2007-4

An open Approach to Contextualising Heterogeneous Cultural Heritage Datasets

John McAuley
Technological University Dublin, john.mcauley@tudublin.ie

James Carswell
Technological University Dublin, james.carswell@tudublin.ie

Funder: T.Arc.H.N.A is a 3 year project partly funded by Culture 2000

2007-04-01

An open approach to contextualising heterogeneous cultural heritage datasets

John McAuley
Dublin Institute of Technology, john@dmc.dit.ie

James D. Carswell
Dublin Institute of Technology, jcarswell@dit.ie

An Open Approach to Contextualising Heterogeneous Cultural Heritage Datasets

John Mc Auley
Digital Media Centre, D.I.T.,
Aungier street,
Dublin, Ireland.
+35314023273
john@dmc.dit.ie

Dr. James Carswell
Digital Media Centre D.I.T,
Aungier street,
Dublin, Ireland.
+35314023264
jcarswell@dit.ie

ABSTRACT

This paper describes a semantic museum application, which aims to present a holistic impression of the Etruscan civilisation. Through the use of a distributed computing paradigm and the CIDOC CRM ontology, the system presents a unified view of a fragmented heritage, while supporting browse and search at a semantic level. Within the cultural heritage world, however, much value is placed on 'context', both in describing and presenting heritage artefacts. From this perspective, a platform built upon the distributed search paradigm, although useful in many respects, does not convey how an artefact sits within a broader setting. Narrative concepts are proposed as a way of reconciling heritage artefacts with their original context. A community of domain experts (i.e. Etruscan archaeologists and heritage professionals) is supported in contributing their knowledge and interpretation through a comprehensive authoring process. Narrative content is then organised according to several broad, hierarchically structured topics known as the 'Sphere of Knowledge' and a domain ontology describing the artefacts and monument of the Etruscan people. Each artefact is consequently represented through the text and associated with broader topics from the 'Sphere of Knowledge'. The artefact is therefore not presented in isolation or with lists of similar artefacts but rather discussed from a broader perspective. In our T.Arc.H.N.A system (**T**owards **Arc**haeological **H**eritage **N**ew **A**ccessibility), annotated narrative content, buttressed by references to real world artefacts, is disseminated to variety of platforms through a semantic web service. The entire approach is developed upon a multi-tiered architecture, allowing for the separation of functionality, yet supporting an open approach to interoperability.

Keywords

Semantic, Ontology, Database, Narrative, Context

1. INTRODUCTION

Cultural heritage artefacts, such as archaeological finds, are normally housed in disparate, often geographically remote museum collections. As such, the typical museum visitor, wishing to develop a deeper understanding of a heritage domain, is often frustrated by collections being fragmented across numerous heritage institutions. In addition, individual museum exhibition space is limited, resulting in artefacts lying in storage and away from the public eye for long periods of time. Increasingly however, museums and heritage institutions are investing resources in digitizing their collections. While much work has been carried out in the area of standards for digital cultural heritage, there is still no commonly agreed consensus on storing and retrieving this "new" digital heritage information.

As with other areas of information management, curators and museum professionals use a variety of approaches and systems to manage their digitised content. The conventional problems that burden the interoperability of heterogeneous datasets are therefore highly significant to the domain of cultural heritage. Further is the difficulty of "context": i.e., museum professionals continually grapple with exhibiting artefacts from a broader perspective. Antecedent approaches have addressed this problem by focusing on presenting a united view of museum collections. Conversely, we suggest capturing the domain expert's interpretation by way of narrative presentations, and conveying a unified and contextualised portrayal of a cultural heritage domain. Although the *TARCHNA* project focuses on Etruscan heritage, the approach is considered general enough to be applied across the sphere of cultural heritage.

This paper gives a brief overview of the *TARCHNA* system architecture, the tiered components, and reasons behind the approach. An explanation detailing the issues of context is provided, plus the proposed solution of using ontologies to describe collections, cultural and narrative concepts. The paper concludes with an example of an archaeologist creating a narrative presentation using the *TARCHNA* system, and a brief discussion illustrating the merits of the approach.

