International Journal of Spatial Data Infrastructures Research, 2009, Vol. 4, 24-51.

A geo-service semantic integration in Spatial Data Infrastructures*

Lorenzino Vaccari[1,2], Pavel Shvaiko[3] and Maurizio Marchese[2]

[2]Dipartimento di Ingegneria e Scienza dell'Informazione,
University of Trento, Italy
email: vaccari/marchese@disi.unitn.it
[3]TasLab, Informatica Trentina S.p.A., Italy
email: pavel.shvaiko@infotn.it

Abstract

In this paper we focus on the semantic heterogeneity problem as one of the main challenges in current Spatial Data Infrastructures (SDIs). We first report on the state of the art in reducing such a heterogeneity in SDIs. We then consider a particular geo-service integration scenario. We discuss an approach of how to semantically coordinate geographic services, which is based on a view of the semantics of web service coordination, implemented by using the Lightweight Coordination Calculus (LCC) language. In this approach, service providers share explicit knowledge of the interactions in which their services are engaged and these models of interaction are used operationally as the anchor for describing the semantics of the interaction. We achieve web service discovery and integration by using semantic matching between particular interactions and web service descriptions. For this purpose we introduce a specific solution, called structure preserving semantic matching. We present a real world application scenario to illustrate how semantic integration of geo web services can be performed by using this approach. Finally, we provide a preliminary evaluation of the solution discussed.

Keywords: Spatial Data Infrastructure, semantic heterogeneity, GIS web services, ontology matching, real world application with GIS web services.

[1] Also affiliated to: Autonomous Province of Trento.

DOI: 10.2902/1725-0463.2009.04.art2

1 INTRODUCTION

The domain of Geographic Information (GI)[2] is experiencing a rapid growth of both computational power and quantity of information, making large spatial data archives available over the Internet. A visionary concept of the integration of geo-information was posed on 1998 by the U.S. vice president Al Gore (Gore, 1998). His "Digital Earth" label became popular for describing a virtual representation of the Earth on the Internet that is spatially referenced and interconnected with the world's digital knowledge archives. In order to put the Al Gore' visions into practice, some preliminary initiatives were born. These initiatives included both the attention to consensus development of standard protocols, tools (e.g., the *Digital Earth Initiative* (IDEW, 2001)) and the development of new technologies (e.g., the *International Digital Earth* SRI project (Leclerc et al., 2001)).

After ten years, the Al Gore's milestone vision is partially implemented by recently available geo-browsers (like Google Earth, Microsoft Virtual Earth, NASA Worldwind and ESRI ArcGIS Explorer) and web applications (like Google Maps, Microsoft Live Search Maps and Yahoo Local Maps). These systems have introduced Geographic Information System (GIS) services to ordinary Internet users, offering them high-resolution aerial imagery with responsive performance (Craglia et al., 2008, Tu and Abdelguerfi, 2006). Moreover, there is an increasing necessity to share this information between different stakeholders (e.g., departments in public administration, professionals, citizens, and GIS expert users) and diverse information systems in order to enable a coherent and contextual use of GI.

This necessity forms the basis for a number of international and national initiatives, to set up global, international, national and regional infrastructures for the collection and dissemination of geographical data, including among others:

- **INSPIRE:** the INfrastructure for SPatial InfoRmation in Europe (INSPIRE) is an European Directive approved in 2007, based on the goal to improve the accessibility, interoperability and affordability of spatial data and information systems in Europe, specifically: "*Experience in the Member States has shown that it is important, for the successful implementation of an infrastructure for spatial information, that a minimum number of services be made available to the public free of charge. Member States should therefore make available, as a minimum and free of charge, the services for discovering and, subject to certain specific conditions, viewing spatial datasets.*" (INSPIRE, 2007).
- **SEIS:** the Shared Environmental Information System (SEIS) is a collaborative

[2] In this paper, we will use the term *geographic information* to group different kinds of geographic objects: geographic services or geo-services, geographic metadata or geo-metadata and geographic data or geo-data.

initiative of the European Commission, the European Environmental Agency (EEA) and the member countries of the Agency. It aims to improve the availability and quality of information needed to design and implement the European Union's environment policy (SEIS, 2008).

- **GEOSS:** the Global Earth Observation System of Systems (GEOSS) is being built by the Group on Earth Observations (GEO). GEOSS seeks to connect the producers of environmental data and decision-support tools with the end users of these products, with the aim of enhancing the relevance of Earth observations to global issues. The ultimate result is to provide a global public infrastructure that generates comprehensive, near-real-time environmental data and information for a wide range of users (GEOSS, 2008).
- **GMES** (Global Monitoring for Environment and Security) is a joint initiative of the European Commission and European Space Agency, adopted by EU Heads of State at the Gothenburg Summit in 2001, and aimed at achieving an autonomous and operational capability in the exploitation of geo-spatial information services by 2008. GMES is the European Union contribution to GEOSS (GMES, 2008).

Moreover, a growing number of public institutions and private companies (GIS agencies) have adopted a GIS to handle their internal geographical information. A number of commercial and open source software packages are available to support such local activities (e.g., ESRI ArcGIS, Intergraph MapInfo, AutoCAD Map 3D, Quantum GIS and GRASS). These products give a complete and powerful set of functionalities to manage geographical information for every GIS agency. Beside this management challenge, the growing number of geographic information providers, the large quantity of the produced GIS data, the availability of high speed networks and sophisticated computer science technologies are creating a heterogeneous set of producers and final users. Typical user roles include:

- Public institutions that require geographic information to support institutional duties (e.g., emergency, health, urban planning, and tourism).
- International, national and regional institutions that coordinate and integrate geographic information provided by different GIS agencies.
- Research institutions that want to analyze the availability and the quality level of geographic information covering a specific study area.
- Private companies that need geographic information in order to create business services and products (geo-marketing).
- Non expert users that need to locate quickly and easily a geographical feature (e.g., address, location name, institution, and business activity).

To support all these kinds of users, and user requests, GIS agencies have started to adopt a Spatial Data Infrastructure model (Groot and McLaughlin, 2000; Nebert D., 2004; Bernard et al., 2005; Masser, 2005). While a GIS is a selfcontained system in which data and software applications are used mainly internally, the SDI goal is to support the interoperability among different kinds of providers and users. The number of SDIs is growing and one of the main challenges is to achieve international cooperation and collaboration in order to support regional, national and international SDI developments. It should allow nations to better address social, economic, and environmental issues. This is also the main goal of the Global Spatial Data Infrastructure (GSDI) association, an inclusive organization of organizations, agencies, firms, and individuals from around the world (GSDI, 2008).

1.1 Interoperability among SDIs

Interoperability issues pervade SDIs, since they can connect heterogeneous organizations and systems. Research in information system interoperability is motivated by the increasing heterogeneity of the computer systems. Heterogeneity in GIS field is not an exception and may be classified as syntactic (differences in data format), structural (differences in schemas) and semantic (differences in intended meaning of terms in specific contexts) heterogeneity (Stuckenschmidt, 2003).

Semantic heterogeneity of GIS can be undertaken by considering their ontological aspects. Heterogeneity of GIS ontologies has been addressed in many works during the last decades, see, e.g., (Nyerges, T. L.,1989) and (Worboys and Duckham, 2005). An ontology is a logical theory accounting for the intended meaning of a formal vocabulary (i.e., its ontological commitment to a particular concept of the world) (Guarino, 1998). However, different geographical data providers may use different application ontologies, so, heterogeneity problems arise when integrating the information from different application ontologies. An ontology typically provides a vocabulary that describes a domain of interest and a specification of the meaning of terms used in the vocabulary. Depending on the precision of this specification, the notion of ontology includes various data and conceptual models (Euzenat and Shvaiko, 2007). The term ontology is used here in a wide sense, and, hence, encompasses, e.g., sets of terms, classifications, database schemas, and thesauri.

