The 13th International Symposium on Virtual Reality, Archaeology and Cultural Heritage VAST (2012)
D. Arnold, J. Kaminski, F. Niccolucci, and A. Stork (Editors)

Building scholar e-communities using a semantically aware framework: Archaia Kypriaki Grammateia Digital Corpus

D. Pitzalis[1,2], E. Christophorou[2]

N. Kyriakou[2]

A. Georgiadou[3] and F. Niccolucci[2]

[1]Université Pierre et Marie Curie, LIP6, Paris, France
[2]The Cyprus Institute, STARC, Nicosia, Cyprus
[3]University of Patras, Department of Philology, Greece

Abstract

Web-based learning communities have developed into a very popular vehicle for sharing information amongst students, researchers and enthusiastic users and are slowly gaining importance in the humanities field. Unfortunately, as data organization and information exchange in such communities are usually unstructured and strongly geo-localized, they generate serious interoperability issues. At the same time, an increasing amount of knowledge based resources is made available on the Web and a lot of effort is put into creating reference ontologies for the Cultural Heritage, actively publishing controlled vocabularies and sharing data across different platforms by using RDF and RDFa. In this context and with the objective to consolidate these two independent efforts, we have developed a platform for advanced structured online collaboration as a framework for advanced e-learning, the Archaia Kypriaki Grammateia Digital Corpus (AKGDC). Our Digital Library supports exchange amongst researchers, educators and students, makes heterogeneous data resources available and easily reusable and suggests semantic relations within the resources. The AKGDC framework is based on a rather comprehensive corpus of the ancient Cypriot literature, the six-volume "Archaia Kypriaki Grammateia", or Αρχαία Κυπριακή Γραμματεία *authored by A. Voskos, K. Michaelides and I. G. Taifacos and published by the Leventis Foundation between 1995 and 2008. The corpus covers the ancient Cypriot literary production of approximately thirteen centuries (from 7th century BC to 5th-6th century AD), and is typically classified by literary genres. This paper describes how the Digital Library has been conceptualized, developed and enriched. We expect our DL to positively impact highly interdisciplinary areas, such as Classical philology, archaeology, epigraphics, history, religion, philosophy, as well as to provide a broad utility service across the scholarly community.*

Keywords: semantic web, intelligent data browsing, ontologies, content management systems, e-learning

1. Introduction

The digital media has offered new methods and tools with which the Greek and Roman antiquity, in its various aspects, has become accessible to large and diverse audiences. One notable example of such a web-based scientific community is the Thesaurus Linguae Graecae, TLG [tlg12], the best-known digital Library in classics, which has as a goal to create a comprehensive digital Library of Greek literature from antiquity to the present era. Also, the Perseus Digital Library [ple12], a hypertextual digital library in the humanities, which was launched in 1987 and got on the Web in 1995, is constantly evolving and expanding. The Perseus Project focuses primarily on Greek and Latin literature providing its users with texts, English translations, as well as short comments and links with dictionaries, maps and images.

Obviously, these two digital resources have changed the topography and the dynamics of Classical Studies perma-

DOI: 10.2312/VAST/VAST12/089-095

nently, thus breaking the ground for the creation of new digital collections and Libraries. The AKGDC (Archaia Kypriaki Grammateia Digital Corpus) project, though much smaller in scale and scope than either the TLG or the Perseus, aims, through the implementation of user friendly features, at the creation of a novel searchable digital library of Ancient Cypriot Literature.

This paper describes how this Digital Library has been conceptualized and developed; it also demonstrates how this new digital medium can be useful for a wide range of users (scholars, non-scholars, students and teachers) and how it will provide service, as an open-access searchable library, to both institutions and individuals who will be able to participate in live discussions and share innovative perspectives and experiences on key research ideas and data.

2. Archaia Kipriaki Grammateia overview

The field of ancient Cypriot literature has recently blossomed as an autonomous body of literary production and has been studied and taught as such over the past twenty years in academic institutions. This has not been the case before, as the Cypriot literature was traditionally taught and studied in conjunction with the rest of the ancient Greek literary heritage.

