UNIVERSITY OF SOUTHAMPTON

Faculty of Engineering and Applied Science

School of Electronics and Computer Science

ABSTRACT

The TRANSLATION Framework for Archaeological Excavation Data:
Transparent Negotiation and Sharing of Local Application Terminologies, Instances and Ontologies. A progress report submitted for continuation towards a PhD

by Leif Isaksen

The Semantic Web is rapidly maturing thanks to consolidation of its technologies and an incipient body of linked data available to the public. Nevertheless, there is a danger that we may throw the baby out with the bathwater. Whilst the utility of bridging separate conceptualizations of a domain through explicit specifications (ontologies) is clear, it does not follow that we want to abandon the original data models. The purpose of this research will be to investigate SW methodologies that loosely-couple local ontologies so that the semantic structures they embody are still accessible to the end-user for comparison and analysis.

The discipline of archaeology provides an excellent case-study in this regard. The fragmentary nature of its sources, and the diverse theoretical approaches of its practitioners render any attempt to establish a universal ‘world-view’ impossible. The result, whilst superficially homogenous, conceals conceptual rifts which may be of great significance. The need to express one’s work in terms of another’s ontology can also give rise to concerns of disenfranchisement that impede user adoption. A successful Semantic Web methodology should first permit contributors to describe their own data in their own (ontological) terms, and then provide the resources by which any user can create (or select) alignments to other Domain or Application ontologies.

This research aims to develop a process by which resource providers are able to publish their data to the Semantic Web in a manner that keeps its semantic origins as transparent as possible via an explicit ontology. Thereafter, alignment should be made possible in a ‘pluggable’ fashion so that alternative combinations of meaning can be explored. Its primary output will be the TRANSLATION Framework: a piece of modular open source software deployed as a case study to explore the benefits of Semantic technologies. This report is intended to provide both computer scientists and archaeologists with a plan of the work to be undertaken.

Contents

Introduction

Since bursting into the public consciousness in 2001 [1], the Semantic Web (SW) has come a long way, with development and consolidation of technologies, several large-scale international conferences annually and an incipient body of linked data available to the public. The possibilities of a machine-readable Web and semantic interoperability are inching closer to reality. Nonetheless, there is a danger, exemplified in many applications of SW technology, that we may throw the baby out with the bathwater. Whilst the utility of bridging separate conceptualizations of a domain through explicit specifications (ontologies) is clear, it does not follow that we want to abandon those original data models altogether. Indeed, various philosophers have argued that understanding concepts can only be done with reference to the framework of meaning in which they are embedded [2,3]. The purpose of this research will be to investigate SW methodologies that keep ontologies loosely-coupled so that the semantic structures they embody are still accessible to the end-user for comparison and analysis.

The discipline of archaeology provides an excellent case-study in this regard. The fragmentary nature of its sources, and the diverse theoretical approaches of its practitioners render any attempt to establish a universal ‘world-view’ impossible. Whilst high-level Domain Ontologies such as the CIDOC CRM may provide a semantic *lingua franca*, they also inevitably mask both the complexities and limitations of the lower-level (frequently implicit) Application Ontologies from which data is derived [4]. The result, whilst superficially homogenous, may conceal conceptual rifts which are of great significance. The need to express one’s work in terms of another’s ontology can also give rise to concerns of disenfranchisement that impede user adoption [5]. A successful Semantic Web methodology should first permit contributors to describe their own data in their own (ontological) terms, and then provide the resources by which any user can create (or select) alignments to other Domain or Application ontologies.

This research aims to develop a process by which resource providers are able to publish their data to the Semantic Web in a manner that keeps its semantic origins as clear as possible via an explicit ontology. Thereafter, alignment should be made possible in a ‘pluggable’ fashion so that alternative combinations of meaning can be explored. Its primary output will be the TRANSLATION[1] Framework: a piece of modular open source software deployed as a case study to explore the benefits of Semantic technologies.

This report is intended to provide both computer scientists and archaeologists with a plan of the research to be undertaken. Section I reviews the literature on the use of SW technology in Cultural Heritage and archaeology, and finishes with a discussion of relevant architectural decisions. Section II gives an overview of the planned research, including the case-study. Section III describes a provisional design for the software and a provisional design for the software. Section IV outlines a timetable for future development. A glossary of technical terms is provided at the back.

---

[1] TRAnsparent Negotiation and Sharing of Local Application Terminologies, Instances and ONtologies

Section I: Literature Review

The Semantic Web

The idea of a 'web of data', in contrast to a 'web of documents', goes back almost to the earliest days of the World Wide Web [6]. The ability to aggregate information across multiple repositories without the danger of conceptual conflation would hugely improve our ability to make meaningful queries that are not determined by arbitrary data structures and boundaries. Although a vast repository of human-readable content is available online, the well known challenges associated with Natural Language Processing and implicit semantics render it largely useless for machine inference. By being explicit about the semantic content of data 'fragments' however, the potential is opened up for automated (or semi-automated) integration and cross-querying, thereby increasing the informatory power of the internet by an order of magnitude.

Research over the last decade has meant that the technologies by which to formalize these semantics, in particular Universal Resource Identifiers (URIs), Resource Description Framework (RDF) along with its SPARQL query language, and the Web Ontology Language (OWL)[7-10], are now well established. High-level services which enable data suppliers or third parties to disseminate, manipulate or 'mash-up' information using these techniques are increasingly available (for example, [11]). This bundle of concepts and technologies is frequently referred to as the 'Semantic Web', but that phrase has come now come to mean both

A. A body of semantically interlinked data hosted across the internet

B. The set of formal and conceptual technologies used to create and embody that data

This is an important distinction because many early successes, most notably in e-Science and medicine (for example, [12]), have related to projects that integrate data held between institutions for research purposes, rather than disseminate it to the general public. It is in large part this lack of publicly accessible semantic data that has restricted further viral growth and it remains true that, as of June 2008, there are still only islands of information available and they apply to limited domains. This is of paramount concern, for although the community of interested users has grown and there is an increasing number of local success stories in applying such technology, it has not yet been sufficient to provoke uptake on a global scale. Whether this will require a so-called 'killer app' or merely a critical mass of users is unclear, but it is at least certain that utility will grow with increased adoption. Developments (such as the Linked Data Initiative [13]) which seek to maximize connectivity across semantic web applications, will play a key role in this process. Moreover, considerable work needs to be done at domain level - in the words of three protagonists of the SW vision, "the ontologies that will furnish the semantics for the Semantic Web must be developed, managed, and endorsed by practice communities"[6].

However, the tools for its creation have matured rapidly and Jim Hendler has noted that two strands of development have emerged: a 'top-down', ontology-driven approach, focused on providing semantically consistent knowledge-bases for a given domain, and 'bottom-up' data-driven techniques concerned with integrating heterogenous information from the web [14]. Both have much to offer, and each must be carefully evaluated and exploited by those tasked with selecting appropriate technologies if they are to mitigate the headaches frequently associated with applying new paradigms to legacy data.

The Semantic Web in Cultural Heritage

The Cultural Heritage sector has been aware of the potential power of the Semantic Web for some time, and the Museums and archives community in particular have been quick to explore it. This has taken the form of both theoretical discussions and practical implementation. The former has generally been through workshops and symposia which tend to generate a combination of excitement and skepticism. The DigiCULT Project [15] brought together a panel of 13 European experts in 2003 to discuss SW development in cultural heritage but the plethora of nascent (and competing) technologies at that stage, along with an apparent dependence on the agent-based applications mentioned in the seminal 2001 *Scientific American* article [1] led to talk of the Semantic Web as a 'Shangri-La" surrounded by a "veil of mystery". Nonetheless, a number of participants concluded that "they would put their money on the Semantic Web" even whilst other contributors maintained that "the heritage sector is likely to be left behind". In a similar vein, the Semantic Web Think Tank project, a series of workshops funded by the AHRC, concluded in 2008 that "There is no coherent answer to the question 'How do I do the Semantic Web?' and almost no information with which to make an informed decision about technologies, platforms, models and methodologies." This appeared to create a gap between the vision and the reality of SW "which critically undermines the ability of the sector to move forward in a clear and constructive way." [16].

Some of this pessimism may be due to a misunderstanding - several references in these this text and meeting reports [17] refer to non-semantic services (Flickr tagging, Google Base, Yahoo Pipes, etc.) normally associated with the broader (and vaguer) banner concept of 'Web 2.0'. Whilst the latter is indeed a confusing and frequently contradictory array of techniques and technologies, the Semantic Web has had a relatively clear road-map ever since its inception [18] and the tools for its creation are reasonably well defined. It is therefore difficult to know whether their concerns at this level are misplaced (a full report will be published in summer 2008). Nevertheless, if this information is not filtering down to the wider community, then such progress is of little value. And it is true that there is still no one-stop-shop or even simple entry point for those who wish to engage with the Semantic Web.