Ontology matching is a plausible solution to the semantic heterogeneity problem faced by information management systems. We can also say that information systems adopt interfaces to exchange information and that *the interfaces between agents, computational and human, are those of web services*. Moreover, *the interface of a service is formally captured by its signature* (Kuhn, 2005). If we consider signatures (name, inputs and outputs) of web services to be graph-like structures, we can use some of-the-shelf ontology matching systems to facilitate the resolution of the semantic heterogeneity problem. In fact, ontology matching is often proposed as a solution to this heterogeneity problem in many applications, including integration of web services.

Since ontologies can be viewed as graph-like structures (Giunchiglia and Shvaiko, 2003), ontology matching can be considered as an operation that takes two graph-like structures, such as web service signatures or descriptions, and produces a set of correspondences between the nodes of the graphs that correspond semantically to each other (Giunchiglia et al., 2007). Then, these correspondences can be used for various tasks, including service discovery, composition and coordination, information retrieval operations, data schema mediation and translation. Thus, matching ontologies enables the knowledge and data expressed in the matched ontologies to interoperate.

Many solutions to the ontology matching problem have been proposed so far (Euzenat and Shvaiko, 2007). In this paper we present and adopt a particular type of ontology matching, namely structure preserving semantic matching (SPSM) (Giunchiglia et al., 2008).

1.2 Contributions and organization of the paper

In this paper we first focus on the open issues of interoperability present in state-of-the-art SDIs. Then, we apply a specific approach to semantically discover and integrate heterogeneous geographic services. This paper is an expanded and updated version of an earlier conference paper (Marchese et al., 2008).The key extensions of the present work include:

- A discussion on the requirements and open issues of interoperability in distributed geographic data and services and an overview of the state of the art in the area of semantic GI integration.
- A detailed description of a geo-service semantic integration use case along with its implementation and preliminary evaluation.

The structure of the paper is as follows. We present SDI geo-data and geo-services interoperability issues in Section 2. We discuss related work in Section 3. In Section 4 we present a real world use case, as well as its formalization in order to support integration of web services shared by different providers. In Section 5 we introduce the structure preserving semantic matching approach, used to support ontology matching between different service providers. In Section 6 we provide a preliminary evaluation of the solution discussed. Finally, the major findings of the paper as well as future work are presented in Section 7.

2 SDI INTEROPERABILITY ISSUES

SDI, like other information technologies, must be implemented in a manner that allows easy interoperability between heterogeneous organizations and systems. The common SDI framework is based on a generic software platform, which supports a variety of geographic dataset types as well as comprehensive tools for data management, editing, analysis, and visualization. Moreover, to share geo-datasets, a number of geographic services have to be provided by the system. Interoperability issues pervade both geographic data and geographic services. In this work we use the following definition for interoperability: *interoperability enables the integration of data between organizations and across applications and industries, resulting in the generation and sharing of more useful information*[3].

In this section, in order to better present and discuss these issues, a separate analysis for geographic data (geo-data, § 2.1) and geographic services (geo-services, § 2.2) is presented. An extended definition and analysis of interoperability in the field of GI is also given in (Kuhn, 2005).

2.1 Geo-data interoperability issues

One of the key services supplied by an SDI is the possibility to retrieve geographical datasets provided by heterogeneous resources. Due to the fact that the logical architecture of an SDI can be based on a set of different data resources, heterogeneous geographical information has to be integrated. Since each geo-data producer adopts internal rules in order to manage its geographical datasets, heterogeneity at the data level arises from a number of different reasons (Friis-Christensen et al., 2005):

- **Different syntax**: geo-datasets are retrieved from different sources that can use different data formats (e.g., ESRI shape files, Mapinfo Files, Oracle geo-DB, PostGIS geoDB, and GRASS files).
- **Different structure**: geographical features can be represented using different geometrical and data schemas. Often the same geographic feature is represented using different geometric features (for instance, roads can be represented using polygons or lines) or multi-temporal techniques (Parent et al., 2006).
- **Different semantics**: interoperability problems due to different semantics are caused by different reasons. *Naming conflicts* occur when classes or attribute types with different semantics are given the same names (homonyms) or when classes or attribute types that are semantically the same are named differently (synonyms).

Moreover, geographical datasets have specific properties, different from other types of data, including (Lemmens et al., 2006):

- **Implicit linking**: in general, explicit relationships must be present to combine information in a meaningful manner. However, geographic information enables linking without explicit references, for instance, via coordinate

---

[3] http://www.esri.com/library/whitepapers/pdfs/spatial-data-standards.pdf [accessed 15 December 2008]

reference systems. For instance, a bridge can be implicitly linked to a river or to a road it crosses.

- **Massive datasets**: compared to general (administrative) information, geo-information can be massive. In case of satellite imagery, for instance, raster data volumes can be huge. For example, satellite images of the COSMO-Skymed mission will provide, in full constellation configuration, an acquisition capacity brought up to 560 GB, roughly corresponding to 1800 standard images, per day[4].
- **Multiple versions**: multiple versions of the same entities on the Earth's surface can differ radically in terms of data model, scale, data generalization, conceptual model, and semantics the data collectors use. The main reason is that data collectors are often represented by different government agencies at different levels (e.g., regional, national and international) in different countries. Specifically, for the case of geo-data integration, we have also scale conflicts and different precisions/resolutions issues. *Scale conflicts* occur when attributes have different units or are represented in varying scales of measures (e.g., in 1999 NASA lost a $125 million Mars orbiter because a Lockheed Martin engineering team used English units of measurement while the agency's team used the more conventional metric system for a key spacecraft operation[5]). *Different precisions/resolutions* occur if requirements for geo-datasets acquisition are different even if they are referred to the same geographical feature. Additional factors have to be also considered like *integration alignment* problem (e.g., data collected at different scales, data corrected using different elevation models, and data produced using different topographic sources).

2.2 Geo-service interoperability issues

Distributed service discovery, composition and coordination are the main research topics in the field of web services. General issues dealing with service integration include:

- **Geo-service discovery**: GIS desktop applications provide to the user a lot of complex functions in order to perform GIS data acquisition, creation, analysis, visualization and mapping. For years these functions were accessible only through the GIS desktop applications, but recently, GIS services have

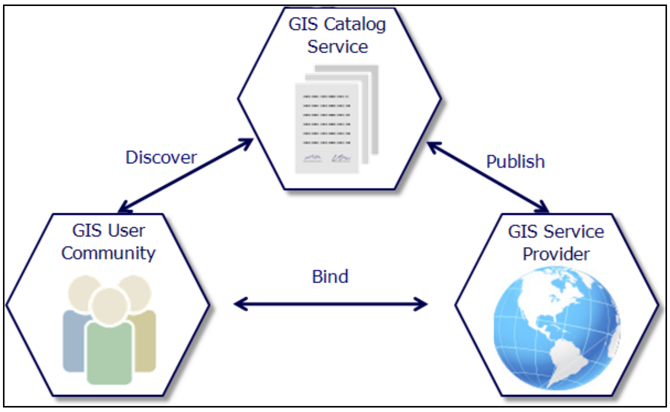
---

[4] http://www1.alcatel-lucent.com/com/en/apphtml/atrarticle/2006q2cosmoskymedtheitaliancontributiontothefrenchitalianintergovernmentalagreementonearthobservationcosmohtmltcm172914881635.jhtml;jsessionid=WW3GKDXA1PEXPLAWFRSHJIFMCYWGQTNS?_DARGS=/common/atr/DATR_table_of_contents.jhtml_A&_DAV=/com/en/appxml/articlepaperlibrary/cosmoskymedtheitaliancontributiontothefrenchitalianintergovernmentalagreementonearthobservationtcm172899121635.jhtml [accessed 23 November 2008]

[5] http://www.cnn.com/TECH/space/9909/30/mars.metric.02/ [accessed 15 December 2008]

became published and available on the web. Distributed Service Oriented Architecture (SOA) is a common framework for modern distributed information systems (Papazoglou, 2003). SOA and Open Geospatial Consortium (OGC)[6] specifications are the base technology used by an SDI in order to provide catalogue services for discovering appropriate data and services for a specific task. Figure 1 shows the three main building blocks in GIS SOA: *(i)* a GIS user community (potential users of GIS services), *(ii)* GIS web services (published by some GIS service providers) and *(iii)* a GIS catalogue service (where available services are published by providers and discovered by users).