2. TARCHNA System Architecture

Typical of enterprise information systems, the *TARCHNA* system is divided into a multi-tiered architecture whereby each tier supports a clear division of labour.

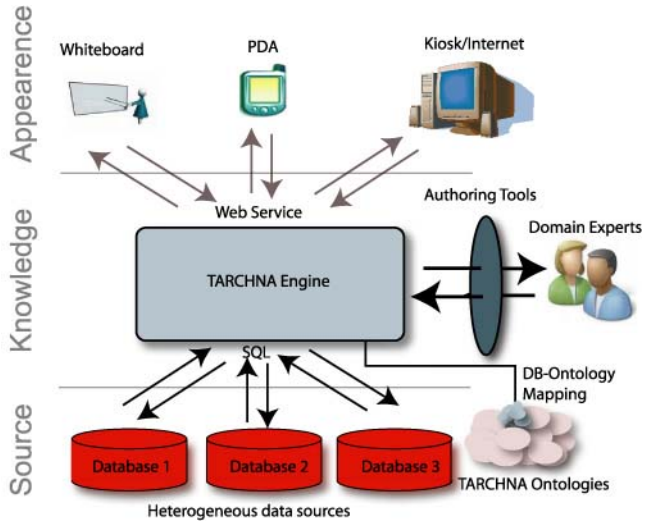

Figure 1. TARCHNA System Architecture

2.1 Source

The source tier consists of several heterogeneous data-sources, each exhibiting a separate (i.e. dissimilar) database schema, and three Ontologies developed upon the CIDOC CRM data standard [1]. The first or *TARCHNA* domain ontology provides a common reference model on which to integrate the different database schemas. This integration takes the form of a "Database to Ontology" mapping, whereby the elements defined by the database schema are expressed in terms of the ontology's concepts. Unlike other efforts, such as ARTISTE [2], the database mappings are held in the ontology. In this way, the mapping information can be accessed in the same way as instances of the ontology classes - thus removing the need for an external procedure to access mapping files or altering the individual database schemas. Two other ontologies, "Sphere of Knowledge" and "Narrative Ontology" also sit in the source tier of the system; each is discussed in later sections. All three ontologies are represented in the RDF formalism and stored in a sesame RDF store [3].

2.2 Knowledge

The knowledge or second tier is the central constituent of the system. It consists of several authoring tools, which support domain experts to develop narrative content, and the *TARCHNA* engine, which guides interaction between the *TARCHNA* domain ontology and the individual databases. The *TARCHNA* engine processes requests expressed as ontological concepts, and converts them into separate SQL statements relevant to each dataset. This process occurs at two different stages: firstly when authors wishing to write a narrative search the system for relevant artefacts on which to base their narrative content; and secondly when a request is accepted from the *TARCHNA* web service, the engine retrieves all narrative content related to a specific artefact, or conversely, retrieves all artefacts related to a specific narrative. The *TARCHNA* engine and authoring tools were developed in the Java programming language with the Protégé Ontology API [4, 5].

2.3 Appearance

The third and final tier acts as the disseminating component of the system, and distributes data (i.e. artefacts) enhanced narrative to several multimedia devices known as virtual wings (VW). The XML web service technology is used as means of interfacing the appearance layer with the virtual wings. It was felt that a service orientated and platform neutral architecture (SOA) supports a clear demarcation between the internal workings of a virtual wing and the overall data model of the system. In this way, new virtual wings may be added without a reliance on proprietary software or adjustments to the system architecture. Virtual wings are currently conceived as operating within three possible spheres:

- Firstly, as contextualised panoramic images. This innovative approach supports a comprehensive way of integrating conceptual models, such as the *TARCHNA* Domain & Narrative Ontologies, into panoramic images. The approach specifies semantic hotspots or trigger points whereby a visitor can query the image and receive information from the semantic model. It offers a new paradigm for accessing and interacting with semantically contextualised multimedia [6].
- Secondly, as handheld interactive tools. As both GPS[1] and PDA[2] technologies evolve into lightweight and economical location aware handheld devices, it has become increasingly possible to develop high bandwidth GPS applications for mobile devices. This VW is thought of as offering visitors a unique opportunity to explore ancient sites through real-time GPS based digital narrative, and thought of as similar to a personal guide.
- Thirdly, as a customised virtual museum. This option consists of narrative rich multimedia based applications operating within a museum space.

