A RETROSPECTIVE ON CLIO: LEGACY, LESSONS LEARNED AND PERSPECTIVES

Panos Constantopoulos[1,3], Costis Dallas[2,3], Martin Doerr[4], Chryssoula Bekiari[4], Ifigenia Dionissiadou[5]
[1] Athens University of Economics and Business, Athens, Greece
[2] University of Toronto, Toronto, Canada
[3] Athena Research Centre, Athens, Greece
[4] Foundation for Research and Technology – Hellas, Heraklion, Crete, Greece
[5] Benaki Museum, Athens, Greece

p.constantopoulos@dcu.gr

Abstract. During a 1990 CIDOC meeting in Athens the idea was born of using a semantic network – based framework for the purposes of museum documentation. Two years later, CLIO, a pioneering system designed and built at the Institute of Computer Science, FORTH, using the Semantic Index System implementation of the Telos knowledge representation language was installed at the Benaki Museum, heavily involved in requirements definition. At that early time, CLIO allowed extremely dense linking of information, access by unlimited chained references, expression of abstract properties and various ways of joint temporal and spatial assignment, as well as the extension and modification of the data schema by the users, thus supporting the easy adaptation of the system to the field of work and the evolution of knowledge.
The significance of the CLIO system lies in that it explicitly acknowledged and addressed the evolving, incomplete and conflicting nature of knowledge, the variety of required documentation structures, and the need for typed associations, as opposed to opaque references, in order to support reasoning over documentation data. It was a precursor to current systems employing Semantic Web technologies. Most importantly, though, it explicitly introduced an ontology and served as the seed from which CIDOC CRM grew.
In this retrospective we mostly discuss the path taken by conceptual and technical developments, the main issues and hurdles encountered in terms of documentation and data curation practices, cognitive overhead, information visualization and information access, and the current challenges in building ontology-driven, semantic-network museum information systems.

Keywords: Cultural Documentation, Semantic Networks, Knowledge Representation, CIDOC CRM

1 Cultural documentation and the CLIO system

The documentation requirements of cultural heritage range from keeping a simple log of objects for the management and care of collections to recording a wide range of scientific and scholarly information aiming to support curatorial work and cultural research. Accordingly, documentation can be coarsely distinguished into administrative, using a fixed set of data mainly to support administrative functions and to provide basic information, and cultural, aiming at maintaining an evolving body of knowledge about objects, pertinent to study, research and publication.

The drive towards higher accountability of museums over the management of their collections (Roberts 1985), the pursuit of standardization and efficiency in managing increasingly large collections (Parry 2007), and the wide availability of microcomputers in the developed world in the 1980s led to the widespread adoption of mostly flat file or hierarchical data base systems by museums, supporting the production of printed inventories and the formulation of queries over basic collection information (Roberts 1993). By the early 1990s, large museums increasingly adopted relational data base management systems, ideally suited for organizing information in table format, value-based search and report generation, and designed to support collections management procedures such as acquisition, loan management, and condition reporting. In tandem, museums such as the National Gallery in London adopted mostly "closed" information systems to provide multimedia information addressed to the public in the form of interactive exhibits, featuring fixed sets of information and navigation paths, and developed *ad hoc* as standalone applications dissociated from the museum documentation system (Hoffos 1992; Dallas 1994).

While addressing the administrative needs of collection care and management, and communication with the public, information systems available to museums in the early 1990s were not intended to support collections-based knowledge work as cultural documentation systems. In museums with heterogeneous collections ranging from fine art to archaeology and ethnographic objects, such as the Benaki Museum in Athens, curators dedicated much of their time in developing and managing knowledge about objects, not just in the form of compiling documentation records, but also through research notes, interpretive materials, exhibition catalogues and scholarly publication. Their work included accessing information not just about objects, but also a broad spectrum of historical and social subjects related to them: places, events, people, ideas and associations with cultural meanings. Much object-related information only made sense if it was clearly connected to its provenance: the date assigned to an artefact, the attribution of a work to an artist, were the result of scientific analysis or scholarly identification, and often the topic of continuing debate between researchers. Even if collection management systems available at the time were adequate for the needs of collections management, their inability to capture such aspects of scholarly information about museum objects, object history and cultural associations made them of little value for scholarly research and curatorial work (Dallas 1992; 1994).