Figure 1: GIS Service Oriented Architecture



- **Geo-service integration** (composition or coordination): after discovering, services can be composed or coordinated to provide complex functionalities. Although at present, the main available web service in GIS is the map request service, the trend is to supply a technological environment that provides a number of stand-alone GIS services. At the moment, the majority of these geo-services exist as single services. In the case of a request for a complex service a manual and static composition of a number of predefined geo-services has to be performed. The future challenge is the (semi)-automatic composition of arbitrary services in order to obtain flexible complex services based on the available web services. In practice, however, chaining geographic services is a nontrivial task, mostly because geographic data have varied differences from other types of data (§ 2.1) and also because individual web service providers use different syntactic structure just as they use different vocabulary to define web service signatures and descriptions.

Moreover, when integrating geo-services from heterogeneous sources, some

[6] http://www.opengeospatial.org [accessed 23 November 2008]

specific issues have to be taken into account (Lemmens et al., 2006):

- **Maps as implicit interfaces**: everyone is familiar with reading maps, so they are a natural human–machine interface for the services interacting with the user and presenting (intermediate) results of geo-information.
- **Geometry based information**: since geo-information is geometry based, geo-service interface has always to take into account the geometric component of the data they provide and process (e.g., the bounding box of a map and the coordinate reference system of geographical layers).
- **Specific topological operations**: it is also possible to apply a whole set of common mathematical tools in geo-services to compute their topological relationships, e.g., to compute the distance between two objects, the buffer around an object, the intersection between different features, and the neighbors of a polygon.

From a technological point of view, SOA for business services and OGC specifications for geographic information represent the reference framework when discovering and integrating available geo-services. However, as in the case of geo-data, also geo-services are defined using implicit or, in the best case, local semantics. At present, no standard notions are used for defining the semantics of a geographic web service. This problem is typically referred to as the need for *semantic interoperability among autonomous and heterogeneous systems*. Semantic heterogeneity (the differences in meaning) problem is an actual challenge for geographic services integration. Currently, geo-information search is performed using mainly string-based techniques.

3 RELATED WORK

The research area of semantic integration of geographic information is relatively young. In fact, even if the concept of geo-service publication is not new, OGC specifications and ISO standards became stable only during the last years. Thereafter, different geo-information providers have started to publish their geo-data and services on the web in a standardized manner. Only recently, the integration of GI became relevant and feasible because of the availability of GIS web services.

In this section, following the structure of the previous one, we divide the presentation and analysis of the related work into two parts: geo-data interoperability (§ 3.1) and geo-service interoperability (§ 3.2).

3.1 Geo-data interoperability

Integrating data from heterogeneous sources is the fundamental task in order to enable value added services. Such a task is complex, especially if the goal is the integration of different geographic datasets. As described in the previous section, geo-data heterogeneity depends on many aspects: acquisition quality aspects (e.g., production process), technical aspects (e.g., different GIS software package) and semantic aspects (related to the diversity of application requirements addressed by different data providers). In the overall task, it is possible to identify two main issues: (*i*) *geo-data integration alignment* and (*ii*) *geo-data semantic heterogeneity*.

The first problem (*geo-data integration alignment*) depends on a number of factors including: different geographic projections, data collected at different scales, corrected using different elevation models, and data production using different topographic sources. Such problems have been clearly identified and addressed by current research. For example, the work in (Chen et al., 2003) proposes a general-purpose geospatial data integration framework to access and retrieve geospatial sources, to accurately and efficiently integrate these sources using dynamically conflation operations in the integration plans, and quickly incorporate new sources that support geo-data standards. Although this issue is an important and complex aspect, we focus, instead, on the second issue: *geo-data semantic heterogeneity*.

In Fonseca et al. (2002) ontologies are used to reduce GI heterogeneity. This work proposed a detailed description of the role of ontologies in geographic data modeling and a solution called ontology-driven geographic information system (ODGIS) that acts as a system integrator. In ODGIS, an ontology is a component, such as a database, cooperating to fulfill the system's objectives. The work suggests an architecture for ODGIS which includes an ontology editor and its embedded translator plus a user interface to browse ontologies.

In the GEOscience Network (GEON) [7] project an interoperability framework has been developed to allow a data provider to register a geographic dataset with one or more mediation ontologies (e.g., standards for data structure and content) and subsequently query the different datasets in a uniform fashion (Nambiar et al., 2006). The system comprises an ontology repository, a dataset registration procedure, and a query rewriting system. Structural and semantic heterogeneities of data sources are resolved using information from the dataset registration procedure and ontologies. Multiple ontologies are supported in the system by allowing users to manually define an articulation between two ontologies which equates some concepts in the source ontology to some concepts in the target ontology. Users are able to switch between ontologies for which an articulation exists. Nevertheless, this system can be adopted only in the case when the user adheres to the community (using the GEON registration procedure).

A specific methodology for geo-ontologies integration was proposed in (Hess et al., 2006), where G-Match, an algorithm and an implementation of a geographic

[7] http://www.geongrid.org [accessed 23 November 2008]

ontology matcher, was presented. The goal is to give a similarity measure between two different geographic ontologies when integrating them. In order to do that, the algorithm considers the features of a concept separately and then gives some weights for each geographical feature (name, attributes, taxonomy, conventional and topological relationships) to compute the overall similarity between two concepts. As the information may be defined in different levels of detail, there is no perfect combination of the weight factors assigned to each concept features. So, some sort of self-adaptation of the weight, depending on the input ontology, has to be performed.

The main focus in Paul and Ghosh (2006) was to integrate diverse spatial repositories for geographic applications using SOA for the discovery and retrieval of geospatial information. The architecture uses a central ontology as metadata information, which acts as service broker. Also here, the system is composed of a domain ontology (a global shared vocabulary) and of the service providers application ontologies that need to adopt the central ontology.

In the Semantic Web-Service Interoperability for Geospatial Decision Making (SWING)[8] project, the issue of GI semantic integration has also been tackled. Below, we mention several works from SWING on geo-data integration:

- The work in Lutz and Klien (2006) presented an ontology based approach to GI retrieval that contributes to the solution of existing problems of semantic heterogeneity and hides most of the complexity of the required procedure from the requestor. Nevertheless, in the proposed approach, it is assumed that a requestor searches for only one source that provides all the required information. Moreover, the data provider has to create and register an application ontology that represents one of the bottlenecks for scalability.
- The problem of generating semantic annotation of geo-data was tackled in Klien (2007). In this work, semantic annotation is understood as making explicit the relationship between a data schema and a domain ontology by defining mappings from elements of the schema to elements in the ontology. Specifically, a strategy for partially automating this process is introduced. It transforms a data schema into an ontology and applies spatial analysis methods during the matching process for exploiting extensional knowledge.
- A similarity-based information retrieval system has recently been introduced by Janowicz et al. (2008). This work proposes an architecture, based on the SIM-DL similarity theory (Janowicz et al., 2007), to support users and systems during information retrieval operations. Use cases for a human web interface, as well as for an SDI system integration workflow and analysis are provided. The proposed architecture includes standard services, such as Web Mapping Service (WMS)[9] and Web Feature Service (WFS)[10] instances,

[8] http://www.swing-project.org/ [accessed 20 November 2008]
[9] http://www.opengeospatial.org/standards/wms [accessed 23 November 2008].

as well as a catalogue service including a feature type catalogue (CS-W & FTC) and a Web Similarity Service (WSS) based on SIM-DL. Both services and the client are assumed to use the same ontology and CS-W needs to store metadata about three types of resources: *(i)* services, *(ii)* data, and *(iii)* feature types.

3.2 Geo-service interoperability

The main technological infrastructure to support web service publication, discovery, selection and composition is based on SOA. This architecture is rapidly becoming the standard in the domain of distributed systems. In the case of geographic information, a SOA framework has been developed by OGC. OGC interoperability specifications mainly approaches technical interoperability among geo-services. The most frequently used are WMS and WFS.

