Semantics - Supportive Element for the Cooperative Evaluation of Geographical and Historical Information

Ashish Karmacharya, Tobias Kohr, Christophe Cruz, Kai-Christian Bruhn, Frank Boochs

Semantics – supportive element for the cooperative evaluation of geographical information and historical texts

Karmacharya A[1]., Kohr T[1]., Cruz C[2]., Bruhn K.-C[1]., Boochs F[1].

[1] Institut i3mainz, am Fachbereich 1 - Geoinformatik und Vermessung
Fachhochschule Mainz, Lucy-Hillebrand str 2, 55128 Mainz
{ashish, kohr, bruhn, boochs}@geoinform.fh-mainz.de
[2] Laboratoire Le2i, UFR Sciences et Techniques, Université de Bourgogne
B.P. 47870, 21078 Dijon Cedex, France
{christophe.cruz}@u-bourgogne.fr

Abstract

The emergence of the Semantic Web and its underlying knowledge technologies has brought changes in data handling. Transferring expert knowledge to machines through knowledge formalization provides us the required support in managing huge datasets like the information in the World Wide Web. We present two of our works in the field of geospatial technology where we implied semantic technologies from the Semantic Web framework not only to achieve higher degree of data integration but also infer semantics to discover new and hidden knowledge. The works are presented in the backdrop of archaeology. Although researches on semantics are active areas in geospatial communities, their initial use is mainly for spatial data integration. We present one of the first works to imply semantics for spatial knowledge discovery through spatial built-ins within SWRL and SPARQL in this paper. The work resembles the standards from Open Geospatial Community (OGC) to define standards for GeoSPARQL.

Zusammenfassung

Die Entstehung des Semantic Web und die damit verbundenen Wissenstechnologien haben Veränderungen in der Datenhandhabung mit sich gebracht. Wissensformalisierung die es erlaubt Fachwissen an Maschinen zu übermitteln, liefert die notwendige Unterstützung zur Verwaltung sehr großer Datensätze, wie die Informationen im World Wide Web. Wir stellen zwei unserer Arbeiten im Feld der Geoinformatik vor, in denen wir semantische Technologien nicht nur einsetzen um einen höheren Grad an Datenintegration zu erreichen, sondern auch um aus Rückschlüssen der Semantik neue Erkenntnisse zu gewinnen. Beide Arbeiten siedeln sich in der Archäologie an. Wenngleich Semantik ein aktives Forschungsfeld in der Geoinformatik ist, konzentriert sich ihr ursprünglicher Nutzen dort größtenteils auf die Integration räumlicher Daten. In diesem Artikel präsentieren wir erste Ansätze Semantik zur raumbezogenen Wissensgenerierung einzubeziehen. Dies wird durch den Einbau räumlicher Parameter in SWRL und SPARQL erreicht. Die Arbeit ist mit Standards des Open Geospatial Consortium (OGC) vergleichbar, welches hierfür mittlerweile GeoSPARQL spezifiziert hat.

KEYWORDS: Knowledge Modeling, Spatial Integration, the Semantic Web, Archaeology, Philology

1 Introduction

Traditional geoinformation systems (GIS) enable the capture, management, analysis and presentation of spatially referenced data. Fields of application are manifold, comprising regional planning, disaster prevention, traffic management, local administration and logistics, just to name a few. In times of spatial data infrastructures (SDI) and open accessible geodata over the Web, the relevance of GIS is steadily increasing. Besides the well-known application fields, GIS also serve a variety of marginal areas, such as archaeology. Here they provide the user distinct views on the data material, which sets the base for domain specific analysis. Considering the heterogeneity, variety and potential incompleteness of data, as well as the complexity of observation, attainable statements are confined however. Nevertheless GIS tools also provide important assistance in this field that will further expand in the future.

A database structure that holds geometrical and attributive feature information, as well as topological relations between features sets the base of the traditional GIS architecture. Based on these data, spatial analyses are performed, graphically visualized, evaluated by the user and (manually) used to infer further knowledge. Deficits of this approach are encountered at two major points: First, a once defined data structure is fixed and not adaptable in retrospect. Second, the information systems cannot contribute to the knowledge retrieval directly.

This leads to drawbacks if the data structure is potentially subject to changes - as i.e. in the case of archaeological excavations, that often encounter unexpected facts - and as far as the interrelationships of interest are multilayered and to be developed in an explorative manner.

Broader flexibility in the creation and more possibilities in the supportive analysis would imply considerable benefits under such circumstances. New developments for the retrieval of the vast information of the Web are promising for this purpose. Currently emerging capabilities of knowledge based processing mechanisms possess significant benefits towards traditional approaches. Data are intelligently and flexibly managed and can therefore be dynamically adapted. This allows modification of the data model where required. Moreover the computer can directly contribute to further knowledge retrieval applying user-defined rules. In particular in complex situations, this can provide significant assistance.

