The 10th International Symposium on Virtual Reality, Archaeology and Cultural Heritage VAST (2009)
K. Debattista, C. Perlingieri, D. Pitzalis, and S. Spina (Editors)

Managing Full-text Excavation Data with Semantic Tools

F. Niccolucci[1], A. Felicetti[2], S. Hermon[1] and K. Nys[3]

[1]STARC, The Cyprus Institute, Cyprus
[2]PIN, Universitá degli Studi di Firenze, Italy
[3]Mediterranean Archaeological Research Institute, Vrije Universiteit Brussel, Brussels, Belgium

Abstract

The paper deals with the management of archaeological data from excavations that cannot be easily processed using traditional forms and relational databases. It proposes a way of preserving the integrity of original information, including spatial relations. It demonstrates that also in this case CIDOC-CRM offers a valid schema that may be easily extended to incorporate geographic elements and relations. The system proposed is applied to a case-study concerning a Bronze Age 10 year-long excavation in Cyprus.

Categories and Subject Descriptors (according to ACM CCS): Information Storage and Retrieval [H.3.5]: Web—based Services

1. Introduction

Documenting archaeological excavations is an issue that does not attract the attention of researchers. It is usually considered an easy exercise by computer scientists, who expect archaeologists to adapt their information to structured schemas as necessitated by database theory. This is the reason that determined the usage of standardized forms in many countries, for example in UK with the MoLAS (Museum of London Archaeological Service, now Museum of London Archaeology) manual [Mus94], [Loc03] for the documentation of (mostly rescue) excavations, and in Italy with the ICCD (Istituto Centrale per il Catalogo e la Documentazione) archaeological forms [ICC] [D'A06]. This approach had the beneficial effect of introducing standardization and organization in archaeological documentation. The induced archaeological practice produced easy manageable structured systems, but it also created a cage where archaeologists often felt uncomfortable, and from where they found several ways of escaping. For example, "Note" fields were often overloaded with the "interesting" stuff, as comments or hints by the archaeologist, which could not find a suitable collocation elsewhere. But, there the information was destined to be lost forever, as in such systems fields are "atomic" by definition [Dat86] and the strict normalization rules of RDBMS forbid the recovery of any sub-field information. Nevertheless, this had little impact on archaeological documentation systems. Also the protests of post-processual archaeologists were taken into no account. This school advocated the use of multimedia and hypertexts to preserve the richness, multi-vocality and subjectivity of the archaeological record [Hod02], proposing documentation systems favoring flexibility but with little or no attention to the core goal of an information system. i.e. efficiency in information retrieval. In a similar way, a paper [Nic02], pointing out the risks of extending the usage of software tools outside the domain where they were born, was much praised and quoted, but had little effect on improving the "accountant informatics", as the above approach was - somehow jokingly - defined there.

The appearance of XML (1998) gave new hope to the advocates of documentation systems more respectful of the richness of information of the archaeological record. Initially proposed for historical sources, it was suggested that XML encoding of (archaeological) texts could provide a way of dealing with legacy data, mostly still on paper, in manuscript or printed form. This was proposed for example, for 19^{th} century archaeological records in [CDN02], and proposed and implemented in the Norwegian Museum Project [HU], [Ore98], [JHOO09] [HOE09]. This project concerned museum records of archaeological finds from University Museums in Norway, and developed an XML encoding system based on TEI [TEI]. The Museum Project concluded successfully with a complete conversion of the dataset from paper records to TEI-encoded digital records.

DOI: 10.2312/VAST/VAST09/125-132

The Museum Project team also undertook the encoding of some old (18th century) archaeological reports in a similar way in the framework of the EU ARENA [ARE] project, which aimed at providing networked access to archaeological records of Europe. Such pioneering work was more appropriate to the sources it dealt with, but had limits in efficiency and did not consider interoperability as a priority goal. Attempts to use XML-native DBMS had no good performance. Above all, such systems did not enable the creation of relations among the elements used for encoding, and allowed no solution for cross- and co-reference.

