HiGeoMes: Distributed Geodatabases in an archaeological joint research project

Frank Boochs[1], Kai-Christian Bruhn[1], Christophe Cruz[2], Ashish Karmacharya[1], Tobias Kohr[1]

[1] Institut i3mainz, am Fachbereich 1 - Geoinformatik und Vermessung
Fachhochschule Mainz, Lucy-Hillebrand-Str 2, 55128 Mainz
{boochs, bruhn, ashish, kohr}@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 project 'HiGeoMes' seeks to connect knowledge about archaeological sites in the Near East with place names in ancient sources of the 2nd millennium BC. The paper describes the digital approach to integrate textual and geographical information by emphasizing the impact of interfaces linking diverse datasets. Several OGC-based Web services provide reviewed archaeological information in a Web GIS with query capabilities. Place-related philological data are stored in an OWL-store to model the topological relations of unidentified sites. An interface integrates a graph-viewer with the Web GIS by invoking both WFS- and REST-services to spatially integrate the datasets. Furthermore, a full integration on the semantic level and emerging potentials are discussed.

KEYWORDS: Archaeological Data Infrastructure, Web GIS, Semantic Integration, Knowledge Management

1   Introduction

The integration of places mentioned in Babylonian and Assyrian texts with the location of known archaeological sites in the Near East is the focus of 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. The project's main goal is to interconnect 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.

Current research in the historical geography of Ancient Mesopotamia in the 2nd millennium BC is largely based on the "Répertoire Géographique des Textes Cunéiformes" (Groneberg 1980), (Nashef 1982), accompanied by the maps of the "Tübinger Atlas zum Vorderen Orient (TAVO)". However, since this research was documented, the number of old Babylonian texts has more than doubled, along with a considerable increase in both quantity and quality of Middle-Assyrian sources. A prevailing compilation of current knowledge also incorporating the archaeological record was published in 2009 (Cancik-Kirschbaum and Ziegler 2009).

Nevertheless, each of the scientific disciplines involved challenges complex methodological questions crucial for interpreting both the philological and the archaeological material. A thorough understanding of the historical geography relies on a combined interdisciplinary analysis of the record, each containing

uncertainties for authoritative conclusions to be appropriately communicated between researchers affiliated at different institutions.

Considering this background, a digital information system is suitable for enabling this knowledge exchange only if it facilitates certain main specifications.

It has to handle the philological and the archaeological record appropriately with a data-model that not only accounts for representing relevant knowledge of each research domain but also allows for sharing and linking the available information. Two aspects inherent in the data of both domains could serve as nodes for this interconnection: space and time. As the Old Babylonian and Middle Assyrian texts already represent a chronological divide of the data sources, space remains as a viable candidate.

Archaeology comprises a vast amount of heterogeneous datasets with relevant absolute geographic references. Standards and technologies developed within international spatial data infrastructure (SDI) initiatives are available to achieve this objective. Although the European SDI initiative (INSPIRE) targets environmental data, previous research (McKeague et al. 2012; Fernández Freire et al. 2012) has demonstrated that the approach can also be applied for culturally protected sites.

In contrast, philological data derived from epigraphic sources commonly reference spatial features as toponyms, which precludes proper geographic modelling. At the same time, textual data contain complex interrelations among themselves. However, knowledge available in the context of this project is usually related to real objects and their environments. These descriptions are of interest to understand the interactions and interrelations of these objects. Consequently, ontologies provide an adequate tool to formally define this knowledge and to make it accessible to either people or to information systems (Grimm et al. n.d.). This knowledge is modelled in a graph data structure which differs from a standard SDI data structure. Since the SDI approach is not adequate for textual data, the integration of an SDI and textual knowledge is of special interest for the HiGeoMes project.

The objectives of the project are threefold. The first one aims at defining interoperable Web services on the archaeological HiGeoMes database. The second one consists of making available the knowledge from the epigraphic ArchiBab (Charpin 2013) database, which is currently not exploitable with a query language. The last one focuses on the ability of the system to integrate the archaeological database and the ArchiBab ontology into one interface. This integration allows us to give a higher degree of interpretation of data and to visualise integrated data.

