data models and then examine if establishment of the domain ontology can refine the shared concept between application ontologies [Klien and Probst 2005] to increase the level of semantic interoperability.

Conceptualization of entities can be visualized and described using software applications such as Protégé. Protégé represents ontologies that define classes, properties, property facets and constraints, instances, and the relationships between them [Knublauch et al. 2004]. We take a classification mapping approach with Protégé by establishing ontologies from schemas used in source and target data [Friis-Christensen et al. 2005]. We propose to establish two application ontologies based on data specifications defined by CMA and EFFIS. Currently, there are no official data specifications defined by CMA or EFFIS for burned areas. Since the majority of attributes contained in both data models are related to forest cover, data schemas for Forest Inventory in Valencia Community by CMA and CORINE Land Cover Classification by EFFIS are used as a guide to classification in Protégé. Studying how data attributes are specified in source and target schemas can describe classes, properties, and relationships. As shown in Figure 9, using a reasoner function (FaCT++) we reclassify source and target attributes to identify equivalent classes and similar classes [Friis-Christensen et al. 2005]. The reasoner based on description logic uses the descriptions of the classes to test if an equivalent or similar class relationship exists between them [Horridge 2009]. The reasoner can also help build the domain ontology shared by application ontologies by detecting inconsistencies, hidden dependencies, redundancies, and wrong classifications [Knublauch et al. 2004].

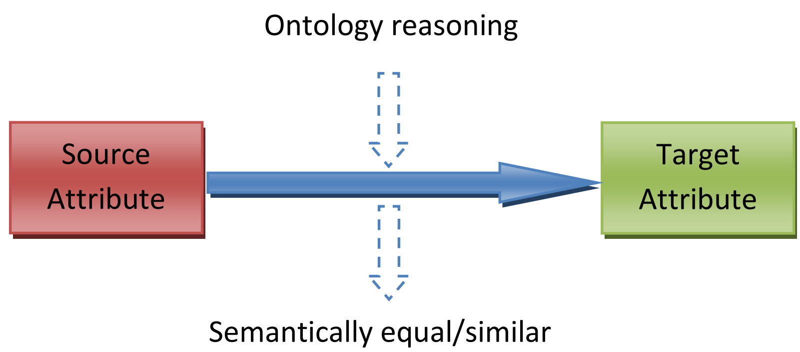


Figure 9. Ontology-based matching approach to semantic interoperability (Step 2).

4.2.3 Ontology-Based Schema Mapping

Following ontology-based attribute matching in Protégé, data transformations are performed by various schema mapping operations from source to target attribute (Figure 10). We apply the following schema mapping operations in the GML feature model suggested by Lehto [2007], in order to generate schema mapping rules at the levels of attributes and attribute values [HUMBOLDT 2009; Schade 2009]:

1. Filtering attributes,
2. Renaming attributes or their values,
3. Reclassification of attribute values by converging or diverging,
4. Merging / splitting attributes values,
5. Changing the order of attributes,
6. Value conversions: spatial generalization and unit conversion of attribute values,
7. Morphing spatial types and data types, and
8. Augmentation of attribute values by interpolation and default.

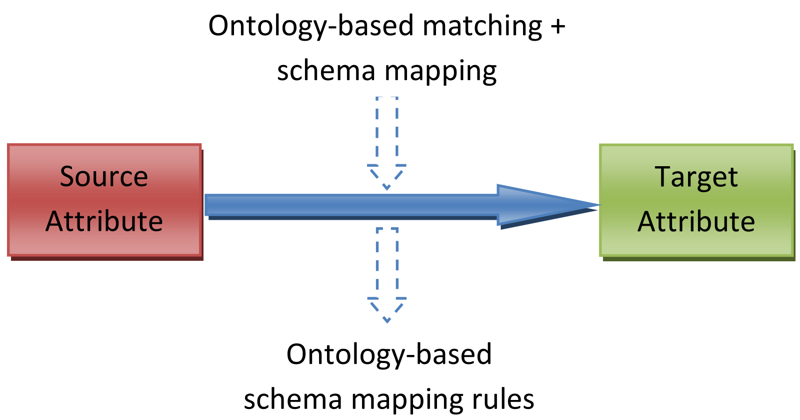


Figure 10. Ontology-based schema mapping rules for schema transformation (Step 3).

5. RESULTS

This chapter presents the results of interoperability testing for geospatial data across regional, national and European administrative levels. Availability of forest cover data and burned area data is compared respectively to indicate the level of syntactic interoperability. For testing schematic and semantic interoperability, matching attributes of burned area data models between CMA and EFFIS and the associated mapping operations are presented. Discussions and recommendations follow in the next chapter.

5.1 Data Availability and Syntactic Interoperability

In this section we compared the current status of data availability for forest cover and burned areas among Valencia Community, Spain, and Europe. By testing syntactic interoperability, we achieved GIS overlay of commonly mapped burned areas between CMA and EFFIS and analyzed the consistency of such geospatial data qualitatively and quantitatively.

5.1.1 Forest Cover Data

Table 6 summarizes syntactic interoperability of forest cover data in the context of format and accessibility via web services between Valencia Community, Spain, and Europe. At European level, EFDAC provides forest cover map at 1:5 000 000 scale [Nunes de Lima 2005] via Map Viewer and WMS in English as a common language across Europe. This information is based on the land cover classification from CORINE 2000 [Nunes de Lima 2005] and downloadable in Shape format via the webpage provided by EEA. Most updated land cover data mapped in 2006 are currently under processing. At national level, in Spain, IDEE provides the forest cover map at 1:50 000 scale via WMS in Spanish as a national standard language. This is based on the Third National Forest Inventory conducted from 1997 to 2006 by MMA. At regional level, Valencia Community in Spain, CMA has its regional SDI which shows forest cover map at 1:10 000, available via WMS in both Spanish and its regional language Valencian.

SDI Level	Regional	National	European
Area	Valencia Community	Spain	Europe
Layer	Forest (Inventario Forestal de la Comunidad Valenciana)	Forest (Mapa Forestal de España)	Pan-European Forest/Non-Forest Map
Access	OGC WMS	OGC WMS	OGC WMS
Format	PNG, JPEG	PNG, GIF, JPEG	PNG, GIF, JPEG, TIFF
Metadata	ISO 19115	ISO 19115	ISO19115
Year	2005	1997-2006	2000
Scale	1:10 000	1:50 000	1:5 000 000
Temporal resolution	Not regulated	Every 10 years	1990, 2000, 2006
Spatial reference system	ED50 / UTM Zone 30N	ED50 / UTM Zone 30N	ETRS89 /ETRS-LAEA
Projection	EPSG:4326 EPSG:23030	EPSG:4326 EPSG:23030	EPSG:3035 EPSG:4326
Language	Spanish or Valencia	Spanish	English
Owner	CMA	MMA	JRC

Table 6. Comparison of forest cover data between regional, national and European levels.

Advanced standard web services at regional, national, and European levels enable the data access to be interoperable within Europe. We can use a GIS software application as a common platform for adding forest cover data from CMA, IDEE, and EFDAC via OGC WMS. Forest cover data from these three sources can be added as layers in the software GUI and the GIS overlay of images from CMA and EFDAC is illustrated in Annex I. Legends of forest cover data are not shown via WMS, however, they can be viewed in Map Viewer on each web portal of CMA, IDEE, and EFDAC. CMA displays forest and non-forest areas in polygon while EFDAC displays forest and non-forest areas in raster. The GIS overlay made it apparent that scales and languages are significantly different even for the most fundamental forest data.