From another perspective, the introduction of CIDOC-CRM as ISO 21127 standard for cultural heritage documentation [CRM] opened new ways of dealing with archaeological data in an interoperable way. CIDOC-CRM has never been really popular among archaeologists, who are scared by its formal structure, and unfortunately has not received the attention it would have deserved. Nevertheless, the extensibility of CIDOC-CRM facilitates mapping current archaeological documentation systems to it. A number of the former have been mapped to a subset of CIDOC-CRM, introducing appropriate extensions when necessary. Among others, see [DMZ06]. In sum, digital archaeological excavation documentation systems may be grouped according to their creation as:

a) Born digital: these systems were created recently (say, from 1990 onwards) using standardized forms, suitable for data input into a relational database. No provision is made for free-text information, except free-text search. CIDOC-CRM compliance of the data model is the exception, ad-hoc structure being the rule, but mapping and conversions are available in most cases.

b) Paper-based, digitized: these systems comprise the vast majority of reports from the beginning of modern archaeology (18th century) until recent time. They include excavation diaries, notes, reports, and so on. A typical approach would preserve the original documents (e.g. as PDF or as texts) linked to a sort of summary records (e.g. introducing a form to be filled for each documentation "unit") to make the dataset searchable. This approach introduces a further filter between the source and the digital document, the interpretation and creation of such summaries, and requires cutting the source into chunks, the units corresponding to the summary records, with an arbitrary choice when to the original source has no "natural" units, for example a long discursive report as typical of the oldest sources. Using no such summary forms would enable only Google-like searches and would exclude any semantic query.

The system proposed in this paper for archaeological documentation belonging to the second group tries to:

i) Preserve the source integrity, making no changes, extraction or summarization in the original text.

ii) Provide efficiency in retrieval, using tools with good performance in searching.

iii) Guarantee interoperability, relying on CIDOC-CRM for encoding.

iv) Enable semantic tools, being based on RDF.

v) Support encoding by non-technical users availing of user-friendly interfaces.

An additional important feature of the present system is the incorporation of spatial information. Traditionally, this takes place by means of GIS, which are based on relational databases for the management of non-geographic information. Any other system not storing archaeological records in a relational database, as those mentioned above, could not be linked to a GIS. Our system incorporates some Python plugins for Quantum GIS [QGI], an open source GIS based on Grass, based on GML (Geographic Markup Language). We rely on the work developed in [DFLP09], which extended CIDOC-CRM to incorporate GML entities in it.

2. The HST Domain Ontology

The entities used for encoding the source belong to a domain ontology that is a subset of CIDOC-CRM, and is created in advance by examining the data. Some entities and relations are taken "as are" from CIDOC-CRM and are not duplicated. This is just a matter of usability: it may be decided to duplicate such entities, for example just to use a label more familiar to the user, and associate the replica with the original CIDOC-CRM entity; or to list them with their original CIDOC-CRM name and no duplication. For example, the entity E53.Place from CIDOC-CRM (all entities whose name begins with E belong to the CIDOC-CRM ontology) is rather intelligible to users and does not need to be duplicated. Others with names formally correct, but sounding difficult to the user or different from their common use, may be renamed and identified with the corresponding CIDOC-CRM concepts. Extension of the domain ontology relative to the case under study is however always possible later on, during encoding, as discussed below.

An important addition to the text annotation consists in a thesaurus and a gazetteer. The thesaurus may be created in advance and used to normalize the terms used in the text, by appending the standard form when necessary, or may be created at the end of the encoding, examining the terminology used in the various instances of the entities, and used in a second iteration to normalize the terminology. This is actually an easy process, because the semantic structure imposed on the text allows the selection of all the instances of a specific entity and the addition of the preferred term from the thesaurus in the relevant annotation.

The annotation tool described below creates an annotation file with pointers to the text, with the annotation (entity+instance+pointer) stored in a MySQL database. Semantic searches are performed in the latter and the system returns the text linked to the query results. RDF triples are managed in a similar way. According to different case studies, it may be decided to add also a simple TEI-like structure to the text, for example to retrieve the paragraph in which the result of a query is located. Images of the original document are stored as PDF scans and are available for inspection.