OGC specifications and SOA technological solutions provide syntactic interoperability and cataloguing of geographic information. Specifically, OGC published the OpenGIS Web Services Common (WS-Common)[11], the OpenGIS Web Processing Service (WPS) [12] and the Catalogue Service (CAT)[13] specifications:

- **WS-Common** specifies parameters and data structures that are common to all OGC Web Service (OWS) standards. The standard normalizes the ways in which operation requests and responses handle such elements as bounding boxes, exception processing, URL requests, URN expressions, and key value encoding.
- **WPS** provides rules for standardizing how inputs and outputs (requests and responses) for geospatial processing services, such as polygon overlay. The standard also defines how a client can request the execution of a process, and how the output from the process is handled. It defines an interface that facilitates publishing of geospatial processes and clients' discovery and of binding to those processes.
- **Catalogue Service** specification supports the ability to publish and search collections of descriptive information (metadata) for data, services, and related information objects. However catalogue services still do not define any method for overcoming the semantic heterogeneity problem described in the previous section.

Some works have already been performed in automatically and syntactically

[10] http://www.opengeospatial.org/standards/wfs [accessed 23 November 2008].
[11] http://www.opengeospatial.org/standards/common [accessed 23 November 2008]
[12] http://www.opengeospatial.org/standards/wps [accessed 23 November 2008]
[13] http://www.opengeospatial.org/standards/cat [accessed 23 November 2008]

locating distributed SDI resources: Skylab Mobilesystems Ltd.[14] uses a form of web crawling to locate WMS servers. Mapdex[15] has a similar solution which is oriented toward ESRI ArcIMS servers in addition to WMS. Mapdex uses Google search API to find possible WMS sites by searching for WMS-specific query strings appended to URLs.

For the geo-service chaining specific case, a syntactic and semantic analysis was also made in (Lemmens et al., 2006). This work develops a methodology that combines service discovery, abstract composition (identifying service chain functionality with the help of conceptual parameters), concrete composition (managing control and data flow among specific services), and execution. The specific application scenario is represented by a Risk Map service chain. The presented approach uses domain ontologies for the different steps in geographic service chaining.

Geo-service integration has been also investigated in the SWING[16] project, in particular:

- The work in Lutz (2005) proposed a methodology for service discovery. This approach uses ontologies describing geospatial operations to create descriptions of requirements and service capabilities. This work investigates how the methodology can be integrated into existing architectures for spatial data infrastructures, and presents a prototypical implementation. This approach currently considers only plug-in or exact matches between signatures in order to limit the number of found services.
- A comparison between Business Process Execution Language (BPEL) (BPEL, 2003) and WSMO approaches has been made in Gone and Shade (2008). This work proposed a semantic web service composition using WSMO as an improvement of BPEL limitations. Moreover, a use case application (ProCon) was developed and implemented in BPEL and in Web Service Execution Engine (WSMX) (Haller and Scicluna, 2005).
- The work in Maué (2008) presented an extensible architecture for a web service catalogue which supports multiple service description standards (schema-based, like WSDL, as well as ontology-based, like WSMO/WSML (Roman et al., 2005)) and discovery tools. The discussion and implementation of the catalogue focuses on geospatial web services. In particular, the implementation of the proposed architecture makes the import and discovery of web services described either with WSDL or the OGC *getCapabilities* operation result. WSMO has been used to describe the service ontology.

[14] http://www.ogc-services.net [accessed 23 November 2008]
[15] http://www.mapdex.org [accessed 23 November 2008]
[16] http://www.swing-project.org/ [accessed 20 November 2008]

The ORCHESTRA[17] project "*designs and implements the specifications for a service oriented spatial data infrastructure for improved interoperability among risk management authorities in Europe".* ORCHESTRA main result is the development of an open architecture based on standards. Within this project the work in Lutz et al. (2007) presented a rule-based description framework (a simple top-level ontology as well as a domain ontology) and an associated discovery and composition method that helps service developers to create such service chains from existing services.

Finally it is worth noting that most of the previous solutions both in geo-data and geo-service integration, employ a single (top-level) ontology. This allows for the reduction of semantic heterogeneity problem to the problem of reasoning within the shared ontology. However, the adoption of a common ontology for the geographic information communities is not practical, because the development of a common ontology has proven to be difficult and expensive (Smits and Friis-Christensen, 2007). In contrast, in our approach, we assume that geo-data and geo-services are described using terms from different ontologies. Therefore, the problem is shifted to the matching of different domain ontologies. In the following section, we focus and contextualize our approach on a geo-service semantic integration scenario.

4 APPLICATION SCENARIO

In this section we introduce an application scenario to be used as a motivating example for the description of our approach. First, we briefly describe the scenario (§ 4.1). Then, we provide its formalization with the Lightweight Coordination Calculus (LCC) (§ 4.2) that is the communication language employed to implement interactions among the actors of our scenario.

4.1 Scenario description

We have analyzed the organizational model of the distributed GIS Agency infrastructure of Trentino. The framework is represented by a number of specialized GIS agencies: civilian protection, urban planning, forestry, roads, agriculture, cadastral, environment, and geologic survey. Each GIS agency is responsible for providing a subset of the geographic information for the local region. To support interoperability among the different GIS agencies the regional information infrastructure is shifting from a traditional GIS to a distributed SDI.

Within the general Trentino SDI management scenario, we focus on the most commonly used specific use case, i.e., *Map Request Service*. Usually, a map service requestor needs to visualize a map of a region with geo-referenced information selected by a user. In this case, the searched map is a composition of

[17] http://www.eu-orchestra.org/overview.shtml [accessed 8 December 2008]

different geographic layers offered by a GIS service provider.

Interactions between a map service requestor and a map service provider can be described in a verbose form as follow:

> *"The service requestor asks a map service provider for the characteristics of the provided services. After having received the service characteristics (e.g., available layers, map extension, and available formats) the service requestor asks for the map service using the information received from the previous step (e.g., asking for specific layers and the bounding box of the map). If available, the service provider returns a map to the service requestor. Finally, the service requestor can ask for the graphic legend that represents the map."*

4.2 Scenario formalization

In order to implement the formalization of the scenario depicted in the previous subsection, we use a peer-to-peer (P2P) infrastructure, i.e., without any central control, in the SDI domain. At the core of our P2P approach is a specific view on semantics of both web service and peer coordination as proposed in Robertson et al. (2007). Peers share explicit knowledge of the "interactions" in which they are engaged and these models of interaction are used operationally as the anchor for describing the semantics of the interaction. Instead of requiring a universal semantics across peers we require only that semantics is consistent (separately) for each instance of an interaction. These models of interactions are developed locally by peers. However, they must be shared and interpreted by peers in order to support interaction coordination.

This approach achieves this by dynamically matching terms in the interaction models to peers service signature/description. This can happen both at design time (i.e., when synthesising different interactions models) and at execution time (i.e., when running them to perform specific tasks) where the proposed approach is capable to capture the semantics emerging from the peers' interactions.

The communication language employed to implement the interactions among peers acting in our peer network is the Lightweight Coordination Calculus (LCC). LCC is a protocol language used to describe interactions among distributed processes, such as agents and web services. LCC was designed specifically for expressing P2P style interactions within multi-agent systems; henceforth, it is well suited for modeling coordination of software components running in an open environment. Its main characteristics are the flexibility, the modularity and the neutrality to the distributed communication infrastructure (Robertson D., 2004).

Interactions in LCC are expressed as the message passing behaviors associated with roles. The most basic behaviors are to send or receive messages, where sending a message may be conditional on satisfying a constraint (pre-condition) and receiving a message may imply constraints (post-condition) on the peer accepting it.

A basic LCC interaction is shown in Figure 2. The peer *A1* playing the role *r1* verifies if it needs the info *X* (pre-condition *need(X)*); if yes, *A1* asks for *X* to the peer *A2* playing the role *r2* by sending the message *ask(X)*. *A2* receives the message, *ask(X)* from *A1* and then obtains the info *X* (pre-condition *get(X)*) before sending back a reply to *A1* through the message *return(X)*. After having received the message *return(X)*, *A1* updates its knowledge (post-condition *update(X)*).