3. Adding Context through Narrative

The heritage domain is to be understood as consisting of expressions, some of which are tangible and others less so. Tangible heritage is embodied in physical objects and artefacts that give an anthropological significance to a society or people. As tangible heritage is considered both representative and metaphorical, its context however remains abstract and intangible [7]. This context is an amalgam of what Svensson calls *knowledge systems* or *life ways*, and relates to an artefact but is not intrinsically part of one [8]. In order to understand the significance of an artefact requires it to be presented within a broader context.

Narrative is proposed as a way of reconciling physical artefacts with their original intention or historic context and, in our case, presenting a holistic impression of Etruscan heritage. The aim is to support a team of domain experts (archaeologists, researchers, etc.) develop narrative presentations, which describe artefacts and their context within Etruscan society. In discussions with several archaeologists, the problem of assigning context to digital artefacts was raised. They suggested that in a cultural heritage setting, an artefact's context can be understood as a combination of its function and role within a specific society.

---

[1] Global Positioning System or GPS is a satellite navigation system.

[2] Personal Digital Assistants or PDA's are versatile handheld personal computers.

From this perspective, artefacts are presented as references to physical objects from the underlying datasets, many of which are accompanied by multimedia illustrations, while their context is woven into the narrative text and buttressed with ontology concepts, representing both function and role, from the Sphere of Knowledge (ontology).

3.1 TARCHNA Ontologies

In summary, the TARCHNA system uses ontologies to define narrative concepts and represent the domain to which they relate. Several distinctions were made to help formalise this process. Firstly, the domain was divided between aspects of tangible heritage, in the form of physical artefacts and monuments, and the broader concepts of Etruscan culture, such as economy, history, and religion. Each was represented by a separate ontology developed upon the CIDOC CRM data standard. The first, eponymously named the TARCHNA Domain Ontology, is a formal definition of Etruscan artefacts and monuments. It functions as an umbrella ontology for the addition of supplementary datasets without the need for replication across repositories, while supporting a faceted search paradigm, and presenting the user with a unified view of a fragmented heritage. The ontology was developed in coordination with a team of archaeologists who have extensive experience of Etruscan antiquity.

The second ontology however is a less formal representation, and describes the broader concepts of Etruscan culture. The 'Sphere of Knowledge' Ontology exhibits weaker semantics by way of hierarchically ordered terms. An explanation of each is provided with a (natural language) scope note. The motivation in using a less formal approach lay with supporting a community of domain experts. It was felt that the community should be involved in, as much as possible, the initial development and continuous refinement of the ontology. In this way the community's knowledge may evolve, and consequently be reflected in the ontology, with the addition of new collections and narrative content. This method was successfully demonstrated by Srinivasan during his work on the Village Voice project where he approached the development of structured knowledge in terms of community participation and mutability. [9] He refers to the concept as fluid ontologies, or 'flexible knowledge structures that evolve and adapt to a communities' interest [10].

The third and final representation is the *TARCHNA* Narrative Ontology. This draws on much of the work by Mulholland and others when formally describing narrative concepts [11-13]. Narrative is thought of as an epistemological container for communicating heritage content. It does this by specifying several properties which tie together concepts from both the *TARCHNA* Domain Ontology and the Sphere of Knowledge in a single narrative presentation. The former describes artefacts by way of direct relations or characteristics of artefacts through indirect relations, while the latter discusses broader domain concepts which often represent the function and role of an artefact.

Table 1. Illustrates the conceptual structure of a TARCHNA narrative and the relation between narrative and domain.