3. The Semantic Archive

The core of the whole system is the Semantic Archive (SAD) developed by extending the MAD [Fel06] technology developed during the EPOCH project and already tested and used in many different applications. It has been now designed to contain the digital version of all the documents edited and annotated with the AnnoMAD tool and the related annotations in a semantic format. The tool is completely written in PHP and take advantage of ARC2 RDF classes for PHP [ARC], an open source set of PHP libraries specifically oriented at the management of semantic information and query features.

With this tool it is is the possibile to store and manage every kind of HTML/XML based resource, including RDF and GML information, and to make them available for every kind of operation or request. The Semantic Archive interface, actually, is able to reply to many requests, supports many communication protocols such as REST, SOAP, JSON, and provides SPARQL query features to retrieve information in a semantic way.

The integration with the annotation tool is perfect and every operation on the Semantic Archive (including document and data creation, modification, deletion and so on) can be also managed remotely by using the features provided by the Flex technology and any communication protocol supported by the system. With the same mechanism it is possible to send normal and semantic queries and receive and show results formatted according with the preferences of users.

The system can also import external semantic resources and to export the whole archive or parts of it in RDF or other formats by using mapping mechanisms.

Geographical data are stored in GML. Geographic concepts may be considered as an extension of CIDOC-CRM, as demonstrated in [DFLP09] and [FL07].

4. System architecture

The proposed system consists of two parts. The first one is a simple GIS storing the spatial structure of the excavation and the relevant features. The second one is an annotation tool and query interface for the management of textual information. Also geographic concepts can be used to encode the source, so probably the geographic layer should be created first.

Fig. 1 illustrates the system architecture.

4.1. The GIS Component

For the implementation of the Geographic Information System, we have chosen Quantum GIS, a very easy to use rather powerful Open Source GIS developed by the Open Source Geospatial Foundation (OSGeo) under the GNU GPL License to visualize, manage, edit, analyse data, and compose printable maps. Our preference for Quantum GIS is related to the multi platform nature of this software (that runs natively on Microsoft Windows, Linux and Mac Os operating systems) and to the extreme flexibility of its modular architecture of core functions and plugins, allowing a high degree of personalization and integration with other frameworks. With minimal effort, it is actually possible to build complex applications on top of QGIS by creating new plugins or extending existing ones to define and implement robust data integration and data exchange operations. The support for GML and other open formats recommended by the Open Geospatial Consortium (OGC) makes QGIS the ideal to be used as a cornerstone for building integrated systems. The most recent release of QGIS available at the time of writing is 1.0.2, aka "Kore".

To create the GIS component it is necessary to digitize the relevant maps. They may be imported into QGIS to create the raster layers on top of which the vector layers providing the geographic information will be defined.

The raster maps are conveniently georeferenced with the built-in georeferencer plugin provided by the QGIS application. Then users can define the vector layer representing the entities having geographic relevance as points or polygons.

Once the creation of the vector layers is finished, geographic data are immediately available for query and publication. QGIS allows to perform many simple or complex spatial queries in the same way other GIS software does. An additional way of connecting geographic information results with data stored into the semantic container (SAD) is based on the possibility of defining specific "actions" related to geographic features, e.g. clicking on a polygon to display the semantic information related to it.

Fig.2 shows the outcome of a GIS search in the case study.

The system also enables the publication of the geographic data via the QGIS GML Publication plugin. We chose GML for the description of the vector layers because of its peculiar features and because its use is widespread among different systems, being it an XML-based grammar originally created by the Open Geospatial Consortium for the modeling, exchange, and storage of geographic information.

The GML grammar makes very easy the description of spatial objects using a variety of coordinate reference systems, geometry, topology, time, units of measure and generalized values to build a set of geographic features that can be considered a digital representation of the real world. Another fundamental reason for our choice is that GML is now the standard for communication among many geographic web applications and has a standardized syntax to describe geographic features, and easily understandable rules and formats for encoding spatial features, such as lines and polygons. The GML Publication plugin that we developed for the QGIS platform, written in Python, makes very easy all the operation of geographic data publication using GML. It allows data to be stored directly into the Semantic Archive as GML features and to be used during the process of annotation from the AnnoMAD tool. A typical GML feature description for one of the HST features, looks like this:

```
<ogr:HTSFeature fid="5061" desc="Wall">
      <ogr:geometryProperty>
        <gml:Polygon>
          <gml:outerBoundaryIs>
            <gml:LinearRing>
              <gml:coordinates>
532.804508218638603,-699.946782206814305
557.576633852669602,-705.046925719703154
552.476490339780867,-733.462011005797649
527.704364705749867,-727.633275562496237
527.704364705749867,-727.633275562496237
532.804508218638603,-699.946782206814305
              </gml:coordinates>
            </gml:LinearRing>
          </gml:outerBoundaryIs>
        </gml:Polygon>
      </ogr:geometryProperty>
    </ogr:HTSFeature>
```