In order to reach this objective, technologically unknown territory needs to be entered in many respects and it is still a long way to comprehensive alternatives of today's GIS. First works already present the potential (Karmacharya, 2011). Yet without complete intelligent GIS, interesting capabilities to facilitate spatial issues are feasible. Geospatial communities have started to give notice to the rising popularity of knowledge technologies with the Semantic Web framework. OGC work on standardizing spatial components in the Semantic Web technologies indicate in this direction.
We at i3Mainz started to notice the rising importance of knowledge in handling large scale data in archaeology in early 2007. Our research project *Räumliches Informationssystem zur Erfassung, Dokumentation und Analyse Industriearchälogischer Objekte* (RIO) which documents the excavation findings in industrial archaeology started to experiment the usage of knowledge in managing these documentations. Though this was our initial project, we experienced the impacts of underlying semantics within information to build knowledge and the impact of knowledge in handling heterogeneous data sources. What was more intriguing was that knowledge technologies not only help in data integration but also provide a base for intelligent systems through their inference capabilities. RIO successfully demonstrated a rule based system that infers the hypothesis and facts to discover new knowledge. We extended our work to include semantics within spatial technology and came out with recommendations of spatial integration mechanism in the Semantic Web framework (Karmacharya, 2011). Those recommendations were one of earliest works in this area. The work of RIO was followed up with another interdisciplinary project: *Die historische Geographie Obermesopotamiens* (HiGeoMes). The project constitutes the works of archaeologists and philologists. One of the core tasks of the project was to locate places mentioned in cuneiform records. Those places possess relative locations but do not have any concrete argument to point out the current day locations. It is not possible to geo-locate these places without help from other datasets. We thus make use of archaeological data and attach semantics to them. Knowledge based data integration between these two datasets provides clue to geo-locate those cuneiform toponyms which were otherwise unlocated.

The following chapters will give an introduction to knowledge management and semantic technologies and discuss their integration with spatial technologies. Subsequently the paper presents feasible capabilities emerging from this approach on two case studies: HiGeoMes and RIO. It then presents the next works in this area that i3Mainz will be engaged in. Finally a conclusion is made.

2 Knowledge Management and the Semantic Web

Knowledge Management in simplest term is the process of identifying, creating, distributing the experiences, expertise and insights possessed within an individual or group or even an organization. Knowledge is commonly distinguished from data and information (Zack, 1999). Data are a representation of an observation or any singular fact kept out of context. Data are meaningless until they are put in context of space or an event. Additionally, unless the relationships between different pieces of data are defined, simply data do not have any significance. Once data are defined in terms of space or events and are defined through relationships, they become information. Information understands the nature of the data but does not provide the reasons behind the existence of data and is relatively static and linear by nature. Information is a relationship between data and, quite simply, is what it is, with great dependence on context for its meaning and with little implication for the future (Bellinger, 2004). Beyond every relationship, arises a pattern which has capacity to embody completeness and consistency of the relations to an extent of creating its own context (Bateson, 1979). Such patterns represent knowledge on the information and consequently on data. The term *Knowledge Management* has wide implications. However, very precisely Knowledge Management is about the capture and reuse of knowledge at different knowledge levels.

Knowledge Management re-evolved with the rise of the Semantic Web. The explosion of information in the World Wide Web (WWW) led to the problem of managing it. It is generally perceived that this vast information could not be managed through human effort only. In some form there should be interference from machines to assist humans manage the information. In order to have machines interfere and assist humans in managing information, it should understand the information first. This would require knowledge formalized from the information. In their paper Berners-Lee, Hendler, & Lassila (2001) have envisaged the next generation of the Web which they call "the Semantic Web". In this Web the information is given with well-defined meaning, better enabling computers and people to work in cooperation. Adding on, the Semantic Web aims at machine-processible information enabling intelligent services such as information brokers, search agents and information filters, which offer greater functionality and interoperability (Decker et al., 2000).

The association of knowledge with the Semantic Web has provided a scope for information management through knowledge management. Since both the technologies use ontologies to conceptualize the scenarios, Semantic Web technology could provide a platform for developments of knowledge management systems (Stojanovi & Handschuh, 2002). The ontologies are core to both the technologies in whichever methods they are defined. The Semantic Web defines ontologies through XML based languages and with the advancements in these languages.

2.1 Knowledge technologies within the Semantic Web

Ontology Language

The term *Ontology* is being used for centuries to define an object philosophically. The core theme of the term remains the same in the domain of computer science; however the approach in defining it has been modified to adjust the domain. Within the computer science domain, ontology is a formal representation of the knowledge through the hierarchy of concepts and the relationships between those concepts. In theory ontology is a *formal, explicit specification of shared conceptualization* (Gruber, 1993) In any case, ontology can be considered as formalization of knowledge representation and *Description Logics* (DLs) (Calvanese et al., 2001); (Baader & Sattler, 2000) provide logical formalization to the Ontologies (Baader et al., 2003). W3C or the World Wide Web Consortium, the main international standard organization for WWW has standardized Web Ontology Language to model ontologies. OWL is actually a family of three language variants of increasing expressive power: OWL Lite, OWL DL, and OWL Full. The standardization of OWL has sparked off the development and/or adaption of a number of reasoners, including FacT++, (Tsarkov & Horroks, 2006) Pellet, (Nguyen & Nguyen, 2010), RACER (Haarslev & Muller, 2001) and HermiT (Shearer et al., 2008), and ontology editors, including Protégé (Protégé) and Swoop, (Kalyanpur et al., 2006). OWL 2 is a new version of OWL, the ontology language which considerably improves the datatype (Motik et al, 2009).

Query and Rule Languages

SPARQL is a query language for RDF triplets of which OWL is syntactically aligned. In this manner SPARQL queries the knowledge within OWL. As a query language, SPARQL is "data-oriented" in that it only queries the information held in the models; there is no inference in the query language itself. SPARQL is able to query OWL ontologies which use RDF graphs to structure it. SPARQL uses FILTERS to limit the solutions to only those which are returned true with the expression. The section presents the syntax of the FILTERS with an example. Generally FILTER comes at the end of any SPARQL expressions.