To address these needs, cultural documentation requires therefore recording the entire range of information that makes up the current knowledge about a set of objects. This includes structured, unstructured and multimedia data and is characterized by a high degree of linking, large variety of references and classifications, broad usage of abstract relations and the need for multiple, mainly referential access. Moreover, this information is incomplete, expected to be enriched but rarely modified, accumulating layers of possibly conflicting information that should be retained, along with its provenance, rather than resolved. No information system at that time addressed these requirements.

In this light, the CLIO system was developed at the Institute of Computer Science, Foundation of Research and Technology - Hellas to fulfill the requirements of cultural, rather than administrative, documentation of collections in the domain of museums and cultural heritage. While initially conceived to address the needs of curatorial knowledge work in the Benaki Museum and the Historical Museum of Crete, CLIO was designed with the broader needs of collections-based cultural documentation in mind. Information was organized in CLIO as a knowledge base according to a specifically designed semantic model (Constantopoulos 1994). The functional kernel of CLIO is the Semantic Index System (SIS), see (Constantopoulos & Doerr 1993), built at the Institute of Computer Science, which in turn features a high performance implementation of the Telos knowledge representation language (Mylopoulos et al. 1990). The construction of CLIO allows extremely dense linking of information, access by unlimited chained references, expression of historical and cultural context as well as of abstract properties, joint temporal and spatial assignment in absolute or relative terms, and recording alternative, possibly conflicting information along with the respective sources. Information is presented in graphical or textual form and an extensible list of predefined queries is offered. A particularly important feature is the uniform treatment of schema and data, enabling the immediate extension and modification of the schema by the users themselves. This information “plasticity” has far reaching consequences with regard to the scope, content and practice of documentation: the knowledge captured by the documentation system evolves dynamically and the information curators are empowered with direct control over its organization. In this sense, CLIO can be characterized as an "open" knowledge base. “Closed” information bases for the public can be generated from it. Finally, CLIO offers an interface to an administrative documentation system. The requirements analysis for the CLIO system has been performed in close cooperation with the Benaki Museum and the Historical Museum of Crete, while the development of the system was partly funded by the STRIDE and ESPRIT programmes in the first half of the 1990s.

2 Knowledge representation

The Telos knowledge representation language (its structural part in particular) has been used for representing knowledge in CLIO. This offers the general mechanisms of attribution, classification and generalization, common in most semantic and object-oriented conceptual data models, and, in addition, it supports the explicit treatment of metaclasses organized in levels of instantiation, and completely equal treatment of entities and relations/properties, resulting in great expressiveness and flexibility. Both entities and relations/properties are essentially Telos propositions with unique identity, implemented as objects in SIS, the engine of CLIO.

Attribution is a general mechanism for representing the properties of an object as relations between the object and the values of the properties, also considered as objects. Classification defines classes of objects on the basis of common properties. An atomic object is declared as instance (member) of one or more classes. Classes are considered both as sets and as objects with unique identity, instances of meta-classes (sets of sets), and so on. Thus, successive levels of instantiation arise: tokens, classes, meta-classes, etc. Generalization defines a subclass-superclass (subset) relation between classes at the same level of instantiation, characterized by the inheritance of properties: when class A is subclass of B, it inherits all the properties of B and is differentiated from it either by having additional properties of its own, or by restricting the value range of certain inherited ones. Classification and generalization are both multiple, i.e. an object can belong to several classes and a class can have several superclasses. This directly supports the development of faceted information structures. As a matter of practice, classes are mainly used to capture the nature, or fundamental defining properties, of objects, while attributes render various relevant properties and relations. Since objects can be entities or properties/relations, this knowledge representation framework enjoys high expressive power. Higher levels of instantiation, i.e. metaclasses, metametaclasses, etc., enable the formulation of modeling templates and of abstract properties which are of interest as such even when instantiated differently in particular cases (e.g., partition, overlapping, creation, measurement interval).