Property	Type	Description
Has title	title	Title of the narrative.
Has text	text	Text of the Narrative
Has author	author	Author of the narrative
Has direct relation	concept	Relation to artefact as represented in the ontology
Has indirect relation	concept	Relation to characteristics as represented in the ontology
Has contextual relation	term	Relation to terms from the 'Sphere of Knowledge'

TARCHNA Narrative is stored as class instances in the narrative ontology. In this way it is abstracted from the underlying datasets, but can still reference database objects via *direct* and in*direct* relations.

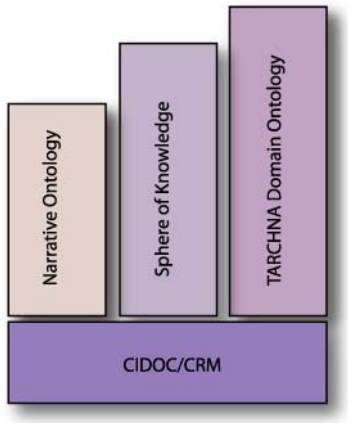

Figure 2. TARCHNA Ontologies.

4. Authoring Scenario

The authoring process, illustrated in Figure 3, takes place in the knowledge tier of the system. Domain experts are provided with a personal narrative space, in which they can add, edit, and delete narrative content. The authoring process is divided into a number of steps, each contributing towards a completed presentation.

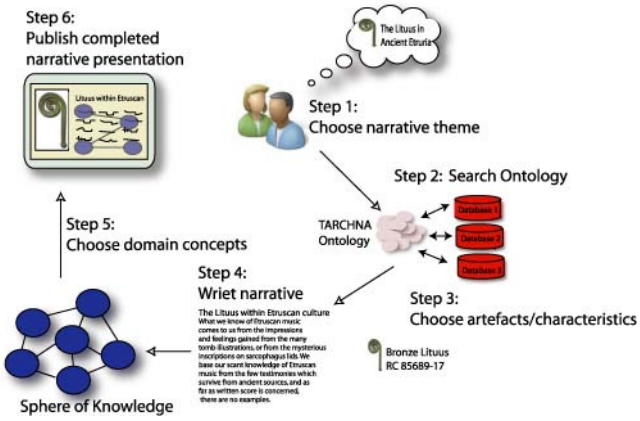

Figure 3. Developing a narrative presentation with the TARCHNA authoring tools.

The following scenario illustrates how a domain expert (*Tom*) contributes narrative content to the *TARCHNA* system:

> *Tom*, an archaeologist working on a dig in Tarquinia (an ancient city in Italy), wishes to discuss Etruscan musical instruments as a whole but would like to focus on the Lituus[3] as an example of the "fruits" of Etruscan culture. He approaches the system with several concepts in mind. Naturally, he wishes to feature the artefact itself**,** and would further like to discuss its role and function within Etruscan society. In step 1 of the authoring process he is asked to choose the theme of his narrative. The theme denotes the nature or broad idea of the text and, from the systems point of view, illustrates the author's interests. For this example, *Tom* wishes to discuss musical instruments and therefore chooses the theme *Finding* from the list presented in Figure 4.

Figure 4. Step 1: choose from a list of several themes on which the author wishes to base their narrative presentation.

> Each theme is supported by a faceted search interface, which uses concepts from the *TARCHNA* Domain Ontology to specify the search criteria across all heterogeneous datasets. In this example, *Tom* is presented with the search interface specific to the theme of *Finding* (Figure 5). The interface highlights a number of characteristics relevant to archaeological findings, such as provenance, inscription, depiction, etc. *Tom* is generally interested in musical instruments, and therefore chooses this concept as the finding class. He further specifies the shape as Lituus and the material as bronze (as is often the case with this type of musical instrument). When satisfied with the choice of criteria, he submits the search to the system.