The six-volume corpus of the Archaia Kypriaki Grammateia (AKG) covers the ancient literary production of the Cypriot literature from the archaic period to the early Byzantine period, with the oldest surviving text probably dating back to the 7th cent. BC, and the most recent to the 5th/6th cent. AD. Evidently, the time span covered in the entire corpus is roughly thirteen centuries. It is the first systematic compilation of ancient Cypriot texts ever edited and commented upon meticulously, and was published in six volumes between 1995 and 2008 by the A.G. Leventis Foundation.

Each separate volume deals with a particular literary genre tracing the various phases of its development within and beyond the regional confines of Cyprus. The first volume focuses on poetry and embraces three representative sub-genres: epic, lyric and dramatic poetry. The second volume provides a compilation of 71 epigrams, most of which (48) were found on tomb stones (sepulchral epigrams), while 22 are classified as dedicatory epigrams. The third volume of the series contains edited fragments from 17 famous Cypriot prose writers (historians, mythographers and paradoxographers, grammarians, orators and sophists), while the fourth comprises the work of Cypriot medical writers and provides invaluable information on the development of medical science in Cyprus. The last two volumes are dedicated to philosophy; volume five deals exclusively with the work of Zenon of Kition, a major figure in Stoic philosophy, while the sixth volume contains the work of other Cypriot philosophers.

Overall, 65 different writers are listed in the six-volume edition, and the ancient text associated with them runs up to a total of 517 pages. Interestingly, this number represents the 14.3% of the entire page count in the printed edition. The remaining 85.7% of the printed version is taken up by lengthy introductions to each author, which place the textual material in a broader context, translations of the ancient texts in Modern Greek, critical apparatus (with different manuscript readings), extensive commentary on each individual text, images that are directly connected to the context of the texts, detailed bibliography, lists of abbreviations, and indices of ancient Greek and Latin words and phrases. The bulk of the printed text, therefore, forms the supporting background material of the ancient text and constitutes, at the same time, a major innovative feature of the Digital Corpus; this quite sizeable portion of the supporting searchable text of the AKG corpus distinguishes, in fact, significantly our collection from all the others, which are primarily text-centric, lacking any contextual information.

2.1. Digital Corpus in the web – State of the Art

Ancient Greek and Latin textual resources have been available to scholars in digital format through Digital Libraries since the mid 90s. The creation of a searchable digital library of the ancient Cypriot literature and its online availability will be a valuable tool and supplement to other large-scale digital libraries, such as the ones mentioned earlier. The AKGDC aims to attract multidisciplinary users at all levels of education, although it will be of particular interest to the specialist audience/readership, such as classicists, linguists, historians, archaeologists and philosophers. Moreover, its broad scope of themes can also offer textual resources for medical scientists and art historians, as well.

Nevertheless, the AKGDC targets also a wider audience which is generally interested in engaging with particular aspects of the ancient Cypriot literature. Modern Greek translations of the ancient texts will contribute immensely towards this goal by encouraging non-experts to delve into the rich material of the digitized corpus. At a later stage, the AKGDC will also offer English translations, as well, and will therefore gain more popularity among non-Greek speakers too. The main goal of the AKGDC is to be user friendly, but at the same time an intelligent tool in the hands of the user.

The AKGDC is distinguished by its innovative capabilities. Comparing its features with those offered by other reference digital libraries in the field, such as the TLG or the Perseus Digital Library mentioned earlier, one can readily notice the novelty of the contextual background provided by the AKGDC. More specifically, the Online TLG digitized collection contains over 10,000 works associated with 4,000 authors and is constantly updated and improved with texts and new features, such as the Ancient Greek Dictionary (LSJ)and the Lexicon for the Homeric Dialect (R.J. Cun-

liffe); these digital dictionaries are embedded in the TLG, whereby the user can check every ancient Greek word and its meaning over time. The Perseus Project, on the other hand, contains about 143 authors, Greek and Latin. The supporting searchable text of AKGDC (87,5%), which provides the contextual environment of the Cypriot literature, is either unavailable in the TLG, or only partially available in the Perseus.