In parallel to these theoretical debates, a number of organisations have been tackling more concrete problems, with perhaps the most important development being the establishment of a Domain Ontology known as the Conceptual Reference Model (or CRM) created by the Comité International pour la Documentation des Musées (CIDOC) [19]. The CIDOC CRM is intended to cover the full spectrum of Cultural Heritage knowledge - from Archaeology to Art History - and has recently been extended to cover the production of literary and musical entities in more depth. Although currently incorporating 84 entity types and 141 property types it is remarkably compact and efficient given its extremely broad scope. It also has an inherently epistemological flavour in order to deal with the innate uncertainty of information about the past [20]. It received ISO standard status in 2006 (ISO 21127:2006) and is now the dominant Domain Ontology in Cultural Heritage. The greatest challenge in mapping legacy datasets to the CIDOC CRM however, has been the considerable mental leap required of both museum creators and their technical staff to map their datasets to such an abstract conceptualisation. Although CIDOC have had a number of successes in mapping legacy data to the ontology[21,22], and encourage an adaptive approach which restricts and/or extends the ontology, the process generally requires extensive collaboration between curators, IT professionals and CRM experts [23]. The recent release of the AMA tool [24] may help to ease this process, but there are still too few case-studies for a formal evaluation of its utility. A recent article by its chief architect also provides some interesting insights into its future direction [25].

Further to this central development, several project groups have been successful at producing end-to-end semantic systems:

MuseumFinland/CultureSampo [26,27] The first major deployment of Semantic Web technologies in the Museum sector was the MuseumFinland project’s provision of tools and services so that Finnish museums could present their collections online through a common semantic web interface. A Domain Ontology was developed for the project and contributing institutions make their data available as XML. This is then mapped to the Domain Ontology although the approach is reflexive and new concepts can be added globally where required. A website allows users to search and browse cultural artefacts in a ‘follow-your-nose’ fashion via their properties. Coming second in the 2004 Semantic Web Challenge, it also forms part of a wider initiative by the Finnish government to semantically enable public web services. A follow up project, the CultureSampo portal, extended the MuseumFinland ontology in order to represent events and processes. This enables the embedding of artefacts in narratives, helping to provide greater contextualisation.

ARTISTE/SCULPTEUR/eCHASE [23,28,29] Building on the previous ARTISTE project which provided cross-archival search capabilities for high-profile galleries using RDF, SCULPTEUR used an ontology-driven approach to provide adaptive search and visualisation mechanisms for 2D and 3D objects. Datatypes include digital images, 3D models, associated metadata, free text documents and numerical tables. Museum databases were mapped to the CIDOC CRM (with several extensions) and an ontology of the system components enabled dynamic interface modification to suit the heterogeneous nature of the data returned. Key amongst these navigation widgets was the mSpace browser [30]. Machine interoperability was created using a Z39.50-based Search & Retrieve Webservice protocol. The ensuing eCHASE project, funded by the European commission, was more directly focussed on the creation of a toolset and framework by which third parties could both contribute and draw multimedia entities from a semantically integrated network of repositories across Europe, chiefly with the aim of catalysing increased exploitation of otherwise moribund resources.

MultimediaN E-culture/CHIP@STITCH [31,32] MultimediaN E-culture is prototype system that brings together multiple online cultural heritage repositories in the Netherlands. Winner of the Semantic Web Challenge 2006 at the International Semantic Web Confence 2006, it is very much aimed at public users and non-technical researchers with a generic browser, ‘/facet’, that enables user to explore the databases along any facet - whether artist, genre, period or otherwise [33]. Application Ontologies were developed for individual databases as well as the Getty thesauri (AAN, TGN & ULAN) [34] and then aligned by hand. The CHIP browser, drawing on ontology mappings from the CATCH STITCH project, as well as MultiMediaN E-culture, combines data from the Rijksmuseum's ARIA database with RDF from the IconClass and Getty thesauri. The results are used to provide an automated Artwork Recommender that creates suggestions based on users 'rating' other examples as well as personalised tours of the museum that can be downloaded to a handheld device.

Contexta/SR [35] Contexta/SR, developed at Federico Santa Maria Technical University, Chile, also federates heterogeneous cultural heritage repositories, choosing to adopt a more diverse array of services and ontologies, many of which (ISAD, FOAF, and the W3C Geo Ontology) were developed outside the cultural heritage sector. In their view this creates an important natural interface to other SW repositories which is lacking from the CIDOC CRM. XSLT templates transform XML data dumps into RDF which is then 'ingested' into several central repositories via a predominantly manual filtering process. This provides a core dataset in which it is hoped that an 'ecosystem' of applications can grow, 'mashing-up' heritage items with contextual information elsewhere on the web. An important aspect of this work is their use of resolvable URIs in order to support the Linked Data Intiative [13].

Whilst all of these projects and others (see, for example [21,36,37]) have remained confident that semantic approaches can prove highly beneficial, they also generally underline the difficulty of mapping legacy datasets to ontologies - especially those with a high degree of complexity. Visualisation tools for RDF must also be carefully thought through in order to avoid drowning users in superfluous information. Although these early adopters have done a great deal of pioneering spade-work, it is clear that challenges remain. Two in particular provide an agenda for this work. The first is the conceptual challenge of mapping an implicit ontology instantiated in a relational database to a different explicit domain ontology instantiated in RDF. This has typically been done using XSLT applied to an XML dump which requires the simultaneous application of high level technical and domain expertise. The second issue is that the rationale behind individual elements of the mapping is difficult to document, and alternative mappings are inhibited by the need to repeat this difficult process. A process which separates this task into two stages – viz, *mapping* to a bespoke local ontology, followed by *alignment* with a domain ontology – would greatly mitigate these problems.

The Semantic Web in Archaeology

In contrast to the Museums community, the Semantic Web has seen less uptake in archaeology to date, despite promising beginnings as early as 2001 [38]. This is mainly due to a traditional focus within the archaeological computing community on analysis and recording rather than dissemination. The 'bleeding edge' nature of current SW technology also puts it out of the range of many field practitioners whose experience is frequently limited to relational database and/or spatial data management. Funding streams for large scale restructuring of datasets are also less prevalent, with a greater emphasis on initial recording, especially in countries with a heavy reliance on developer-funded rescue archaeology. Nevertheless, the potential benefits to the archaeological field are significant. The great majority of archaeological data is now rendered digital at some stage within the recording process but it has recently been argued that there has been an "information explosion" with the result that "field archaeology has drifted out of control" and now overwhelms those tasked with interpreting it [39]. Much of this is due to the complexity of combining siloed resources. In contrast, a regrettable lack of pressure to concentrate on the outward-facing aspects of their work has allowed many academic archaeologists to remain protective of their data, especially where it is not already published in traditional media [40].

Whilst a demonstration of the potential advantage of SW technologies could have a major impact on attitudes to recording, storage and dissemination strategies, it needs to be undertaken with consideration for the specific needs of the community. Foremost amongst these are low budgets for IT infrastructure and training which could constitute a major hurdle in ontology development [41]. Secondly, content is frequently of a different nature to the multimedia-centric concerns of museums, consisting more typically of widely divergent 'contexts' comprising bundles of numerical and categorical values. SKOS services for commonly used thesauri may be particularly useful in this regard [42]. The need to perform large scale statistical analyses over these entities also means that 'follow-your-nose' faceted browsers (such as /facet, mSPace, Longwell or Autofocus [33,30,43,44]) are of limited (if still important) application. Instead, it may be necessary to build interfaces that are context dependent, in order to visualise the varying types of data (contexts, locations, typologies, finds etc.) effectively. As the success of any web-based approach requires widespread uptake, it will therefore be necessary to reduce these barriers to entry as much as possible, even if this reduces the potential efficacy of the technology to some degree in the short-term.

The following projects highlight some of the main work undertaken in this space so far:

VBI-ERAT-LUPA [22] An early example of the use of the CIDOC CRM in the archaeological sphere, the VBI-ERAT-LUPA project aggregated data from several large databases of roman finds into a central triplestore using an XML-based export process. The approach appears to have been successful but was undertaken with minimal fanfare and access to the data is still via a fairly traditional website that does not emphasise any potential gain from merging the resources (users must access each via a separate web-form). If the data is available as resolvable RDF however, it could provide an extremely important contextualising dataset for other archaeological SW projects.