3 Ontology of CLIO

Recognizing the futility of trying to design an all-encompassing conceptual model for cultural documentation, capable of adequately serving the information needs of scientific research in different domains, was the driving force behind the definition of a parsimonious, general model for CLIO, specific to no particular domain but amenable to specialization as needed. This led the design team to explicitly adopting an ontological approach resulting in a model that was a radical departure from the prevailing relational documentation models of the time. The CLIO ontology includes concepts of matter, location, occurrence, quantity, mankind, spiritual creation and naming, as well as relations between them, and is presented in detail in (Christoforaki et al. 1992). The CLIO model categories subsume those of the CIDOC/ICOM fine arts documentation standard at the time of its conception. Below we provide a very brief review.

The concepts of matter account for the kinds and the structure of physical objects, the elements of their appearance, the distinction into natural and artificial, composition or construction materials, special information concerning fine art objects and museum objects, descriptions of style and ornamentation, and kinds of tools. The concepts of location define absolute and relative location, orientation and various topological relations. The concepts of occurrence address the fundamental notions of event as a point in time and space, and existence as a set of events delimited by a beginning and an end, and include categories of events of particular significance such as activity, use, creation, etc. The concepts of quantity distinguish between physical, arithmetic and numismatic quantities, elaborate their kinds, and define ways of measurement. The latter are distinguished on one hand into absolute, defined with respect to conventional coordinate systems (e.g., chronology, location), and relative, expressing magnitudes in given measurement systems (e.g., time, distance, area, volume, weight); and on the other into exact, yielding a single value, and approximate, yielding a range of values. The concepts of mankind concern the description of persons and of manifestations of organization, such as group, membership, institution. The concepts of spiritual creation essentially differentiate the product of spiritual creation from its physical embodiments. Finally, a system for naming and identification is defined.

4 Implementation and applications of CLIO

The structure of the CLIO system together with its functional incorporation in an integrated museum information system are shown in Fig. 1. Such an integrated system offers combined administrative and cultural documentation functionality by interfacing respective autonomous systems. Administrative and multimedia data are acquired and managed by the administrative documentation system, using a relational data base management system to store them. Part of the administrative data is also imported into the cultural documentation system for purposes of object identification and matching certain basic information. Furthermore, multimedia data are retrieved from the common store for presentation by the cultural documentation system. Details about the architecture of the integrated museum information system, designed at the Institute of Computer Science, FORTH, and of CLIO, as part of it, can be found in (Theodoridou et al. 1992). In the sequel the structure and function of CLIO are briefly presented.

CLIO consists of three subsystems: the semantic index (knowledge base), the user interface and the communication interface. The semantic index stores and manages information organized according to the knowledge representation model (see Sections 2 and 3), using SIS. The communication interface comprises the data transport system, performing batch import of selected data from the administrative to the cultural documentation system, and the retrieval of multimedia data. Finally, the user interface mediates the functions of the system to the user, thus determining its perceived behaviour. The user interface is highly adaptable without reprogramming. The functions of the user interface are distributed to three subsystems: data entry, information retrieval and browsing, and multimedia display. Most important, from a functional point of view, are the first two.

Data entry is carried out through a special form offering guidance and vocabulary support and control. The data entry form is dynamic, in that it adapts to the current task automatically. The data entry function uniformly addresses both the entry of data conforming to a given schema and the modification of schema information. Thus, the modification of an information structure is a simple task subject, of course, to authorization constraints.