Figure 2: LCC fragment; double arrows (⇒, ⇐) indicate message passing; single arrow (←) indicates constraint satisfaction

```
a(r1,A1)::
        ask(X) ⇒a(r2,A2) ←need(X) then
        update(X) ←return(X) ⇐ a(r2,A2)
a(r2,A2)::
        ask(X) ⇐ a(r1,A1)
        return(X) ⇒a(r1,A1) ←get(X)
```

The constraints embedded into the protocol express its semantics and could be written as first-order logic predicates (e.g., in Prolog) as well as methods in an object-oriented language (e.g., in Java). The characteristic of modularity allows separating the protocol from the peer engineering. While performing the protocol, peers can therefore exchange messages, satisfy constraints before (after) messages are sent (received) and jump from one role to another so that a flexible interaction mechanism is enabled still following a structured policy, which is absolutely necessary for team-execution of coordinated tasks.

As example of the implementation and usage of the interaction model language, Figure 3 shows LCC code related to the map requestor (GIS Agency Service Requestor, *ga_sr*) and to the map provider (GIS Agency Service Provider, *ga_sp*) roles briefly described in § 4.1.

i. **GIS Agency Service Requestor.**

The GIS agency service requestor, *R*, asks the GIS service provider, *P*, for the characteristics of the provided services (*requestCapabilities()*). After that, the service requestor, *R*, waits until the service provider returns the list of the available services (*AvailableServices*), the list of the available layers (*AvailableLayers*), the format of the returned file (*Format*), and the geographic extent of the available services (i.e., the coordinates of the map extension: *Xmin_ME, YMin_ME, Xmax_ME, YMax_ME*). Then the map requestor, *R*, asks the service provider for a map (*requestMap(Version, Layers, Width, Height, Format, XMin_BB, YMin_BB, XMax_BB, YMax_BB)*). Before asking the map it selects some geographic layers from the list of the available ones (*selectLayers(AvailableLayers, Layers)*), it defines the map dimension (*needMap(Width, Height)*) and it selects the bounding box of the map within the available geographic extension (*selectBoundingBox(Xmin_ME, YMin_ME, Xmax_ME, YMax_ME, Xmin_BB, YMin_BB, Xmax_BB, YMax_BB)*). Finally, the map requestor, *R*, requests the legend representation of the selected layers (*requestLegend(Layers)*).

Figure 3: LCC code for the Map Request service scenario

```
a(ga_sr, R) ::
    requestCapabilities() ⇒ a(ga_sp, P) then
    returnCapabilities(AvailableServices, AvailableLayers, Format,
            XMin_ME, YMin_ME, XMax_ME, YMax_ME, Version) ⇐ a(ga_sp, P) then
    requestMap(Version, Layers, Width, Height, Format,
            XMin_BB, YMin_BB, XMax_BB, YMax_BB) ⇒ a(ga_sp, P)
                ← selectLayers(AvailableLayers, Layers) ∧
                  needMap(Width, Height) ∧
                  selectBoundingBox(XMin_ME, YMin_ME, XMax_ME, YMax_ME,
                                    XMin_BB, YMin_BB, XMax_BB, YMax_BB)
    returnMap(Map) ⇐ a(ga_sp, P) then
    requestLegend(Layers) ⇒ a(ga_sp, P) then
    returnLegend(Legend) ⇐ a(ga_sp, P)

a(ga_sp, P) ::
    ( requestCapabilities() ⇐ a(ga_sr, R) then
      returnCapabilities(AvailableServices, AvailableLayers, Format,
            XMin_ME, YMin_ME, XMax_ME, YMax_ME, Version) ⇒ a(ga_sr, R)
            ← getCapabilities(Version, AvailableServices, AvailableLayers,
                  Format, XMin_ME, YMin_ME, XMax_ME, YMax_ME) ) or
    ( requestMap(Version, Layers, Width, Height, Format,
            XMin_BB, YMin_BB, XMax_BB, YMax_BB) ⇐ a(ga_sr, R) then
      returnMap(Map) ⇒ a(ga_sr, R)
            ← getMap(Version, Layers, Width, Height, Format,
                  XMin_BB, YMin_BB, XMax_BB, YMax_BB, Map) ) or
    ( requestLegend(Layers) ⇐ a(ga_sr, R) then
      returnLegend(Legend) ⇒ a(ga_sr, R) ← getLegend(Layers, Legend) )
```

ii. **GIS Agency Service Provider.**

The GIS agency service provider, *P*, waits for one of the following requests: a request for its characteristics (*requestCapabilities*), a request for a geographical map (*requestMap*) or a request for a graphic legend (*requestLegend*). After receiving one of them, it performs, respectively, the following actions:

- It builds its capabilities (*getCapabilities(MapFile, Version, AvailableServices, AvailableLayers, Format, Xmin ME, YMin ME, Xmax ME, YMax ME)*) and returns (*returnCapabilities(…)*) to the requestor: *(i)* the list of available services (*AvailableServices*), *(ii)* the list of geographic datasets managed by the server (*AvailableLayers*), *(iii)* the file format of the returned services (*Format*), *(iv)* the geographic bounds of the available services (*Xmin_ME, YMin_ME,Xmax_ME, YMax_ME*), and *(v)* the software version (*Version*) of the provided service.

- It builds a digital map (*getMap(…)*) following the features defined by the service requestor, *R*, and returns it to the service requestor (*returnMap(Map)*).
- It builds a legend of the requested layers (*getLegend(Layers, Legend)*) and returns it to the service requestor (*returnLegend(Legend)*).

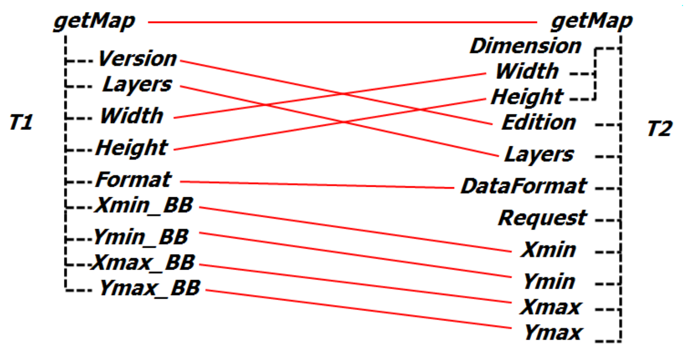
In the following section we will use the ***getMap*** constraint (bolded in Figure 3 to facilitate the presentation) as the motivating example to the structure preserving semantic matching approach.

5 STRUCTURE PRESERVING SEMANTIC MATCHING APPROACH

In our scenario, peers are selected at run time and these can change every time. Selection of peers is based on the similarity of the capabilities of the peer (e.g., the web service operations it publishes or requests) to the constraints of the LCC role the peer wants to play. Let us suppose that we want to match a constraint on a role, such as: *getMap(Version, Layers, Width, Height, Format, Xmin_BB, YMin_BB, Xmax_BB, YMax_BB)* (*T1* in Figure 4, see also bolded in Figure 3) with the capabilities of a peer willing to play that role, such as: *getMap(Dimension(Width, Height), Edition, Layers, DataFormat, Request, Xmin, Ymin, Xmax, Ymax)* (*T2* in Figure 4). These can also be viewed as web service operation descriptions which in turn, can be represented as tree-like structures.

As shown in Figure 4, the first description requires the second argument of *getMap* function (*Layers*) to be matched to the fourth one (*Layers*) of the *getMap* function in the second description. The value of *Version* in the first description must be passed to the second web service operation as the *Edition* argument. Moreover, *Request* (this parameter indicates which web service operation is being invoked) in *T2* has no corresponding term in *T1*.

Figure 4: Two web service descriptions (dashed lines) and correspondences (lines) between them



In our scenario, since there is no a priori semantic agreement (other than the interaction model), the ontology matching component is needed to automatically make semantic commitments between roles of interaction models and peers web service description.