CIDOC CRM-EH/STAR [45,46] In 2004 the English Heritage Research and Standards Group undertook a project in conjunction with CIDOC to develop a CRM extension that would adequately describe the wide variety of archaeological datasets which they hold in-house as well as permit integration with excavation data from County Councils. This is an important step forward for archaeologists, as the CRM was developed with a focus on the needs of the museums and archive community in mind. Following this work, the STAR project, run in partnership with the Hypermedia Research Unit, University of Glamorgan, is now developing tools to help third-party organisations map their resources to this ontology. These include SKOS services that are capable of providing archaeological thesauri (such as those provided at [47]), and a demonstrator of several archaeological databases aggregated into RDF following the CRM-EH ontology.

ArcheoServer [48] Undertaken by the Department of Informatics, Systems and Communication, University of Milan-Bicocca, the Archaeoerver project has developed an ontology and navigation and editing tools for an e-library on pre- and proto-history in Italy. It is now starting work on the integration of excavation datasets in the Po Valley region. Interestingly, it has selected not to use the CIDOC CRM, but instead to 'roll its own' ontology using the NavEditOW ontology editor [49]. This may in part be due to a relative degree of homogeneity in such datasets, as prescribed by Italian law and similar work has produced an extremely complex ontology elsewhere [50].

Archaeology Platform (@PL) [51] This project, developed by the Vienna University of Applied Science, is attempting to create a generic peer-to-peer server framework that specifically utilises the CIDOC CRM for integrating excavation data. The project scope was presented at CAA 2008 and it is still in mid-development cycle, but the intention is to open-source the technology after the initial phase, possibly in October. If this approach is successful it could provide a key architectural component for further work in this field, but without a concrete implementation and case study to date, no further evaluation is possible.

Virtual Environment for Research in Archaeology [52] The VERA project, based at the University of Reading, has mainly focused on SOA techniques to increase the utility of the Integrated Archaeological Database (IADB) developed by the University of York. As a further goal they now intend to explore approaches for integrating the IADB with other archaeological datasets. They are currently developing a peer-to-peer server framework based on Tycho [53] that shares data between the IADB and the Vindolanda database hosted at Oxford. The intention is to incorporate RDF technology so that it can be used as the basis for a broader archaeological data-integration system.

These projects and others (see [54-59]) demonstrate the interest that SW technology is beginning to excite in some corners of the discipline, but also the necessity of a significant IT resource and knowledge-base at this stage of its evolution. However, awareness is growing and with many of the key technologies already freely available, the discipline is now well positioned to join the Museums community in restructuring its dissemination methods. In order to encourage more archaeologists to participate however, it will be crucial not only to demystify the technology but also to demonstrate the concrete benefits of participation to both the overall community and the contributors themselves.

One of the principle means by which to do this will be to stress its potential relevance for multivocality within the discipline [5]. This can be at many levels, from the use of different theoretical frameworks by separate research institutions to indigenous communities expressing local/traditional interpretations of the material record. Whereas the World Wide Web (and especially the Web 2.0 phenomenon) has greatly enhanced the participatory aspects of global knowledge sharing, it is poor at abstracting information from context: data *aggregation* (e.g. RSS feeds) often loses vital contextual information [25]. In contrast, the Semantic Web has the capability to maintain such information, but is frequently used to *homogenise* data, once again losing the context from which much of its original meaning is derived. In order to have a truly multivocal archaeological knowledge space, the original semantic context needs to be easily accessible. It is this goal which the current research seeks to attain.

Architectures

As we have seen above, the wide range of work undertaken in the domain during this initial period of development in SW technologies has led to considerable diversity in their approaches. This makes architectural decisions difficult as there are still no well-established and documented methodologies for the full life-cycle of a SW project. Nonetheless, we can discern several key trends emerging from which to choose, each with their own exemplars. With this information in hand, we can evaluate the key technologies they have used in order to develop a suitable approach for a particular project.

The first distinction is between processes which centralise the data (e.g. MuseumFinland, Contexta/SR and UBI-ERAT-LUPA) and those which keep it distributed (e.g. MultiMediaN, @PL, VERA). Whilst the former approach has many advantages in terms of simplicity in maintenance and was used frequently in early projects, there are a number of problems with it. Generally speaking, any methodology which seeks to integrate data from separate institutions which regularly update their information will have to implement an architecture that leaves them in full control of their data. This could be done using either a master-slave model (with a centralised control server), or as a peer-to-peer network. Although the former is a simpler architecture, it requires that one institution take responsibility for the whole framework, whilst the latter is more in keeping with the provision of a 'web of data' that has no single point of failure. Whilst de-centralising data is not necessary in order separate out individual ontologies (these can still be held in the same triple-store) it clearly lends itself to a more distributed semantic structure as well.

The second consideration is whether data should be exported to an RDF store prior to querying and integration, or whether it can be mapped dynamically on-the-fly. There are currently few if any cultural heritage applications of the second approach but it is beginning to become more common elsewhere with DBpedia [60] as a notable example. In terms of maintenance these approaches are more or less equivalent. Dynamic systems have the advantage of providing a 'live-update', so that information entered into a relational database does not need to be regularly updated, but they are dependent on a mapping server such D2R Server[61] or DartGrid [62] which can provide a SPARQL endpoint for querying. This may also provide restrictions on inferencing capabilities. Dumping the data on the other hand, requires either that users remember to export their data to the RDF store or that a polling mechanism does this regularly. As this process can be resource intensive it may also cause unwelcome performance issues. Finally, in either case there will be a requirement for specific components to remain permanently 'live', whether it be the database and server, or the triplestore. As the priorities of different contributors are like to vary, it may not be possible to find a one-size-fits-all solution, and a unified peer-to-peer architecture may need to incorporate facilities for both approaches.

Section II: Research Objective

Overview

The ultimate objective of the research is to *develop a semantic framework for archaeological knowledge integration that maintains ontological transparency between data sources.*

Whilst previous work in this field is increasingly successful at combining complex resources, this still involves reinterpreting legacy data systems within a generic ontology and important information is being lost in this process. The framework will therefore keep sources separate, and provide tools for each to expose its data in accordance with an Application Ontology that reflects its internal structure. Alignment between ontologies will be 'pluggable' thereby allowing users to experiment with alternative interpretations.

There are a number of benefits to this approach:

- Information generated form different methodological approaches can be integrated losslessly
- Different alignments between ontologies can be experimented with where these are contended
- Original ontologies themselves can be browsed by users
- By separating mapping from alignment, the process is made conceptually simpler.

This does include some extra costs however:

- Each source repository will have to develop and maintain its own ontology along with a default alignment to a Domain ontology.
- There may be added complexity in inferencing and querying.

Deliverables will be:

- An open source implementation forming the semantic framework for a specific case study.
- A guide that details the framework in a modular fashion so that other projects can implement it, incorporating new developments where appropriate.

The Framework will be specifically deployed amongst participating organizations in the Roman Ports in the Western Mediterranean project (see below).

Use Cases

In order to better assess the utility of a given framework, three particular Use Cases will be considered.

Use Case 1: Integration

To encourage participation and increase uptake, both in the current project and amongst potential adopters, it will be necessary to provide facilities that make integration of local datasets to the framework as painless as possible. It is envisaged that a participating contributor will have access to an easily downloadable open source software package that can be installed on an off-the-shelf Linux or Windows box. Once installed it will provide tools that permit them to

- Create a Application Ontology that makes explicit the semantics of their dataset
- Map their database to the Application Ontology
- Align the Application Ontology with other Application or Domain Ontologies.

In the first instance this is likely to require considerable assistance, but a Wiki Knowledge-base will be created in order to aid future participants. General technical documentation will also be available along with links to appropriate database connectivity drivers where appropriate. Once mapped, data will be made available to the framework either dynamically or by export to a triplestore component of the server. Contributor databases are likely to contain a mixture of both public and restricted data, so it will need to be possible to specify whether elements of the mapping are 'accessible' or 'inaccessible'. Inaccessible data will not be made available to the network.

Use Case 2: Browsing

In order to grant instant access to the data in human-readable form, as well as provide some of the concrete benefits of a Semantic Web approach, a Web Application will be developed. The primary function will be to contextualise entities within the RDF datagraph. Using a combination of widgets to provide, e.g., faceted browsing, node browsing, webmapping, timelining and automated graph plotting, a user should be provided with information relevant to any given entity (find, archaeological context, period, location, etc.). Using a follow-your-nose paradigm, clicking on an associated link will reconfigure the application appropriately in order to display the new entity and its own contextual information. An important addition to traditional semantic web browsing will be inclusion of an ontology browser, that allows users to also visualize the semantic structure(s) of the dataset(s) to which a given entity belongs.