This paper is composed of 5 sections. After the background of the key issues is illustrated, section 2 presents the Archaeological Data Infrastructure of HiGeoMes. Chapter 3 describes the semantic approach towards epigraphic data. In part 4, the integration of both approaches is discussed, before we conclude with a resume and outlook of the project.

Background

Corns and Shaw (2010) give a broad overview on the suitability of the SDI concept for the syntactically interoperable dissemination of archaeological data. The implementation of interoperability within an SDI,

however, also presents several semantic challenges (Janowicz et al. 2010). Most semantic problems arise due to the lack of suitable content descriptions. To design an SDI with interoperability capabilities, it is necessary to add proper semantics to data, metadata and services.

Currently, most of the available information on the Web is provided as syntactic blocks, which require human experts to determine their semantics. The needed semantics can be provided by ontologies, allowing for reasoning by software inference mechanisms. In the Artificial Intelligence field, an ontology is an engineering mechanism, constituted by a vocabulary and a set of assumptions regarding this vocabulary. This information is specified using a first order logic theory. Using an ontology, it is possible to specify a concept, to define a number of restrictions that apply to a certain concept and create a hierarchy of concepts with subsumption relationships among them (Guarino 1998).

Knowledge formalization and knowledge discovery have previously been used in archaeology. For instance, previous work has integrated heuristics with knowledge discovery tools within an intelligent search, and combined them with a genetic algorithm framework (Lazar and Raynold 2001). Data was used to build models of the real world. Tools were tested on several large-scale archaeological data sets generated from an archaeological survey of the Valley of Oaxaca in Highland Mesoamerica. The "Archaeotools" project (Jaffrey et al. 2009) integrates huge sets of structured database records of archaeological sites with information extracted from literature reports and other unstructured sources. It advocates the benefits of well-defined ontologies and thesauri. In another project, researchers successfully integrated spatial inference through user defined spatial rules to classify unclassified excavated objects in industrial archaeology (Karmacharya et al. 2010).

2   Archaeological Data Infrastructure

The Archaeological Data Infrastructure (ADI) of HiGeoMes is built upon existent standards and technologies of an SDI. It leverages OGC Web services to facilitate the exchange of collected archaeological information among project partners and to ensure external researchers syntactically interoperable access to it.

2.1   Technical Architecture

The architecture follows the principle of a Service Oriented Architecture (SOA). Components are loosely coupled, self-describing, stateless and standardised, ensuring a high degree of independence and reusability. Data dissemination is based on the "publish-find-bind" paradigm, in which a provider publishes resource metadata in a catalogue, a consumer finds this metadata, and then binds the resource in a final step.

[Bruhn_HiGeoMesGeodatabases_fig1.jpg]

Fig. 1: Architectural sketch of the Archaeological Data Infrastructure

Technically the architecture consists of three tiers: The data level stores vector-geometries of archaeological sites and raster images, Web services publish the data and a Web GIS client acts as an

entry point for visualisation and simple analysis. An OGC-compliant Web Feature Service (WFS) enables direct access on data, while Web Map Services (WMS) perform visualisation tasks. Web Map Tile Services (WMTS) provide a standardised method to cache generated graphics for increased performance. The application harnesses a major asset of the SOA concept by integrating Web services from other providers. Google and OpenStreetMap serve cartographic layers and Geonames provides a Gazetteer. All components are implemented with open source products, and will be presented in more detail in this chapter.

2.2 Archaeological Data within HiGeoMes

HiGeoMes examines the 2nd millennium BC in Upper Mesopotamia. The archaeological database that is maintained by the project partners at the Universität Mainz comprises of 706 sites. One quarter of the sites indicate a high confidence level based on the excavation reports. Information on other sites is based on around 30 surveys that are of lower certainty.

The project orientates the capture of archaeological data on three basic requirements: sites need to be locatable, datable and possess a bibliographic reference to facilitate review and further research. Both digital and analogue sources are investigated in this regard. Direct integration of digital sources has remained impossible with relevant sources, because interfaces that would allow access to them are still widely missing. Even reuse of digital material is usually restricted to different thematic or technological constraints: some sources do not fit into the geographical, e.g. MEGA Jordan (Myers and Dagity 2012) or temporal scope, e.g. Pleiades (Simon et al. 2012), and other sources only comprise of the most prominent sites. Technically, some sources do not provide better usability than analogue material (TAY Project 1998). Nevertheless, data needs to be derived from such sources because of their thematic quality. The most widely used digital source for locating sites is the ANE Placemarks (Pedersén 2012). The geometrically accurate locations are adopted for nearly half of the HiGeoMes sites, whereas no other information is included in the data.