Another very important operation that can be performed from the QGIS interface, is the possibility to export the entire project towards a WMS server, for example MapServer, from which clients can build customized maps. QGIS native capabilities allow also to reimport the whole project from the WMS server into another instance of QGIS in case of a brand new installation of the software.

4.2. The AnnoMAD Flex Application

The second tool deals with text data. They are digitized by expert archaeologists, who can understand the meaning of the original source content and solve any interpretation problem. The process works in a similar way with printed or typewritten sources and with manuscripts.

Text encoding may take place at the same type of digitization or at a second stage. For this activity an annotation tool is used. Fig. 3 shows the annotation manager for a record belonging to the pilot set of the key study and the aspect of the annotation interface in an example: parts of texts are selected in the main window and assigned to the appropriate entities listed on the right.

The AnnoMAD Application is composed of an Annotation Tool and a Query Interface for retrieving semantic information. The Annotation Tool is not just a tool for creating annotation but it is designed to build integrated semantic archives using the annotation paradigm. The tool interface has been developed using the Adobe Flex [FLE], a software development kit versatile and powerful for the development and deployment of cross-platform rich internet applications based on the Adobe Flash platform. Thanks to the Flash nature of the Flex framework, applications are perfectly cross platform and cross browser, since the Flash plugin, available for any kind of browser, allow the execution of any Flex application and solves the browser compatibility bugs that usually affect Javascript and Ajax applications.

Flex also provides developers with a wide range of communication protocols for the creation of real client/server applications. SOAP and JSON are the most important supported communication protocols, but many others can be used to interface Flex applications with online services and databases. Flex provides also a strong internal mechanism of communication with the most important web framework and technologies, such as PHP, JSP and ASP.

Moreover Flex applications can be deployed both as web or standalone applications thanks to Adobe Air, another interesting framework that acts like a Java Virtual Machine and allows the execution of Flex compiled software as local applications in a platform independent flavour.

4.3. The AnnoMAD Annotation Tool

The application we have developed can be used as a local application or accessed from remote. In any case the user will not perceive the difference since every Flex application, once loaded, acts as a normal system application, eliminating unnecessary page reloads and refresh operations typical of the classical web applications.

The AnnoMAD provides many features for the encoding of textual archaeological documentation and it can be considered the main access to the semantic data creation and management.

Mainly user can perform two operations: create and edit a brand new document, intended to be a digital copy of the paper version, or load an existing one already created or existing. The main panel looks like an online text editor and many editing features are also provided to format the document for publication. The panel gives also the possibility to define HTML links between two web resources, a very useful feature to create internal relations among documents. As a final result of the editing operations the user gets a standard HTML version of the document, ready to be published and shared over the web. It also also possible to export the current loaded document to PDF using the "Create PDF" button to save a local PDF copy of it.

The "Browse Original Document" loads and shows (if it exists) the scanned version of the original document to assist users during the editing operations.

What makes the applications totally different from any other online editing tool is the possibility to put a conceptual layer on top of the formatted text, either during the editing of the text itself or after the document has been finalized. Annotation editing works the same way of the other editing features: the user has the possibility to select a portion of the text and to assign an annotation to it by using the annotation features provided on the right part of the interface.

An annotation, in our view, is a piece of semantic information that can be attached to a whole document or a portion of it, in order to create a semantic description of it, to specify its meaning in a formal way and to establish relations with other pieces of annotated texts. For this purposes the user is provided with a rich set of concepts mainly derived from CIDOC-CRM ontology and from the geographic information coming from the archaeological GIS (QGIS).