Another innovative feature of the AKGDC is the text itself in conjunction with the critical apparatus. The text offered in the printed edition is, in many cases, different from the text we find listed in either the TLG or the Perseus. This is the result of a number of emendations and revisions of the ancient evidence introduced to the text by the authors of the printed volumes. As a consequence, the ancient texts in the AKGDC, replenished by the different manuscript readings provided by the editors in the critical apparatus, offer an updated and enriched version of the outdated editions of the texts which are often used in the other two digital libraries, together with a repository of alternative readings. Such varied textual information will then encourage the user to develop different ways of reconstructing a text and enhancing interpretation, processes which are less feasible through the other digital platforms.

Furthermore, as technology provides opportunities for innovations, the AKGDC project aims to go beyond the confines of the printed edition with the implementation of additional semantic features, which will enable scholars to pursue various lines of inquiry. Towards this goal, the AKGDC is offering hyperlinks of individual words with explanatory comments, clusters of concepts and updated bibliographical references, which are in turn linked to major University libraries and Library catalogues. Also, with the creation of a lemmatised database, the AKGDC is planning to provide links with major dictionaries (such as the LSJ and the Lewis Short), whereby the user can retrieve the full lemma of a word and its translation and uses across the ages.

Timelines 1 of writers and historical periods, maps indicating the ancient toponyms which feature in the texts, along with their modern counterparts, will also be part of the features provided by the AKGDC [ple12]. Maps and timelines will strengthen the temporal and geographical setting [MMB08] of Archaia Kypriaki Grammateia, while material such as 3D models and multimedia, will contribute to the visual enhancement of users' experience.

2.2. Use in education

Apart from individual use, the AKDGC can also be further implemented in the curricula of secondary and tertiary education. Already, the gradual publication of the AKG volumes has triggered the introduction of ancient Cypriot literature University courses, both undergraduate and graduate, into the University of Athens (Department of Philology). Also, an anthology of Cypriot literature, culled from

Figure 1: *AKGDC timeline*

the six printed volumes, has been recently introduced to the curriculum of the secondary education in Cyprus. There is no doubt that the digital version of the printed volumes will be a welcome and challenging addition to the teaching of ancient Cypriot literature at all levels, providing thereby alternative methods of learning and interacting. [Sha06].

3. The Archaia Kipriaki Grammateia Framework

Platforms such as the ones described previously require both human effort and technological infrastructure to assure their growth. However, as these specialized online libraries have evolved and gained popularity only within their specific communities, no infrastructure has been developed to allow data sharing, interoperability or reuse. In particular, in such platforms, data and information are very often organized and structured in many different ways without a common format for data exchange. Moreover, the use of the Semantic Web to exchange information among communities in a machine-readable format remains a challenge [BLHL01], although a lot of effort is put into Cultural Heritage institutions to create and share resources that are now becoming available, as W3C Resource Description Framework (RDF) triples (http://www.w3.org/TR/rdf-primer/).

CIDOC-CRM [CDG*06], LIDO [lid12] and VRA [vra12] are just some examples of the most widely used ontologies in the Cultural Heritage domain. The European Cultural Heritage Digital Library Europeana project [Pur09] that publishes circa 5,689,387 images, 3,059,966 texts, 373,110 audio files and 73,635 video files using a semantic framework has also released its own version [DGH*10] as well as the ongoing Michael Culture Aggregator [mic12], both of them acting as "aggregators" of client initiatives. From a technical point of view, one of the main objectives of AKGDC is to annotate the ancient text, the bibliography, the places and the peoples with terms and identifiers from these metadata schemas and resources, and to produce and serve linked data on the Semantic Web.