Because of the restricted availability of useful digital material, conventional sources such as excavation reports, survey reports and compilations remain a major base for archaeological research within HiGeoMes. For instance, the "Atlas of Pre-Classical Upper Mesopotamia (APUM)" served as a starting point for the archaeological database with approximately 50 sites (Anastasio et al. 2004). Other important analogue sources include "Bibliographie der archäologischen Fundstellen und Surveys in Syrien und Libanon" (Lehmann 2002), "Tübinger Atlas des Vorderen Orients (TAVO)" (Hartmann 1980), "Répertoire Géographique Des Textes Cuneiformes", (Zadok and Röllig 1985), and "The Imperial Landscape of Ashur"(Altaweel 2008).

The archaeological data can be modelled geographically. This entails the capacity to visualise and analyse the information in a spatial context. In fact, correct localisation presents the most important component of archaeological data. HiGeoMes keeps the geographic model and the archaeological data separately, linked through a common HiGeoMes ID; archaeological sites are stored in a geo-dataset with basic information, whereas comprehensive archaeological data are stored in an extensive database. This modular approach implies the advantage that the light geo-dataset can be used without the complex database involved. At the same time the data can be regarded in its entirety.

Currently, the geo-dataset forms part of the archaeological data infrastructure, where it is implemented in PostgreSQL 9.1 (PostGIS 1.5.3). The locations are captured with the help of Quantum GIS and Google Earth. Capture of archaeological data is conducted through a standalone desktop application. Based on the existing reference, an integration of the database within the ADI is possible in the future. The archaeologist can decide which data should be disseminated by the infrastructure and which data should not. Additionally, different access rights can be granted in the service layer (2.3).

The resulting HiGeoMes data presents a large collection of geo-referenced sites attached with detailed archaeological information. Collecting and investigating conventional sources, however, is a very time consuming task. Moreover, this work is done repeatedly, because many sources already consist of derived information. A large friction loss occurs in the simple process of reusing existing information. Digital information technologies possess the capacity to reduce this loss and invest more time to gain new knowledge. Nevertheless, the lessons learned in HiGeoMes show that an efficient use of digital technologies for the dissemination of archaeological data is still rare. The next chapter discusses the approach HiGeoMes takes towards this challenge.

2.3   Dissemination via Geo Web services

Web services present syntactically interoperable interfaces to provide data and functionality on the Web. The OGC explored this approach relatively early (1997) to develop such standards for spatial data (Zhao and Di 2011). During the next decade, SDI initiatives evolved on all administrative levels with the common aim to improve dissemination of spatial data. Based on the SOA concept, these initiatives act as a major driving force to further push OGC efforts. While the European initiative (INSPIRE) originally focuses on environmental data, today many other domains have recognized the capabilities of OGC Web services and adopted them (http://www.opengeospatial.org/domain). In archaeology, the concept is not only convenient because archaeological data usually possess spatial references, but also for the particular capabilities the service abstraction layer implicates; metadata allow for the description of the data's complex characteristics and to increase their traceability. Likewise, it is possible to protect potentially sensitive data via control mechanisms or algorithms for intentional falsification. HiGeoMes demonstrates the advantages the approach contributes and discusses the shortcomings it still implies.

Archaeological sites collected at the University of Mainz are disseminated via a Web Feature Service (WFS), allowing standardized access to the data in Geography Markup Language (GML). This enables other components to consume the data and conduct analysis, such as search queries executed by the Web GIS client (2.4). While the WFS currently only exposes the geo-dataset with basic information, it can be extended to provide more specific data of the archaeological database.