Use Case 3: Developer API

The power of the Semantic Web lies predominantly in its potential for future developers to repurpose information in as-yet-unthought-of ways. To quote the JISC Common Repository Interfaces Group (CRIG), "The coolest thing to do with your data will be thought of by someone else" [63]. A SPARQL endpoint will provide the minimal functionality for data access, but a web API will also be implemented in order to provide information about entities, ontologies, thesauri, framework implementation information and utility services.

Key Design Criteria (in order of priority):

Added value

The framework must demonstrably increase the research potential of the data by enabling human and computational inferences to be made about data and its context which were not previously possible due to either technological or resource limitations.

Ease of use

Whilst the initial process of dataset integration will inevitably require a considerable level of technical knowledge, it is important to keep maintenance levels low and provide a user interface that enables non-technical users to engage with the data.

Robustness

The framework must show a reasonable level of robustness and structural integrity and ideally should have no single point of failure. Core components should not require regular maintenance, assuming deployment in a stable environment.

Generic

Although it is important to ensure fitness-for-purpose with regard to the case study, this research will also provide an important exemplar for future work on archaeological data integration. As such, it should use generic and open source components wherever possible, in order to improve reusability.

It is difficult to benchmark such criteria in any absolute sense, but evaluation will be done in the form of regular (pre-, mid- and post-project) questionnaires to project partners. They will be asked to comment on all the above areas, and indicate points for special attention. Potential points of failure can then be addressed correspondingly. Regular dialogue will also be maintained with technical representatives of each partner institution. All interaction with project partners will be done in consultation with the university's Ethics Committee.

Case Study for Implementation: Roman Ports in the Western Mediterranean

The Roman Ports in the Western Mediterranean project [64], directed by Prof. Simon Keay and Dr. Graeme Earl (British School at Rome/University of Southampton), is an investigation into the relationship of Portus (the principal port of Rome in the Imperial era) to ports in the Western Mediterranean basin. The principal methodology is to look at the co-presence of ceramics and marble at a range of key port sites as a means of gauging fluctuating trans-Mediterranean connections during the Roman period. Source data comprises harbour and shipwreck excavation databases from a variety of academic and research institutions. As an international endeavour, requiring the synthesis of large quantities of data with heterogeneous format but restricted scope, it provides an ideal opportunity to work through the issues specific to the archaeological community in deploying semantic web technologies. In particular, the wide range of theoretical approaches, including contested typologies and recording methodologies, make this a challenge that is expressly suited to the techniques described above.

An important element of the project consists of annual themed workshops held at the British School at Rome. At the first of these, held in March 2008 [65], an agreement was reached between a number of institutional resource directors to mutually contribute data to a common framework which is to be developed specifically for the purpose (See Data Sources, below). Following a presentation by the candidate at the workshop, it was further agreed to use this PhD research as the foundation of that framework with data contributors providing appropriate local technical knowledge where required. The framework should permit all data providers with access to the complete combined dataset by means of a common web interface. Ultimately this should also be made available to the general public although this may not necessarily be in the first instance.

Data Sources

The following catalogue is a brief description of resources likely to be contributed. It is hoped that all resources can be utilised but if serious technical or legal issues arise (lack of local technical support, licensing issues, etc.) then individual repositories may be dropped. Integration of three or more datasets will be considered the minimum sufficient for successful completion.

University of Southampton, UK [66]
OS: Windows
DB: ARK (MySQL)
Contents notes: Data from the ongoing Portus excavation. Records are currently few but are expected to increase considerably over the next two seasons
Contact: Graeme Earl

Archaeological Data Service (ADS), UK [67]
OS: Windows
DB: MS Access
Contents notes: Amphora typologies. Approx. 300 records
Contact: Michael Charno

ASMOSIA, Catalan Institute of Classical Archaeology, Spain [68]
OS: Windows
DB: FileMaker
Contents notes: Marble finds. 4 tables w/ approx. 5,500 records
Contact: Anna Gutierrez

Institute for Archaeological and Monumental Heritage, Italy [69]
OS: Windows

DB: MS Access
Contents notes: All ceramics finds in Sicily dated from Late Hellenistic to Late Roman period
Contact: Daniele Malfitana

Institute for Studies of Ancient Culture, Austrian Academy of Sciences [70]
OS: Windows
DB: FileMaker
Contents notes: Amphora finds from Ephesus. Approx. 15,000 records
Contact: Tamas Bezeczky

University of Leuven (K.U.Leuven), Belgium [71]
OS: Windows
DB: MS Access
Contents notes: ICRATES database. Database recording large number of ceramics deposits and their publications. It contains 6 tables and approx. 25,000 records
Contact: Jeroen Poblome

Other
Other potential datasets include those from the University of Aix-en-Provence, the University of Oxford, the University of Seville, the University of Cadiz, and the Society for Libyan Studies.

Section III: Work To Date

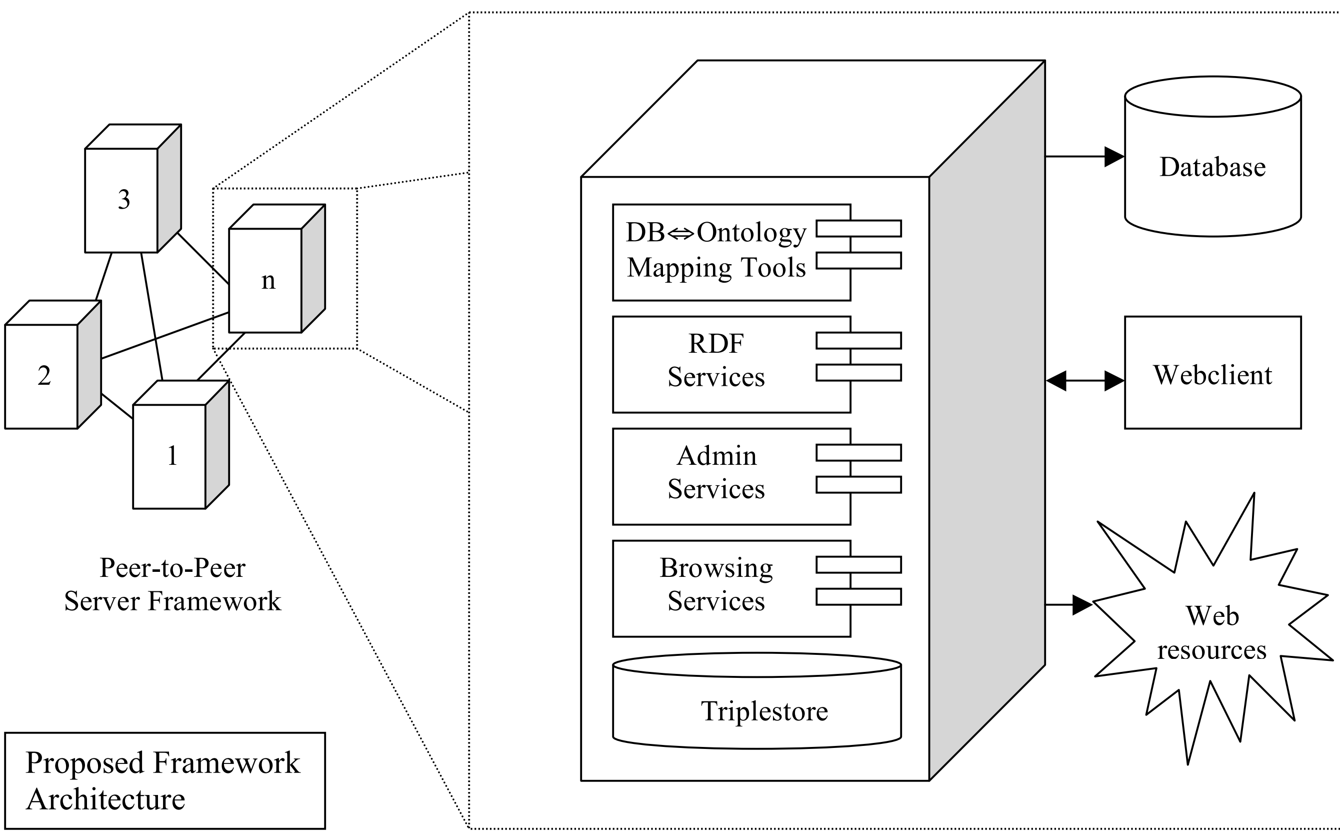
Work undertaken so far has predominantly focused on:

1. Reviewing the literature on similar work and relevant technologies
2. Establishing a case study
3. Developing a provisional schema for the framework architecture
4. Experimenting with a variety of technologies which are likely to form components of the framework

For (1) and (2) see Sections I & II, above.