Wikis are nowadays very popular knowledge management tools, but wikis do not fully support structured search and knowledge reuse. Reversely, Semantic wikis do address

these requirements and as such, can be used for collaborative annotation and knowledge management systems development [ODM*06]. In fact, semantic wikis allow users to make formal descriptions of resources by annotating the pages that represent those resources. Where a regular wiki enables users to describe resources only in natural language, a semantic wiki provides additional functionalities to describe resources in a formal language: by using metadata over the ordinary wiki content, the semantic wiki users can achieve improved retrieval, information exchange, and knowledge reuse. As such, although both wikis constitute a useful technology for focused annotation, they do not efficiently support community/social tools.

Such capabilities are part of the AKGDC framework which is more oriented to facilitate the collaboration, the fruition of knowledge and discussion among the specialists and enthusiasts in the field. Our framework is built on top of the already semantically enable Content Management System Drupal version 7 [dru12]. The framework has been designed to enhance the existing digital content and enable the researcher to make use of knowledge and annotation available on Semantic Web technologies using the CIDOC-CRM and other ontologies.

Drupal is one of the most frequently used open source content management system available today. It is built in PHP and, in our case, it lies on a relational database where it stores its content. Not only is it very popular, but it is also easy to use and administer. Furthermore, Drupal is conceived in a pluggable way making it highly extensible and suitable for our special needs. In the next chapter, we describe our software design methodology, modules and architecture.

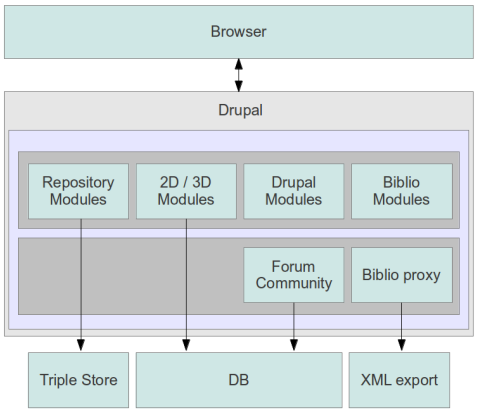

Figure 2: *System architecture of the AKGDC Framework.*

3.1. Interface methodology

Being a tool for non IT experts, our framework has to be designed with close interactions with the final users. Our design methodology is composed of four basic sequential steps, each of them providing a cycle of interaction: an analysis phase, a design phase, a prototyping phase and a development phase; the final goal is to understand and satisfy the user's goals, tasks and concerns at the early stages of the process.

Through this process we found that in order to become a primary destination for our intended audience our framework must include:

- A rich source of information via curated bibliography, publications and texts.
- Access to a trustworthy source, identical to the physical copy for its contents as well as its presentation.
- The possibility to add multimedia such as images, videos, audio (i.e. lessons or lectures) and 3D.
- New ways of presenting data, such as geolocalisation, timeline and browsing by subject.
- Groups with appropriate access control.
- A complete and referenced bibliography.

While taking into account the user's needs, we evaluate alternatives based on content presentation, navigation and hierarchy, interaction and ergonomy.

Figure 3: *Example of a complete page*

The result is, as shown in figure 3, a site that follows the structure of the six original books, presenting on the same page the content in ancient Greek, with hermeunetical notes and translations in modern Greek (in the future, English translations will be added). Once all the text is available in digital form, then every page will host a space for discussion, where the community members will be able to share comments about the text.

3.2. The AKG data enrichment

There are two different levels of semantics in the semantic web:

- passive semantics, the one embedded into the HTML code.
- proactive semantics, where the authors can inject a range of special tags into the HTML content that carry metadata, providing explanation on significant pieces of content for the evaluating software (i.e. Homer).