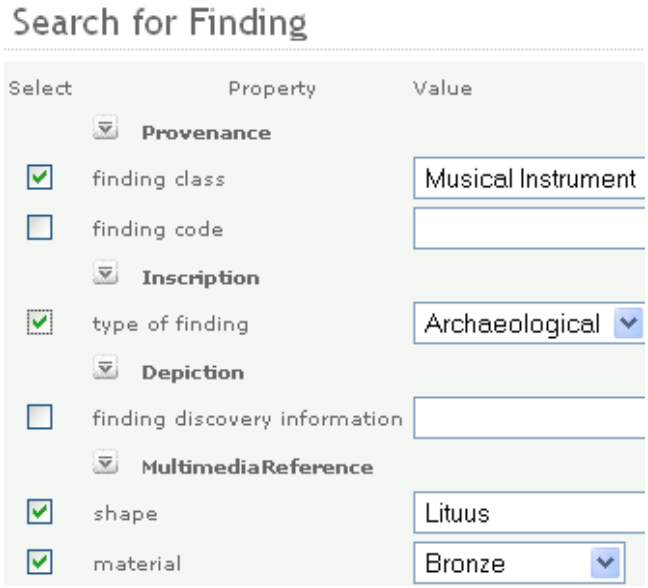

Figure 5. Step 2: the author is presented with a faceted search interface specifying a number of characteristics related to Etruscan findings. The interface is made up of concepts from the underlying domain ontology.

> At this point the *TARCHNA* engine (Figure 1) receives the search criteria and translates the ontology concepts into individual SQL statements relevant to each of the heterogeneous datasets. This process involves querying the *TARCHNA* Domain Ontology (stored in RDF) for the mapping information applicable to each database. The resulting RDF triples are transformed into SQL statements (Figure 6) and each database is queried.

```
SELECT `Findings`.`ClassID` AS `FindingClass` ,
`Findings`.`Type_of_Finding` AS `FindingType` ,
`Findings`.`Object` AS `FindingShape` ,
`Findings`.`MaterialID` AS `FindingMaterial` ,
`Findings`.`FindingID` AS `FindingID` FROM `Findings`
WHERE `Findings`.`ClassID`="Musical Instrument" AND
`Findings`.`Type_of_Finding`="Archaeological" AND
`Findings`.`Object`="Lituus"
AND `Findings`.`MaterialID`="Bronze"
```

Figure 6. SQL query generated, from (TARCHNA domain) ontology concepts, by the TARCHNA engine. The properties chosen in the previous figure are highlighted in yellow.

> The results are returned to the TRAHCNA engine, where they are correlated into a single resource and sent back to the author, as demonstrated in Figure 7.

[3] The Lituus was both a crooked staff, usually held by influential individuals, and an L-shaped wind instrument. Although it functioned as a musical instrument, its role was often during religious rituals.

Figure 7. Step 3: the results of a search for describing the wind instrument Lituus. *Indirect Narrative* allows the author to write about the characteristics of the Lituus without reference to an actual artefact; conversely, *Direct Narrative* supports the author in discussing an actual artefact from the system.

The results of *Tom's* search are divided under the headings of *Direct* and *Indirect Narrative*. While *Direct Narrative* discusses artefacts with reference to specific database objects, an *Indirect Narrative* discusses characteristics of artefacts as represented by concepts within the *TARCHNA* Domain Ontology (e.g. Shape: Lituus, or Material: Bronze).

There are several incentives motivating this approach. Firstly, authors are supported in discussing artefacts from a general perspective, without relying on reference to a specific database object. This could amount to a discussion on Etruscan musical instruments, without explicit artefact references, but with an indication as to the shape of Lituus for example and consequently to any artefact of that type. It is suggested that the approach could draw on a more active participatory role from the reader, as the narrative acts as a gateway to further exploration of, in this case, Etruscan musical instruments. Secondly, the concept of indirect narrative supports collections that may be added to the TARCHNA system at a later date. For example, let's consider artefacts with the shape of a Lituus that are discussed by an indirect narrative. If a new collection is added and, following the mapping procedure, there are new artefacts of shape Lituus present, those artefacts are immediately associated with that indirect narrative. Thirdly, an author wishing to contribute to the system's content is not discouraged from doing so by the absence of a particular artefact, and is instead proffered with the opportunity to contribute, albeit from a more general perspective.