5.1.2 Forest Fire Data

As forest fire data in Spain at national level derives from autonomous regions, a comparison is made in Table 7 directly from regional to European level. At European level, EFFIS at JRC provides burned area data in English via Map Viewer for the fire history of member states. Although the fire history is well updated, the spatial resolution is low when one attempts to focus on a region such as Valencia Community. The minimum fire size mapped is coarse due to the low spatial resolution. In the regional SDI provided by CMA, the forest fire theme in general is not freely accessible to the public. Moreover, the up-to-date information is not available on the Intranet since burned areas from the last summer 2009 are not inputted into the SDI yet. However, it must be noted that the quality of spatial resolution is much higher than the burned area data provided by EFFIS, which is why fires as small as 0.05 hectare can be detected.

SDI Level	Regional	National	European
Area	Valencia Community	Spain	Europe
Layer	Forest Fires (incendios)		Burned Area
Access	ArcIMS (internal)		Map Viewer
Format	PNG, (Shape upon request)		OWS, (Shape upon request)
Metadata	ISO 19115		ISO 19115
Year	1993-2007		1987-2009
Spatial resolution	20 m		250 m
Minimum fire size	0.05 ha		50 ha
Temporal resolution	annually		annually
Spatial reference system	ED50 / UTM Zone 30N		ETRS89 /ETRS-LAEA
Projection	EPSG:23030		EPSG:3035
Language	Spanish or Valencia		English
Owner	CMA		JRC

Table 7. Comparison of forest fire data between regional and European levels.

5.1.3 Burned Area Calculations

This section presents the results of quantitative analysis for burned areas mapped by CMA and EFFIS. Figure 11 shows an example of the overlay image areas between CMA data and EFFIS data. It is apparent that CMA data shows more detailed mapping along the boundary of polygon A. Another scale issue illustrates that some small burned areas (polygons B, C, and D) mapped by CMA are missing in the EFFIS data layer. As indicated in Table 7, this is due to the high spatial resolution and the minimum fire size adopted by CMA at regional level.

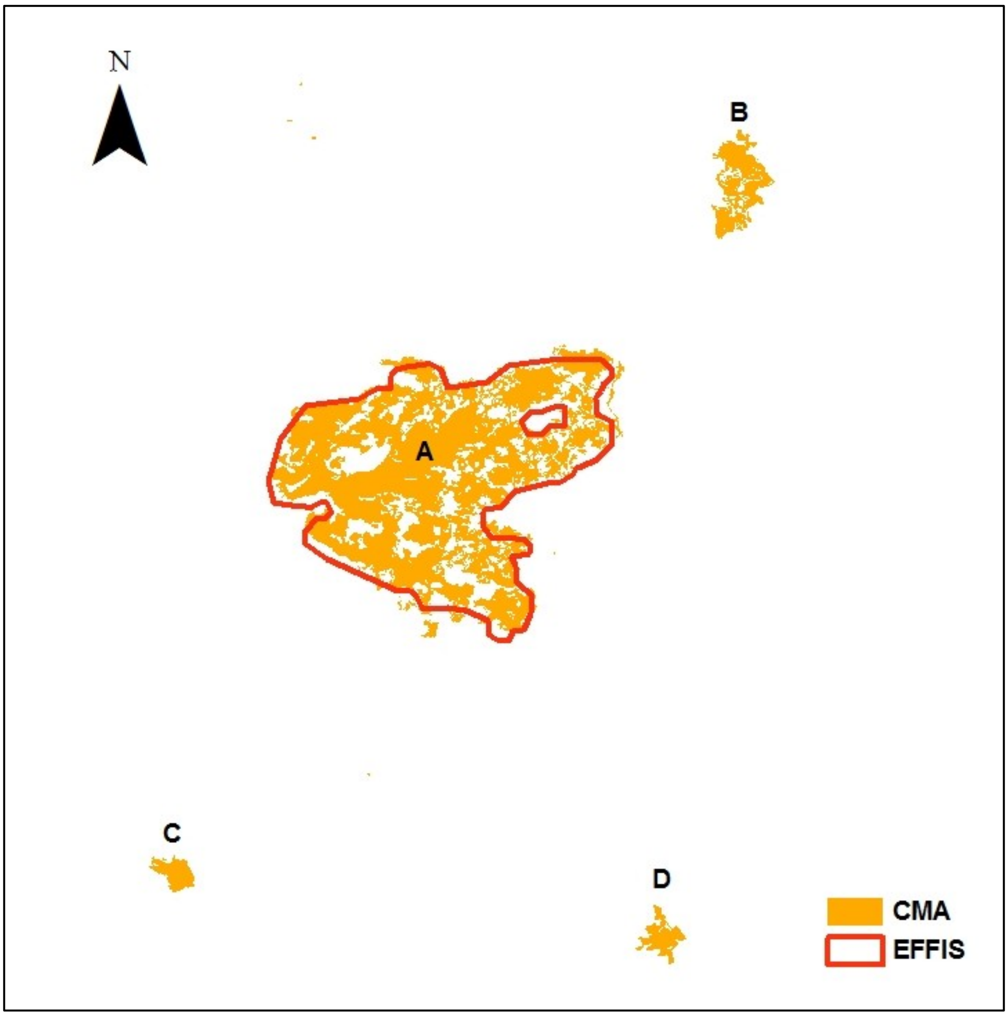


Figure 11. GIS overlay image of burned areas in Valencia Community mapped by CMA (solid polygons) and EFFIS (solid lines) at scale 1:20 000.

To quantitatively compare the discrepancy caused by map scales between the two data models, burned area calculations are summarized and compared in Table 8. While EFFIS only mapped three fires in Valencia Community in 2007, CMA mapped a number of small fires, resulting in mapping a larger total area. The largest fire they both recorded refers to the same fire (polygon A in Figure 11), however, there is a difference of 2266 hectares in burned area between them. We found that CMA and EFFIS use different mapping schemas for representing burned areas, where CMA only maps forest cover burned (excluding non-forest land cover) while EFFIS maps land cover burned (including non-forest land cover).

	CMA	EFFIS
Number of Fires	372	3
Minimum Size (ha)	0.05	89.01
Maximum Size (ha)	5860.00	8125.56
Total Area (ha)	8524.88	8315.20

Table 8. A summary of burned areas in Valencia Community mapped by CMA and EFFIS.

In addition, GIS overlay analysis of the two data layers enabled us to calculate the intersected area as a common burned area mapped by both CMA and EFFIS (Table 9). The commonly mapped burned area was 5565 hectares, which accounts for 67% of the total area mapped by CMA and 65% of the total area mapped by EFFIS. In other words, 33% of the total area mapped by CMA was not detected as burned by EFFIS, and 35% of the area total area mapped by EFFIS was not considered as burned by CMA. These discrepancies were caused by scales, spatial resolutions, and mapping schemas.

Total area mapped by CMA	Total area mapped by EFFIS	Common area mapped by CMA and EFFIS	Area mapped as burned by CMA but not by EFFIS	Area mapped as burned by EFFIS but not by CMA
8525	8315	5565	2960	2750
67%	65%		33%	35%

Table 9. Quantitative overlay analysis of burned areas in hectares mapped by CMA and EFFIS.

5.2 Ontology-Based Schema Mapping

This section shows outcomes of testing interoperability on the schematic and semantic level. Matching attributes are presented following name-matching approach and ontology-based matching approach. In addition, schema mapping operations are identified corresponding to matching attributes.

