Data Management

The 5th International Symposium on Virtual Reality, Archaeology and Cultural Heritage VAST (2004)
K. Cain, Y. Chrysanthou, F. Niccolucci, N. Silberman (Editors)

Arcada – An Archaeological Database Management System on the Internet

P. Staynov[1], N. Genov[1], N. Spasov[2]

[1]New Bulgarian University - Department of Informatics, Sofia, Bulgaria
[2]New Bulgarian University - Department of Archaeology, Sofia, Bulgaria

Abstract

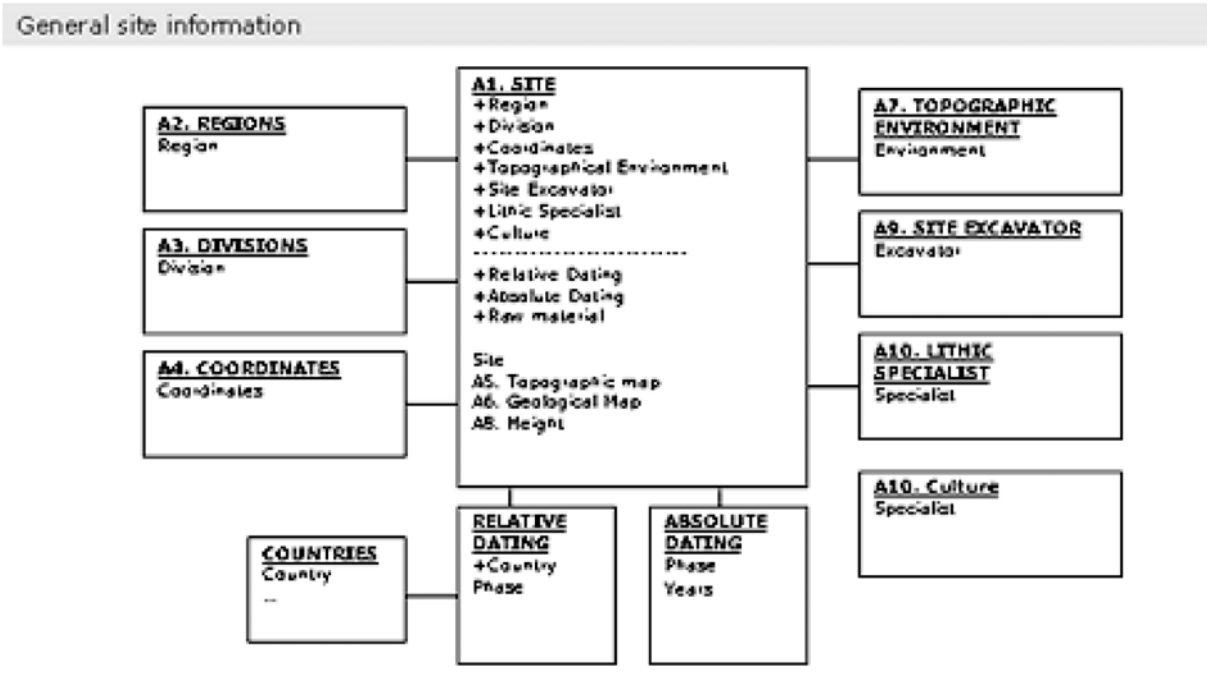
Arcada is a database management system developed by the departments of Informatics and Archaeology of the NBU. The existing implementations of the system concern different lytic assemblages from prehistoric sites on the territory of Bulgaria, Greece and Turkey. Arcada has been connected to a GIS in order to permit the presentation of data on a digital map of the region.

The main purposes of Arcada are:

- to allow the storage of data about archaeological artifacts permitting further quantitative processing;
- to avoid mistakes, ambiguities and misinterpretation of data by the imposition of data representation standards;
- to allow the exchange of data provided by different researching teams;
- to allow the common processing of data coming from different archaeological sites;
- to allow the spatial representation (on a digital map) of gathered and processed data.