Web Map Services (WMS) portray raster data sets of interest: Soviet ordnance maps (Wikipedia, s.v. "Sowjetische   Generalstabskarte,"   last   modified   June   8,   2013, http://de.wikipedia.org/wiki/Sowjetische_Generalstabskarte)   were   processed   to   display   relevant information such as water fountains, Corona satellite imagery (Scardozzi 2008; Fowler 2012) to show the region before human impact expanded to a broader scale (e.g. dam constructions), and a layer generated from SRTM data serves as a visual base map. As opposed to WFS, WMS do not serve the actual data, but

only a visualization of them (raster and vector). The restrained access makes WMS very popular among data providers who do not want to uncover their data entirely, but are able to serve a generated map in an interoperable way. WMS render Web capable image formats such as PNG or JPEG, which present lighter formats than the original raster data (often TIFF, etc.). The generating process takes time, however, potentially resulting in poor usability on the client side. This time can be eliminated by (pre-)caching the generated WMS output in form of rendered tiles that can be served directly upon request. While the approach might entail drawbacks for dynamic data, it is ideal to provide fast static map services (Weskamm 2011). The emerged standard of a Web Map Tile Service (WMTS) is used for all raster data within HiGeoMes.

Currently HiGeoMes offers four OGC-conform Web services that are implemented with MapServer (Vatsavai et al. 2006) and MapCache:

- WFS with archaeological sites
- WMS/WMTS with Russian maps
- WMS/WMTS with Corona images
- WMS/WMTS with SRTM base map

These Web services are invoked by the Web GIS client for data delivery, as they can be accessed by any GIS tool or Web application that supports OGC standards. HiGeoMes integrates the Google Maps API (Miller 2006), OpenStreetMap (Neis et al. 2010) and the Geonames gazetteer (Geonames 2013) from external interfaces. More specialized data like the Corona satellite images and Russian maps need to be downloaded, processed and set up in proper Web services before they can be used in a Web infrastructure. It should be obvious that the more initiatives provide their data over interoperable interfaces, the more information can be combined without additional processing efforts.

2.4   Web GIS client

The Web GIS serves as a first entry point into the ADI of HiGeoMes. While the data services can be consumed by any OGC-compliant GIS client, the Web GIS creates a graphical user interface (GUI) to the infrastructure for a specific use case: it provides a research tool for archaeological researchers to explore, visualise and analyse HiGeoMes data. This use case requires the application to be directed towards domain experts, who are also technically untrained. A Web GIS is suitable to fulfil these needs, because functional access can be restricted. For future developments, this implies the possibility to identify further target groups with different use cases and provide diverse customized views on the data and capabilities.

[Bruhn_HiGeoMesGeodatabases_fig2.jpg]

Fig. 2: Buffer selection of archaeological sites in the Web GIS client

The GUI is arranged into four regions, with the centre region displaying all data of the services within a map. A tool bar enables navigation, basic GIS functionality and querying for locations by their HiGeoMes ID or modern name. The latter tool invokes the Geonames gazetteer service (Geonames 2013). On the left, the data service layers can be (de-)activated in a tree structure. The bottom region supplies a sortable view on attribute data of the archaeological sites. A synchronized selection tool highlights rows in the table with corresponding locations on the map simultaneously. The right region offers further functionalities that can

be applied on the archaeological sites. These tools can search locations by modern or prehistoric names, list and highlight sites within a specified perimeter of a selected site or measure and highlight direct distances between sites.

Technically the client is completely implemented in HTML, CSS and JavaScript, making it portable to PCs with a common browser. The ExtJS library helps to build a GUI and arrange the different components within it (Orchard et al. 2009). OpenLayers brings geospatial capabilities to the application (Jansen and Adams 2010). Built upon these libraries, GeoExt facilitates the development effort offering capabilities such as measuring tools and a geonames search out of the box.

An internal survey among project partners showed that the majority perceive the Web GIS as easy to use and plan its regular use in the future. While searching and measuring capabilities were popular among participants, the combined geographic (map) and thematic (table) views were their favourite functionality. Besides the archaeological sites, the Google satellite imagery was very popular. Additionally, more specific layers, such as the Corona data, seemed useful.

3   Semantic approach towards epigraphic sources

The emergence of the Semantic Web in early 2000 has proposed inclusion of semantics for better machine-humans collaboration for handling huge amount of information that exists on the Web (Berners-Lee et al. 2001). The intention was to give information a well-defined meaning in order to make it understandable to machines. In the Semantic Web, emphasis is given to the meanings of words rather than to words themselves so that the information provided by them could be used as knowledge formalized through semantic interpretations.