5.2.1 Name Matching Attributes

Firstly, source data attributes expressed in Spanish are manually translated into English using a multi-language dictionary (Table 4). Based on the names of source attributes translated into English, they are manually matched to the names of target attributes in the FME, as illustrated in Figure 12. There are four matching attributes from to target data models:

- NUMPARTE → ID
- PROVINCIA → Province
- MUNICIPIO → Commune
- FECHA → FireDate
- SUP_TOTAL → Shape_HA

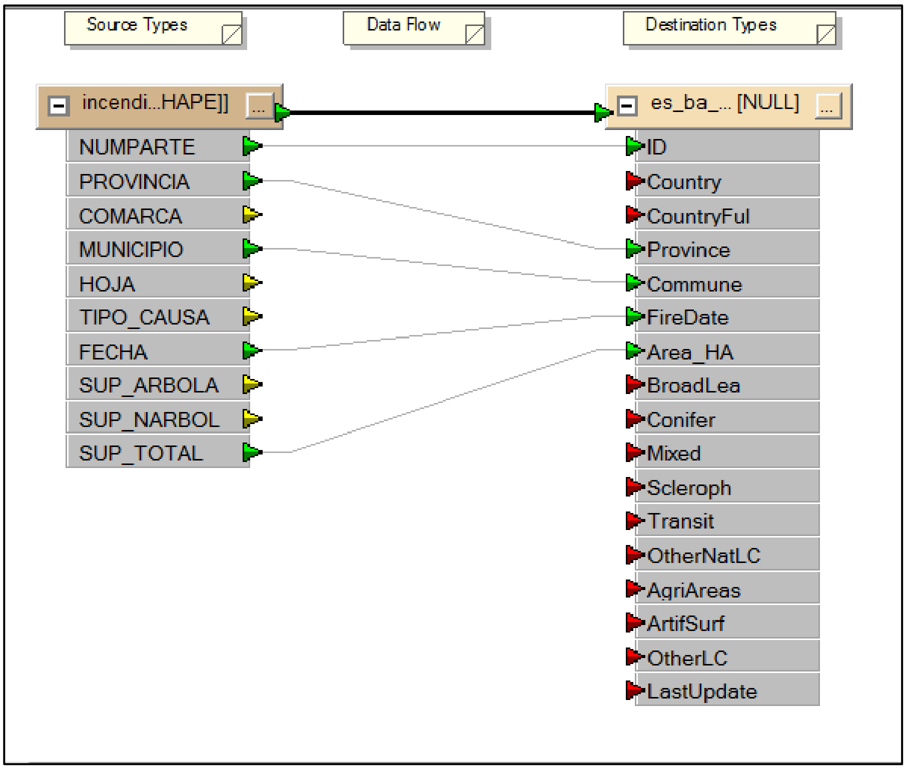


Figure 12. Name matching attributes from source to target data using the FME.

5.2.2 Ontology-Based Matching Attributes

Ontology reasoning aided to assess consistency between definitions used for name-matching attributes in the previous section. The five name-matching attributes are further tested on the semantic level using the reasoner in Protégé, which infers equivalent classes and the class hierarchy based on the definitions of classes described in application ontologies. Figure 13 illustrates the result of ontology-based matching attributes between the two application ontologies from source to target. Four matching attributes were identified:

- NUMPARTE = ID
- PROVINCIA = Province
- MUNICIPIO = Commune
- FECHA = FireDate
- SUP_TOTAL ≠ Shape_HA → SUP_TOTAL ≈ superclassOf(Shape_HA)

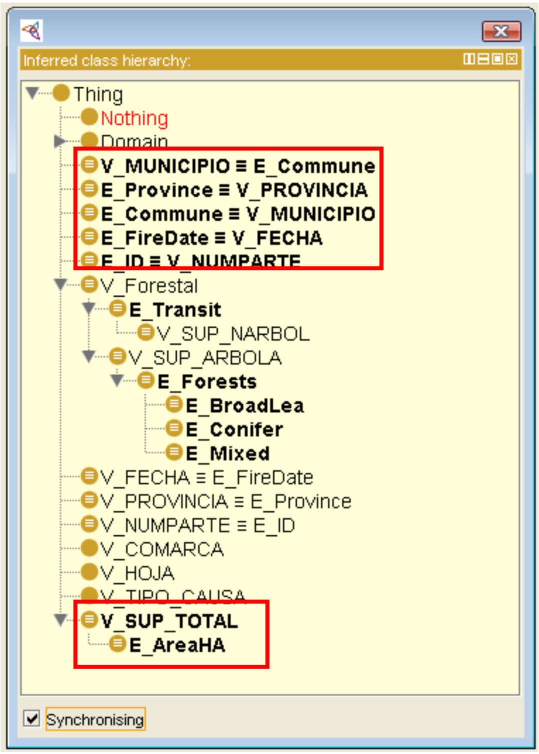


Figure 13. Equivalent and similar classes inferred by reasoner in Protégé for name-matching attributes.

'NUMPARTE' and 'ID' were inferred as equivalent classes as they both have an object identifier in number. 'PROVINCIA' and 'Province were also interfered as equivalent according to the administrative level specified by European Union (i.e. NUTS[21] Level Code 3). In the same manner, 'MUNICIPIO' and 'Commune' that belong to NUTS Level Code 5 were inferred as equivalent. 'FECHA' and 'FireDate' were confirmed as equivalent classes as they both have the date of a fire event. On the other hand, the reasoner did not infer 'SUP_TOTAL' and 'Shape_HA' as equivalent classes. This is because 'SUP_TOTAL' is defined by the total forest area burned while 'Shape_HA' is identified by the total land area (forest and non-forest)

[21] Nomenclature of Territorial Units for Statistics. Eurostat. URL: http://simap.europa.eu/codes-and-nomenclatures/codes-nuts/index_en.htm (last accessed on December 7th 2009).

burned. Nevertheless, both attributes share the common class description of having the forest area burned, which resulted in that 'SUP_TOTAL' is superclass of 'Shape_HA'.

Some source attributes are not matched to the target attributes on the schematic level for two reasons. One is that the matching attributes simply do not exist. In such case, those source attributes may be lost after schema mapping [HUMBOLDT 2009]. For example, 'TIPO_CAUSA' in source data does not have any matching candidates related to the type of cause in target data. Another reason is that some source attributes have matching candidates but the definitions of attributes used in source and target data models are not known. Examples of those source attributes in our use case are related to forest cover types.

The application ontology based on the source data model is based on forest cover classification in Valencia Community defined by CMA while the application ontology based on the target data model follows CORINE land cover classification defined by EEA. Criteria to define forest classes include tree type, tree height, and canopy cover closure. Using the reasoner in Protégé, equivalent and similar classes can be reclassified. There are no equivalent classes found by ontology reasoning due to the complexity of forest type definitions in both source and target classifications. However, the following similar classes are inferred in the same manner as shown in Figure 14:

- SUP_NARBOL ≈ subclassOf (Transit)
- SUP_ARBOLA ≈ superclassOf (Forests: Conifer/BroadLea/Mixed)

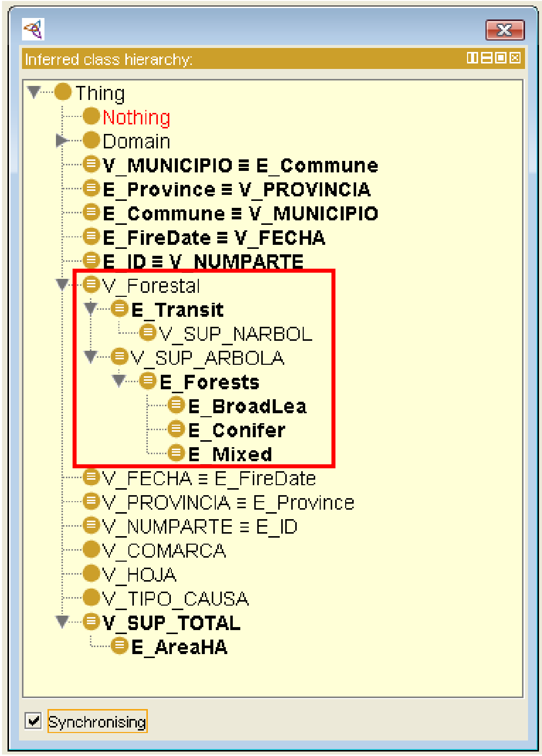


Figure 14. Similar classes inferred by reasoner in Protégé for attributes based on forest types.