Arcada is destined to the following kinds of archaeological artifacts:

- Normally it can be used with any kind of archaeological artifacts.
- The modules developed so far concern assemblages of stone artifacts, but the open architecture of the system allows the users having the necessary rights to set it for any other kind of material.
- The only limitation could result from analytical approach which stands at the bases of the system – the attribute analysis



Arcada's basic characteristics are:

- client/server architecture
- multilingual interface
- different levels of access
- multimedia features
- support of numerous input and output data formats
- offline/online mode of functioning

Arcada's basic modules

The system comprises two main types of modules:

- **general modules –** dedicated to general archaeological data – site description, absolute and relative dating, institutional context, copyrights, etc.
- **specific modules** – concerning the different artifact assemblages and collections and the different descriptive and analytical approaches

All the modules are open in manner – the users of the system can either add, or modify the already entered attributes and characteristics, depending on the rights they have been assigned with.

Arcada is based on the following technology :

On the server side :

- MySQL server
- PHP scripts
- XML parser and DOM

On the client side :

Offline version

- Browser
- XSL

Online version

- DOM
- Browser

Online access – levels of access (roles)

The system allows the definition of different roles depending on the access and action permissions.

Typical roles:

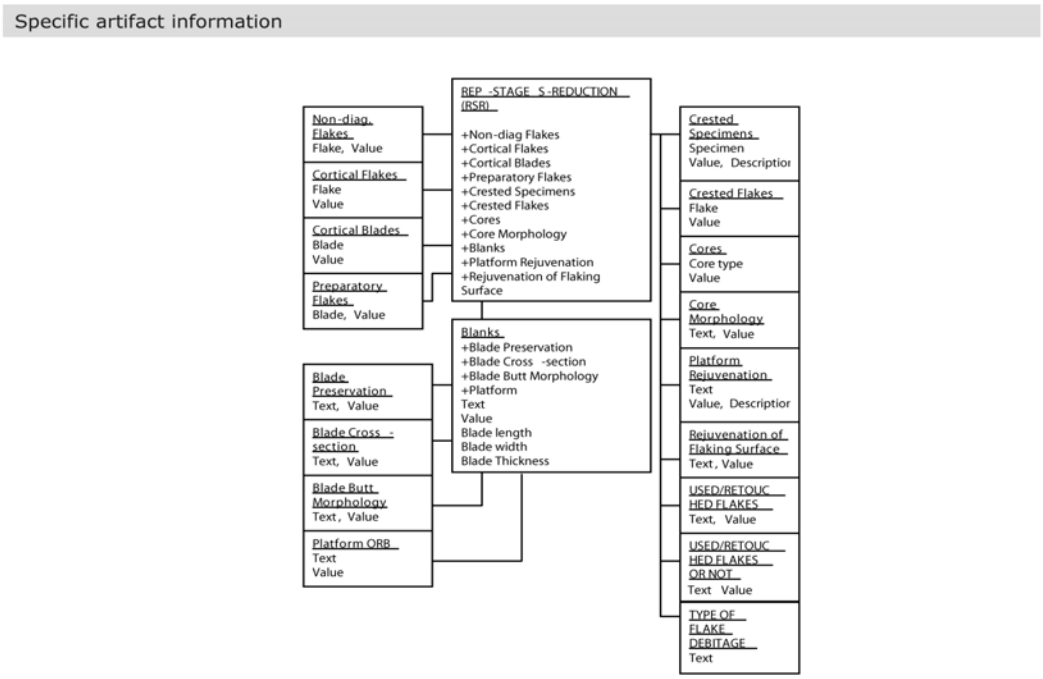
Super-administrator – he is the main administrator of the system; he defines and assigns roles (especially with administrator's rights), creates modules and tables. He has not the possibility to edit and delete data entered by the other users of the system.

Administrator – he defines roles (operators of different levels), creates modules and tables, links tables and enters/edits/deletes data in both the master and the slave tables.

Operator – enters/edits/deletes data in the slave tables.