The Semantic Web provides a common framework that allows data to be shared and reused across applications, enterprises and community boundaries (Decker et al. 2000). At the core of the framework lies knowledge and knowledge representation tools such as ontologies. Conventionally, ontologies have been used as tools to represent knowledge and the Semantic Web framework maintains it. Web Ontology Language (OWL) is that standard language recommended by W3C to model and represent knowledge in the Semantic Web framework.

The missing geographic link within epigraphic data is one of the key factors in achieving interoperability with other data. The vast amount of information in the dataset and the complexity in information patterns limits the dataset to be used through a conventional database system. Ontologies can express complex information through their axioms and theorems. In addition, expressing information logically through ontology helps machines understand information and thus can act upon it. This is vital for managing vast and complex datasets as in the epigraphic database.

3.1   ArchiBab

ArchiBab is an online service that allows users to query a database that contains information of Babylonian and Middle-Asyrian texts (Charpin 2013). It contains translations of clay tablets written in cuneiform. The textual ArchiBab database is regularly updated and it is close to be completed. Currently

the database contains more than 31,000 texts corresponding to paleo-babylonian archives. ArchiBab allows conventional queries using keywords. It also allows temporal queries, by using keywords referring to dates, for example "7/vii during the reign of Zimri-Lim". The database also contains a textual description for identified keywords in the Babylonian texts. Possible keywords are names of people, toponyms, kings, objects, etc. Thanks to ArchiBab, users can have access to textually modelled information. However, ArchiBab is unable to provide an overview of the knowledge contained in the texts. One of our main objectives is to extract the knowledge stored in the texts and develop tools that would link it to graphic representations of the Babylonian routes. An important limitation of ArchiBab is that because the information is stored as texts, it is not possible to use a standard query language to extract knowledge from it. We propose the use of constraints and rules to extract knowledge from ArchiBab texts, while focusing on toponyms and their environment. Some of the ancient toponyms have a modern equivalent in the archaeological database. Ancient routes used by caravans, merchants or troops are described in clay tablets. However, through using only textual information it is difficult to obtain a spatial representation of the routes. To accomplish this goal, we use the extracted knowledge referred to toponyms. Two processes were defined to visualize ArchiBab data. The first one consists of extracting all the knowledge that can be extracted from the ArchiBab database. Thus, an OWL file is generated from this process, which allows the computation of the ontology consistency, and to query the knowledge base using a standard query language such as SPARQL. Once the knowledge is queryable, the second process consists of generating data for its visualization, and more specifically the route map. The result of this process can be a JSON file, which can be easily integrated to geospatial data.

[Bruhn_HiGeoMesGeodatabases_fig3.jpg]

Fig. 3: Presentation of the two main processes, extraction of knowledge to an OWL file, production of a Web interface to visualize the routes, and the toponyms and the environment.

3.2   Graph visualization

Visual Data Exploration eases the search of relevant data pieces in large volume datasets. It has always been important to include humans in the data exploration process (Keim 2002). Humans allow their perceptions to explore into the data to formulate hypotheses and conclusions when it could be explored visually. The Information seeking mantra (Shneiderman 1996) discusses three steps in "Visual Data Exploration": overview, zoom and filter. We implement this mantra for exploring data in the ArchiBab database.

[Bruhn_HiGeoMesGeodatabases_fig4.jpg]

Fig. 4: Graph visualization of the routes

ArchiBab is a large database with its own hidden graph to represent implicit relationships among data pieces. The graph consists of a set of objects extracted from the database which are called nodes. The connections or relationships between these objects are called edges. An example of such a graph can be a

network defining the interrelation between toponyms and their routes. Such interrelations define domain knowledge. In this project, we use an OWL ontology and use graph visualization to explore the interconnections and to define its associated semantics, which also model the knowledge. We use a Javascript Infovis Toolkit (Belmonte 2011) to generate the graph and the corresponding functionalities. Graphs are extracted from the ontology and rendered with the help of the Javascript Infovis Toolkit. Figure 4 shows the map of routes which are deduced from the relationships defined in the ontology. On the right of the figure, a file describes the environment of a selected station of the route. In addition, a route can be selected by a select box on the left of the graphical interface. The main intension is to provide broader picture of the semantic connections between different toponyms within the Archibab database.