CORINE defines forests as land with a canopy cover of greater than 30% [Nunes de Lima 2005] while Spanish Forest Inventory defines forests ('forestal' in Spanish) as land with a canopy cover of greater than 5% [MMA 2009b]. Spanish forests are further categorized into sub types by canopy cover where Valencia Community defines wooded ('arbolado' in Spanish) forests with a canopy cover of greater than 20%. Therefore, the reasoner in Protégé inferred that the source attribute 'SUP_ARBOLA' is superclass of the target attribute 'Conifer[22]' (or 'BroadLea' or 'Mixed').

As demonstrated above, it is often the case that source and target attribute are not equivalent between application ontologies. Establishing the domain ontology can be a solution to identify an equivalent class defined by a common concept between application ontologies. One possibility is to establish the domain ontology based on forest cover classification in a larger scale than Europe. For example, FAO Forestry defines forest as land with a canopy cover of greater than

[22] In Valencia Community, the majority of forest cover is occupied by sub type 'Conifer' [SIOSE 2009].

10% in the Global FRA 2005 [FAO 2004]. When such forest cover classification at global level is introduced as the domain ontology, the reasoner in Protégé reclassifies source and target attributes again shown in Figure 15. With the domain ontology, the source 'SUP_ARBOLA' and the target 'Conifer' both belong to their superclass 'G_Forest' based on the shared concept of having a canopy cover of greater than 10%.

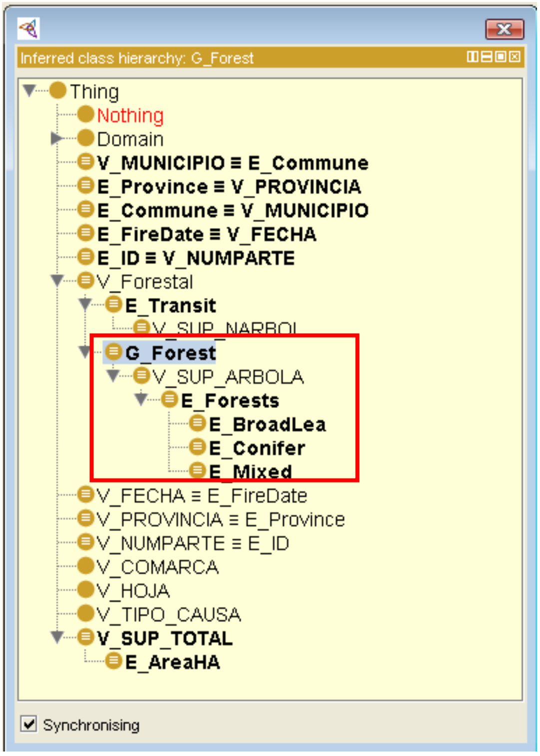


Figure 15. Establishment of the new shared concept between application ontologies based on the domain ontology.

The name matching approach on the schematic level was not sufficient, thus we revised the Spatial ETL to add new matching attributes identified by the ontology-based matching approach. As shown in Figure 16, 'SUP_ARBOLA' was mapped to 'Conifer' and 'SUP_NARBOL' was mapped to 'Transit'.

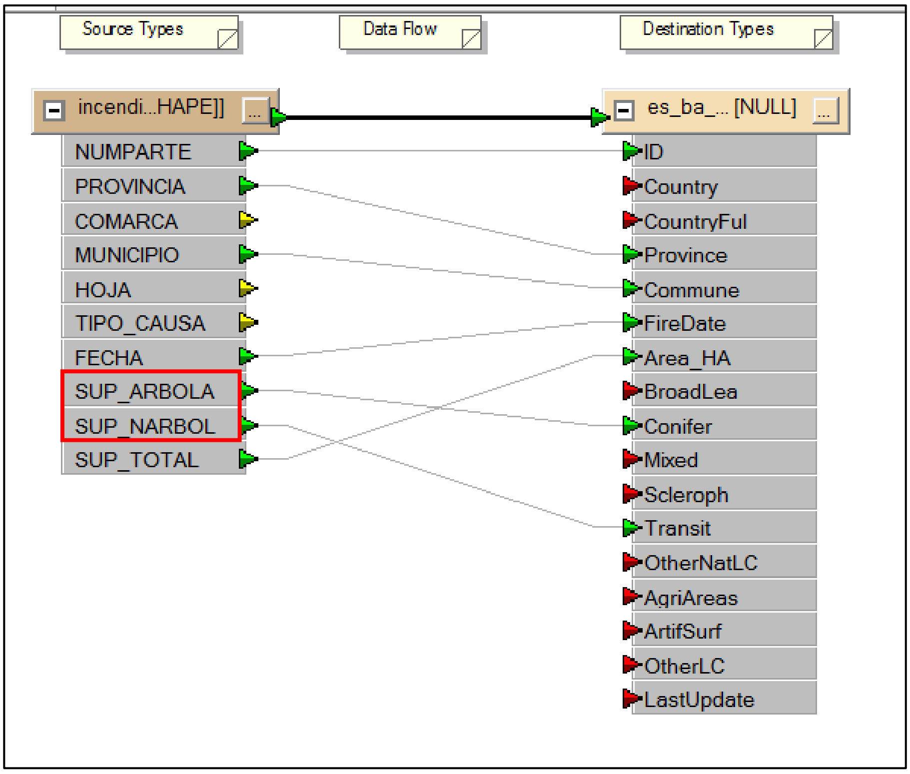


Figure 16. Ontology-based matching attributes mapped in the FME.

5.2.3 Schema Mapping Operations

To map matching attributes from source data to target data, various types of mapping operations are often required for transforming source to target attributes (Chapter 4.2.3). Some mapping operations such as augmentation can be applied for manipulating target attributes in cases where matching attributes are not found. Table 10 summarizes the types of schema mapping operations that we applied for transformation of matching attributes and manipulation of target attributes [Lehto 2007; Schade 2009; Chunyuan et al. 2010, forthcoming].

Source Attribute	Target Attribute	Rename	Change Order	Convert Value	Morph	Augment
NUMPARTE: decimal	ID: integer	x	x		x	
	Country: character					x
	CountryFul: character					x
PROVINCIA: character	Province: character	x	x			
MUNICIPIO: character	Commune: character	x	x			
FECHA: character	FireDate: character	x	x			
SUB_TOTAL: character	Area_HA: integer	x	x	x	x	
	BroadLea: decimal					
	Conifer: decimal					
SUB_ARBOL: character	Mixed: decimal	x	x	x	x	
	Scleroph: decimal					
SUP_NARBOL: character	Transit: decimal	x	x	x	x	
	OtherNatLC: decimal					
	AgriAreas: decimal					
	ArtifSurf: decimal					
	OtherLC: decimal					
	LastUpdate: character					x

Table 10. Schema mapping operations to transform source data attributes to target data attributes.

Based on the type of mapping operations identified for each attribute in Table 10, we added more transformations in the FME (Figure 17). Renaming and changing order of attributes are automatically operated once source and target attributes are mapped. As the workflow of mapping operations is illustrated in Figure 17, morphing was applied to 'SUP_ARBOLA', 'SUP_NARBOL', and 'SUP_TOTAL' to change their data type from character to decimal. This transformation was necessary for the following mapping operations of converting their values from hectares to percentages. Finally, augmentation was applied for target attributes 'Country', 'CountryFul', and 'LastUpdate' by adding known values from source data and metadata. Mapping rules to perform this transformation are saved in the FME mapping file and the example of 'Convert Value' operation is shown in Annex II.