The information retrieval and browsing subsystem returns views of the knowledge base in graphical or textual form (see Figs. 2, 3). Each browsing step takes an item, called query target, as point of reference. Queries may involve objects directly related to the query target through specific relation types (1st order queries, Fig. 2), or express recursive searches generating chains of objects and relations rooted at the query target (recursive queries, Fig. 3), or address combinations of attribute- or classification- based predicates (search card, Fig. 4). General or frequent queries are offered in lists of predefined queries for ease of use. These are easily configured to match the terminology and requirements of the field of work. Most notably, predefined queries can be abstract expressions with context-sensitive semantics depending on the situation (e.g., "main characteristics", Fig. 3). A session history module prevents disorientation and facilitates repeating previous transactions.

Due to the implementation of the underlying SIS engine, the execution of recursive queries in CLIO is particularly efficient (in fact, two orders of magnitude faster than on relational systems of those days), thus CLIO stood out in its time for its capacity to exploit graph traversal at a realistic scale. The predefined recursive queries of CLIO include: Classification Tree, Get Main Characteristics Of, Get Parts Of Object, Get Parts of Kind, Get Temporal Bounds, Get Spatial Bounds, Get History Of. Navigation through the semantic network of

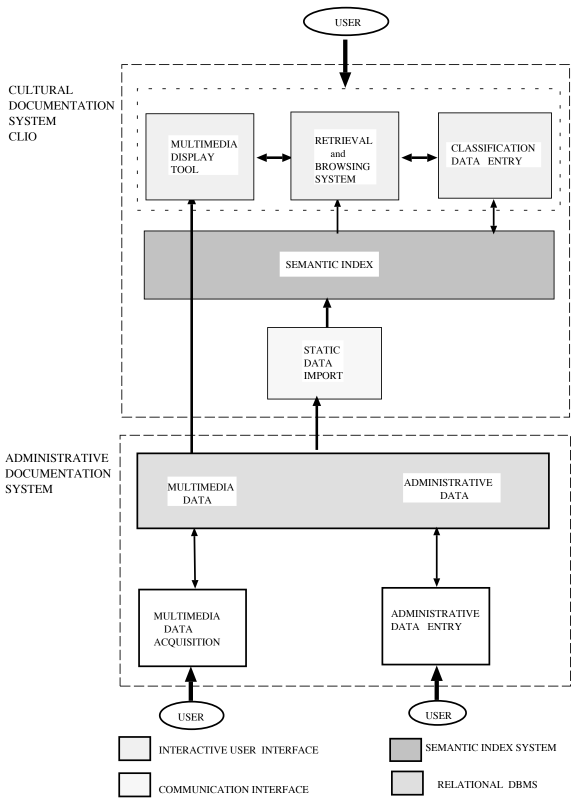

Fig. 1. Integrated Museum Information System including CLIO

the CLIO knowledge base is driven by the successive choices of query target. Referential access to information was found to match a large part of the users' information seeking behaviour. This is well matched by the then unique feature of CLIO to equally manage information at the categorical and the factual levels.