Returning to the example, it can be seen from Figure 7 that, in this instance, *Tom's* search yields a reference to a bronze Lituus with the uid *RC 85689-17*. *Tom* decides therefore to concentrate on a *direct narrative* discussing the value of the Lituus within Etruscan culture. He chooses the Lituus reference (as illustrated in Figure 7) and proceeds directly to writing his text.

Figure 8 illustrates the *TARCHNA* Narrative Page. The author is presented with a Title Box (1), the Sphere of Knowledge or listing of terms broadly fitting the Etruscan domain (2), the artefact chosen as the corollary of the previous search (3), and a larger Text Box (4). *Tom* enters the title 'The Lituus within Etruscan culture' into the title box, and begins to write his text.

Figure 8. Step 4: narrative authoring screen, comprising of title (1), text (4), chosen artefact (3), and the Sphere of Knowledge - lightweight Etruscan ontology (2).

Having completed his text, *Tom* chooses the terms from the Sphere of Knowledge which best describes his narrative content. Again, the sphere of knowledge is a less formal ontology illustrating the broader aspects of Etruscan culture. Each term, from A*rt and Artefacts* to *Environment and Landscape*, represents the top level of the ontology. By clicking on a term, the author expands the ontology tree and a more specialised branch of the hierarchy is displayed. In this example, *Tom* is discussing a particular type of musical instrument, therefore the term *Customs* is chosen, followed by the more specialised term of *music and musician* depicting the artefact's function. However, a Lituus had a different role in Etruscan society, it was often used during votive offerings and other religious rituals, and as a result *Tom* expands the term of religion choosing both *Rituals in a sacred context* and *Offering* (Figure 9).

Figure 9. Step 5: expanding the ontology branches, *Tom* chooses the terms which best fit his narrative content.

When satisfied, *Tom* saves the finished product into the *TARCHNA* system. The narrative is comprised of title, text, a direct relation to the artefact lituus, and a reference to the function, *music and musicians*, and the role *rituals* and *offering*. Once saved it is available for dissemination by the *TARCHNA* web service.

Figure 10. Step 6: the completed narrative is stored in the author's personal narrative space.

5. Discussion

This paper introduced a unique way of presenting geographically disparate heritage collections. The key advantages of which are listed below:

5.1 An open approach to interoperability

Separating responsibility between tiers serves a very specific purpose by way of semantic interoperability, and promoting new and exciting ways of accessing cultural heritage information. A key principle behind this approach was that the system rely on no single data model, therefore databases can be added or removed with the minimal of effort. In this way, amendments to a data source do not impact the underlying semantic structure, and through the TARCHNA web service, developers are encouraged to invent new ways of exploring the narrative content.

5.2 A collected view of distributed heritage

Heritage collections are often distributed across several, geographically remote, museum databases. By separating functionality between tiers and mapping collections into single umbrella ontology, the *TARCHNA* system presents a collected view of a distributed heritage. While users are supported searching multiple datasets, data replication is avoided, and cultural institutions retain tutelage over digitised collections.

5.3 Enhanced data dissemination with contextualised narrative content

Heritage professionals have acknowledged the importance of "context" when presenting artefacts to the general public. By foregrounding artefacts within a narrative backdrop, it is proposed that objects are considered from a broader contextual perspective.

6. Conclusion

The *TARCHNA* system presented in this paper proposes a novel way of contextualising heterogeneous datasets through the construction and presentation of knowledge intensive narrative. The system hinges on an open approach to information by promoting a clear separation of source, knowledge, and appearance. The multi-tiered architecture, while supporting semantic integration of heterogeneous datasets and avoiding data replication, provides a platform independent way to interact with and disseminate knowledge based narrative.

Currently, the system is being used by a number of archaeologists developing a suite of narrative discussing varying aspects of Etruscan heritage. While the approach was developed to support cultural institutions to amalgamate artefacts and present a holistic understanding of a specific heritage, it is not proprietary to subject matter or domain. The multi-tiered architecture supports data integration at both the procurement and dissemination stages, while the knowledge layer exploits narrative as a unifying platform, and presents both knowledge and data in an engaging format.