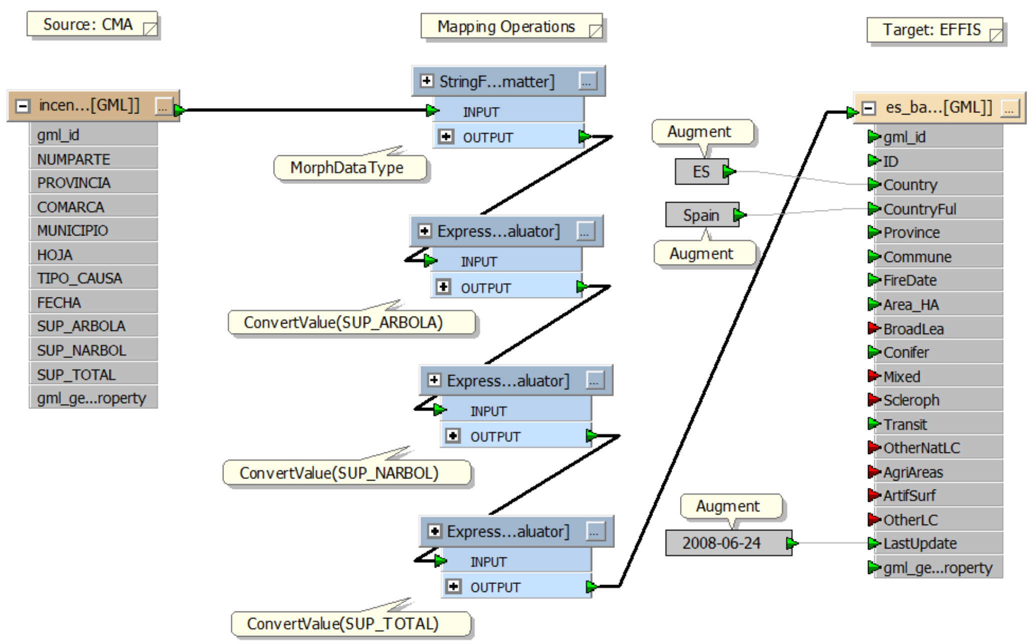


Figure 17. Mapping operations added in the FME to perform transformations from source to target attributes.

6. DISCUSSIONS AND RECOMMENDATIONS

In this chapter, we discuss issues found in the results and suggest solutions to increase interoperability on syntactic, schematic and semantic levels.

6.1 Issues of Syntactic Interoperability

In our use case, the overview of available geospatial data related to forestry showed how heterogeneous they are from regional, national to European level in the context of data standards and accessibility. At service level, forest data from Europe and member states (Spain in this case) are syntactically interoperable at least for the most fundamental theme, forest cover maps via OGC WMS. EFDAC is attempting to be in compliance with INSPIRE where implementation of OGC web services is required. Therefore, EFDAC including EFFIS is expected to accelerate the process of implementation for such standard web services. Syntactic interoperability of forest fire data is more difficult to achieve than forest cover data. Burned area data provided by EFFIS is easily accessible via Map Viewer although the WMS link is not available yet. Burned area data provided by CMA is only available via ArcIMS in Spanish or Valencia (regional language) for the internal use, thus it is not accessible to the public. To increase syntactic interoperability, the link of ArcIMS should be publicly available as the WMS link for forest cover data is publicly available in the same SDI. Not only allowing the link of ArcIMS to be public, it also can be standardized to OGC web services since only few software applications support ArcIMS.

At client level, commercial software applications such as ArcGIS allow layers to be added via web services such as ArcIMS and OGC WMS, WFS and WCS, which increase the chance of achieving syntactic interoperability among heterogeneous data. However, it may not be affordable for some users to buy commercial tools. As an alternative, a number of open source GIS software applications are currently available in various languages and communities. Some of them support various vector and raster formats as well as standard web services. Examples of such applications include uDig[23] and OpenJUMP[24]. Another open source software application, gvSIG, developed by the Valencia government supports ArcIMS additionally.

As more geospatial data are standardized to OGC web services, the level of syntactic interoperability can be increased at service level. At client level, more formats and web services should be supported by software applications.

6.2 Issues of Schematic Interoperability

On the schematic level, we address the issue of harmonizing source and target data models. Data structure at attribute level was easily manipulated by FME to harmonize the two data models.

To identify matching attributes, name matching approach was quick and simple when source and target attributes matched linguistically. Some attributes such as 'Country' may be an easy example for name matching since it is well defined at the administrative level. 'Date' can be another easy name matching example within Europe where the international standard is applied (i.e. Gregorian calendar) [Sumrada 2003]. However, some countries like China use their own traditional calendar systems [Sumrada 2003].

The drawback of using name matching is that we can easily misinterpret the meanings of attributes. For example, schemas used to map burned areas are fundamentally different between CMA and EFFIS. The total burned area ('SUP_TOTAL') mapped by CMA only includes forest cover while the total burned area ('Area_HA') mapped by EFFIS includes forest and non-forest cover. The total burned area by CMA may have been underestimated due to excluding non-forest cover burned. If there were some source attributes that indicate non-forest burned area in the data model, we could have transformed the data by recalculating the attribute value of the total burned area.

Name matching approach may not assure if the definition of each attribute is the same among different communities and languages. When we studied

[23] uDig. Refractions Research. URL: http://udig.refractions.net (last accessed on January 23th 2010).
[24] OpenJUMP. Vivid Solutions Inc. URL: http://jump-pilot.sourceforge.net (last accessed on January 23th 2010).

classification of forest types used by EFFIS and CMA, we found that 'Forests' by EFFIS refers to land with a canopy cover of greater than 30% while 'Forestal' (linguistically equivalent to 'Forests') by CMA refers to land with a canopy cover of greater than 5%. Therefore, there are cases where names ('Forests' and 'Forestal') match, but the definitions are inconsistent.

As discussed above, heterogeneous data models were structurally harmonized by FME. Linguistic approach did not always provide sufficient information to identify some matching attributes, which led us to apply ontologies to schema mapping as a solution to this issue.

6.3 Issues of Semantic Interoperability

In comparison with the name matching approach, the ontology-based matching approach enabled us to identify more matching attributes from source to target data. However, establishment of application ontologies was time-consuming for defining each class with associated properties. Even with the aid of application ontologies based on source and target data specifications, identifying a common concept between them remained difficult for some attributes such as forest types. In our use case, we tested the concept of forest defined by FAO Forestry to establish the domain ontology based on the nested forest information system described in Figure 3. This is one way to establish the domain ontology derived from existing forest type specifications at global scale. As a more sophisticated approach to establishing the domain ontology, we could create the new domain ontology derived from various application ontologies. This approach may involve investigating application ontologies from other member states than Spain to reach common ground which they all can commit to [Klien and Probst 2005].

Heterogeneous geospatial data can be harmonized between application ontologies where semantics of terms used in source and target applications match easily. In cases where semantic common ground cannot be reached at application level, the domain ontology may enable semantic matching. Establishment of the domain ontology only by domain experts, who are knowledgeable about forestry, may not be sufficient. The domain ontology can be improved by collaboration with various participants, including philosophers who can guide domain experts with a foundational ontology, ontology engineers who are experienced with knowledge software applications (e.g. Protégé), and service providers who develop forestry data models from different communities [Klien and Probst 2005; Gruber et al. 2006; Schade 2009].