Offline functioning

1. Download of a data scheme from the server or creation of a new scheme by means of a special application (HTML format – browser). The scheme is stored on the PC in XML format
2. Visualisation fo the scheme in HTML format (XML – XSL – HTML)
3. Modification of the scheme (if needed) by the creation of new tables, fields or the definition of new values (HTML format – browser).
4. Entering/editing/deleting data (HTML format – browser).
5. Saving the modifications (HTML interface – DOM – XML file)script
6. Uploading of the scheme filled wit data on the server (XML – DOM – PHP script – MySQL DB



The 5th International Symposium on Virtual Reality, Archaeology and Cultural Heritage VAST (2004)
K. Cain, Y. Chrysanthou, F. Niccolucci, N. Silberman (Editors)

Heritage and Information

I. Rizzo[1], A. Mignosa[2]

[1]Department of Economics and Quantitative Studies (DEMQ)
University of Catania (I), rizzor@unict.it
[2]Erasmus Univeristy - Rotterdam (NL), mignosa@fhk.eur.nl

Abstract

The aim of this short paper is to sketch, with an economic perspective, the main features of the decision-making processes underlying heritage production and consumption. This paper proposes a possible guideline which, if further developed, would lead to the design of the information flows models needed to support the decision-making processes. The model would also serve to develop the architecture and procedures for the information system management.

It is widely recognized that the conservation [WB94] and enhancement of cultural heritage can be a relevant input for a sustainable economic and social development. In fact, these activities, in addition to the potential economic benefits deriving from tourism, may expand the cultural capital of a country, may increase the value that nationals attribute to their historic heritage and, therefore, their willingness to pay for consuming heritage services. For this to happen, though, some conditions, concerning the decision-making process related to the production and consumption of heritage and the distribution of information within it, have to be satisfied. It is worth noting, as it will be outlined later, that the decision-making process may occur according to different patterns, depending on the actors and the specific heritage issues involved. In what follows, given the limited scope of this paper, a possible guideline to address this issue is only sketched. If further developed, this analysis would lead to the design of the information flows models needed to support the decision-making processes. The model would also serve to develop the architecture and procedures for the information system management.

According to the economic literature in the heritage sector the market "fails" because it is not able to ensure an *efficient* outcome, i.e. an outcome which maximizes society's well being. Among the various causes of 'market failure' here attention is concentrated on the asymmetrical and incomplete information characterizing the heritage decision-making process about heritage, its causes and its remedies. Such a process is complex because it results out of the interaction of many actors – public, private, no profit

– with various flows of exchanges – both monetary and real – occurring among them. For example, central and local governments exchange grants or in-kind services with private cultural institutions and they exchange services with single consumers as well as with large audience; donors and sponsors exchange funds with institutions. These exchanges are usually characterized by trade-offs between conflicting goals [Tri03] – for instance, between preservation and access to heritage, between the appreciation of regular consumers and the expectations of the occasional ones – and by the asymmetrical distribution of information among the actors involved.

In economics, this is the well-known *principal-agent* paradigm; i.e. a relationship in which the *principal* delegates power to the agent to act on his/her behalf and information is asymmetrically distributed in favour of the *agent*. The *principal-agent* paradigm occurs within the public, private and no profit sectors and in the relationships among them. For instance, in the public sector, at each level of government, politicians are the *agents* of citizens and, at the same time, they are the *principals* of the bureaucrats who have been delegated authority to produce public policies [MR01]. Alternatively, Governments rather than directly producing cultural services may delegate their production to private or no profit institutions with various financial arrangements taking place. In all these decision-making processes monitoring is needed to induce *agents* to act according to the *principals'* preferences.

The information produced and processed by the *agents* (archaeologist, museum curators, art historians, specialists) plays a strategic role. "*[Cultural heritage] become identified as heritage goods usually by archaeologists and historians who have obtained some form of official recognition or public acceptance of their status as experts in determining their artistic or historical significance*" [Pea94]. Being based on subjective evaluation, information is not easily translated by the *principal* into objective data. This asymmetrical distribution of information can affect negatively the outcome of the decision-making processes in terms of society's well-being because each *agent* is

likely to pursue his/her own utility. Therefore, the process is likely to be supply-oriented rather than demand-oriented unless suitable monitoring procedures, i.e. tailored on each decision-making process, are introduced. The outcome of the decision-making process, in fact, will depend on the institutional design, i.e. on the incentives imposed on the various *agents* and on the possibility to control them. The production and distribution of information, then, is crucial to make the process working and informatics can be strategic both to produce and to distribute information. Moreover, the use of new technologies allows for a wider range of consumption, though related to different experiences.