4   Linking places and sites

Toponyms in epigraphic data are not geo-localized. This makes an SDI impossible for the distribution of such data, unless a mechanism is engineered to geo-localize the toponym sites. One of the most sought after mechanisms to provide spatial signatures to a spatial 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 in a relational database and epigraphic data in OWL each have their own structure. 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. 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 such as an SDI are in need of semantic enablement or a semantic extension to accomplish this task (Janowicz et al. 2010). With this mantra, an architecture is designed to use semantics for data integration (as seen in figure 5).

[Bruhn_HiGeoMesGeodatabases_fig5.jpg]

Fig. 5: System Architecture

The HiGeoMes architecture (see fig. 2) illustrates a mechanism that integrates the archaeological data shared through the 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 geolocalisation 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 modelled information is possible. However, 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).

[Bruhn_HiGeoMesGeodatabases_fig6.jpg]

Fig. 6: Graph Visualization describing routes of the queried toponym within WebGIS client

5   Conclusion & Outlook

The paper motivates the use of available SDI technologies for the dissemination of archaeological data and for providing digital research tools for scientists in the humanities. It presents the strengths of the concept for spatially referenced data and reflects on the drawbacks it still implies. A semantic approach is presented for modelling and querying epigraphic data without spatial reference. Motivated by the demand from humanities scholars, the work presents an integration of the two different approaches. The outcome has received high acceptance among archaeologists and philologists, who maintain two different views on their data and are able to explore them in a geographical and in a topological context respectively. The mode of information exchange within the research group is ready to be expanded to other scholars or for general public use, because the gateway and data interfaces follow standardised, mature and established protocols. At the end of the project the full list of sites will be openly available comprising a subset of the attributes via WFS and derivative file formats. The raster data will be moved into a more sustainable server environment and can be consumed in any GIS-environment capable of communicating with OGC-compliant WMS. Archaeological as well as philological research in the project area will have a common, citable and quality assured set of data on disposal on which further investigations with similar or different scope can rest upon.

From the information-technology perspective we regard the current solution as a proof of concept that indicates further potential. The integration implies several shortcomings, however, that will be addressed in the future. Currently the connection between the different sources is only established on a visual level. For the next step, we need to achieve integration on the semantic level (Fig. 5). Therefore, it will be important to further investigate semantic interfaces within the data sources. In this context, the involvement of OGC Web services such as the WFS will be of particular interest. Data generated during the HiGeoMes project is of value to the archaeologists and researchers working in Upper Mesopotamia and will be made available under a suitable free licensing scheme.

Such epigraphic and archaeological databases consist of numerous unanswered issues that need to be dealt with in future research. One issue is sharing a common name by multiple toponyms and multiple locations. Another issue are the missing links between toponyms in epigraphic datasets and modern day locations. This makes it impossible to geo-localize these toponyms. Every philologist and archaeologist has his or her own method to answer these unanswered issues. This expert knowledge needs to be documented and analysed in a future system. One must take into account that these unanswered issues are not possible to resolve independently through separate independent datasets. The issues within epigraphic data can only be solved if other datasets (like archaeological data in this case) are taken into account, for example. A ready to integrate prototype application will be developed for archaeologists and philologists to enter their expert knowledge through rules within the system. These rules combine different knowledge sources for reasoning. This way they will infer semantics within the archaeological dataset and ArchiBab dataset in order to provide answers to unanswered issues. The prototype application will use inference capabilities of the knowledge technology with the Semantic Web framework to infer expert knowledge through their hypothesis and facts for such answers.

The semantic enabled application for collaborative management of archaeological and philological datasets through expert knowledge possesses huge potential. Besides, integrating different datasets semantically to provide answers through knowledge discovery, the application has potential to automate the identifications of toponyms described with their relative geography by matching them to their corresponding absolute geography. This concept of geo-localizing toponyms through spatial mapping with existing spatial dataset can be emulated on other similar projects.