7. CONCLUSIONS AND FUTURE WORK

Geospatial data harmonization from regional level to European level was investigated, with a use case in forest fire data derived from Valencia Community in Spain and Europe. To harmonize heterogeneous data among different communities, languages, and administrative scales, we tested interoperability on the syntactic, schematic and semantic levels.

For testing syntactic interoperability, we studied a common platform in the context of data formats and accessibility via web services. To answer our research question whether forest fire data from EFFIS and member states (CMA in our use case) are syntactically interoperable, we found that standard web services need to be implemented in all administrative scales to achieve interoperability at service level. At client level, we found that GIS software applications that support various formats and standard web services can increase the chance of achieving interoperability. Thus, our findings supported the hypothesis A that establishing standard web services and common tools can increase the chance of achieving syntactic interoperability between multiple geospatial data derived from different sources. In addition, we achieved syntactic interoperability at client level and analyzed the GIS overlay to answer another research question whether there are any scale issues of forest fire data from different sources. We conclude that there are significant discrepancies in the total burned areas mapped by EFFIS and CMA due to the difference in scales.

For testing schematic and semantic interoperability, we took the ontology-based schema mapping approach to transforming a regional data model to a European data model on the conceptual level, with combined techniques of a Spatial ETL tool and an ontology modelling software application. The FME enabled various types of data transformation from source to target attributes to achieve schematic interoperability. Ontological modelling in Protégé helped identify a common concept between the source and target data models, especially in cases where matching attributes were not found at the schematic level. More specifically, application ontologies were established by studying forest cover classifications and definitions of each application, combined with the domain ontology, to reach common ground between applications and achieve a higher level of semantic interoperability. These findings are answers to our research question of how forest fires data can be transformed and mapped into common schemas and semantics across administrative scales. Finally, we support the hypothesis B that the regional data model can be transformed to the European data model on the semantic level when common schemas and concepts are identified.

Our methodology for testing interoperability suggested available tools such as ArcGIS software application on the syntactic level, FME on the schematic level, and Protégé on the semantic level. These existing tools were appropriate to explore our research questions and support the hypotheses, however, our approach could be improved by testing other available tools. On the syntactic level, an open source GIS application gvSIG would perform as well as ArcGIS to deal with various data formats and web services (including ArcIMS). Another open source Spatial ETL application Spatial Data Integrator would replace FME for most of transformations on the schematic level. On the semantic level, WSML based on logic programming would be implemented as an alternative to OWL based on description logic in Protégé.

There are opportunities for future work related to our use case. Those include schema transformations of feature components by FME to represent geographic elements in the GML model. We transformed source and target attributes and their values, however, they are only part of the components which construct the GML model. The OGC WFS specification requires GML as a standard format to exchange geospatial data. Thus, transformation of the GML model from CMA schema to EFFIS schema by the FME server can be tested for publishing and downloading via OGC WFS.

On the semantic level, the new domain ontology can be created to redefine 'forest' as common ground according to the level of abstraction. This may involve the introduction of a foundational ontology such as DOLCE to improve quality and efficiency of the methodology. We can also survey applications that include all the member states of EU and compare them with EFDAC and FAO Forestry in the context of forest cover. To find optimal common ground it may require top-down and bottom-up approaches in the ontology architecture between foundational and domain ontology levels as well as domain and application ontology levels. The level of abstraction can be explored for semantic matching by adjusting the range of a shared concept 'forest' to be more flexible or restrictive. We may introduce more applications outside EU to explore the level of abstraction for redefining 'forest' to a global scale.

Additionally, we may investigate how schema mapping rules are executed in the complete process of schema translation from source to target data so that the regional data can be inputted into the forest fire model developed by EFFIS. It would be practical further research to test the mapping rules generated by FME. The FME is one means of generating mapping rules programmed by the software specification, which is not standardized. Thus, it may require another rule language to reuse and exchange those mapping rules that can be processed by other execution tools. We may also investigate how ontologies saved as RDF or OWL format in Protégé can be used as mapping rule language.

REFERENCES

Abadie, N. 2009. Schema Matching Based on Attribute Values and Background Ontology. *Proceeding of 12th AGILE International Conference on Geographic Information Science*, Hannover, Germany.

Antoniou, G. and Van Harmelen, F. 2008. *A semantic Web primer*. MIT Press.

Beckmann, O., Blut, C., Deelmann, T., Michels, H., Osmanov, A., Roth, M., Weerasinghe, W.M.T.H., Wilden, M.,Schade S. *Getting inspired?* Geoinformatik 2009, Osnabrueck, Germany.

Berners-Lee, T., Shadbolt, N., and Hall, W. 2006. The Semantic Web Revisited. *IEEE Intelligent Systems*, May/June 2006. IEEE Computer Society.

Bishr, Y. 1998. Overcoming the Semantic and Other Barriers to GIS Interoperability. *International Journal of Geographical Information Science*, 12 (4), 299–314.

Boschetti, L., Roy, D., Barbosa, P., Boca, R., Justice, C. 2008. A MODIS assessment of the summer 2007 extent burned in Greece. *International Journal of Remote Sensing*, 29 (8), 2433 – 2436.

Breitman, K., Casanova, M.A., Truszkowski, W. 2007. *Semantic Web: Concepts, Technologies and Applications*. NASA Monographs in Systems and Software Engineering.

Chunyuan, C., S. Schade and T. Gudiyangada. 2010 (forthcoming). Schema Mapping in INSPIRE - Extensible Components for Translating Geospatial Data. *Proceeding of 13th AGILE International Conference on Geographic Information Science*, Guimarães, Protugal.

CMA (Conselleria de Medio Ambiente, Agua, Urbanismo y Vivienda). 2007. ISO19115 Metadata: Incendios 2007. Generalitat Valenciana.

Díaz,L., Granell, C., Gould, M. 2009. Spatial Data Integration over the Web. In *Handbook of Research on Innovations in Database Technologies and Applications: Current and Future Trends,* edited by Ferraggine, V. E., Doorn, J.H., and Rivero, L.C. Information Science Reference, Hershey.

Donaubauer, A.,Straub, F., and Schilcher, M. 2007. mdWFS: A Concept of Web-enabling Semantic Transformation. *Proceeding of 10th AGILE International Conference on Geographic Information Science*, Aalborg, Denmark.

Donaubauer, A., Fichtinger, A., Schilcher, M., and Straub, F. 2006. Model Driven Approach for Accessing Distributed Spatial Data Using Web Services - Demonstrated for Cross-Border GIS Applications. *Proceeding of the 23rd International FIG Congress*, Munich, Germany.

Duchesne, P., Maué, P., and Schade, S. 2008. *Semantic annotations in OGC standards*. OGC Discussion Paper.

EuroGEOSS. 2009. *D.3.1: report on user requirements for the EuroGEOSS Forestry operating Capacity*. URL: http://www.eurogeoss.eu/Documents/EuroGEOSS_D3-1.pdf (Retrieved on January 13th 2010).

ESRI. 2009. *ArcGIS Desktop Documentation*. Redlands, California. ESRI Press.

FAO (Food and Agriculture Organization of the United Nations). 2004. *Global Forest Resource Assessment Update 2005: Specification of National Reporting Tables for FRA 2005 (Terms and Definitions).* Working Paper 81, Forestry Department. Rome, Italy.

FAO (Food and Agriculture Organization of the United Nations). 2009. *Global Forest Resource Assessment.* URL: http://www.fao.org/forestry/fra/en (Retrieved on October 9th 2009).

European Commission. 2003. *Action 22003: Monitoring the forests in Europe (FOREST)*. Land Management and Natural Hazards Unit. URL: http://ies.jrc.ec.europa.eu/index.php?page=action-22003 (Retrieved on November 7th 2009).