Relevant information is highly heterogeneous and in many cases it is not available as such but it needs to be produced: for instance, public grants should be oriented toward the more efficient institutions – because this is in the society's interest – but to do so adequate measures of performance have to be constructed. From a methodological point of view, this means for each decision-making process to design a systematic information flow based on:

- identification of relevant actors;
- identification and definition of the output or of the mix of outputs;
- identification and measurement of the relevant inputs;
- estimation of the production function;
- identification of data collection timing and procedures;
- evaluation of the performance (parametric and nonparametric systems);
- identification of systems to distribute information (next to the actors more directly involved in the process, information should be offered to the general public, as citizens, being taxpayers, are the actual *principal* of any activity concerning heritage which is publicly supported).

The outcome of the above process constitutes the basis to design a set of incentives to stimulate the *agents* to improve their performance in line with the *principals'* preferences. The distribution of information provides already an incentive for good performance so long as the *agent* obtains utility from reputation, fame and, more in general, public appreciation. However, in economics, great emphasis is put on economic incentives – such as career advancements, fund assignment, salary increases and fringe benefits of various kinds – to be related to the *agent's* performance.

So far, the attention has focused on the supply side. Further information issues have to be addressed when the consumption of heritage is at stake, in order to exploit it as an input of sustainable economic and social development. The rationale lies on some peculiarities of heritage itself, stressed in the economic literature: it is an 'experience good' – i.e. people need to have information about heritage to demand for it – as well as an 'addictive good' – contrasting the conventional economic law of 'decreasing marginal utility', people's appreciation increases by consuming it. As a result, information is strategic to satisfy the various and diversified demand for heritage of cultural tourists and, eventually to enlarge it. Cultural tourists, in fact, can be distinguished according to the depth of experience sought and it is possible that even non-cultural tourists may include the visit to a heritage site in their trip (combined-motivated tourists). The matter is complicated because different actors, with different objectives, are involved. Consumers' demand is affected by experts' preferences. The information provided by each institution, even if satisfactory from the experts' point of view, may fulfil only the institution's objective without considering the spillover effects – on other institutions, on the creation of new demand and/or on the enlargement of the existing one. Lack of coordination may result in congestion of outstanding heritage items and/or in the under-use of minor ones; it may also prevent the creation of cultural districts. Attention has to be paid to the identification of different segments of demand (using methods such as contingent valuation, travel cost analysis, etc.) so that specific cultural output can be constructed to satisfy them. Eventually, it is necessary to find a balance between the objectives of those responsible for the management of heritage and the actors of the tourism industry in order to maximize the beneficial impact of heritage tourism on local development.

Such an approach should lead to the provision of highly diversified information flows. Next to the satisfaction of various tourist demands, it would also increase residents' appreciation of heritage, possibly increasing their willingness to pay for consuming heritage. In this perspective, new technologies and the Internet are ideal tools to allow for a wider range of consumption, related to different experiences. Internet and digital technologies, in general, would decrease information and transaction costs that are particularly relevant in the case of cultural heritage and this would somehow reduce the gate-keeping role of experts, since people, having more chances to get the information necessary to demand for heritage, would be able to decide independently from experts' influence.

References

[MR01] Mazza I., Rizzo I.: (2001), Scelte collettive e beni culturali. In Valentino P., Mossetto G. (eds), *Museo contro museo* (2001). Giunti, pp. 41–54.

[Pea94] Peacock A.: *A future for the past: the political economy of heritage*. The David Hume Institute, 1994.