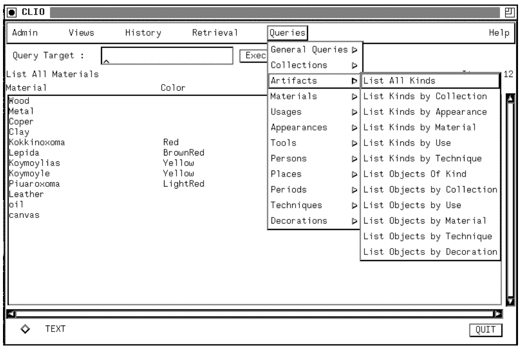

Fig. 2. 1st order query, text view

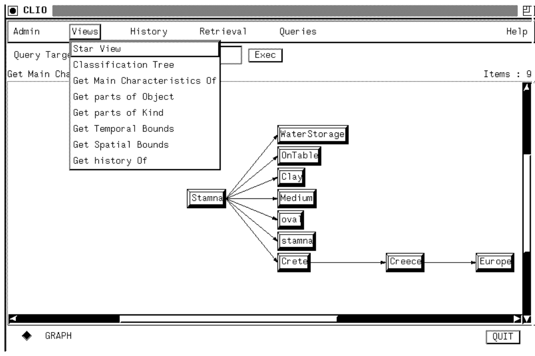

Fig. 3. Recursive query, graph view

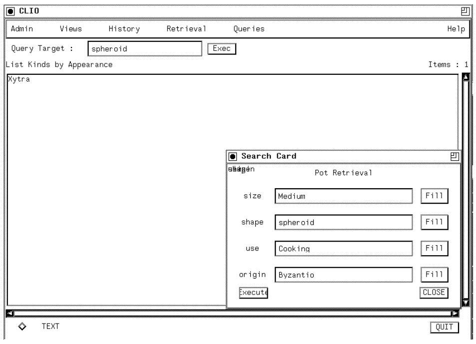

Fig. 4. Multicriteria query, search card

In line with the architecture presented above, an integrated museum information system comprising an administrative documentation module (MITOS) and a cultural documentation module (CLIO) was introduced in the Benaki Museum in 1993 through the STRIDELL and POLEMON programmes. It took a lot of meticulous preparation for the Documentation Department of the Benaki Museum to endorse the documentation approach required by the new system. The various kinds of objects had to be classified on the basis of common properties, which had to be accurately defined. Subclasses were defined with respect to inherited properties, instead of the purely scholarly - often vaguely perceived - previous criteria. The methodology of attribution, classification and generalization had to be comprehended. New concepts were adopted regarding the differentiation between kinds and instances of museum objects, but also the broadening of the meaning of an event, as anything to be analyzed in time and space; in the world of art history mostly historical events are being encountered. The semantic approach and the uniform treatment of schema and data were a real challenge for museum collections curators, who were used to think and work rather differently (Dallas & Dionissiadou 1993). Conceptual analysis of the information and its presentation in graphical form advanced the way of treating cultural information and proved to be fruitful for the future of the documentation practices followed by the museum ever since. One of the most important benefits of the semantic approach regarded the clear distinction of a museum object from its conceptual context, iconographic, linguistic or symbolic. The methodology of differentiating the product of intellectual creation from the physical embodiment was immensely useful and prevented the museum from spending valuable time and resources upon mixing up dissimilar data. The CLIO-MITOS system operated in the museum until 2000. Without an evolution lifecycle comparable to commercial applications, the system lagged behind in terms of the necessary improvements in order to operate in high availability and fulfill the museum's primarily managerial needs. Nevertheless, its influence has been outstanding and marked the further development of the Benaki Museum's information systems.

5 From CLIO to CIDOC CRM and beyond

Over many years, CIDOC and the CIDOC Documentation Standards Working Group (DSWG) have engaged in the creation of a general data model for museums, with a particular focus on information interchange. Until 1994 the product of these activities had been the CIDOC Relational Data Model. The CLIO system, with its ontological, "plastic" information model and semantic navigational information access, was proposing a compelling alternative for scientific documentation to the stiff and bounded relational systems as well as to the semantics-unaware hypermedia systems of its time. After a presentation of CLIO at the CIDOC meeting in Washington, D.C., USA in 1994, discussions within CIDOC to take a different approach were accelerated (Doerr & Dionissiadou 1994). Indeed, in an interim meeting in March 1996 in Crete, the DSWG decided to engage in an object-oriented approach in order to benefit from its expressive power and extensibility for dealing with the necessary diversity and complexity of data structures in the domain. This effort resulted in 1999 in the first complete edition of the *"CIDOC Conceptual Reference Model" (CRM)*, the outcome of the intensive voluntary work of many contributors, which draws its origins in the CLIO information model (Crofts et al. 1999). In 2006 CIDOC CRM became ISO standard 21127:2006, updated in 2014.