Friis-Christensen, A., Schade, S., and Peedell, S. 2005. Approaches to solve schema heterogeneity at European Level. *Proceeding 11th EC-GI & GIS Workshop,* ESDI: Setting the Framework, Alghero, Sardinia, Italy.

Friis-Christensen, A., Luts, M., Ostländer, N., and Bernard, L. 2007. Designing Service Architectures for Distributed Geoprocessing: Challenges and Future Directions. *Transactions in GIS*, 11(6), 799–818.

Goodchild, M., Egenhofer, M., Fegeas, R., & Kottman, C. 1999. *Interoperating geographic information systems*. Norwell, Kluwer Academic Publishers.

Gruber, A., Westenthaler, R., and Gahleitner, E. 2006. Supporting domain experts in creating formal knowledge models (ontologies). *Proceedings of I-KNOW'06. 6th International Conference on knowledge management*, Graz, Austria.

Guarino, N. 1998. Formal ontology and information systems. *Proceedings of 1st International Conference on Formal Ontologies in Information Systems (FOIS)*, Trento, Italy.

HUMBOLDT 2008. *Concept of Application-Specific Harmonised Data Models Humboldt*: *Deliverable A7.1-D1*. HUMBOLDT Community Project. URL: http://www.esdi-humboldt.eu/.../818-a7_1d1_concept_of_application-specific_harmonised-tud-001-final.pdf (Retrieved on October 17th 2009).

HUMBOLDT. 2009. HUMBOLDT Data Harmonisation Tools in Action. *Workshop 11th GSDI World Conference on Spatial Data Infrastructure Convergence*, Rotterdam, Netherland.

INSPIRE. 2003. *Consultation Paper on a forthcoming EU Legal Initiative on Spatial Information for Community Policy-making and Implementation*. URL: http://www.ec-gis.org/docs/F6017/INSPIRE-INTERNETCONSULTATIONPHASEII.PDF (Retrieved on October 7th 2009).

INSPIRE. 2004. *INSPIRE scoping paper*. Joint Research Centre, European Commission. URL: http://www.ecgis.org/inspire/reports/inspire_scoping24mar04.pdf (Retrieved on October 7th 2009).

INSPIRE. 2007. *Directive 2007/2/EC of the European Parliament and of the Council of 14 March 2007 establishing an Infrastructure for Spatial Information in the European Community*. Joint Research Centre, European Commission. URL: http://eur-lex.europa.eu /LexUriServ/LexUriServ.do?uri= OJ:L:2007:108:0001:0014:en:PDF (Retrieved on November 3th 2009).

INSPIRE. 2008. *Technical Report: SOAP HTTP Binding Status - Survey on OGC and ORCHESTRA Specifications Relevant for the INSPIRE Network Services*. Joint Research Centre, European Commission. URL: inspire.jrc.ec.europa.eu/ reports/.../SOAP_binding_survey.pdf (Retrieved on November 17th 2009).

INSPIRE. 2009a. *INSPIRE Data Specification on Protected Sites* (Version 3). INSPIRE Thematic Working Group.

INSPIRE. 2009b. *INSPIRE Data Specification on Transport Networks* (Version 3). INSPIRE Thematic Working Group.

JRC (Joint Research Centre). 2008. *ENFIN Cost Action E – 43*. Carcavelos, Portugal.

JRC (Joint Research Centre). 2009. *UNECE/FAO ToS Forest Fire Meeting* Report. Geneva.

Klausen, F.M. 2006. *GeoXSLT: GML processing with XSLT and spatial extensions.* Department of Informatics, University of Oslo.

Klien, E. and Probst, F. 2005. Requirements for Geospatial Ontology Engineering. *Proceeding of 8th AGILE International Conference on Geographic Information Science*, Estoril, Portugal.

Knublauch, H., Fergerson, R.W., Noy, N.F., and Musen, M.A. 2004. *The Protégé OWL Plugin: An Open Development Environment for Semantic Web Applications.*

Lehto, L. 2007. Schema translations in a Web service based SDI. *Proceeding of 10th AGILE International Conference on Geographic Information Science*, The European Information Society: Leading the way with geo-information, Aalborg, Denmark.

Lehto, L. and Sarjakoski, T. 2004. Schema Translations by XSLT for GML-Encoded Geospatial Data in Heterogeneous Web-Service Environment. *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, 35 (4), 177-182.

Iliffe,J. and Lott, R. 2008. *Datums and Map Projections For Remote Sensing, GIS and Surveying* (2nd ed.). Technology & Engineering.

Madhavan, J., Bernstein, P.A., and Rahm, E. 2001. Generic Schema Matching with Cupid. *Proceedings of the 27th International Conferences on Very Large Databases*.

Masser, I. 2005. *GIS Worlds: Creating Spatial Data Infrastructures*. Redlands, California. ESRI Press.

Masolo, C., Borgo, S., Gangemi, A., Guarino, N., and Oltramari, A. 2003. *D18 - Ontology Library* (final version). Deliverable of the WonderWeb Project.

MMA (Ministerio de Medio Ambiente y Medio Rural y Marino). 2009a. Biodiversity: Legislation and Agreements. URL: http://www.mma.es/portal/secciones/ biodiversidad/legislacion_convenios (Retrieved on September 27th 2009).

MMA (Ministerio de Medio Ambiente y Medio Rural y Marino). 2009b. Biodiversity: National Forest Inventory. *Inventario Forestal Nacional. Descripción de los Códigos de la Base de Datos de Campo*. URL: http://www.mma.es/portal/ secciones/biodiversidad/inventarios/ifn/ index.htm(Retrieved on November 23th 2009).

MMA (Ministerio de Medio Ambiente y Medio Rural y Marino). 2009c. Biodiversity: Protection Against Forest Fires. *Incendios Forestales en España 2009: Avance Informativo*. URL: http://www.mma.es/portal/secciones/ biodiversidad /defensa_incendios (Retrieved on October 17th 2009).

Nebert, D.D. 2004. *Developing Spatial Data Infrastructures: The SDI Cookbook*. GSDI-Technical Working Group.

Nunes de Lima, M.V. 2005. *CORINE Land Cover updating for the year 2000: IMAGE2000 and CLC2000 Products and Methods*. European Commission.

OGC (Open Geospatial Consortium). 2007. Geography Markup Language (GML) Encoding Standard - Version 3.2.1. The Open Geospatial Consortium.

OGC (Open Geospatial Consortium). 2009. Open GIS Consortium. Standards. The Open Geospatial Consortium.

Percival, G. 2003. *OGC Reference Model*. OGC 03-040. URL: https://portal.opengeospatial.org/files/?artifact_id=3836 (Retrieved on December 15th 2009).

Phillips, A., Williamson, I., Ezigbalike, C. 1999. Spatial Data Infrastructure Concepts. *Australian Surveyor*, 44 (1), 20-28.

Horridge, M. 2009. *A Practical Guide To Building OWL Ontologies Using Protégé 4 and CO-ODE Tools Edition 1.2*. The University Of Manchester.

Rahm, E. and Bernstein, P. 2001. A survey of Approaches to Automatic Schema Matching. *Very Large Data Bases Journal*, 10 (4), 334–350.

Safe Software. 2009. *FME® Workbench transformers Quick-reference booklet* [Software].

Schade, S. 2009. *Ontology-Driven Translation of Geospatial Data*. AKA, Heidelberg, Germany.

SIOSE (Sistema de Información de Ocupación del Suelo en España). 2007. *Anexo I: Guía de Comprobación en Campo*. Ministerio de Fomento.