Provisional Architecture

The intended framework, called TRANSLATION (TRAnsparent Negotiation and Sharing of Local Application Terminologies, Instances and ONtologies), will be of a peer-to-peer nature in order to have no single point of failure and encourage open web dissemination. Each server will contain a variety of component modules that provide the necessary resources for participants to map, align, disseminate and browse their data, as well as access that hosted by others (see diagram). In order to take account of differing circumstances, there may be several options open to contributors as to the manner in which their information is made accessible to the system.



Proposed Framework Architecture

Peer-to-Peer Server Framework

Servers should be capable of both resource discovery and messaging. This will provide the foundations for higher-level querying of resources. The proposed framework is Tycho, developed by the University of Reading [53].

DB⇔Ontology Mapping Tools

A number of mapping tools are available which could incorporated. Application Ontologies can be created with the Protégé editor [72] and an alignment may be undertaken using a system such as OntoMediate [73]. Alignment must be possible over a web interface in order to permit external users to try alternative alignment descriptions. D2RQ [74] may provide a suitable language for mapping to RDF, but there are currently no visual editors available.

RDF Services

If data is held within the triple store, many of these are likely to be available by default (see below). If the data needs to be read dynamically then the mapping must be interpreted via a server such as D2R or DartGrid. These both provide SPARQL end points, and D2R also makes entity URIs resolvable which is important for linking data.

Additional services could also provide an API for querying meta-information about the RDF (which ontologies are used, etc).

Admin Services

An API for server configuration and potentially for distributing updates, patches, etc.

Browsing Services

Web server functionality to provide a default user interface to the data. This is likely to be widget/AJAX based, using a variety of Web 2.0-style APIs to contextualise the data. Possible candidates for incorporation include SIMILE Timeline & Timeplot [11], GoogleMaps [75], and the mSpace faceted browser [30]. The JUNG Framework [76] may be used as a simple ontology browser.

Triplestore

The primary role of the triplestore is as a temporary place to hold triples harvested from other servers whilst querying. It may also function as a persistence layer for RDF, if not mapped dynamically from the relational database. Several triplestores have begun to reach a reasonable degree of maturity, including Jena [77], Sesame [78] and the Southampton University 3store [79], each with a variety of access methods. The API and technical specifications of each will be considered and selected as appropriate.

Database

This is the database provided by the contributor. DBC drivers will need to be provided to cover common formats – MS Access, FileMaker and MySQL are likely to be prevalent. Access to the data will be read-only in order to ensure security of the data.

Webclient

Browsing and Admin services should be accessible using a standard web browser. Firefox 3 and IE7 will be used for testing.

Web resources

Various web resources may need to be called, in particular RDFS and/or OWL representations of domain and top-level ontologies and SKOS thesauri [80]. A discovery protocol and registry will also be needed in order to communicate with other servers in the framework.

Experimentation

D2RQ Platform

The D2RQ Platform [81] is a suite of technologies developed by Berlin Free University comprising D2RQ, a declarative language for mapping relational databases to RDF, and D2R Server which can publish the data to the Web. An experiment was undertaken to map an example archaeological dataset (taken from [82]) stored in a PostgreSQL database, to the CIDOC CRM ontology. Results were mixed: although straightforward mappings are relatively easy to edit in the mapping file, complex mappings (in which intermediary nodes expressed in an ontology are not expressed in the DB) are not. If local Application Ontologies are used however, it may not be a problem. It does provide the advantage of providing several methods of accessing the data, including a SPARQL endpoint, resolvable URIs and dumping to a triplestore.

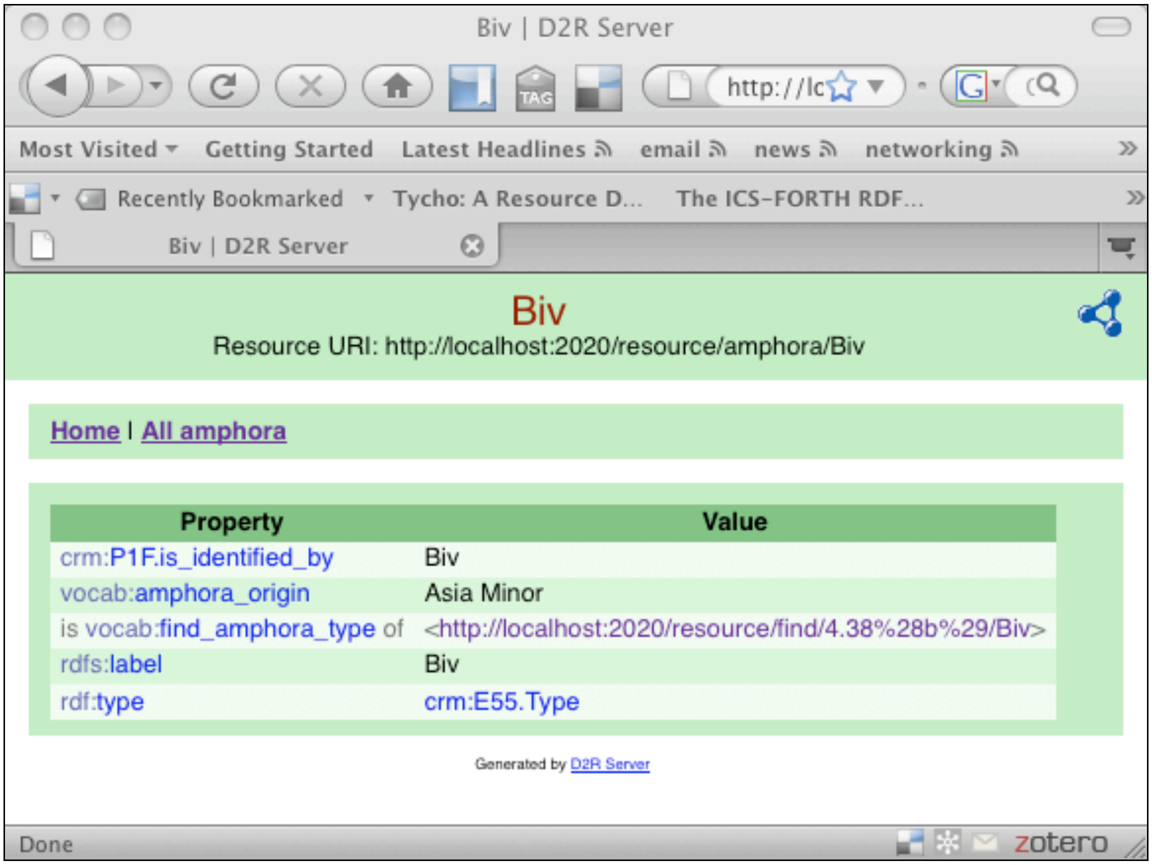


Figure 1. An amphora type represented in CIDOC CRM RDF using D2R Server

DartGrid

DartGrid [83] was also trialed as means of mapping relational data. The Eclipse-based Semantic Mapping Tool provides a reasonably straightforward means to visually associate relational tables with an RDFS ontology. It also has the possibility to introduce blank-node relations in order to deal with complex mappings. On the other hand, an attempt to import the CIDOC CRM was less successful as only entites (and no properties) were accessible within the tool. It currently also only supports connection to mySQL and Oracle databases. Java libraries with which to interpret the mapping are available, but currently no source code (described as 'Will be available soon!' since July 2006).

AMA

The Archive Mapper for Achaeology [84], developed by EPOCH, is a LAMP-based tool for mapping database schemas expressed in XML to the CIDOC CRM (or other ontologies) . It was released in June 2008 and is currently available both as a download and an online demonstrator, but as documentation is limited to a single conference paper [24], evaluation proved difficult. It is also not clear what system is intended to interpret the generated mappings. Nevertheless, as an ongoing project with the specific intention of mapping archaeological information to a Domain Ontology, it is still worth further investigation and positive results have been reported in [85].

Protégé

Protégé [72] is the most commonly used ontology-building tool. Experiments in producing application ontologies that reflect relational database structures and saving them as RDFS proved successful and relatively pain-free.

JUNG

The Java Universal Network/graph Framework [76] is a set of Java libraries for visualizing graphs. A demonstration application was built during the early stages of research in order to investigate whether a Network Analysis approach might be suitable for Typology alignment. Whilst the experiment proved unsuccessful (alignment requires human interpretation), the JUNG framework looks extremely useful for RDF visualization and has been used in other projects to powerful effect [86].