[Riz02] Rizzo I.: Heritage conservation: the role of heritage authorities. In Rizzo I., Towse R. (eds), *The economics of the heritage: a study in the political economy of culture in Sicily* (2002). Edward Elgar, pp. 31–47.

[Tri03] Trimarchi M.: Principal-agent analysis. In Towse R. (ed.),*A handbook of cultural economics* (2003). Edward Elgar, pp. 373–378.

[WB94] World Bank: *Cultural heritage in environmental assessment.* EA Sourcebook Update, 8 (1994).

The 5th International Symposium on Virtual Reality, Archaeology and Cultural Heritage VAST (2004)
K. Cain, Y. Chrysanthou, F. Niccolucci, N. Silberman (Editors)

Ontology-based Narrations from Cultural Heritage Texts

E.C. Mavrikas[1], E. Kavakli[1], N. Nicoloyannis[2]

[1] CILab, Department of Cultural Technology and Communication, University of the Aegean, Greece
[2] Laboratoire ERIC, Université Lumière-Lyon 2, France

Abstract

This paper proposes the use of an ontology-based, decentralised semantic repository for the classification and consultation of cultural heritage documents, and outlines the way in which semi-automatic natural language processing and discourse methods can assist scholarly community members in the task of integrating the meaning and utility of cultural heritage information, and forming value judgements.

Categories and Subject Descriptors (according to ACM CCS): C.2.4 [Computer Communication Networks]: Distributed Systems-Distributed applications, H.3.1 [Information Storage and Retrieval]: Content Analysis and Indexing-Linguistic processing

1. Introduction

The electronic publishing of cultural heritage documents is becoming common practice among heritage scholars and institutions. Despite this uptake, the current downstream communication pattern of the Internet does not leave room for proper analysis, structuring, search and exploration of the body of knowledge contained within such texts.

This project researches a semantically consistent framework for the online presence of cultural heritage documents, building upon a domain ontology that indicates the type of knowledge sought after, a participatory structure model based on *networks and flows* [Hod01], and a storytelling interface. The project drive is to unlock the *web of ideas and perspectives* [SMD00] that remain implicit in text, and encourage new individual and collective interpretative positions [PM03].

2. Conceptual design

The proposed semantic repository aims to accomodate an annotated collection of textual documents close to their point of creation and their point of use on the Internet [MNK04]. In the context of an online pre-defined or ad hoc community of contributors and users, the repository pools together distributed cultural heritage information sources connected in a schema-based peer-to-peer network infrastructure.

Shared documents within each information source (peer) are analysed and annotated for content and structure following semi-automatic, ontology-based natural language processing (NLP) methods. Relational metadata resulting from this one-time process are used to classify the information source thematically into one or more queriable concept clusters.

As concept clusters are populated and evolve over time, an ongoing multi-document summarisation [SBBW00] process identifies the most representative document passages within each cluster, and creates a narrative structure. The combined concept cluster narrations are presented to the repository contributors and users as a *structured progression* [RAB*03] through the entire collection.

3. Choice of technology

The CIDOC Conceptual Reference Model (CRM) [DS04] is a formal conceptualisation of the cultural heritage domain; an RDFS or OWL conversion of the CIDOC CRM can provide the semantic support for ontology-based cultural heritage document analysis and annotation tasks, with ontology instances encoded in RDF and XML as relational metadata, and structural information encoded using selected Text Encoding Initiative (TEI) tags [GR01]. The schema-based peer-to-peer network infrastructure introduced by the Edutella project [NWQ* 02] supports the ontology-based clustering of peers, and may be adapted to self-organise with respect to a simplified version of the CIDOC CRM. The GATE

[CMB*00] and Ellogon [PKP*02] text engineering platforms offer an integrated development environment (IDE) suitable for the modular design of NLP toolsets for repository contributors and users -including a structured progression adaptation tool and its associated visualisation component -and their evaluation.