CLIO has proved a catalyst for the advancement of the idea of ontology-based, extensible scientific documentation systems, as well as of the underlying knowledge representation formalism and methodology and SIS, the system it was implemented on. Following the development of CLIO, SIS was used to implement a series of other cultural information systems for documentation and education applications in cultural heritage organizations and, notably, the terminology management system SIS-TMS, commercially exploited by ICS-FORTH until recently, which is capable of handling large vocabularies such as the Getty vocabularies Art and Architecture Thesaurus, Thesaurus of Geographic Names and Union List of Artist Names. CIDOC CRM which, as explained, sprang out of CLIO, enjoys increasing uptake, e.g. by the British Museum, and a set of mutually harmonized extensions, covering museum collections, history, archaeology, library data and the recording of scientific observations in various disciplines including archaeology, art conservation, geology and biodiversity. While most users now employ an OWL version, the master copy is still maintained and validated in Telos by ICS-FORTH on SIS. In the above applications, there is recently an increasing need for using attributes of attributes and metamodelling, both of which can be done nicely in Telos, but are not among the prominent features of the currently established ontology languages RDFS and OWL.

6 Where have all the flowers gone?

The significance of the CLIO system lies in that it explicitly acknowledged and addressed the evolving, incomplete and conflicting nature of knowledge, the variety of required documentation structures, and the need for typed associations, as opposed to opaque references, in order to support reasoning over documentation data. It was a precursor to current systems employing Semantic Web technologies. Most importantly, though, it explicitly introduced an ontology and served as the seed from which CIDOC CRM grew (Doerr & Dionissiadou 1998). The CLIO system itself, however, enjoyed rather limited deployment.

The decade following the development of CLIO, other SIS applications and the definition of CIDOC CRM witnessed the emergence of Semantic Web technologies, notably RDF, RDFS, and OWL. The development of these technologies regrettably ignored several previous developments in knowledge representation models, languages and systems, including the family of Telos. Consequently, these older technologies, although more refined and in certain respects more effective, proved precocious, a pattern not uncommon among innovative systems.

On the other hand, the requirements for information management at the categorical and factual levels that had to be addressed by CLIO triggered the development of innovative human-computer interaction functionality. In particular, the chief HCI design challenges were: the hybrid graph-based and text-based treatment of network structures, the effective concealment/exposition of multi-step processes depending on the context, and the isomorphic treatment of the categorical and the factual levels of information. The first two challenges are also facing the wider class of Semantic Web - based information systems. The third challenge is shared with meta-modelling environments, the possible solutions being delimited by the underlying data model and language.

7 In perspective

In what concerns documentation systems, the basic philosophy underlying the architecture of Fig. 1 is still valid. Services are much more distributed, often cloud-based, adhering to linked open data and FAIR principles. Data are acquired along with their metadata in repositories and various registry and semantic integration services enable multiple semantic chains of access. Multi-faceted annotation is achieved using curated thesauri. The underlying conceptual structures must follow development principles conformant with those governing the development of the CLIO ontology and CIDOC CRM. Therefore the legacy of CLIO, carried by CIDOC CRM, is essentially methodological. On the data management side, the challenge remains for the Telos school of thought to influence the current developments of graph databases.

The CLIO system provided an elegant, parsimonious and expressive framework for the symbolic representation of aspects of artefact knowledge drawn from art historical, ethnographic and archaeological scholarship (Dallas 2009). Thirty five years after its pioneering deployment of CLIO in the Benaki Museum and the Historical Museum of Crete, the vision of an information system able to capture, index and allow the curation of evolving cultural knowledge about cultural heritage objects, their historical contexts and associations, remains current. While major developments in the field of cultural heritage documentation are driven by information integration, linking and standardization across networked collections and systems, integrated support of knowledge work such as introduced by CLIO remains still an open challenge. A key priority for future cultural information systems is their ability to express and accommodate research questions based on the epistemic discourse, domain knowledge, and object representations of the human sciences (Dallas 2009; Benardou et al. 2010). Perhaps today is a good time to leverage the affordances and innovations of this early experiment in museum documentation, towards the development of cultural information systems capable of truly supporting curatorial work and cultural research in the contemporary museum environment.