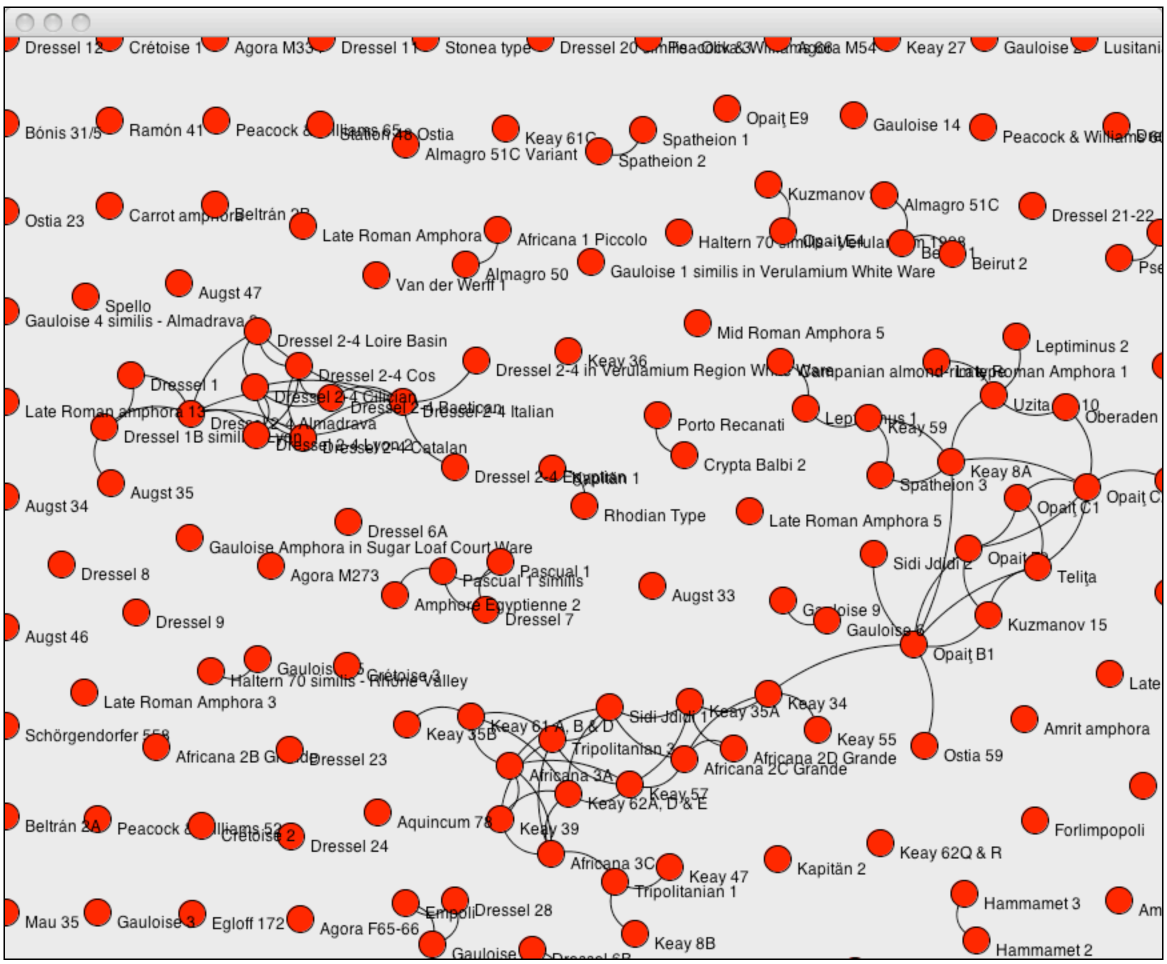


Figure 2. Amphora types associated by facet correspondence using JUNG.

Section IV: Future Work

Develop demonstrator. – 3rd Q. 2008

Development of a demonstrator that maps two databases to Application Ontologies and aligns them with a domain-level ontology (probably the CIDOC CRM-EH) to allow cross-walking with a SPARQL endpoint. Mapping will require the establishment of a specific workflow and toolkit.

Develop Zero Feature Release (ZFR) of Peer-to-Peer network – 4th Q. 2008

Integration of tool set into servlet-based modules on a single server. Testing with datasets hosted on separate servers, with external access to Domain Ontologies and thesauri.

Develop Prototype Interface – 1st Q. 2008

Web Server functionality to be integrated. RDF entities should be contextualised via associated data represented in appropriate format (tabular, mapped, graphed, etc.). Essential admin services should be available via a web interface.

Minithesis: 'A Model of Semantic Web integration for archaeological datasets'

As well as describing the work undertaken and future direct of research, the minithesis should enable both archaeological and technical specialist to evaluate its utility for integrating archaeological data. It will also form the basis for a presentation at a workshop in Rome for participants in the Roman Ports in the Western Mediterranean project who wish to contribute data.

Integration of third party datasets – 2nd-4th Qs. 2009

Collaboration with data contributors via series of workshops in order to enable them to integrate data into the network. Feedback will be through webfora and a wiki, in order to rapidly build a knowledge base of development issues and user tips. These will also provide input for a process manual.

Further Development – 3rd-4th Qs. 2009

Dependent on user needs, further features may be integrated, including alternative mapping mechanisms, interface widgets, and stability and performance optimization.

Write Up – 1st-3rd Qs. 2010

Write thesis, along with Cookbook and/or User Manual for other developers to apply the framework or similar approaches to other archaeological integration projects.

References

[1] T. Berners-Lee, J. Hendler, and O. Lassila, “The Semantic Web,” *Scientific American*, May. 2001; http://www.sciam.com/article.cfm?id=the-semantic-web.

[2] M. Foucault and A. Sheridan, *Archaeology of Knowledge*, Routledge, 2002.

[3] W.V. Quine, “Two Dogmas of Empiricism,” *The Philosophical Review*, vol. 60, Jan. 1951, pp. 20-43.

[4] E. Kansa, “A community approach to data integration: Authorship and building meaningful links across diverse archaeological data sets,” *Geosphere*, vol. 1, Oct. 2005, pp. 97-109.

[5] R. Boast, M. Bravo, and R. Srinivasan, “Return to Babel: Emergent Diversity, Digital Resources, and Local Knowledge,” *The Information Society*, vol. 23, 2007, p. 395.

[6] N. Shadbolt, T. Berners-Lee, and W. Hall, “The Semantic Web Revisited,” *IEEE Intelligent Systems*, vol. 21, 2006, pp. 96-101.

[7] “Web Naming and Addressing Overview (URIs, URLs, ...)”; http://www.w3.org/Addressing/.

[8] “Resource Description Framework (RDF) / W3C Semantic Web Activity”; http://www.w3.org/RDF/.

[9] “SPARQL Query Language for RDF”; http://www.w3.org/TR/rdf-sparql-query/.

[10] “OWL Web Ontology Language Overview”; http://www.w3.org/TR/owl-features/.

[11] “SIMILE Project”; http://simile.mit.edu/.

[12] K. Wolstencroft et al., “A Little Semantic Web Goes a Long Way in Biology,” *The Semantic Web – ISWC 2005*, 2005, pp. 786-800; http://dx.doi.org/10.1007/11574620_56.

[13] "Linked Data"; http://linkeddata.org/.

[14] J. Hendler, "Semantics and the Network Effect: A little semantics goes a long way," Redmond, Washington, USA: 2007; http://research.microsoft.com/workshops/SemGrail2007/Papers/JimH_Position.doc.

[1] *DigiCULT Issue 3: Towards a Semantic Web for Heritage Resources*, DigiCULT, 2003; http://www.digicult.info/pages/pubpop.php?file=http://www.digicult.info/downloads/ti3_high.pdf.

[16] R. Parry, N. Poole, and J. Pratty, " Semantic Dissonance: Do We Need (And Do We Understand) The Semantic Web?," *Museums and the Web 2008*, Montreal, Canada: 2008; http://www.archimuse.com/mw2008/papers/parry/parry.html.

[17] "UK Museums and the Semantic Web"; http://culturalsemanticweb.wordpress.com/.

[18] "Semantic Web roadmap"; http://www.w3.org/DesignIssues/Semantic.html.

[19] N. Crofts et al., "Definition of the CIDOC Conceptual Reference Model. Version 4.2.4," Jan. 2008; http://cidoc.ics.forth.gr/docs/cidoc_crm_version_4.2.4_March2008_.pdf.

[20] M. Doerr, "The CIDOC conceptual reference module: an ontological approach to semantic interoperability of metadata," *AI Mag.*, vol. 24, 2003, pp. 75-92.