According to [Doe00], [PLP*06] and [BSM*05], we can represent one of our record in CIDOC-CRM as:

akg_node
E84.Information_Carrier "*panel*" →
P70B.is_documented_in → E31.Document "*AKGDC*"
panelId *T1*
E84.Information_Carrier "*panel*" →
P48F.has_preferred_identifier → E42.Identifier "*T1*"
title *T1*
E84.Information_Carrier "*panel*" → P102F.has_title →
E35.Title "*T1*"
E35.Title "*T1*" → P2F.has_type → E55.Type "*Main Title*"
→ P72F.has_language → E56.Language "*El*"
→ P73F.has_translation → E35.Title "*T1*"
E35.Title "*T1*" → P72F.has_language → E56.Language "*En*"
creator
E84.Information_Carrier "*panel*" →
P108B.was_produced_by → E12.Production "*Main Homer work*"
E12.Production "*Main Homer work*" →
P14B.carried_out_by → E21.Person "*Homer*"
E21.Person "*Homer*" → P14.1B.in_the_role_of →
E55.Type "*Writer*"
dataEntry 31/03/2012
E42.Identifier "*T1*" → P37B.was_assigned_by →
E15.Identifier_Assignment "*Recording of T1*"
E15.Identifier_Assignment "*Recording of T1*" →
P4F.has_time-span → E52.Time-Span * →
P82F.at_some_time_within → E61.Time_Primitive
"31/03/2012"
ownerEntry *The mapper*
E15.Identifier_Assignment "*Recording of*" →
P14F.carried_out_by → E21.Person "*The mapper*" **uri**
http://akg.cyi.ac.cy/1vol/ΜΕΡΟΣ Α/ΚΙΝΥΡΑΣ/T1
D13.Digital_Information_Carrier "*panel*" →
P48F.has_preferred_identifier → E42.Identifier
"http://akg.cyi.ac.cy/1vol/ΜΕΡΟΣ Α/ΚΙΝΥΡΑΣ/T1"
→ P2F.has_type → E55.Type "*URI*"

This metadata schema can be represented in RDFa as:

```
<!DOCTYPE html PUBLIC "-
//W3C//DTD XHTML+RDFa 1.1//EN"
"http://www.w3.org/MarkUp/DTD/xhtml-rdfa-
2.dtd">
<html xmlns="http://www.w3.org/1999/xhtml"
  xmlns:dc="http://purl.org/dc/elements/1.1/"
  xmlns:crm="http://www.cidoc-
crm.org/rdfs/cidoc_crm_v5.0.4.rdfs#"
  >
<head>
  <title>T1</title>
  <meta property="crm:P102F.has_title" con-
tent="T1" />
  <meta
    property="crm:P48F.has_prefered_identifier"
    content="T1" />
</head>
<body>
  <h1>T1</h1>
  <h2 about="http://live.dbpedia.org/page/Homer"
    typeof="crm:E21.Person"
    property="crm:P108F.has_produced"
    content="T1"><span>Hom.</span> &Lambda; 19-
28</h2>
...
```

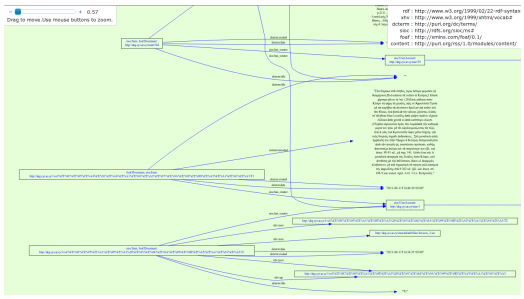

Figure 4: *Semantic graph example of a RDFa page*

4. Digital technology

Our main goal is to develop an innovative framework presenting new and challenging research findings in a more structured way. To reach that objective, the knowledge base has to be structured in a machine interpretable format to enable both data reuse and information exchange using the already described Semantic Web technology in addition to other technicalities.