SIOSE (Sistema de Información de Ocupación del Suelo en España). 2009. *Segmentación Territorial Basada en el Proyecto SIOSE en la Comunidad Valenciana*. Conselleria de Medi Ambient, Aigua, Urbanisme i Habitatge.

Stuckenschmidt, H. 2003. *Ontology-Based Information Sharing in Weakly Structured Environments*. PhD dissertation, Vrije Universiteit Amsterdam, Amsterdam, Netherlands.

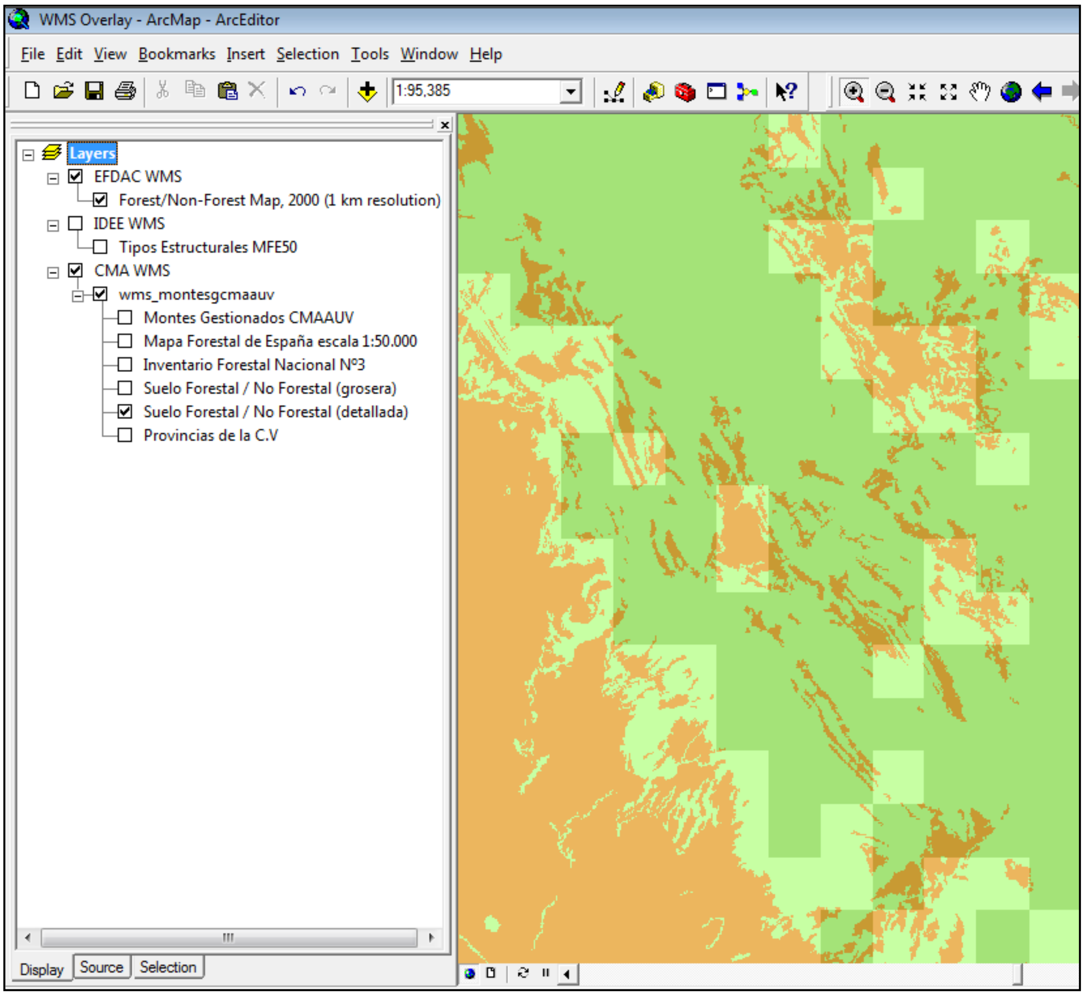
Sumrada, R. 2003. Temporal Data and Temporal Reference Systems. *Multi-Dimensional Approaches and New Concepts in SIM.*

Vaccari, L., Shvaiko, P., and Marchese, M. 2009. A geo-service semantic integration in Spatial Data Infrastructures. *International Journal of Spatial Data Infrastructures Research*, 4, 24-51.

Visser, U., Stuckenschmidt, H., Wache, H., and Voegele, T. 2001. Using environmental information effciently: Sharing data and information from heterogeneous sources. In *Environmental Information Systems in Industry and Administration*, edited by Rautenstrauch, C., 41-73. Idea Group.

W3C. 2006. *XML Schema Datatypes in RDF and OWL*. World Wide Web Consortium.

ANNEXES



Annex I. GIS overlay of images in regional and European scales via OGC WMS in ArcMap. CMA displays forest and non-forest areas in polygon while EFDAC displays forest and non-forest areas in raster.

```
#! <TRANSFORMERS>
#! <TRANSFORMER
#! IDENTIFIER="13"
#! TYPE="ExpressionEvaluator"
#! VERSION="1"
#! POSITION="1121 -252"
#! ORDER="23"
#! PARMS_EDITED="true"
#! ENABLED="true"
#! >
#! <OUTPUT_FEAT NAME="OUTPUT"/>
#! <XFORM_ATTR ATTR_NAME="Conifer" IS_USER_CREATED="false"
FEAT_INDEX="0"/>
#! <XFORM_PARM PARM_NAME="EXPRESSION" PARM_VALUE="(@Value
(SUP_ARBOLA)/@Value(SUP_TOTAL))*100"/>
#! <XFORM_PARM PARM_NAME="XFORMER_NAME" PARM_VALUE="ConvertValue"/>
#! <XFORM_PARM PARM_NAME="VAL_ATTR" PARM_VALUE="Conifer"/>
#! </TRANSFORMER>
#! <TRANSFORMER
#! IDENTIFIER="34"
#! TYPE="ExpressionEvaluator"
#! VERSION="1"
#! POSITION="1138 -615"
#! ORDER="24"
#! PARMS_EDITED="true"
#! ENABLED="true"
#! >
#! <OUTPUT_FEAT NAME="OUTPUT"/>
#! <XFORM_ATTR ATTR_NAME="Transit" IS_USER_CREATED="false"
FEAT_INDEX="0"/>
#! <XFORM_PARM PARM_NAME="EXPRESSION" PARM_VALUE="(@Value
(SUP_NARBOL)/@Value (SUP_TOTAL))*100"/>
#! <XFORM_PARM PARM_NAME="XFORMER_NAME" PARM_VALUE="ConvertValue_2"/>
#! <XFORM_PARM PARM_NAME="VAL_ATTR" PARM_VALUE="Transit"/>
#! </TRANSFORMER>
#! <TRANSFORMER
#! IDENTIFIER="37"
#! TYPE="ExpressionEvaluator"
#! VERSION="1"
#! POSITION="1150 -907"
#! ORDER="27"
#! PARMS_EDITED="true"
#! ENABLED="true"
#! >
#! <OUTPUT_FEAT NAME="OUTPUT"/>
#! <XFORM_ATTR ATTR_NAME="Area_HA" IS_USER_CREATED="false"
FEAT_INDEX="0"/>
#! <XFORM_PARM PARM_NAME="EXPRESSION" PARM_VALUE="(@Value (Conifer)
+@Value (Transit))"/>
#! <XFORM_PARM PARM_NAME="XFORMER_NAME" PARM_VALUE=" ConvertValue_3"/>
#! <XFORM_PARM PARM_NAME="VAL_ATTR" PARM_VALUE="Area_HA"/>
#! </TRANSFORMER>
#! </TRANSFORMERS>
```

Annex II. Example of schema mapping rules for converting attribute values from source data model to target data model using FME transformers.