Once the portion of text has been selected, indeed, it is possible to use the top right panel in order to chose which set of entities should be used for the annotation. The user can actually decide to use the concepts provided by the HST Domain Ontology, or to choose entities among a subset of the CIDOC-CRM ontology containing the most relevant entities of this ontology (Place, Actor, Objects and so on) or among the full CIDOC-CRM set of classes. Finally it is possible to use the spatial information coming out from the GIS activities, a very interesting feature that offers a straightforward mechanism to link together geographic and semantic information.

After the entity to use for the annotation has been chosen, the user can directly assign it to the text fragment by using the "Annotate" button. A popup will then come out and the user will have the possibility to give a name to the annotation and to add additional comments. The "Ok" button will finally create the annotation and its details will be shown in the "Annotations" panel.

The "Annotation" panel, containing the list of all the created annotations for the current document, gives also the possibility to edit existing annotations, to remove them and to select them for the definition of relations.

Relations assignment can be made by using "CIDOC-CRM Relations" panel. Once an annotation is selected from the "Annotations" panel, a list of all available properties will populate the "Properties" combo box

It is also possible to define relations with annotations related to other documents. The "Recall" button can be used to recall previously defined annotations stored in the semantic archive and to make it available according with the range of the chosen property. Finally, the "Add Event" button gives the possibility to define subsumed events not explicitly described within the text, and to use them to extend the relationship scope.

The "Save to DB" button will store both the edited document and its annotations into the semantic archive. A complete list of stored documents and the related annotations can be browsed by using the semantic db browser, accessible through the "Browse Semantic DB" button.

4.4. The AnnoMAD Query Interface

The AnnoMAD tool comes also with a set of query features accessible through the "Query Interface" button. The query interface provides mainly 3 types of query:

- **Simple query.** This is just a free text search into the whole archive that looks for a certain string or expression and returns results accordingly. The simple query can be though assisted by the HST Gazetteers able to provide geographic names, and by an internal thesaurus, based on the instances of entities defined both using the GIS system and the Annotation tool.
- **Semantic Query Builder.**From this panel it is possible to specify entities and relationships using all the classes and properties provided by the HST Ontology and by CIDOC-CRM. Additionally, the query can be refined by adding a string for domain or range entities. The "+" and "-" buttons can be used to add more relations to query for. Results of the query builder are visualized in real time every time that one condition changes.
- **Faceted Browser Semantic Navigator.** This panel is based on the faceted browser paradigm and allows user to navigate and slice the RDF graph stored in the semantic container by defining starting points (just clicking on the entity for witch the user is looking for) and proceede by restricting the search portion of the RDF data by selecting the related properties. Also for this "navigation" system, the results become immediately available as soon as the user selects one or more entities or properties. Results are shown in the "Query Results" panel.

Once the user has found what they were looking for and the results are visualized into the result panel, it is possible from here to visualize the geographic features, if any, concerning the resulting entities. If geographic information is available for a certain entity, indeed, the "Show Geographic Info" button will become clickable allowing the visualization of the spatial information related to the selected entity. If the user has the Quantum GIS software installed, the "Quantum GIS" button will send the user back to the GIS together with the GML information of the chosen entity in order to show geographic information and all the related GIS features. Finally, the "MapServer Map" button will send the GML code to the MapServer online application to create a snapshot of the map related to the chosen entity.

4.5. Final considerations on the System Architecture

With this system, it is possible to mix conceptual and spatial search parameters. They both operate on the semantic database, which includes the geographic concepts stored in GML. The outcomes may be displayed as lists or with maps [FL07].

One of the strong points of this system is its estensibility. It is possible to start with a draft ontology, and if later additions are needed they can be implemented with no problem, and also be propagated backwards using simple search tools. For example, assume that "Site_Area" is not included among the initial entities. Let us suppose that at some point, during the use of the system, it is decided that it would be preferable to have such an entity "Site_Area", not included in the initial ontology. It is sufficient to text-search the part already encoded for the occurrences of the possible instances of "Site_Area" (say, "port", "stoa", "house" and "sanctuary") and re-encode them accordingly. The new entity might correspond, for example, to the CIDOC-CRM E53.Place class.