Semantic Web Rule Language (SWRL) is a logic programme which infers the knowledge base to derive a conclusion based on the observations and hypothesis. The SWRL as the form, antecedent -->consequent, where both antecedent and consequent are conjunctions of atoms written a1 ^ ... ^ an. Variables are indicated by using the standard convention of prefixing them with a question mark (e.g., ?x). URI references (URIrefs) are used to identify ontology elements such as classes, individual-valued properties and data-valued properties. For instance, the following rule asserts that one's parents' brothers are one's uncles where parent, brother and uncle are all individual-valued properties.

$$hasParent(?x, ?p) \wedge hasBrother(?p, ?u) \rightarrow hasUncle(?x, ?u)$$

3 The Background

The basic tasks of a GIS system can be broken down into five groups (which we like to call 5Ds), **data acquisition**, **spatial data management**, **database management**, **data visualization** and **spatial data analysis** (Jones, 1997). Most archaeological data such as artifacts, features, buildings, sites or landscapes, have spatial and aspatial attributes that can be explored by GIS. These attributes include the spatial location that informs about the local or global context concerning the pieces of information, and the morphology that defines the shape and the size of an object.

In recent years, the rapid growth in data acquisition techniques - that are applied in archaeology - has made some limitations visible in traditional standalone GIS system. It is not only the volume of data any more. The diversity in data collected play equal role in current archaeological projects. Nevertheless, for many of the archaeological projects an information system is still either a GIS system or a 3D modeling system. Applications like ArchaeoCAD from ArcTron and PointCloud from Kubit rely heavily on the geometry of the objects excavated. Correspondingly, research projects like 3D MURALE (Cosmas et al., 2001) and GIS DILAS (Wüst et al., 2004) lean heavily towards spatial data management focusing on extending GISs to fit in archaeological data management. One could thus argue spatial components play a major role in an archaeological project. However, with rapidly diversifying data patterns through modern sophisticated data acquisition techniques have exposed the incompatibility of current GISs to accommodate them. Even data structures like the 3D point cloud which possess spatial dimension finds itself extremely difficult to succor with existing GISs. This is even more prominent with aspatial data structures like multimedia datasets. Most of the archaeological projects today use 3D point clouds to document their findings for future reconstructions. Likewise, they use document files to document their processes of excavations and other multimedia to record their steps. All these need to be taken care of in order to have a comprehensive analysis process. 5Ds process within any GISs should thus need to take aspatial data structure into account. Data integration is an important issue within current information management systems in archaeological projects.

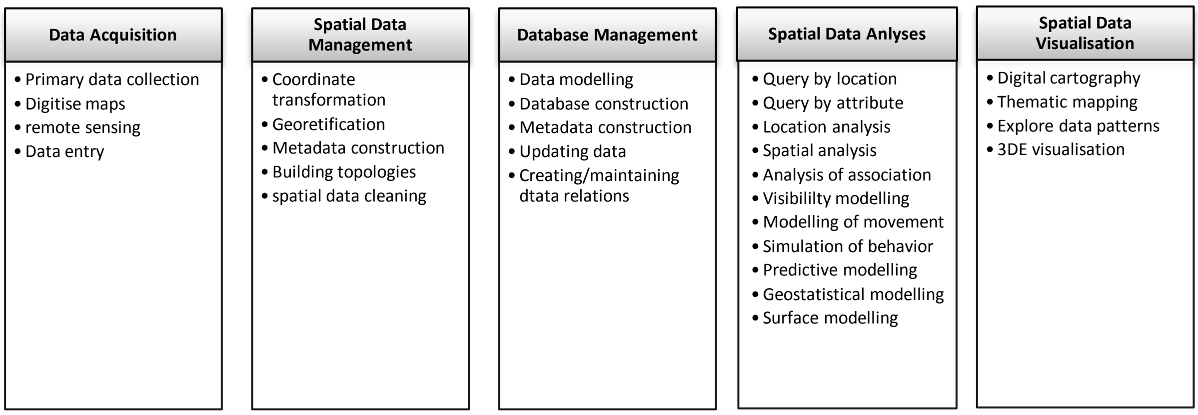

Figure 1. Five main groups of tasks performed by a GIS (Zack, 1997)

Like other disciplines, archaeology does not only produce very heterogeneous data within one project, but in numerous ventures, which can have very diverse spatial, temporal and thematic scopes. At the same time very close interrelations among projects exist, making it essential to connect their information in order to gain new knowledge. In the field of geo-informatics, Spatial Data Infrastructures (SDIs) have emerged to enable the dissemination of spatial data over the Web. International initiatives like INSPIRE urge their introduction on all administrative levels. This trend acts as a driving force for the specification of relevant standards and the development of corresponding technologies. Becoming the leading standardization instance, the Open Geospatial Consortium (OGC) has defined a stack of Web service specifications generally complying with OGC Web service (OWS) Commons. These standards and technologies are leveraged far beyond the initial INSPIRE directive, including the dissemination of archaeological data. They have achieved a means for information exchange that is syntactically interoperable which is a big accomplishment in itself. The actual meaning of the information remains unreadable to a machine however. This restrains discovery, integration and automated processing of the information (Janowicz et al., 2010). With the current trend of information explosion a next level of interoperability is aspired. Even with syntactical interoperable systems, the user needs to know in-depth knowledge about the data structures of the data sources s/he is evoking. This is not always possible. In order to have higher degree of interoperability, data structures of one data source should understand other data structures of other data sources. This will require attaching meaningful semantics to each data source. It would facilitate common understanding between different datasets irrespective to their structures and sources. Data sources in a distributed environment could be thus connected through a common semantic and syntactic model and syntax.