This scenario poses additional constraints on conventional ontology matching. Specifically, we need to compute the correspondences holding among the full graph structures and preserve certain structural properties of the graphs under consideration. Thus, the goal here is to have a *structure preserving semantic matching* operation. This operation takes two graph-like structures and produces: (*i*) one-to-one correspondences between semantically related nodes of the structures preserving a set of structural properties of the graphs being matched, namely that functions are matched to functions and variables to variables, (*ii*) only in the case if the graphs globally correspond semantically to each other, e.g., $graph_1$ is *0.65* similar to $graph_2$, according to some measure.

The approach outlined next follows the work in Giunchiglia et al. (2008). We briefly report it here for completeness and discuss it with the help of an example from our scenario. The matching process is organized in two steps: (*i*) node matching and (*ii*) tree matching. Node matching solves the semantic heterogeneity problem by considering only labels at nodes and domain specific contextual information of the trees. To match nodes, SPSM approach uses the *S-Match* system as proposed in Giunchiglia et al. (2007). Technically, two nodes *n1* and *n2* in trees *T1* and *T2* match if: *c@n1 R c@n2* holds based on S-Match. Where *c@n1* and *c@n2* are the concepts at nodes *n1* and *n2* *and* $R \in \{=, \sqsubseteq, \sqsupseteq,$ "*not related*"$\}$.

In particular, in semantic matching as implemented in the *S-Match* system the key idea is that the relations (e.g., $=, \sqsubseteq$) between nodes are determined by (*i*) expressing the entities of the ontologies as logical formulas and (*ii*) reducing the matching problem to a logical validity problem. Specifically, the entities are translated into logical formulas which explicitly express the concept descriptions as encoded in the ontology structure and in external resources, such as WordNet (Miller, 1995). This allows for a translation of the matching problem into a logical validity problem, which can then be efficiently resolved using sound and complete state of the art satisfiability (SAT) solvers (Giunchiglia et al., 2005). Notice that the result of this stage is the set of correspondences holding between the nodes of the trees.

Tree matching, in turn, exploits the results of the node matching and the structure of the trees to find if these globally match each other. We are mainly interested in approximate matching, since two web service descriptions may only rarely match perfectly in open environments, see (Giunchiglia et al., 2008) for details. Technically, two trees *T1* and *T2* approximately match if there is at least one node $n1_i$ in *T1* and node $n2_j$ in *T2* such that: *(i)* $n1_i$ approximately matches $n2_j$, *(ii)* all ancestors of $n1_i$ are approximately matched to the ancestors of $n2_j$. Notice that the horizontal order of siblings is not preserved being not a desirable property for the data translation purposes.

The implementation of approximate structure preserving semantic matching is based on (*i*) the theory of abstraction and (*ii*) the tree-edit distance. Specifically, the work in Giunchiglia and Walsh (1992) categorizes the various kinds of abstraction operations, including:

- ***Predicate (Pd)*:** Two or more predicates are merged, typically to the least general generalization in the predicate type hierarchy, e.g., $Height(X)$ + $Dimension(X) \rightarrow Dimension\ (X)$. We call $Dimension(X)$ a predicate abstraction of $Height(X)$, namely $Dimension(X) \sqsupseteq_{Pd} Height(X)$. Conversely, we call $Height(X)$ a predicate refinement of $Dimension(X)$, namely $Height(X) \sqsubseteq_{Pd} Dimension(X)$.
- ***Domain (D)*:** Two or more terms are merged, typically by moving constants to the least general generalization in the domain type hierarchy, e.g., $Xmin\_BB$ + $Xmin \rightarrow Xmin$. We call $Xmin$ a domain abstraction of $Xmin\_BB$, namely $Xmin \sqsupseteq_D Xmin\_BB$. Conversely, we call $Xmin\_BB$ a domain refinement of $Xmin$, namely $Xmin\_BB \sqsubseteq_D Xmin$.
- ***Propositional (P)*:** One or more arguments are dropped, e.g., $Layers\ (L1) \rightarrow Layers$. We call $Layers$ a propositional abstraction of $Layers(L1)$, namely $Layers \sqsupseteq_P Layers(L1)$. Conversely, $Layers(L1)$ is a propositional refinement of $Layers$, namely $Layers(L1) \sqsubseteq_P Layers$.

Let use consider the following example: ($Height(H)$) and ($Dimension$). In this case there is no abstraction/refinement operation that makes those first order terms equivalent. However, consequent applications of propositional and predicate abstraction operations make the two terms equivalent: $Height\ (X) \sqsubseteq_P Height \sqsubseteq_{Pd} Dimension$.

Then, the key idea is to use abstractions/refinements as tree-edit distance operations in order to estimate the similarity of two tree structures. Tree-edit distance is the minimum number of tree-edit operations, namely node *insertion*, *deletion*, *replacement*, required to transform one tree to another (Valiente, 2002). We want to: (*i*) minimize the editing cost, i.e., computation of the minimal cost composition of abstractions/refinements, (*ii*) allow only those tree-edit operations that have their abstraction theoretic counterparts. Thus, as an initial hypothesis (to be further analyzed), we assign the same unit cost to all operations that have their abstraction theoretic counterparts, while operations not allowed by definition of abstractions/refinements are assigned an infinite cost.

Finally, the global similarity between two trees is computed as follows:

$$TreeSim = 1 - \frac{\min \sum_{i \in S} n_i \cdot Cost_i}{max(T1, T2)}$$

where, S stands for the set of the allowed operations; n_i stands for the number of i-th operations necessary to convert one tree into the other, $Cost_i$ defines the cost of the i-th operation. The cost is normalized by the size of the biggest tree.

For the example in Figure 4, *TreeSim* would be *0.54*. Then, based on some predefined threshold (e.g., *0.5*) it is decided whether the trees under consideration are similar enough, namely if *TreeSim* is lower than the threshold, those trees are considered as not similar. If the similarity score exceeds a given threshold, the correspondences connecting the nodes of the term trees are further used for data translation.

6 PRELIMINARY EVALUATION

The problem of integration of web services on the basis of the capabilities that they provide has recently received a considerable attention. As pointed in § 2.2, integration of services includes their discovery and their composition/coordination. However, services are often defined using implicit or, in the best case, local semantics. Thus, we specifically evaluate, using a geo-service coordination scenario, a framework to support unsupervised or semi-supervised service discovery and chaining between service operations that are not identical to the one required in a formally described interaction model.

To evaluate the proposed framework, a *map request* scenario has been used. This scenario was implemented within the OpenKnowledge[18] project as a specific interaction model. We notice here that, even if the map service is being standardized by WMS, it could happen that some map service providers do not comply with the WMS specification (e.g., Yahoo Map Image API[19]). Our scenario provides an example of such a situation. We can use the proposed framework to integrate different kinds of geo-services whose signatures are not standardized (e.g., ESRI ArcWeb Services and GRASS web services) or whose descriptions are given through standard specifications like the OGC Web Processing Service (WPS)[20] and OGC's Sensor Web Enablement (SWE)[21] standards.

Ontology matching between service operations is done through the SPSM solution implemented in Java. SPSM builds up a correspondence between each element of the service signature to each element of an interaction constraint. In the case of non-perfect matching, there may be elements in either the ability or the constraint that remain unmatched, and the matches that do exist may not be between entities that are semantically identical. Nevertheless this enables the

[18] www.openk.org [accessed 23 November 2008].
[19] http://developer.yahoo.com/maps/rest/V1/ [accessed 10 December 2008]
[20] http://www.opengeospatial.org/standards/wps [accessed 23 November 2008].
[21] http://www.opengeospatial.org/ogc/markets-technologies/swe [accessed 23 November 2008].

service' users to use its own abilities to satisfy constraints to the highest degree possible.

As the work in Giunchiglia et al. (2009) indicates, it takes around one year to build a large scale evaluation dataset[22]. We have already started the process of building a specific dataset for the SDI/GIS domain and an extensive evaluation constitutes one of the key directions of our future work. Here, we only report preliminary quality evaluation results based on the motivating example of Figure 4 and ca. 50 of similar test cases we have acquired so far. The reference results for these problems were established manually. Then, the results computed by the SPSM solution have been compared with the reference results.