The same semi-automatic mechanism might also be implemented to facilitate the encoding. It depends on the digitization process, in fact: if the digitized text is available before of the encoding, such semi-automatic processing may be helpful. If, on the contrary, digitizing and encoding proceed together, the advantage of such a mechanism is less relevant. A similar procedure may be applied to properties, searching the text for the occurrence of entities belonging to the property domain.

As a final consideration, co-referencing and cross-referencing may be easily resolved with this approach. It is sufficient to define the appropriate properties for co-reference/cross-reference resolution, for example P48.has_preferred_identifier, and annotate the text accordingly.

5. The case study

The system originates from, and has been tested on, a case study that has been proposed to us, concerning an important archaeological investigation in Cyprus, a Late Bronze Age town, situated in a field west of the mosque of Hala Sultan Tekke and the Larnaca Salt Lake. This site was excavated by two British expeditions at the end of the 19^{th} century, yielding a rich set of objects of gold, silver, bronze, faience, ivory and pottery. Some tombs were investigated by the Cypriot Department of Antiquities in the '60s, under the direction of Prof. Vassos Karageorghis. Since 1971, excavations were directed by the Swedish Prof. Paul Åström [Åst07]. Since 2001, Prof. Karin Nys of VUB (Vrije Universiteit Brussel) has been the Assistant Director and from 2009 she is in charge of all the archaeological research of this Late Bronze Age site. The town extends about 600 x 400 m. The house complexes generally consist of a courtyard surrounded by rooms, often with a paved platform, raised above the level of the street, in front of the entrance. They were used for living, as workshops and for commercial activities. The city also has remains of two buildings that have been interpreted as sanctuaries. An important area was the port on the current salt lake, in the past probably a lagoon with a natural or man-made access to the sea. The city has been populated from the 16^{th} century BC to the 12^{th} century BC, with some revival during the Hellenistic period. The finds are very rich and witness intense trade relations. The city economy was based on the copper industry, typical of this period in Cyprus. Data collected include also bio-environmental information, and a study on the palaeo-environment is under way.

Due to the premature death of Prof. Åström, all the records of the most recent excavations have not yet been published. They consist of notes taken by him and the trench masters using "forms" for finds and features, which were filled rather freely with their observations, frequently regardless of the headings present in the form. Prof. Nys, now in charge of publishing these records, has asked us to collaborate in the design and implementation of a computer system to manage the records. A joint documentation project has been setup between MARI (Mediterranean Archaeology Research Institute) of VUB and STARC. In the design, the peculiar needs of this case study have been taken into account: records are more similar to free text than to structured forms; geographical information is paramount; state-of-art tools are to be used, as semantic tools, standards and thesauri.

Due to the handwritten nature of the notes, the contribution of the archaeological team working with him - Prof. Nys *in primis* - is indispensable, because only they can decrypt the handwriting and abbreviations. Prof. Nys support has also been necessary during all the design stage of the documentation system.

So far, we have developed a pilot system with the Hala Sultan Tekke data and tested it on a limited number of records, with excellent results. The bulk of the digitization work is expected to require about a year, and will start in Autumn 2009. A draft ontology has been prepared and mapped to CIDOC-CRM.

6. Conclusions and further work

The system described in this paper may help in solving a great deal of problems with ancient sources and with documentation that was not born (or conceived) as digital. It requires only the creation of the domain ontology and its mapping onto CIDOC-CRM, a task in which our team has an extensive experience [DMZ06] [FL07] [ND06] [FM08]

Above all, it requires plenty of human work for the digitization. In the Hala Sultan Tekke case, this is necessary in any case and so does not require further effort. In other cases, it may be cumbersome to annotate manually the source, even with the help of the semi-automatic tool described above. In the Museum Project we heard that it was carried out by temporarily unemployed personnel, supported by a grant of the Norwegian government for socially useful works done by staff in job mobility. Fortunately for the workers, but unfortunately for us, this opportunity is infrequent. Therefore it will be necessary to find a way to provide such manual work. Perhaps, social networking may offer a solution. It is possible to add a wiki structure to the annotation tool to take advantage from the collaboration of many annotators. This leads to considerations about quality control that go beyond the scope of the present paper, but a sort of versioning may be easily implemented in our system.