Bibliography

Bekiari, Chryssoula and Ifigenia Dionissiadou. 1994. "CLIO, MITOS: Greek Action for Computer Applications and Multimedia Technology". Report at the information day organized by the Hellenic VALUE Relay Centre during the exhibition INFOSYSTEM HI TECH '94, Oct. 1994, Thessaloniki, Greece.

Benardou, Agiatis, Panos Constantopoulos, Costis Dallas, and Dimitris Gavrilis. 2010. "Understanding the Information Requirements of Arts and Humanities Scholarship: Implications for Digital Curation." *International Journal of Digital Curation* 5 (1): 18–33.

Christoforaki, Maria, Panos Constantopoulos and Martin Doerr. 1992. "CLIO: An Object-Oriented Model of Cultural Data". Report ICS- FORTH.MUIS.92.1. Institute of Computer Science, FORTH, Nov. 1992.

Constantopoulos, Panos. 1994. "Cultural Documentation: The CLIO System". Technical Report FORTH-ICS/TR115. Institute of Computer Science, FORTH, Jan. 1994.

Constantopoulos, Panos and Martin Doerr. 1993. "The Semantic Index System - A brief presentation". Institute of Computer Science, FORTH, May 1993.

Crofts, Nicholas, Ifigenia Dionissiadou, Martin Doerr, Mathew Stiff (editors). 1999. "Definition of the CIDOC Object-Oriented Conceptual Reference Model (version 2.1)". Published by the CIDOC Documentation Standards Working Group, Sept. 1999.

Dallas, Costis J. 1992. "Information Systems and Cultural Knowledge: the Benaki Museum Case". *Computers and the History of Art Journal,* 3.1, 7-15.

Dallas, Costis. 1994. "A New Agenda for Museum Information Systems." In *Pre-circulated papers, Problems and Potentials of Electronic Information in Archaeology seminar,* edited by Seamus Ross, 251–64. London: British Library. http://eprints.rclis.org/archive/00007978/.

Dallas, Costis. 2009. "From Artefact Typologies to Cultural Heritage Ontologies: Or, an Account of the Lasting Impact of Archaeological Computing." *Archeologia e Calcolatori*, no. 20: 205–222.

Dallas, Costis J. and Ifigenia Dionissiadou. 1993. "Representing Cultural Information in the Benaki Museum *MITOS* System". CIDOC/ICOM Annual Conference, Ljubljiana, Slovenia, Sept. 1993.

Doerr, Martin and Ifigenia Dionissiadou. 1994. "Mapping of Material Culture to a Semantic Network". *Sourcebook of the 1994 CIDOC and MCN 1994 Joint Annual Meeting*, Washington DC 1994, pp 31-37.

Doerr, Martin and Ifigenia Dionissiadou. 1998. “Example of the CIDOC Model: Annotation to the Model of the Epitaphios GE34604 of the Benaki Museum”. Published by the CIDOC Documentation Standards Working Group, Oct. 1998.

Hoffos, S. 1992. *Multimedia and the Interactive Display in Museums, Exhibitions and Libraries,* Library and Information Research Report 87, British Library: London.

Mylopoulos, John, Alex Borgida, Matthias Jarke and Manolis Koubarakis. 1990. “Telos: Representing Knowledge about Information Systems”. ACM Trans. on Information Systems, 8/4, Oct. 1990, pp. 352-362.

Roberts, D. Andrew. 1985. *Planning the documentation of museum collections,* Museum Documentation Association: Cambridge.

Parry, Ross. 2007. *Recoding the Museum: Digital Heritage and the Technologies of Change*. Museum Meanings. London: Routledge.

Roberts, D. Andrew. 1993. “Documentation Practice, Systems and Standards in European Museums.” In D. Andrew Roberts (ed.), *European Museum Documentation: Strategies and Standards*, Museum Documentation Association: Cambridge, 9-28.

Theodoridou, Maria, Panos Constantopoulos, Chryssoula Bekiari and Martin Doerr. 1992. “Architecture of the Museum Information System”. Report ICS- FORTH.MUIS.92.2. Institute of Computer Science, FORTH, Nov. 1992 (in Greek).