The importance of semantics in achieving higher level of interoperability is widely felt in the geo-spatial community. OGC is taking forward its efforts on existing web service standards for publishing geospatial data through attaching semantics to the services. Reports by Houbie et al. (2012) point out the importance of semantically annotating the OGC standards publishing geodata in web services like WFS. They identify three levels that are in need of semantic annotations: Service metadata, data models or process descriptions and data instances. Annotations enrich OWS data within the existing interface that can be used by semantically enabled and non-enabled consumers. The approach can enhance the aforementioned shortcomings in data discovery, integration and automated processing.

Besides providing semantic annotations within the current standards, OGC is also moving towards standardizing spatial extensions to the Semantic Web technologies. GeoSPARQL is the latest example of such activities. It defines vocabularies to represent geospatial data in RDF, and defines spatial extension to the SPARQL query language for processing geospatial RDF data (Perry & Herring, 2012). Roth et al. (2013) pursue to provide linked data from OGC Web service output to harness capabilities of Semantic Web technologies. This has the potential to perform semantic analysis on the data and integrate them efficiently with non-spatial information. Moreover mapping schemas entail the capacity to link data sources with each other independent of their format.

4 Case Studies

The rapid growth of knowledge technologies through the Semantic Web framework is widely felt after its introduction in 2001. Though the technology entered our domains of research only in 2007, we have made a huge leap forward by today. We present two case studies that provide insight of our efforts to evaluate and experiment with these technologies. At i3Mainz we primarily focus our researches in spatial technologies whether that is geoinformatic science or measurement techniques in geometric data as point cloud. It is thus obvious that we use knowledge technologies to collaborate with spatial technologies for managing and interpreting data we generate. The research projects presented in this section highlight the research activities in converging the spatial and semantic technologies for managing information through knowledge management.

4.1 HiGeoMes

The three-year bilateral joint research project *HiGeoMes - Die historische Geographie Obermesopotamiens im 2. Jt. v. Chr.: Interdisziplinäre Forschungen*, funded by the French ANR and the German DFG since 2011 focuses on the integration of places mentioned in Babylonian and Middle-Assyrian cuneiform texts with the location of known archaeological sites in the Near East. The projects main goal is to connect the documented settlement sites with place names in the written sources to better understand political, social and environmental developments in Upper Mesopotamia in the 2nd millennium BC.

HiGeoMes uses a spatial data infrastructure (SDI) based on OGC Web services to facilitate the exchange of collected archaeological information among project partners, and to ensure external researchers syntactically interoperable access. Two distinct data patterns are collected and managed: 1) the archaeological data collected at the University of Mainz are shared via a Web Feature Service (WFS) along with Web Map Services (WMS) portraying relevant raster imagery. A Web GIS client acts as an entry point for visualization and simple analysis of the site locations, 2) epigraphic data generated through cuneiform records are commonly not yet associated with known locations, which prevents absolute geographic modeling. However these places often entail some relative geographic information, such as the proximity to a river or to another place. Consequently, textual data contain complex (spatial) interrelations.

A major challenge of the project is extracting information patterns from the epigraphic data and structuring it. This requires in-depth knowledge about the epigraphic data and its original source: cuneiform records. A mechanism is developed for close collaboration with the domain experts to extract semantics from the textual information and store it inside OWL ontology. The ontology thus defines knowledge of domain experts and possesses semantics of epigraphic data. This is completely in contrast to the archaeological data where the data are stored in a database system and shared through OGC compliant standards. It is thus clear that conventional approach of integrating and sharing data is not possible with the existing SDI structure. The other issue of interest within HiGeoMes project is to geocode the places in epigraphic data. So the data integration is even more relevant for the project.

One of the most sought after mechanisms to provide spatial signatures to aspatial dataset is to couple it with equivalent spatial databases. Such heterogeneous data integration is a leading subject of research today (Cruz, 2004), (Cruz et al., 2004), (Tanasescu, et al., 2006). However, it is not straightforward in our case. Archaeological data that are stored in a relational database and distributed through an SDI have their own structure and epigraphic data that are stored and expressed through OWL have their own structure respectively. Hence, it seems unlikely that a syntactic integration would resolve the data integration issue. SDIs though designed for spatial data integration concentrate on syntactic integration. They have broad limitations when the data structures are this diverse. In order to have integration among such data structures, one data source should understand the other data source. Semantics of the datasets thus play a major role in data integration. Syntactic approaches provided by standards, metadata and infrastructures like SDI are in need of semantic enablement or a semantic extension to accomplish this task (Janowicz et al. 2010). With this mantra architecture is designed to use semantics in data integration (as seen in fig 2).