4.1. Images and 3D models

The use of High Resolution images and maps will enhance the user's experience allowing him to interact in a different manner with the ancient text. Utilities like timeline, maps, 3D objects and graphs will empower the users with a new approach to and a new "vision" of the text. Many of the objects mentioned into the hardcopy of the AKG are available at the Nicosia Museum of Archaeology and some of them have already been acquired and semantically annotated in

3D and 2D at the Cyprus Institute and will then be added to the digital corpus.

4.2. Semantic and Taxonomies

The semantic enrichment of data can be done manually, coding RDFa into the actual XHTML format, or using some interactive tool. The encouraging results described in [kha11] motivated us to plan the integration of a similar tool into our AKG framework, with the necessary modification to support CIDOC-CRM RDFa notation.

While this represents the most important component and central feature of the corpus, it is still at an early stage. The development will lead to the extraction of terms and concepts in the form of a taxonomy. Such taxonomy will include terms and concepts found in the printed volumes, e.g. medical instruments, concepts about love or death etc, which will enable semantic searching in ancient Greek. Such classification of terms under clusters can enable a more scientific approach to the texts that goes beyond the reading and interpretation of the textual resources helping the user to contextualize and identify where and in what frequency these terms appear within the texts.

4.3. Bibliographic references

What is more, in the AKGDC framework the user can view the full bibliographical records that were used by the authors of the volumes. This constitutes an important tool for researchers in the field and is usually not easily available. Furthermore, the bibliographic record of the AKGDC is linked to external major libraries, such as the Library of Congress [loc12] and the Solo Library of the University of Oxford, as well as Google Scholar, where the user can access additional bibliographic information and full texts. The bibliographical references are accessible to the user in different ways and options such as sorted by the author's name in alphabetical order or by the publication date of the source.

5. Conclusion and future works

AKGDC's creation and presence on the Web is unquestionably a major challenge from both a technological and an educational point of view. We believe that the AKGDC fulfills the criteria of becoming an evolving and expandable digital library by incorporating new material and tools on the way.

At its current stage, the AKGDC contains fixed material, which is treated in a rather homogeneous way by the three editors, and is drawn from a fixed collection of volumes. The example of other digital libraries, apart from the TLG and the Perseus Project, has shown that the incorporation of additional material can easily contribute to their expansion [aph12]. In the case of the AKGDC, new material from the works of Christian Cypriot writers, or Byzantine writers,

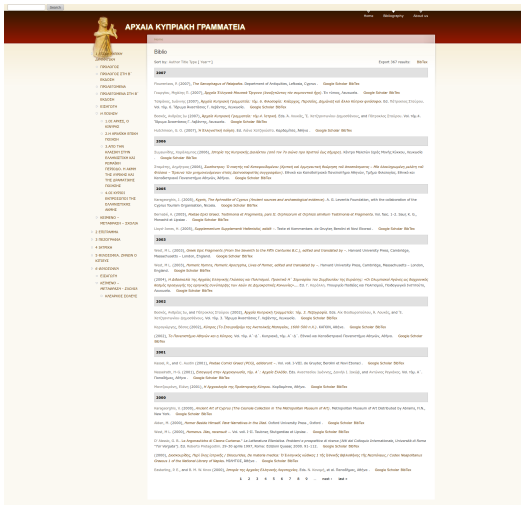

Figure 5: *Part of the bibliography, containing more than documented 340 sources*

for instance, can easily be embedded in the project providing therefore the user with a much wider spectrum of writing from later periods.

There are, of course, some issues at stake that should be appropriately addressed once the AKGDC gets on the Web, in order to make it popular and secure its future sustainability. Such issues are related to the evaluation of its utility and practicability by multidisciplinary users from all levels of education.

The creation of a complete, coherent, semantic framework is still a challenge that can be primarily achieved with the input and the help of a trained community. The RDFa looks very promising, although it is a very new technology, not wisely used in this field.

Acknowledgments.

This work was supported by grants from the A.G. Leventis Foundation.