Leaving apart such practical considerations, we would like to note that we are going to use the present system in another project concerning medieval archaeology and architecture in the Aegean Islands. In this case, sources are much different from each other and encompass archaeological records, historical sources and direct observations. It seems that a general-purpose tool as the present one, which is indifferent to source inhomogeneity, is the most appropriate one.

7. Acknowledgements

The case study was made possible thanks to the collaboration with Prof. Karin Nys.

The present work is partially supported by the European Project 3D-COFORM, contract number 231809.

References

[ARC] ARC2 RDF PHP Libraries. http://arc.semsol.org.

[ARE] The ARENA Project. http://ads.ahds.ac.uk/arena.

[Åst07] ÅSTRÖM P. E. A.: *Hala Sultan Tekke 1-12*. (SIMA 45: 1-12), Göteborg, Jonsered, Sävedalen, 1975-2007.

[CDN02] CRESCIOLI M., D'ANDREA A., NICCOLUCCI F.: XML Encoding of Archaeological Unstructured Data. In *Archaeological Informatics: Pushing the Envelope. Proceedings of CAA2001* (2002), Burenhult G., (Ed.), Archaeopress, Oxford, pp. 267–275.

[CRM] CIDOC-CRM. http://cidoc.ics.forth.gr.

[D'A06] D'ANDREA A.: *Documentazione archeologica, standard e trattamento informatico*. Archaeolingua, Budapest, 2006.

[Dat86] DATE C.: *An Introduction to Database Systems*. Addison-Wesley, Reading, 1986, pp. 241–243.

[DFLP09] D'ANDREA A., FELICETTI A., LORENZINI M., PERLINGIERI C.: Spatial and non-spatial archaeological data integration using MAD. In *Layers of Perception. Proceedings of CAA2007* (2009), Posluschny A., Lambers K., Herzog I., (Eds.), Archaeolingua, Budapest.

[DMZ06] D'ANDREA A., MARCHESE G., ZOPPI T.: Ontological Modelling for Archaeological Data. In *VAST: 7th International Symposium on Virtual Reality, Archaeology and Intelligent Cultural Heritage* (2006), Ioannides M., Arnold D., Niccolucci F., Mania K., (Eds.), Eurographics Association, pp. 211–218.

[Fel06] FELICETTI A.: MAD: Managing Archaeological Data. In *The e-volution of Information and Communication Technology in Cultural Heritage* (2006), Ioannides M., Arnold D., Niccolucci F., Mania K., (Eds.), Archaeolingua, Budapest, pp. 124–131.

[FL07] FELICETTI A., LORENZINI M.: Open Source and Open Standards for Using Integrated Geographic Data on the Web. In *VAST: 8th International Symposium on Virtual Reality, Archaeology and Intelligent Cultural Heritage* (2007), Arnold D., Niccolucci F., Chalmers A., (Eds.), Eurographics Association, pp. 63–70.

[FLE] Adobe Flex. http://flex.org.

[FM08] FELICETTI A., MARA H.: Semantic Web, Digital Libraries and the future of Cultural Heritage. In *VAST: 9th International Symposium on Virtual Reality, Archaeology and Intelligent Cultural Heritage* (2008), Ashley M., Hermon S., Proença A., Rodriguez-Echavarria K., (Eds.), Eurographics Association, pp. 117–123.

[Hod02] HODDER I.: *The Archaeological Process: An Introduction*. Wiley-Blackwell, Oxford, 2002.

[HOE09] HOLMEN J., ORE C.-E., EIDE Ø.: From XML-tagged Acquisition Catalogues to an Event-based Relational Database. In *Beyond the Artifact. Digital Interpretation of the Past. Proceedings of CAA2004.* (2009), Niccolucci F., Hermon S., (Eds.), Archaeolingua, Budapest, pp. 79–83.

[HU] Holmen, J. and Uleberg, E.: SGML-encoding of archaeological texts. http://www.dokpro.uio.no/engelsk/text/getting_most_out_of_it.html.

[ICC] ICCD: Normativa. http://www.iccd.beniculturali.it/Download/Normativa.

[JHOO09] JORDAL E., HOLMEN J., OLSEN S., ORE C.-E.: From XML-tagged Acquisition Catalogues to an Event-based Relational Database. In *Beyond the Artifact. Digital Interpretation of the Past. Proceedings of CAA2004.* (2009), Niccolucci F., Hermon S., (Eds.), Archaeolingua, Budapest, pp. 79–83.