As match quality measures we have used *F-measure*. It varies in the [0-1] range and the version computed here is the harmonic mean of *precision* (the measure of correctness) and *recall* (the measure of completeness), namely that each of these was given equal importance (Euzenat and Shvaiko, 2007). While computing F-measure we considered only if the trees match globally. Based on our previous experience we used here a cut-off threshold of *0.5*. As a performance measures we have used *time* and *main memory*. It estimates how fast SPSM solution when matching trees fully automatically is. All these tests have been performed on a standard laptop: Core Duo CPU - 2Hz, with 2 GB of RAM, with the Windows Vista operating system, and with no applications running but a single matching system.

For the example of Figure 4, the matching algorithm identified *10* node-to-node correspondences, namely *6* equivalence and *4* abstraction/refinement relations. These were further aggregated into a similarity score of *0.54*, which in turn is higher than the selected cut-off threshold of *0.5*, and, therefore, the two trees globally match as expected by human inspection.

For the ca. *50* test cases we ran, we have obtained an average Precision of *0.98*, an average Recall of *0.82* and an average F-measure of *0.71*. The average execution time was *46*ms. The quantity of main memory used by the algorithm during matching did not rise more than 3Mb higher than the standby level. These results look encouraging, especially for what concerns the execution time, which

[22] The key issue here is how to build a set of *reference correspondences* or *reference alignments* against which the results produced by ontology matching systems are to be compared. Notice that the number of possible correspondences grows quadratically with the number of entities to be compared. The work in Giunchiglia et al. (2009) gives an example with web directories, such as Google, Yahoo and Looksmart, each of which has about 10^5 entities. This means that construction of *reference alignments* would require the manual evaluation of 10^{10} correspondences. In our case the situation is slightly easier since we do not have several large tasks, but many small matching tasks, though the number of reference correspondences anyhow grows quadratically with the number of entities to be compared. See Shvaiko and Euzenat (2008) for an overview of the current challenges in the ontology matching field.

if confirmed by more extensive evaluation, would allow for a run-time usage of the semantic matching approach. In any case, further extensive large-scale evaluation is needed.

7 CONCLUSIONS AND FUTURE WORK

In this paper we presented the requirements, the scenarios and open issues of interoperability and the state of the art in the area of distributed geographic data and service integration. We focused our investigation on a semantic interoperability approach in order to integrate geo-services. We presented a semantic heterogeneity problem scenario in SDI, which includes geo-service integration provided by the GIS agencies and the formalization of a map service request using a particular coordination language, i.e., LCC, for the description of the interactions of the involved peers. We then discussed an automatic SPSM ontology matcher used as a solution to the semantic heterogeneity problem between different implementations of required geo-services. We applied the matching algorithm to this scenario and evaluated it with encouraging results, especially with reference to the efficiency indicators. Currently, the matching solution is elementary in the sense that it provides means to match web services available in the LCC interaction models. However, to run an interaction model, a peer should know which interaction model it wants to execute and with which peers it will be interacting. The ultimate goal is to provide a unifying framework based on interaction models that are mobile among peers, being a mechanism for web service coordination and enabling ad hoc peer coalition formation, as required by hastily formed networks (Denning, 2006). In turn, future work on the approximate SPSM proceeds at least in the following directions:

- Conducting extensive and comparative evaluation, including other kind of GIS web services like the ones available from the OGC specifications[23] and the GRASS package[24].
- Extending the matching approach to semantic geo-data integration, for dealing with fully-fledged GIS ontologies (e.g., to match INSPIRE themes classification[25] to user data classifications).
- Add different kinds of thesauri, e.g., the multilingual GEMET[26] and AGROVOC[27] thesauri to support multilingual semantic matching.

---

[23] http://www.opengeospatial.org/standards [accessed 23 November 2008]
[24] http://geobrain.laits.gmu.edu:8099/axis/services [accessed 23 November 2008]
[25] http://inspire.jrc.ec.europa.eu/reports/ImplementingRules/inspireDataspecD2_3v2.0.pdf [accessed 24 November 2008] and http://www.eionet.europa.eu/gemet/groups?langcode=it [Eionet multilingual thematic classification, accessed 24 November 2008]
[26] http://www.eionet.europa.eu/gemet [accessed 23 November 2008]
[27] http://www.fao.org/aims/ag_intro.htm [accessed 23 November 2008]

ACKNOLEDGEMENTS

We are grateful to Fausto Giunchiglia, Mikalai Yatskevich, Paolo Besana, Juan Pane and Fiona McNeill for many fruitful discussions on the structure preserving semantic matching.

This work has been partly supported by the FP6 OpenKnowledge project FP6-027253 (http://www.openk.org). The second author appreciates support from the Trentino as a Lab (TasLab, www.taslab.eu) project of the European Network of the Living Labs at Informatica Trentina S.p.A.

REFERENCES

Bernard, L., Craglia, M., Gould, M. and W. Kuhn (2005). “Towards an SDI research agenda”, *Proceedings 11th EC-GI & GIS Workshop, June 29 - July 1 2005, Alghero, Italy*, pp.147-151.

BPEL (2003). Business Process Execution Language for Web Services, version 1.1 at http://www.oasis-open.org/committees/download.php/2046/BPEL%20V1-1%20May%205%202003%20Final.pdf [accessed 21 November 2008].

Chen, C.-C., Thakkar, S., Knoblock, C.A. and C. Shahabi (2003). “Automatically Annotating and Integrating Spatial Datasets", *Proceedings 8th International Symposium on Spatial and Temporal Databases (SSTD), Vol. 2750, Springer-Verlag, July 24-27 2003, Santorini Island, Greece*, pp. 469-488.

Craglia, M., Goochild, M. F., Annoni, A., Câmara, G., Gould, M., Kuhn, W., Mark, D., Masser, I., Maguire, D., Liang, S. and E. Parsons (2008). Next-Generation Digital Earth. *International Journal of Spatial Data Infrastructures Research*, Vol. 3:146-167. http://ijsdir.jrc.ec.europa.eu/index.php/ijsdir/article/view/119/99

Denning, P. J. (2006). Hastily formed networks. *Communication of the ACM,* 49(4):15-20.

Euzenat, J. and P. Shvaiko (2007). Ontology matching. Springer-Verlag.

Fonseca, F. T., EgenHofer, M. J., Agouris P. and G. Câmara (2002). Using Ontologies for Integrated Geographic Information Systems. *Transactions in GIS*, 6(3): 231-257.

Friis-Christensen A., Schade S. and S. Peedell (2005). “Approaches to solve schema heterogeneity at European Level”, *Proceeding 11th EC-GI & GIS Workshop, ESDI: Setting the Framework, June 29-July 1 2005, Alghero, Sardinia, Italy.*

GEOSS (2008). The Global Earth Observation System of Systems at http://www.earthobservations.org/geoss.shtml, [accessed 18 November 2008].

Giunchiglia F. and T. Walsh (1992). A theory of abstraction. *Artificial Intelligence*, 57(2-3): 323-389.

Giunchiglia F. and P. Shvaiko (2003). Semantic matching. *The Knowledge Engineering Review*, 18(3): 265–280.

Giunchiglia, F., Yatskevich, M. and E. Giunchiglia (2005). “Efficient semantic matching”, *Proceedings 2nd European Semantic Web Conference (ESWC), May 29- June 1 2005, Heraklion, Greece,* pp. 272–289.

Giunchiglia F., Yatskevich M. and P. Shvaiko (2007). Semantic matching: Algorithms and implementation. *Journal on Data Semantics*, IX:1-38.

Giunchiglia, F., McNeill, F., Yatskevich, M., Pane, J., Besana, P., and P. Shvaiko (2008). Approximate structure-preserving semantic matching. *Proceedings 7th Conference on Ontologies, Databases, and Applications of Semantics (ODBASE)*, November 11-13 2008, Monterey, Mexico, pp. 1234-1253.

Giunchiglia F., Yatskevich M., Avesani P. and P. Shvaiko (2009).”A large scale dataset for the evaluation of ontology matching systems”. *The Knowledge Engineering Review Journal*, 24(2), to appear.