6   Acknowledgements

7   References

Altaweel, M. *The Imperial Landscape of Ashur: Settlement and Land Use in the Assyrian Heartland.* 2008.

Anastasio, S, M Lebeau, and M Sauvage. *Atlas of Preclassical Upper Mesopotamia.* Brepols, 2004.

Belmonte, N G. *JavaScript InfoVis Toolkit.* 2011. http://thejit.org (accessed June 13, 2013).

Berners-Lee, T, J Hendler, and O Lassila. "The Semantic Web." *Scientific AmericaN*, May 2001: 34-43.

Cancik-Kirschbaum, E, and N Ziegler. *Entre les fleuves - I; Untersuchungen zur historischen Geographie Obermesopotamiens im 2. Jahrtausend v. Chr.* PeWe-Verlag, 2009.

Charpin, D. *ArchiBab.* 2013. http://www.archibab.fr/en (accessed June 2013, 17).

Corns, A J, and R Shaw. "Cultural Heritage Spatial Data Infrastructures (SDI)-Unlocking the Potential of our Cultural Landscape Data." *In Proceeding of 30th EARSeL Symposium.* Conference Publication, 2010. 1 - 8.

Cruz, I F. "Geospatial Data Integration." *ADVIS Lab, Department of Computer Science, University of Illinois, Chicago* (ADVIS Lab, Department of Computer Science, University of Illinois), 2004.

Cruz, I F, W Sunna, and A Chaudhry. "Semi-Automatic Ontology Alignment for Geospatial Data Integration." *In Proceeding of GIScience.* Maryland: LNCS, 2004. 51 - 66.

Decker, S, S Melnik, F van Harmelsen, D Fensel, M Klein, J Broekstra, M Erdmann, I Horrocks. "The semantic web: The roles of XML and RDF." *IEEE Internet Computing*, 2000: 63 - 74.

Fernández Freire, C, I Del-Bosque, E Pérez-Asensio, A Uriarte, J Vicent, P Fábrega, and Parcero "Proposal for a Cultural Heritage Application Schema within the INSPIRE Directive." *in Proceeding of AGILE2012.* Avignon: Conference Publication, 2012.

Fowler, M J. "A review of the CORONA Atlas of the Middle East (Beta)." *Aerial Archaeology Research Group News (AARGnews)* (Aerial Archaeology Research Group News (AARGnews)), 2012: 24-29.

Geonames. *Geonames.* 2013. http://www.geonames.org (accessed June 17, 2013).

Grimm, S, P Hitzler, and A Abecker. "Knowledge Representation and Ontologies -- Logic, Ontologies and Semantic Web Languages ." n.d.

Groneberg, B. "Die Orts- und Gewässernamen der altbabylonischen Zeit." *unter Benutzung der Sammlungen von J.-R Kupper und W.F. Seemans und unter Mitarbeit von M. Stol bearbeitet.* Wiesbaden: L. Reichert, 1980.

Guarino, N. "Formal ontology and information systems." *In Proceeding of FOIS 98.* Conference Publication, 1998. 3 - 15.

Hartmann, K P. "Tubinger Atlas des Vorderen Orients." In *Untersuchungen zur Sozialgeographie christlicher Minderheiten im Vorderen Orient (.* 1980.

Jaffrey, S, J Richards, F Ciravegna, S Waller, S Chapman, and Z Zhang. "The Archaeotools project: faceted classification and natural language processing in an archaeological context." *Philosophical transaction of Royal society: A mathematical, Physical and Engineering Science* , 2009: 2507 - 2519.

Janowicz, K, S Schade, A Bröring, C Keßler, P Maué, and C Stasch. "Semantic enablement for spatial data infrastructures." *Transactions in GIS*, 2010: 111 - 129.

Jansen, M, and T Adams. *OpenLayers: Webentwicklung mit dynamischen Karten und Geodaten.* Open Source Press, 2010.

Karmacharya, A, C Cruz, F Boochs, and F Marzani. "Integration of Spatial technologies and Semantic Web Technologies for Industrial Archaeology." *In Proceeding of WEBIST.* Conference Publication, 2010. 75 - 80.

Keim, D A. "Information visualization and visual data mining." *IEEE TRANSACTIONS ON VISUALIZATION AND COMPUTER GRAPHICS* 8, no. 1 (2002): 1 - 8.