[21] N. Crofts, "Combining data sources – prototype applications developed for Geneva's department of historical sites and monuments based on the CIDOC CRM," 2004; http://cidoc.ics.forth.gr/docs/st_petersburg_combining_data_sources_.doc.

[22] M. Doerr, K. Schaller, and M. Theodoridou, "Integration of complementary archaeological sources," *Proceedings of Computer Applications and Quantitative Methods in Archaeology Conference 2004*, Prato, Italy: 2004; http://www.ics.forth.gr/isl/publications/paperlink/doerr3_caa2004.pdf.

[23] M. Addis et al., “New Ways to Search, Navigate and Use Multimedia Museum Collections over the Web,” *Museums and the Web 2005*, Vancouver, Canada: 2005; http://eprints.ecs.soton.ac.uk/10909/.

[24] O. Eide et al., “Encoding Cultural Heritage Information for the Semantic Web. Procedures for Data Integration through CIDOC-CRM Mapping,” *EPOCH Conference on Open Digital Cultural Heritage Systems (2008)*, D. Arnold et al., ed., 2008, pp. 1-7.

[25] M. Doerr and D. Iorizzo, “The dream of a global knowledge network\—A new approach,” *J. Comput. Cult. Herit.*, vol. 1, 2008, pp. 1-23.

[26] E. Hyvönen et al., “MuseumFinland--Finnish museums on the semantic web,” *Web Semantics: Science, Services and Agents on the World Wide Web*, vol. 3, Oct. 2005, pp. 224-241.

[27] E. Hyvönen et al., “CultureSampo―Finnish Culture on the Semantic Web: The Vision and First Results,” *In: K. Robering (ed.): Information Technology for the Virtual Museum. LIT Verlag, Berlin.*, 2007.

[28] M.J. Addis et al., “The eCHASE System for Cross-border Use of European Multimedia Cultural Heritage Content in Education and Publishing,” *AXMEDIS 2006: 2nd International Conference on Automated Production of Cross Media Content for Multi-channel Distribution*, Leeds, UK: 2006; http://eprints.ecs.soton.ac.uk/14265/.

[29] M. Addis, P. Lewis, and K. Martinez, “ARTISTE image retrieval system puts European galleries in the picture,” *Cultivate Interactive*, Jun. 2002; http://eprints.ecs.soton.ac.uk/7413/.

[30] M.C. schraefel et al., “The evolving mSpace platform: leveraging the semantic web on the trail of the memex,” *Proceedings of the sixteenth ACM conference on Hypertext and hypermedia*, Salzburg, Austria: ACM, 2005, pp. 174-183; http://portal.acm.org/citation.cfm?doid=1083356.1083391.

[31] J. van Ossenbruggen et al., “Searching and Annotating Virtual Heritage Collections with Semantic-Web Techniques,” *Museums and the Web 2007*, 2007; http://www.archimuse.com/mw2007/papers/ossenbruggen/ossenbruggen.html.

[32] L. Aroyo et al., "CHIP Demonstrator: Semantics-Driven Recommendations and Museum Tour Generation," *The Semantic Web*, 2008, pp. 879-886; http://dx.doi.org/10.1007/978-3-540-76298-0_64.

[33] M. Hildebrand, J. van Ossenbruggen, and L. Hardman, "/facet: A Browser for Heterogeneous Semantic Web Repositories," *The Semantic Web - ISWC 2006*, 2006, pp. 272-285; http://dx.doi.org/10.1007/11926078_20.

[34] "Getty Vocabulary Program (Research at the Getty)"; http://getty.edu/research/conducting_research/vocabularies/.

[35] H. Astudillo, P. Inostroza, and C.A.L. Moncada, "Contexta/SR: A multi-institutional semantic integration platform," *Museums and the Web 2008*, Montreal, Canada: 2008; http://www.archimuse.com/mw2008/papers/astudillo/astudillo.html.

[36] A. Aiello et al., "An Experimental Ontology Server for an Information Grid Environment," *International Journal of Parallel Programming*, vol. 34, Dec. 2006, pp. 489-508.

[37] Ye Liang, Hong Bao, and Hongzhe Liu, "Hybrid Ontology Integration for Distributed System," *Software Engineering, Artificial Intelligence, Networking, and Parallel/Distributed Computing, 2007. SNPD 2007. Eighth ACIS International Conference on*, 2007, pp. 309-314.

[38] H. Alani, "Spatial and Thematic Ontology in Cultural Heritage Information Systems," Jul. 2001; http://eprints.ecs.soton.ac.uk/6147/.

[39] R. Bradley, "Bridging the Two Cultures – Commercial Archaeology and the Study of Prehistoric Britain," *The Antiquaries Journal*, vol. 86, 2006.

[40] G. Mantegari et al., "Towards a Web-based Environment for Italian Prehistory and Protohistory," *Digital Discovery: Exploring New Frontiers in Human Heritage. CAA 2006. Computer Applications and Quantitative Methods in Archaeology. Proceedings of the 34th Conference, Fargo, United States, April 2006*, Fargo, USA: Archaeolingua, 2007.

[41] H. Alani et al., "Building a Pragmatic Semantic Web," *IEEE Intelligent Systems*, vol. 23, May. 2008, p. 61—68.

[42] D. Tudhope and C. Binding, “Toward Terminology Services: Experiences with a Pilot Web Service Thesaurus Browser,” *Assist Bulletin*, Jul. 2006; http://www.asist.org/Bulletin/Jun-06/tudhope_binding.html.

[43] “Longwell - SIMILE”; http://simile.mit.edu/wiki/Longwell.

[44] “AutoFocus Overview”; http://www.aduna-software.com/technologies/autofocus/overview.view.

[45] P. Cripps et al., *Ontological Modelling of the work of the Centre for Archaeology*, Centre for Archaeology, Fort Cumberland: English Heritage, 2004; http://cidoc.ics.forth.gr/docs/Ontological_Modelling_Project_Report_%20Sep2004.pdf.

[46] D. Tudhope and C. Binding, “STAR Project, Hypermedia Research Unit, University of Glamorgan”; http://hypermedia.research.glam.ac.uk/kos/STAR/.

[47] “INSCRIPTION”; http://www.fish-forum.info/inscript.htm.

[48] A. Bonomi et al., “A System Supporting Users of Cultural Resource Management Semantic Portals,” *AI*IA 2007: Artificial Intelligence and Human-Oriented Computing*, 2007, pp. 757-764; http://dx.doi.org/10.1007/978-3-540-74782-6_66.

[49] A. Bonomi et al., “NavEditOW – A System for Navigating, Editing and Querying Ontologies Through the Web,” *Knowledge-Based Intelligent Information and Engineering Systems*, 2007, pp. 686-694; http://dx.doi.org/10.1007/978-3-540-74829-8_84.

[50] A. Aiello, M.M. Furnari, and F. Proto, “ReMuNaICCD: a formal ontology for the ICCD,” *Applied Ontology*, vol. 3, 2006.

[51] F. Eckkrammer, T. Eckkrammer, and R. Feldbacher, “CIDOC CRM in data management and data sharing: Data sharing between different databases,” Budapest, Hungary: 2008.

[52] “VERA”; http://vera.rdg.ac.uk/.

[53] "Tycho: A Resource Discovery Framework and Messaging System for Distributed Applications"; http://www.acet.reading.ac.uk/projects/tycho/index.php.

[54] K. Byrne, "Tethering Cultural Data with RDF," *Proceedings of Jena Users Conference (JUC2006)*, Bristol, UK: 2006; www.ltg.ed.ac.uk/np/publications/ltg/papers/Byrne2006Tethering.pdf.

[55] J.M.F. González, A.P. Márquez, and E.C. Cuenca, "Bases for the creation of ontology in the context of Archaeology," *The world is in your eyes - Proceedings of the XXXIII Computer Applications in Archaeology Conference: Tomar March 2005*, Tomar, Portugal: 2007.

[56] J. Holmen, C. Ore, and O. Eide, "Documenting two histories at once: Digging into archaeology," *Enter the Past. The E-way into the four Dimensions of Cultural Heritage. CAA 2003, Computer Applications and Quantitative Methods in Archaeology. BAR International Series 1227*, Vienna, Austria: Archaeopress, Oxford, 2004.

[57] E.C. Mavrikas, N. Nicoloyannis, and E. Kavakli, "Cultural Heritage Information on the Semantic Web," *Engineering Knowledge in the Age of the SemanticWeb*, 2004, pp. 477-478; http://www.springerlink.com/content/y63g57tj05wqydtd.