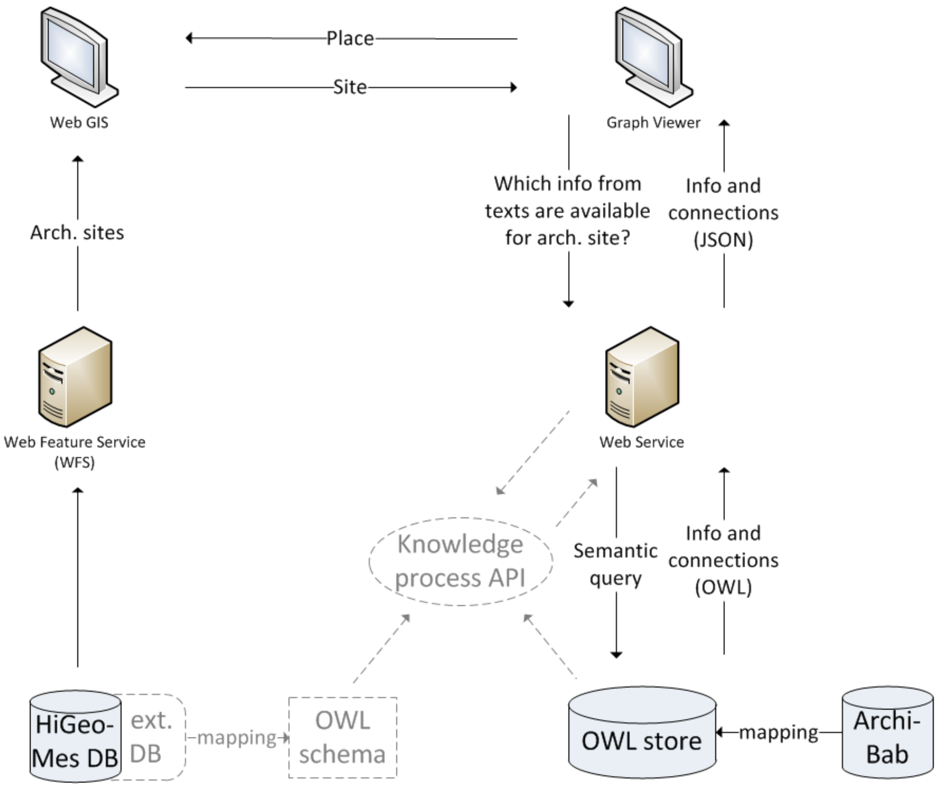

Figure 2. HiGeoMes Architecture

The HiGeoMes architecture (see fig. 2) illustrates a mechanism that integrates the archaeological data shared through SDI and epigraphic data present in the OWL ontology. HiGeoMes makes the resources of integrated data available via a web service interface. The current approach establishes the connection to this additional information on a visual level: The WebGIS client hosting the spatial dataset from the archaeological data is capable of making smart queries to related toponyms in the epigraphic OWL ontology. These smart queries yield detailed knowledge on the queried sites. A graph visualization to represent evoked knowledge about the queried sites presents the semantic relation of the site with respect to other places. This is independent to geolocalization of the places. In parallel, the user can locate places from texts on the map that have been associated with archaeological sites. An integrated visual exploration of geographically and semantically modeled information is possible. However, in order to discover new knowledge the different data sources need to be connected on the semantic level in the future (see section 5 and dotted features in figure 2).

4.2 RIO

Räumliches Informationssystem zur Erfassung, Dokumentation und Analyse Industriearchälogischer Objekte or simply RIO is the pioneer project in the direction of knowledge management and semantic technology at i3Mainz. The project implements the concept behind the Semantic Web and its underlying knowledge technologies within the backdrop of industrial archaeology. The project's initial site was a 200 hectares Krupp area in Essen belt, west-north Germany but was extended to other industrial archaeological sites later. The area was used for steel production in the early 19th century. The site was destroyed in the Second World War and was never rebuilt. However lately ThyssenKrupp has started building its new headquarter on the site. One of the major problems in any industrial archaeology is that the excavation site is available for limited duration and RIO was not an exception. The site was available for a very short time, which had significant effects on the management of the excavated objects. The objects are recorded as soon as they are excavated and these records are stored in a repository in their respective data formats. There was no clear data structure defined and hence there was not much possibility of data analysis. Next problem was the amount and diversity of data patterns collected during the excavation. The sophistication in technology not only led to generate huge volumes of data but also generate diverse data patterns. Table 1 provides glimpses of diversity in patterns of data collected.

Laser Scanning	Images	Old Archives	ArcGIS database
Resolution: 0.036 degrees (6 mm on 10 m) Bunker, Oven and Wall scanned	Aerial Image of the site with a resolution of 10 cm. Images taken of findings (Mostly without reference system)	Floor plans Notes and sketches (old and during excavation)	Buildings archaeologists are interested on

Table 1. Diversity of data collected

RIO complements the principles of the Semantic Web within the project. Data collected during the excavation are attached with their semantics through the process of semantic annotation. In this manner a foundation is laid to make all data understand each other irrespective to their structures. An ontology reflecting the knowledge of archaeologists is designed. This ontology schema not only provides a knowledge structure to define semantics of the objects excavated but also provides a mechanism to annotate data and documents related to the respective objects. Accordingly, the data and documents are provided with their semantics. This ontology schema provides a base for knowledge formalization with which it is logically possible to relate an object detected in a point cloud to the object identified in an image or a CAD document.

A web based application prototype ArchaeoKM (Karmacharya et al., 2010) was developed during the project. It is a hybrid system which uses the potentials of current database technologies (esp. spatial processing) and knowledge technologies. ArchaeoKM provides a semantic platform for formalizing and managing their knowledge. It also provides a virtual platform for data integration between different data patterns through their semantic relations. It thus attempts to document the expert archaeological knowledge from archaeologists and facilitates to manage their knowledge. It complements the 5Ds processing steps of a GI-System (see fig. 1) with its own 4Ks: **knowledge acquisition, knowledge management, knowledge visualization and knowledge analysis.**

The most important achievement of ArchaeoKM under the RIO project was however laying a foundation for spatial integration with the Semantic Web framework. It is probably one of the earliest projects which worked on integrating spatial and semantic technologies together. Though the project primarily focuses on facilitating archaeologists to manage their spatial data through knowledge management techniques, it can be implemented in other areas of spatial data management.