4. Related work

There are some analogies between the ontology-based NLP toolsets currently developed within this project,

and other projects exploring ontology-based annotation, multi-document summarisation and discourse structure composition. MnM [VVMD* 02] and Melita [CDWP02] are both Java-coded, ontology-based annotation tools that may be adapted and integrated into GATE or Ellogon modules. ScholOnto [SMD00] is an ontology-based digital library server supporting scholarly interpretation and discourse. Finally, the Artequakt [AKM* 03] and Topia [RAB*03] projects process Semantic Web formats into an intermediary discourse structure followed by a final presentation, using bottom-up -narrative templates -and top-down -composition level generation -approaches respectively; these two approaches have recently been combined [GBvOH03].

5. Acknowledgements

This paper discusses research work undertaken by the first author in partial fulfilment of the requirements for the PhD degree of the University of the Aegean and the Université Lumière-Lyon 2, following a joint thesis supervision agreement between the two institutions (cotutelle de thèse).

References

[AKM*03] Alani H., Kim S., Millard D.E., Weal M.J., Hall W., Lewis P.H., Shadbolt N.R.: Automatic Ontology-based Knowledge Extraction from Web Documents. *IEEE Intelligent Systems*, 18(1) (Jan-Feb 2003), 14–21.

[CDWP02] Ciravegna F., Dingli A., Wilks Y., Petrelli D.: Adaptive Information Extraction for Document Annotation in Amilcare. In *Proceedings of SIGIR 2002* (2002).

[CMB*00] Cunningham H., Maynard D., Bontcheva K., Tablan V., Wilks Y.: Experience of using GATE for NLP R and D. In *Proceedings of COLING 2000* (2000).

[DS04] Doerr M., Stead S.: The CIDOC Conceptual Reference Model, a Tool for Integrating Cultural Information. In *Proceedings of CAA 2004* (2004).

[GBvOH03] Geurts J., Bocconi S., van Ossenbruggen J.R., Hardman L.: *Towards Ontology-driven Discourse: From Semantic Graphs to Multimedia Presentations*. Tech. rep., CWI, 2003. INS-R0305.

[GR01] Gibson M., Ruotolo C.: Beyond the Web: TEI and the eBook Revolution. In *Proceedings of ACH 2001* (2001).

[Hod01] Hodder, I.: Archaeology and Global Information Systems. *Internet Archaeology*, (6) (1999).

[MNK04] Mavrikas E.C., Nicoloyannis N., Kavakli E.: Cultural Heritage Information on the Semantic Web. In *Proceedings of EKAW 2004* (2004), vol. 3257 of *LNAI*, Springer, pp. 477–478.

[NWQ* 02] Nejdl W., Wolf B., Qu C., Decker S., Sintek M., Naeve A., Nilsson M., Palmér M., Risch T.: Edutella: a P2P Networking Infrastructure based on RDF. In *Proceedings of WWW2002* (2002).

[PKP* 02] Petasis G., Karkaletsis V., Paliouras G., Androutsopoulos I., Spyropoulos C.D.: Ellogon: a new Text Engineering Platform. In *Proceedings of LREC 2002* (2002).

[PM03] Papadopoulos D.C., Mavrikas E.C.: Peer-to-Peer Ways to Cultural Heritage. In *Proceedings of CAA 2003* (2003).

[RAB* 03] Rutledge L., Alberink M., Brussee R., Pokraev S., van Dieten W., Veenstra M.: Finding the Story -Broader Applicability of Semantics and Discourse for Hypermedia Generation. In *Proceedings of Hy-pertext'03* (2003).

[SBBW00] Stein G., Bagga A., Bowden Wise G.: Multi-document Summarization: Methodologies and Evaluations. In *Proceedings of TALN 2000* (2000).

[SMD00] Shum S.B., Motta E., Domingue J.: ScholOnto: An Ontology-based Digital Library Server for Research Documents and Discourse. *International Journal on Digital Libraries*, 3(3) (2000), 237–248.

[VVMD* 02] Vargas-Vera M., Motta E., Domingue J., Lanzoni M., Stutt A., Ciravegna F.: MnM: Ontology-driven Semi-automatic Support for Semantic Markup. In *Proceedings of EKAW 2002* (2002), vol. 2473 of *LNCS*, Springer, pp. 379–391.