7. ACKNOWLEDGMENTS

T.Arc.H.N.A is a 3 year project partly funded by Culture 2000 (2004 -1488/001 -001 CLT-CA22) that started in September 2004. Seven project partners are involved; the University of Milan in Italy, the Computer Science Department at Claude Bernard University in Lyon, France, the Institute of Archaeology/ Art Collections at University of Bochum, Ruhr, Germany, the Dublin

Institute of Technology and University College Dublin both based in Ireland, the Department of Classical Archaeology at the University of Warsaw, Poland and the Aristotle University of Thessaloniki situated in Greece. Further information is available from the project website [15].

8. REFERENCES

[1] Crofts, N., et al., *Definition of the CIDOC Conceptual Reference Model*. 2005, ICOM/CIDOC. p. 94.

[2] Allen, P., R. Vaccaro, and G. Presutti, *ARTISTE: An integrated Art Analysis and Navigation Environment.* Cultivate: interactive, 2000(1).

[3] Broekstra, J., A. Kampman, and F.v. Harmelen. *Sesame: A Generic Architecture for Storing and Querying RDF and RDF Schema*. in *International Semantic Web Conference*. 2002. Sardinia, Italy.

[4] Gennari, J.H., et al., *The evolution of Protege: an environment for knowledge-based systems development.* Int. J. Hum.-Comput. Stud., 2003. **58**(1): p. 89-123.

[5] Noy, N.F., R.W. Fergerson, and M.A. Musen, *The Knowledge Model of Protege-2000: Combining Interoperability and Flexibility*, in *Proceedings of the 12th European Workshop on Knowledge Acquisition, Modeling and Management*. 2000, Springer-Verlag.

[6] Mazzoleni, P., et al. *Towards a contextualized access to the cultural heritage world using 360 Panoramic Images*. in *Software Engineering and Knowledge Engineering* 2006. San Francisco.

[7] A.K.Das. *Intangible Heritage: Dynamics of Conservation in Museum of Living History*. in *ICTOP - International Committee for the Training of Personnel - Annual Conference (ICOM)*. 2004. Seoul, Korea.

[8] Svensson, T.G. *Knowledge and context - the social life of objects* in *Museums and Intangible Heritage ICOM general conference*. 2004. Seoul, Korea.

[9] Srinivasan, R. *Village Voice: An Information-based Architecture for Community-centered Exhibits*. in *Museums and the Web*. 2003. Toronto Ontario.

[10] Srinivasan, R. and J. Huang, *Fluid ontologies for digital museums.* Int. J. Digit. Libr., 2005. **5**(3): p. 193-204.

[11] Mulholland, P. and T. Collins. *Using Digital Narratives to Support the Collaborative Learning and Exploration of CH.* in *IEEE International workshop on Presenting and Exploring Heritage on the Web (PEH'02) in conjunction with the 13th International Conference and Workshop on Database and Expert Systems Applications (DEXA 2002)*. 2002. Aix-En-Provence, France. .

[12] Mulholland, P., T. Collins, and Z. Zdrahal. *Story Fountain: Intelligent support for story research and exploration*. in *Intelligent User Interfaces (IUI'2004)*. 2004. Madeira, Portugal.

[13] Mulholland, P. and Z. Zdrahal. *Knowledge Support for Story Construction, Exploration and Personalisation in Cultural Heritage Forums.* in *14th International Conference and Workshop on Database and Expert Systems Applications (DEXA 2003)*. 2003. Prague, Czech Republic.

[14] Brooks, K.M. *Do Story Agents Use Rocking Chairs?* in *Proceedings of the fourth ACM international conference on Multimedia*. 1997. Boston, Massachusetts, United States: ACM Press New York, NY, USA.

[15] TARCHNA. *T.Arc.H.N.A. Towards Archaeological Heritage New Accessibility*. 2005 [cited; TARCHNA project website]. Available from: http://www.tarchna.org/.