A Spatial Extension Ontology Schema (SEOS) was designed in the process. SEOS constitutes axioms and theorems for spatial functionalities and is extended within the ontology schema. The spatial axioms and theorems are presented through DL concepts and roles. These DL components are populated through a semantic measurement mechanism which extracts spatial signatures of the measured objects. The geometries are stored as a spatial data type in PostgreSQL database system (http://www.postgresql.org) with its spatial extension PostGIS (http://www.postgis.net). Now the populated objects are geocoded and can be used in spatial analysis.
SEOS has provided a foundation where non spatial data types can participate in spatial analysis or spatial queries. It is now possible to have a smart query like "*provide me the chimneys stored in CAD document* ***xyz.dwg*** *which is* ***touching*** *ovens stored in point cloud file* ***pc.xyz***". This example illustrates two different data structures that are integrated and queried through the spatial function ***touch.*** However, the spatial integration has far more benefits than the query of finding chimneys. It is possible to define spatial rules to infer the knowledge possessed in the populated ontology (we call it knowledge base or KB from now). ArchaeoKM proposes spatial built-ins for SWRL which could combine with other built-ins to infer the KB. One such example is illustrated below.

Figure 3 (a) illustrates the remains of buildings and remains of the rail track in the excavation site. A hypothetical GIS analysis of determining the buildings that might have used the rail track to get supplies of the raw materials during the production period is presented as an example. In order to draw the example closer to the case study presented here, the location map is a section of the industrial archaeology site that produces steel. Likewise, the machine halls represented through instances `MH_1` and `MH_2` consists of machines and one of those machines processes iron ore (one of the raw materials needed to produce steel). So, the GIS analysis should determine the machine hall that uses the railway track to receive its supply for processing in steel production. We present an alternate to perform spatial analysis through spatial inference. A backend inference engine manipulates the spatial operations and functions through a database system to provide the result. A SWRL rule with spatial built-ins developed during the project to infer the KB:

> feat:RailTrack(?x) ^ feat:MachineHall(?y) ^ feat:Furnace(?f) ^ att:hasRawMaterial(?y, att:ore_iron) ^ spatialswrlb:Buffer(?x, 50, ?z) ^ spatialswrlb:Intersection(?y, ?z, ?res) ^ spatialswrlb:Touches(?y, ?f) → feat:hasSupplyLine(?y, ?x)

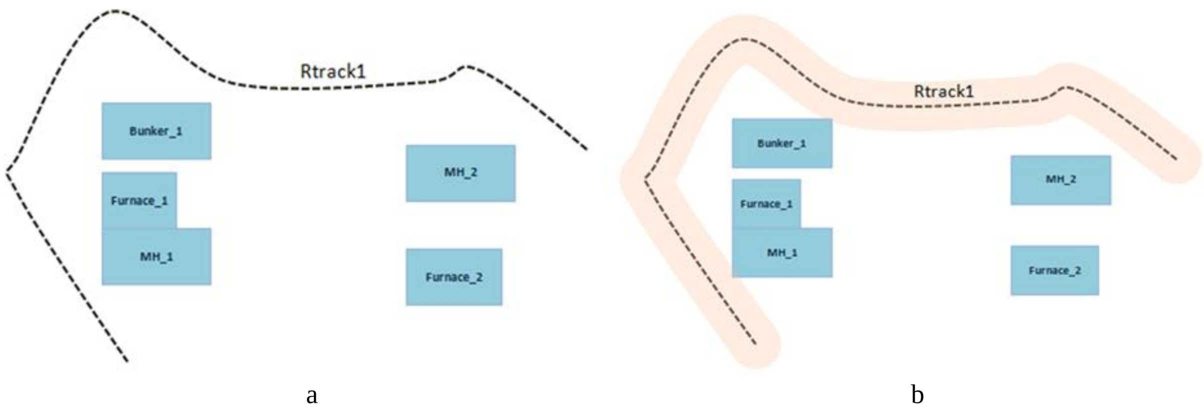

Figure 3. a) Excavation scenario from the ground b) a hypothetical GIS analysis result

The result of this inference is given in figure 3 (b).

ArchaeoKM also proposes spatial filters for SPARQL. The following example demonstrates its use to query a famous wine ontology. This highlights the principle of SEOS that allows the computation of spatial data on any ontology whether it is under development or currently existing. The wine ontology is selected for several reasons: 1) the wine ontology appears frequently in the literature as an example to define tutorials 2) this ontology is far from the industrial Archaeology which underlines the universal principle of the ontology adjustment for spatial processing.

```
SELECT  ?region
WHERE
{
        vin:CoteDOrRegion rdfs:type   vin:Region
        ?region        rdfs:type   vin:Region
        SPATIAL_FILTER [buffer (?buffer,200000,vin:CoteDOrRegion)]
        SPATIAL_FILTER [intersection (?res,?buffer,?region)]
}
```

The above example queries all the regions within 200 km of (CoteDOrRegion Noteworthy, the filters (as buffer and intersection) within the keyword SPATIAL_FILTER (see above example) were proposed later in similar context within OGC GeoSPARQL (reference) specifications.

5 What Next….

Conscious efforts have been made in semantically enriching geospatial data by the geospatial communities. We continue our trend on advocating the advantages of semantically attached geospatial data. Though the global efforts on semantic attached spatial data are driven through the motivation of data integration, we believe such efforts have far more implications. One of the major potentials in our view is knowledge discovery. ArchaeoKM through the RIO project has provided a glimpse of success in discovering knowledge through a rule based inference mechanism. We intend to carry this success story forward and implement knowledge discovery in the next edition of HiGeoMes.