[Loc03] LOCK G.: *Using computers in archaeology*. Routledge, London, 2003, pp. 86–89.

[Mus94] MUSEUM OF LONDON ARCHAEOLOGICAL SERVICE: *Archaeological site Manual*. MoLAS, London, 1994.

[ND06] NICCOLUCCI F., D'ANDREA A.: An Ontology for 3D Cultural Objects. In *VAST: 7th International Symposium on Virtual Reality, Archaeology and Intelligent Cultural Heritage* (2006), Ioannides M., Arnold D., Niccolucci F., Mania K., (Eds.), Eurographics Association, pp. 203–210.

[Nic02] NICCOLUCCI F.: XML and the future of humanities computing. *ACM Applied Computing Review 10*, 1 (2002), 43–47.

[Ore98] ORE C.-E.: Making multidisciplinary resources. In *The Digital Demotic: A Selection of Papers from DRH97* (1998), Burnard L., Deegan M., Short H., (Eds.), Publication 10, Office for Humanities Communication, King's College, London.

[QGI] Quantum GIS. http://www.qgis.org.

[TEI] Text Encoding Initiative. http://www.tei-c.org.

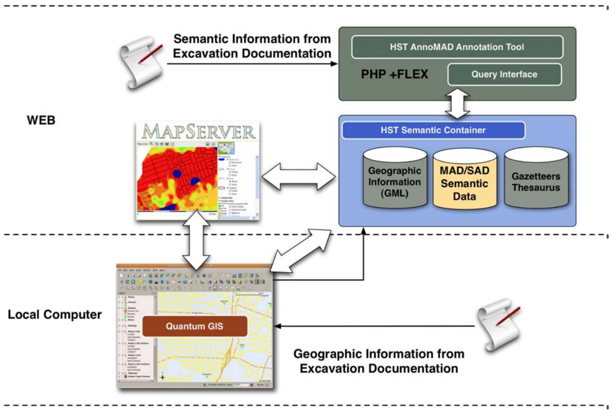

Figure 1: *System Architecture*

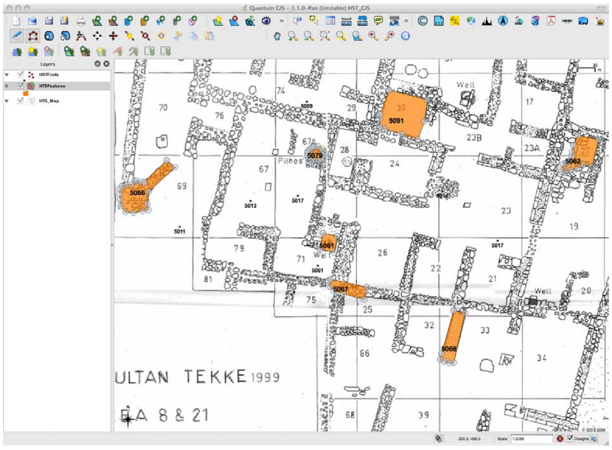

Figure 2: *GIS interface*

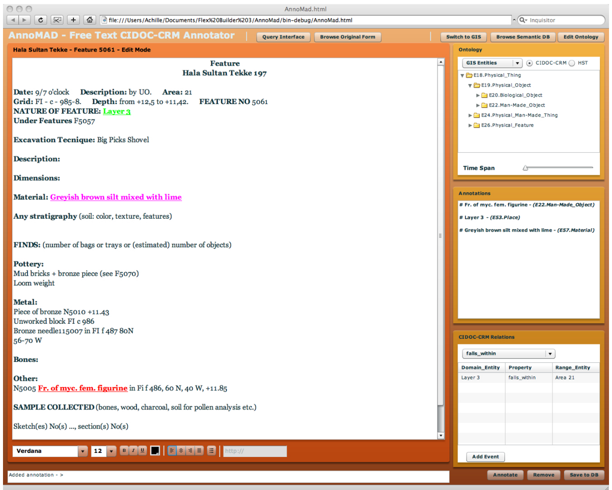

Figure 3: *Annotation Interface*