GMES (2008). Kopernicus – Global Monitoring for Environmental and Security (GMES) at http://www.gmes.info/, [accessed 18 November 2008].

Gone, M. and S. Shade (2008). “Towards semantic composition of geospatial web services using WSMO in comparison to BPEL”. *International Journal of Spatial Data Infrastructures Research*, Vol. 3, 192-214. http://ijsdir.jrc.ec.europa.eu/index.php/ijsdir/article/view/72/103

Gore, A. (1998). The Digital Earth: Understanding our planet in the 21st Century. *Photogrammetric Engineering and Remote Sensing vol. 65(5), p. 528.*

Groot, R. and J. McLaughlin (2000). *Geospatial Data Infrastructure: Concepts, Cases and Good Practice*. Oxford University Press.

GSDI (2008). Global Spatial Data Infrastructure at http://www.gsdi.org/, [accessed 18 November 2008].

Guarino, N. (1998). “Formal ontology and information systems”, *Proceedings 1st International Conference on Formal Ontologies in Information Systems (FOIS), June 6-8 1998, Trento, Italy*, pp. 3-15.

Haller, A. and J. Scicluna (2005). WSMX Choreography. WSMX Working Draft at http://www.wsmo.org/TR/d13/d13.9/v0.1 [accessed 20 November 2008].

Hess, G.N., Iochpe, C. and S. Castano (2006). "An Algorithm and Implementation for GeoOntologies Integration". *Proceedings 8th Symposium on GeoInformatics*, November 19-22 2006, Campos do Jordao, Brasil, pp.129-140.

IDEW (2001). The Big Picture: Digital Earth and the Power of Applied Geography in the 21st Century at http://www.digitalearth.gov/BigPicture.doc, [accessed 18 November 2008]

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 (INSPIRE). *Official Journal of the European Union*. L108:pp. 1-14.

Janowicz, K., Kebler, C., Schwarz, M., Wilkes, M., Panov, I., Espeter, M. and B. Baeumer (2007). "Algorithm, Implementation and Application of the SIM-DL Similarity Server". *Proceedings 2nd International Conference on GeoSpatial Semantics (GeoS)*. November 29-30, Mexico City, Mexico, pp.128-145.

Janowicz, K., Wilkes, M. and M. Lutz (2008). "Similarity-Based Information Retrieval and Its Role within Spatial Data Infrastructures". *Geographic Information Science,* 5266, pp. 151-167.

Klien, E. (2007). A Rule-Based Strategy for the Semantic Annotation of Geodata. *Transaction in GIS,* Blackwell Publishing, Volume 11(3), pp. 437-452.

Kuhn, W. (2005). Geospatial semantics: why, of what, and how? *Journal on Data Semantics*, *Special Issue on Semantic-based Geographical Information Systems*, 3534, pp. 1-24.

Leclerc, Y., Reddy, M., Iverson, L., and A. Heller (2001). The GeoWeb - A New Paradigm for Finding Data on the Web. *Proceedings 20th International Cartographic Conference (ICC)*, August 6-10 2001, Beijing, China.

Lemmens, R., Wytzisk, A., de By, R., Granell, C., Gould, M. and P. van Oosterom (2006). Integrating Semantic and Syntactic Descriptions to Chain Geographic Services. *IEEE Internet Computing, 10(5)*:pp. 42-52.

Lutz, M. (2005). "Ontology-Based Service Discovery in Spatial Data Infrastructures". *Proceedings 2nd Workshop on Geographic Information Retrieval, November 4 2005,* Bremen, Germany, pp. 45-54.

Lutz, M. and E. Klien (2006). Ontology-based retrieval of geographic information, *International Journal of Geographical Information Science,* 20(3): 233–260.

Lutz, M., Lucchi, R., Friis-Christensen, A., and N. Ostländer (2007). A rule-based description framework for the composition of geographic information services. *Geospatial Semantics, 4853*:pp. 114--127.

Marchese, M., Vaccari L., Shvaiko P. and J. Pane (2008). "An application of approximate ontology matching in eResponse". *Proceedings 5th International Conference on Information Systems for Crisis Response and Management (ISCRAM)*, May 4-7 2008, Washington DC, USA, pp. 294-304.

Masser, I. (2005). GIS Worlds: Creating Spatial Data Infrastructures. ESRI Press.

Maué, P. (2008). An extensible semantic catalogue for geospatial web services. *International Journal of Spatial Data Infrastructures Research*, 3, pp. 168-191.

Miller, A. G. (1995), WordNet: A lexical database for English. *Communications of the ACM*, 38(11): 39–41.

Nambiar U., Ludscher Ludäscher B., Lin K. and C. Baru, (2006). "The GEON portal: accelerating knowledge discovery in the geosciences". Proceedings 8th ACM International Workshop on Web Information and Data Management (WIDM), November 10, 2006. Aarlington, Virgina, USA, pp. 83-90.

Nebert, D. (2004). *Developing Spatial Data Infrastructures. The SDI CookBook, version 2.0*. GSDI-Global Spatial Data Infrastructure.

Nyerges, T. L. (1989). Schema integration analysis for the development of GIS databases. *International Journal of Geographical Information Systems*, 3(2):153-183.

Papazoglou, M.P. (2003). "Service-Oriented Computing: Concepts, Characteristics and Directions", *Proceedings 4th International Conference on Web Information Systems Engineering (WISE)*, December 10-12 2003, Rome, Italy, pp. 3-12.

Parent, C., Spaccapietra, S. and E. Zimányi (2006). *Conceptual Modeling for Traditional and Spatio-Temporal Applications, The MADS Approach*. Springer-Verlag, Berlin.

Paul, M. and S. K. Ghosh (2006). "An Approach for Service Oriented Discovery and Retrieval of Spatial Data". *Proceedings International Workshop on Service Oriented Software Engineering (IW-SOSE'06)*, May 27-28 2006, Shangay, China, pp. 84-94.

Robertson, D. (2004). "A lightweight coordination calculus for agent systems". *Proceedings 2nd International Workshop Declarative Agent Languages and Technologies*, July 19, 2004, New York, pp. 183-197.

Robertson, D., Walton, C., Barker, A., Besana, P., Chen-Burger, Y., Hassan, F., Lambert, D., Li, G., McGinnis, J., Osman, N., Bundy, A., McNeill, F., van Harmelen, F., Sierra, C. and F. Giunchiglia (2007). *Models of interaction as a grounding for peer to peer knowledge sharing*. LNCS Advances in Web Semantics.

Roman, D., Keller, U., Lausen, H., de Bruijn, J., Lara, R., Stollberg, M., Polleres, A., Feier, C., Bussler, C. and Dieter Fensel (2005). Web Service Modeling Ontology. *Applied Ontology*, 1(1): 77 – 106.

SEIS (2008). Towards a Shared Environmental Information System (SEIS). Communication from the Commission to the Council, the European Parliament, the European Economic and Social Committee and the Committee of the Regions, at http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:52008DC0046:EN:NOT [accessed 18 November 2008].

Shvaiko P. and J. Euzenat (2008). Ten challenges for ontology matching. *Proceedings 7th Conference on Ontologies, Databases, and Applications of Semantics (ODBASE)*, November 11-13 2008, Monterey, Mexico, pp. 1164-1182.

Smits P. C. and A. Friis-Christensen (2007). “Resource discovery in a european spatial data infrastructure”. *IEEE Transactions on Knowledge and Data Engineering*, 19(1):85-95.

Stuckenschmidt, H. (2003). Ontology-Based Information Sharing in Weakly Structured Environments. PhD thesis, Vrije Universiteit Amsterdam (NL).

Tu, S. and M. Abdelguerfi (2006). “Web Services for Geographic Information Systems”, *IEEE Internet Computing, 10(5)*:pp. 13-15.

Valiente, G. (2002). *Algorithms on Trees and Graphs*. Springer.

Worboys, M. and M. Duckham. (2005). GIS a computing perspective, 2nd ed. CRC Press.