Spatial data mining which deals with the process of discovering interesting and previously unknown but potentially useful patterns from large datasets has been around for some time. Spatial data mining and knowledge discovery have emerged as a leading research domain that handles challenges and limitations to manage vast and diverse data which conventional geospatial technologies struggle to address (Andrienko & Andrienko, 1999), (Shekhar & Huang, 2001). Guo & Mennis (2009) list out theoretical and applied researches in spatial data mining and knowledge discovery. Spatial data mining benefits from both classical data mining techniques and traditional spatial analysis processes to integrate and further develop the analysis mechanisms for large and complex data structures (Guo & Mennis, 2009). We intend to take this effort forward to execute spatial knowledge discovery through usage of knowledge technologies. Usage of knowledge technologies within spatial data mining is however not new and researches like (Hwang, 2004) have highlighted the benefits of using ontologies as a means to customize algorithms for different purposes.

The problem with the data structure in the HiGeoMes project is not straightforward spatial. As mentioned two distinct data types are observed: data generated from cuneiform records and archaeological data collected through excavations. We have built a model to integrate these two differently structured data patterns to allocate geo-locations to toponyms in cuneiform data. However, problems still exist in geocoding each place in cuneiform data due to lack of information (or better understanding of data). It is almost practically impossible to analyze each piece of information because of its volume and diversity. The next edition of HiGeoMes will bring on the hypothesis and facts from archaeologists and philologists and formulate individual knowledge models for the respective domain. This will extend current HiGeoMes knowledge schemas to fit in the formulated hypothesis. A semantic middleware will create the semantic bridge between these knowledge schemas. In this context we will especially consider OGC standards and regard what role interfaces like WFS can play. As discussed they build a stable infrastructure for the syntactically interoperable dissemination of spatial HiGeoMes data that need semantic enablement. The semantic middleware not only acts as a facilitator for data integration but will also lay a foundation for knowledge discovery.

What differentiates this with a typical spatial knowledge discovery is that we do not always have and consider relying on complete spatial packages. The data from cuneiform records are partially geocoded in the current state leaving most part with no spatial signatures. In other hands, archaeological data are supported through strong spatial presence. We have two scenarios here: first cuneiform data supported by strong a semantic pattern but no spatial contents except relative spatial references (e.g. toponyme ***xyz*** is ***close to*** river which ***flows through*** a hill top is a military base) and second archaeological data supported through strong geocoded spatial signatures with limited semantics (except the facts collected on ground and few hypothesis made by archaeologists). We intend to use whatever limited semantics archaeological data possesses to relate them to semantics from the data of cuneiform records to discover new insight knowledge. In short, the hypothetical knowledge from the facts discovered in the excavated site will be related to see its relevance in the philological data from cuneiform records. This relevance will be deducted through inferring hypothetical rules against the facts in both datasets. It leads to discovering both spatial and non-spatial knowledge. An obvious benefit is that it will provide geocoded spatial signatures to places in cuneiform datasets which were not located previously. Following up the above example, excavation in modern day ***XYZ*** with geo-coordinates ***Xx, Yy, Zz*** discovered mass weapons and ammunitions and this excavation site has a hill top in some distance and a reminisce of a probable river through it, then it could be inferred that modern day ***XYZ with its coordinate value is a candidate location for toponyme xyz*** from the cuneiform record.

6 Conclusions

In this paper we presented our works on implications of semantics in geospatial technology through semantic technologies. ArchaeoKM: a prototype application to facilitate industrial archaeologists to manage their information through the knowledge management tools was one of the earliest works on including geospatial activities within the Semantic Web framework. Though there have been a few research works benefiting from the semantic technologies in geospatial communities, they mainly focused on data interoperability or data integration between different data sources. ArchaeoKM also roots itself in this same framework of achieving data integration among heterogeneous data sources to start with. However, it builds up to facilitate archaeologists to formulate domain rules to discover knowledge from the facts and hypothesis upon which these rules are built in. In this process, the paper highlighted spatial extension for SWRL, a rule language within the framework, to facilitate spatial inferences through rules during the process. The extension constitutes spatial built-ins which could be combined with other built-ins within the formulated rules. Likewise, it also presented spatial filters for the query language of SPARQL which were echoed through OGC standards for GeoSPARQL. The work was followed by the interdisciplinary project of HiGeoMes. The project requires the usage of semantics to suggest the relative geography of the places mentioned in epigraphic data through the absolute geography of the archaeological sites. It is a fitting example to demonstrate how the limitations of one datasets could be overcome by other datasets through semantically integrating those datasets.

We intend to continue our efforts on convergence of these two technologies. The benefit of semantics is widely felt in the geospatial community and the latest efforts are visible. We follow this trend and continue providing our inputs on these global efforts through our upcoming projects like the next edition of HiGeMes. The next edition of HiGeoMes not only focuses on data interoperability through semantics in OGC compliance standards but also concentrates itself in spatial knowledge discovery. This will bring the much needed reinforcement in this next dimension of knowledge management techniques where semantics could be implied to discover hidden knowledge from the existing facts and hypothesis. We have already demonstrated this in relatively small cases within ArchaeoKM, but we want to re-emphasis it with the next edition of HiGeoMes.

7 References

Andrienko, G. L. & Andrienko, N. V.: Data Mining with C4.5 and Interactive Cartographic Visualization. User Interfaces to data intensive systems, pp. 162 – 165, 1999.

Baader, F., Horrocks, I. & Sattler, U: Description Logics as Ontology Languages for the Semantic Web, 2003.

Baader, F. & Sattler, U.: An overview of tableau algorithms for description logics. Scotland, LNAI, 2000.