References

[aph12] Inscriptions of aphrodisias project [online]. Nov 2012. URL: http://insaph.kcl.ac.uk/project/aphrodisias.html. 6

[BLHL01] BERNERS-LEE T., HENDLER J., LASILLA O.: The semantic web. *Scientific American 284*, 5 (May 2001), 34–43. URL: http://www.sciam.com/print_version.cfm?articleID=00048144-10D2-1C70-84A9809EC588EF21. 3

[BSM*05] BOEUF P. L., SINCLAIR P., MARTINEZ K., LEWIS P., AITKEN G., LAHANIER C.: Using an ontology for interoperability and browsing of museum, library and archive information. In *International Council of Museums 14th Triennial Meeting* (2005), ICOM COmmittee for Conservation. URL: http://eprints.soton.ac.uk/264790/. 5

[CDG*06] CROFTS N., DOERR M., GILL T., STEAD S., STIFF M. (Eds.): *Definition of the CIDOC Conceptual Reference Model*. ICOM/CIDOC CRM Special Interest Group, Oct. 2006. URL: http://cidoc.ics.forth.gr. 3

[DGH*10] DOERR M., GRADMANN S., HENNICKE S., ISAAC A., MEGHINI C., DE SOMPEL H. V.: The europeana data model (edm). In *World Library and Information Congress: 76th IFLA General Conference and Assembly* (Aug 2010). 3

[Doe00] DOERR M.: Mapping of the dublin core metadata element set to the CIDOC CRM. *ICS-FORTH, Technical Report 274* (2000). 5

[dru12] Drupal content management system [online]. Nov 2012. URL: http://www.drupal.org. 4

[kha11] Rdface - the rdfa content editor, 2011. 6

[lid12] Lido - lightweight information describing objects [online]. Nov 2012. URL: http://www.lido-schema.org/. 3

[loc12] Library of congress [online]. Nov 2012. URL: http://www.loc.gov/. 6

[mic12] Michael culture project [online]. Nov 2012. URL: http://www.michael-culture.eu. 3

[MMB08] MARTINS B., MANGUINHAS H., BORBINHA J.: Extracting and exploring the geo-temporal semantics of textual resources. In *Proceedings of the 2008 IEEE International Conference on Semantic Computing* (Washington, DC, USA, 2008), ICSC '08, IEEE Computer Society, pp. 1–9. URL: http://dx.doi.org/10.1109/ICSC.2008.86, doi:10.1109/ICSC.2008.86. 3

[ODM*06] OREN E., DELBRU R., MÖLLER K., VÖLKEL M., HANDSCHUH S.: Annotation and navigation in semantic wikis. In *SemWiki* (2006), Völkel M., Schaffert S., (Eds.), vol. 206 of *CEUR Workshop Proceedings*, CEUR-WS.org. 4

[ple12] The pleiades project [online]. Nov 2012. URL: http://pleiades.stoa.org/. 1, 3

[PLP*06] PITZALIS D., LAHANIER C., PILLAY R., AITKEN G., RUSSELL A., SMITH D. A., SINCLAIR P. A. S., ADDIS M. J., LOWE R., HAFEEZ S., LEWIS P. H., MARTINEZ K., M. C. SCHRAEFEL: Semantically exposing existing knowledge repositories: a case study in cultural heritage. In *The first international conference on Semantics And digital Media Technology (SAMT 2006)* (2006). Event Dates: 6-8 December 2006. URL: http://eprints.soton.ac.uk/264270/. 5

[Pur09] PURDAY J.: Think culture: Europeana.eu from concept to construction. *The Electronic Library 27*, 6 (2009), 919–937. 3

[Sha06] SHARIFABADI S. R.: How digital libraries can support e-learning. *The Electronic Library 24*, 3 (2006), 389–401. 3

[tlg12] Thesaurus linguae graecae [online]. Nov 2012. URL: http://www.tlg.uci.edu/. 1

[vra12] Vra core - virtual resources annotation [online]. Nov 2012. URL: http://www.loc.gov/standards/vracore/schemas.html. 3