Lazar, A, and R G Raynold. "Heuristic Knowledge Discovery for Archaeological Data Using Genetic Algorithms and Rough Sets." In *Heuristic and Optimization for Knowledge Discovery*, by A Hussein, C S Newton and R Sarkar, 263 - 278. Idea Group Pub, 2001.

Lehmann, G. *Bibliographie der archäologischen Fundstellen und Surveys in Syrien und Libanon.* Rahden/Westf: Leidorf Pub., 2002.

McKeague, P, A Corns, and R Shawa. "Developing a Spatial Data Infrastructure for Archaeological and Built Heritage." *International Journal of Spatial Data Infrastructures Research*, 2012: 38 - 65.

Miller, C C. "The Google Maps mashup as GIS/2." *Cartographica: The International Journal for Geographic Information and Geovisualization* 41, no. 3 (2006): 187 - 199.

Myers, D, and A Dagity. "The Middle Eastern Geodatabase for Antiquities (MEGA): An Open Source GIS-Based Heritage Site Inventory and Management System." *Change Over Time: An International Journal of Conservation and the Built Environment*, 2012: 32 - 57.

Nashef, K. *Die Orts- und Gewässernamen der mittelbabylonischen und mittelassyrischen Zeit.* Wiesbaden: L Reichert, 1982.

Neis, P, D Zielstra, A Zipf, and A Strunck. "Empirische Untersuchungen zur Datenqualität von OpenStreetMap-Erfahrungen aus zwei Jahren Betrieb mehrerer OSM-Online-Dienste." *In Proceeding of Angewandte Geoinformatik 2010.* Salzburg: Conference Publication, 2010.

Orchard, L M, A Pehlivanian, S Koon, and H Jones. *Professional JavaScript Frameworks: Prototype,YUI, ExtJS, Dojo and MooTools.* WROK, 2009.

Pedersén, O. "Ancient Near East on Google Earth: Problems Preliminary Results, and Prospects." *In Proceedings of the 7th International Congress on the Archaeology of the Ancient Near East.* London: Harrassowitz Verlag, 2012. 385 - 393.

Scardozzi, G. "Old high resolution satellite images for landscape archaeology: case studies from Turkey and Iraq." *In Proceeding of SPIE 7110, Remote Sensing for Environmental Monitoring, GIS Applications, and Geology.* Conference Publications, 2008.

Shneiderman, B. "The eyes have it: a task by data type taxonomy for information visualizations." *Proceedings of IEEE Symposium on Visual Languages* . Conference Publications , 1996. 336 - 343.

Simon, R, E Barker, and L Isaksen. "Exploring Pelagios: a visual browser for geo-tagged datasets." *In Proceeding of International Workshop on Supporting Users' Exploration of Digital Libraries.* Paphos, Cyprus: Conference Publication, 2012. 23 – 27.

Tanasescu, V, A Gugliotta, J Domingue, L Gutiérrez-Villarías, R Davies, M Rowlatt, M Richardson. "A Semantic Web Services GIS based Emergency Management Application." *The 5th International Semantic Web Conference (ISWC06).* Athens, Georgia: ISWC06, 2006.

TAY Project. *TAY: The Archaeological Settlements of Turkey.* August 3, 1998. http://tayproject.org/enghome.html (accessed June 17, 2013).

Vatsavai, R R, S Shekhar, T E Burk, and S Lime. "Umn-mapserver: A high-performance, interoperable, and open source web mapping and geo-spatial analysis system." *Geographic Information Science*, 2006: 400 - 417.

Weskamm, J. "Beschleunigung von WebMapServices mit Hilfe verschiedener Proxies – Wenn's mal wieder schnell gehen muss." *In Untersuchung der Nutzung von OpenStreetMap Daten zur Darstellung von TMC Verkehrsmeldeinformationen; FOSSGIS.* Heidelberg: Conference Publication, 2011.

Zadok, R, and W Röllig. *Répertoire géographique des textes cunéiformes.* Reichert, 1985.

Zhao, P, and L Di. *Geospatial Web Services: Advances in Information Interoperability.* Information Science Publishing, 2011.