Bateson, G.: Mind and Nature: A Necessary Unity (Advances in Systems Theory, Complexity, and the Human Sciences). s.l.:Hampton Press, 1979.

Bellinger, G.: Knowledge Management — Emerging Perspectives, http://www.systems-thinking.org/kmgmt/kmgmt.htm, 2004, last access 13/12/ 2010.

Berners-Lee, T., Hendler, J. & Lassila, O.: The Semantic Web. Scientific American, May, pp. 34-43, 2001.

Calvanese, D., Giacomo, G. D., Lenzerini, M. & Nardi, D.: Reasoning in expressive description logics. In: A. Robinson & A. Voronkov, eds. Handbook of Automated Reasoning. s.l.:Elsevier Science Publishers, pp. 1581-1634, 2001.

Cruz, I. F.: Geospatial Data Integration. ADVIS Lab, Department of Computer Science, University of Illinois, Chicago, 2004.

Cruz, I. F., Sunna, W., & Chaudhry, A.: Semi-automatic ontology alignment for geospatial data integration. In: Geographic Information Science (pp. 51-66). Springer Berlin Heidelberg, 2004.

Decker, S. et al: The semantic web: the roles of XML and RDF. IEEE Internet Computing , pp. 63-74, 2000.

Gruber, T. R.: A Translation Approach to Portable Ontology Specifications. Knowledge Acquisition, 5(2), pp. 199 – 220: 1999.

Guo, D. & Mennis, J: Spatial data mining and geographic knowledge discovery — An introduction. Computers, Environment and Urban Systems, pp. 403 – 408, 2009.

Haarslev, V. & Muller, R.: RACER system description. Automated Reasoning, pp. 701 – 705, 2001.

Houbie, F., Duchesne, P., & Maué, P.: Semantic annotations in OGC standards. OGC, 2012.

Hwang, S.: Using Formal Ontology for Integrated Spatial Data Mining. Assissi, Italy, Lecture Notes in Computer Science, Springer, pp. 1026 – 1035, 2004.

Janowicz, K., Schade, S., Bröring, A., Keßler, C., Maué, P., & Stasch, C: Semantic enablement for spatial data infrastructures. In: Transactions in GIS, 14(2), 111-129, 2010.

Jones, C.: Geographical Information Systems and Computer Cartography. s.l.:Addison Wesley Longman Ltd, 1997.

Kalyanpur, Parsia, Sirin, Grau, & Hendler: SWOOP: A web ontology editing browser. Web Semantics: Science, Services and Agents of World Wide Web, pp. 144 – 153, 2006.

Karmacharya, A: Introduction to a spatial layer within the Semantic Web framework: A proposition through web application ArchaeoKM, Mainz: FH-MAINZ, 2011.

Karmacharya, A., Cruz, C. & Boochs, F: Spatialization of the Semantic Web. In: Semantics-Advances in Theories and Mathematical Models. s.l.:Intech, 2011.

Karmacharya, A., Cruz, C., Boochs, F. & Marzani, F.: ArchaeoKM: Realizing Knowledge of the Archaeologists. Cotbus, Conference Publication, pp. 102 – 108, 2010.

Motik, Grau, Horrocks, Wu, Fokou, & Lutz: OWL 2 Web Ontology Language Profiles, http://www.w3.org/TR/2009/REC-owl2-profiles-20091027/, 2009, last accessed 28/10/ 2010.

Nguyen, D. T. & Nguyen, T. P.-M.: Pellet reasoner based evaluation of OVL (Ontology for Vietnamese Language). s.l., Conference Publications, pp. 180 – 184, 2010.

Perry, M. & Herring, J.: OGC GeoSPARQL - A Geographic Query Language for RDF Data, s.l.: Open Geospatial Consortium, 2011.

Perry, M. & Herring, J.: OGC GeoSPARQL - A Geographic Query Language for RDF Data, s.l.: OGC, 2012.

Protégé, n.d. Protégé.: http://protege.stanford.edu last accessed 26/09/2012].

Roth, M., Bröring, A. & Maué, P: Converting SDI Contents to Linked Data with Geographic Feature Pipes. Geoinformatik 2013, Heidelberg, 2013.

Shearer, R., Motik, B. & Horrocks, I.: HermiT: A highly-efficient OWL reasoner. Karlsruhe, Conference Publication, 2008.

Shekhar, S. & Huang, Y.: Discovering Spatial Co-location Patterns: A Summary of Results. Redondo Beach, Springer, pp. 236 – 256, 2001.

Stojanovi , N. & Handschuh, S.: A Framework for Knowledge Management on the Semantic Web. Hawaii, 2002.

Tanasescu, V., Gugliotta, A., Domingue, J., Gutiérrez-villarías, L., Davis, R., Rowlatt, M., et al.: A Semantic Web Services GIS based Emergency Management Application . The 5th International Semantic Web Conference (ISWC06). Athens, Georgia: ISWC06, 2006.

Tsarkov, D. & Horroks, I.:. FaCT++ description logic reasoner: System description. s.l., Springer-Verlag, pp. 292 – 297, 2006.

Wüst, T., Nebiker, S. & Landolt, R.: Applying the 3D GIS DILAS to Archaeology and Cultural Heritage Projects Requirements and First Results. In: International Archives of Photogrammetry Remote Sensing and Spatial Information Sciences. s.l.:Natural Resources Canada, pp. 407 – 412, 2004.

Zack, M. H.: Managing Codified Knowledge. Sloan Management Review, 40(4), pp. 45 – 58, 1999.