[58] J. McAuley and J.D. Carswell, "Knowledge Management for Disparate Etruscan Cultural Heritage," *Digital Society, 2008 Second International Conference on the Digital Society*, 2008, pp. 70-74.

[59] O. Missikoff, "Ontologies as a Reference Framework for the Management of Knowledge in the Archaeological Domain," *Enter the Past. The E-way into the four Dimensions of Cultural Heritage. CAA 2003, Computer Applications and Quantitative Methods in Archaeology. BAR International Series 1227*, Vienna, Austria: Archaeopress, Oxford, 2004.

[60] "DBpedia"; http://dbpedia.org/About.

[61] C. Bizer and R. Cygniak, "D2R Server - Publishing Releational Databases on the Semantic Web," Athens, Georgia, USA: 2006.

[62] H. Chen and Z. Wu, "DartGrid III: A Semantic Grid Toolkit for Data Integration," *Proceedings of the First International Conference on Semantics, Knowledge and*

Grid, IEEE Computer Society, 2005, p. 12; http://portal.acm.org/citation.cfm?id=1158335.1158470&coll=&dl=.

[63] "CRIG - DigiRepWiki"; http://www.ukoln.ac.uk/repositories/digirep/index/CRIG.

[64] "The British School at Rome - Roman Ports Project"; http://www.bsr.ac.uk/BSR/sub_arch/BSR_Arch_05Roman.htm.

[65] "Port Networks in the Roman Mediterranean. Workshop at the British School at Rome. 6-8th March 2008"; http://www.bsr.ac.uk/BSR/sub_arch/extra/BSR_RPorts_Workshop08.htm.

[66] "Portus Project"; http://www.portusproject.org/.

[67] "Roman Amphorae: a digital resource"; http://ads.ahds.ac.uk/catalogue/archive/amphora_ahrb_2005/.

[68] "ASMOSIA"; http://www.asmosia.org/index.html.

[69] "ibam-cnr"; http://www.ibam.cnr.it/englishversion/Home.htm.

[70] "Ephesos - Keramikforschung - Amphorenforschung (Institut für Kulturgeschichte der Antike der ÖAW)"; http://www.oeaw.ac.at/antike/ephesos/keramikforschung/amphoren/amphoren.html.

[71] P. Bes and J. Poblome, "A new look at old data: the ICRATES platform," *Old pottery in a new century: innovating perspectives on Roman pottery studies*, Catania, Italy: Bretschneider, 2006, pp. 141-165.

[72] "The Protégé Ontology Editor and Knowledge Acquisition System"; http://protege.stanford.edu/.

[73] "OntoMediate: Ontological Mediation and Semantic Gateways for Domain/Enterprise Translations"; http://www.ecs.soton.ac.uk/research/projects/ontomediate.

[74] C. Bizer and A. Seaborne, "D2RQ - Treating Non-RDF Databases as Virtual RDF Graphs," Hiroshima, Japan: 2004; http://sites.wiwiss.fu-berlin.de/suhl/bizer/pub/Bizer-D2RQ-ISWC2004.pdf.

[75] "Google Maps API - Google Code"; http://code.google.com/apis/maps/.

[76] "JUNG - Java Universal Network/Graph Framework"; http://jung.sourceforge.net/.

[77] "Jena Semantic Web Framework"; http://jena.sourceforge.net/.

[78] "openRDF.org: Home"; http://www.openrdf.org/.

[79] S. Harris and N. Gibbins, *3store: Efficient Bulk RDF Storage*, 2003; http://eprints.ecs.soton.ac.uk/7970/.

[80] "Simple Knowledge Organization System (SKOS) - home page"; http://www.w3.org/2004/02/skos/.

[81] "D2RQ - Treating Non-RDF Databases as Virtual RDF Graphs"; http://www4.wiwiss.fu-berlin.de/bizer/d2rq/.

[82] *Excavations at Carthage: The British Mission. Vol. 2 The circular harbour, north side*, Oxford: Oxford University Press, 1994.

[83] "DartGrid -"; http://ccnt.zju.edu.cn/projects/dartgrid/.

[84] "AMA - Archive Mapper for Archaeology"; http://www.epoch-net.org/index.php?option=com_content&task=view&id=222&Itemid=338.

[85] A. Felicetti and H. Mara, "Semantic Maps and Digital Islands: Semantic Web technologies for the future of Cultural Heritage Digital Libraries," *SIEDL 2008: Semantic Interoperability in the European Digital Library*, Tenerife, Spain: 2008, pp. 51-62; mage.ntua.gr/swamm2006/SIEDLproceedings.pdf.

[86] "Resist RKB Explorer"; http://www.rkbexplorer.com/explorer/.

Glossary of Technical Terms

Alignment
The process of connecting concepts between different ontologies.

AJAX (Asynchronous JavaScript and XML)
AJAX is a group of web development techniques chiefly characterised by embedding programming logic within a web page so that it can be executed in the client browser rather than requiring further requests to a server. This is principally in order to improve the performance, and thus usability, of the site.

API (Application Programming Interface)
A set of functions provided by a piece of software that is intended to be used by a computer program so that the functionality can be incorporated in other systems. This is generally in contrast to human-computer interfaces, typically either a Command Line or Graphical User Interface (see below)

GUI (Graphical User Interface)
A visual interface with which a human can interact with a software system, frequently composed of 'widgets' such as buttons, menus or interactive images.

LAMP (Linux-Apache-MySQL-PHP)
A 'stack' of open source technologies which are frequently selected as the basis for Web applications due to their robustness, extensive documentation and free availability. Variations of one or more of these components are common (e.g. deployment on Windows).

Mapping
The process of connecting data held with a relational database to the concepts within an ontology.

Ontology
In Computer Science an ontology is an 'explicit specification of a conceptualisation'. It provides a definition (i.e. the semantics) of entity types and their possible properties and relationships in a theoretical model of a given domain. A Domain (or Core) Ontology provides a common (often complex) model of the sector, whereas an Application Ontology reflects the semantic structure of a given application (often a single database).

OWL (OWL: Web-ontology Language)
A family of knowledge representation languages used to describe ontologies. Endorsed by the World Wide Web Consortium, they are now the preferred method for expressing complex semantic relationships between classes on the Semantic Web.

URI (Universal Resource Identifier)
A unique identifier referring to either an information resource (on the internet) or a non-information resource (in the real world). It is composed of a string of characters

consisting of a protocol and server location with optional additional information such document location and/or other parameters. They may or may not be 'resolvable' i.e. capable of returning a *representation* of the resource over the internet. URIs form the atomic 'words' of the Semantic Web and are used as the expression of well-specified concepts.

RDF (Resource Description Framework)
A method of modelling semantically meaningful statements on the Web via subject-predicate-object expressions (known as triples) composed of URIs (or two URIs and a literal value such as a number or string). Combinations of these statements form a graph over which more complex meanings can be inferred. RDF is format-independent, but frequently expressed in a (restricted) form of XML.

RDFS (RDF Schema)
As RDF provides virtually no constraints on how URIs are combined, RDFS (which is expressed in RDF) provides a small URI vocabulary for ontological concepts such Class and Property that permit the construction of simple ontologies.

Semantic Web
The Semantic Web is an extension of the World Wide Web intended to encode meaning in machine-readable form. This is done by creating subject-predicate-object statements (called triples) composed of URIs, each of which equates to a specific concept. The aggregation of these statements forms a graph which defines the knowledge-base. Graphs are constrained by the use of ontologies which define what types of concept can exist (the nodes of the graph) and the types of relationship that can exist between them (the arcs).

SKOS (Simple Knowledge Organization System)
A standard for expressing Knowledge Organisation Systems (KOSs), such as thesauri, taxonomies, glossaries and other classification schemes in terms of RDF, in order to make them available to the Semantic Web.

SOA (Service-oriented Architecture)
A software architecture composed of independent software 'services' (which may or may not be distributed on different servers) that offer discrete pieces of functionality with a clearly defined API. This 'loose coupling' is intended to allow for greater flexibility and adaptability when maintaining and/or developing the framework.

Triplestore
A data repository for storing RDF triples.

XML (eXtensible Markup Language)
XML is a general purpose specification for creating custom markup languages. Data expressed in XML is 'well-formed' if it conforms to certain syntactical rules and 'valid' if conforms to structural rules defined in a separate document (an XML Schema or Document Type Definition).

SPARQL (SPARQL Protocol and RDF Query Language)
A language for expressing queries across an RDF graph.

XSLT (eXtensible Stylesheet Language – Transformations)
A language for transforming XML documents into other XML documents, generally either so that they are valid against a different schema, or to alter the data content in some fashion.