Semantic Web in Cultural Heritage After 2020

Konstantinos N. Vavliakis[1,2], Georgios Th. Karagiannis[2], and
Pericles A. Mitkas[1,3]

[1]Department of Electrical and Computer Engineering
Aristotle University of Thessaloniki, GR541 24, Thessaloniki, Greece
[2]Ormylia Foundation - Art Diagnosis Centre, GR630 71, Chalkidiki, Greece
[3]Information Technologies Institute, CERTH, GR570 01, Thessaloniki, Greece
kvavliak@issel.ee.auth.gr,g.karagiannis@artdiagnosis.gr,mitkas@auth.gr

Abstract. In this paper we present the current status of semantic data management in the cultural heritage field and we focus on the challenges imposed by the multidimensionality of the information in this domain. We identify current shortcomings, thus needs, that should be addressed in the coming years to enable the integration and exploitation of the rich information deriving from the multidisciplinary analysis of cultural heritage objects, monuments and sites. Our goal is to disseminate the needs of the cultural heritage community and drive Semantic web research towards these directions.

Keywords: Semantic Web, Cultural Heritage, CIDOC-CRM

1 The Multidisciplinarity of Data in Cultural Heritage

Data management for the cultural heritage field needs to accommodate an assortment of different information types relevant to the identification, description, interpretation, aesthetic appeal, technical operations, condition assessment, and historical background of art objects, monuments, and historical sites. Data are produced and processed by a wide range of scientists, like engineers, physicists, chemists, archaeologists, historians, restorers, librarians, etc., who employ a variety of methods and techniques.

Conservation science, for example, can shed light on various aspects of an art object, by revealing its internal structure/stratigraphy, which, in turn, provides important information related to the materials and the technique used for the creation of the object. The identification of materials leads to a better understanding of the creation phase of the artifact, while materials are also strongly connected with the artist, the creation period and the technique used. Proper identification, however, can be hindered by the deterioration or alteration of materials due to the passage of time, as well as due to previously unsuccessful restoration or conservation attempts.

For the analysis of art objects a variety of cross-domain techniques are used [5], [7], [8], including: Advanced physicochemical techniques, like spectroscopy

(visual, near-IR, mid-IR), multispectral imaging [3] and ellipsometry, that provide surface information and information from the underlayers; Unilateral nuclear magnetic resonance that provides information about the structural stability of the object; X-ray fluorescence that provides elemental analysis information from the under-layers. Other microsampling techniques that acquire information from the depth profile are also employed, like Raman analysis, which examines the inorganic materials of the objects, Fourier-transform infrared spectrometry that provides information about the organic materials, and optical coherence tomography that provides depth profiling images from the paint layers. Laser induced breakdown spectroscopy is also used, providing rapid elemental analysis results.

All this information combined (data about creation, restoration, consolidation, overpainting layers, aesthetic impression, etc) comprises the current state of the artwork, which is the result of a series of past states. Figure 1 depicts some of the techniques used and the information they provide. The management of this information is a tedious task, while efficient knowledge extraction using all these (possibly interlinking) data sources constitutes a formidable open research challenge for the next years.
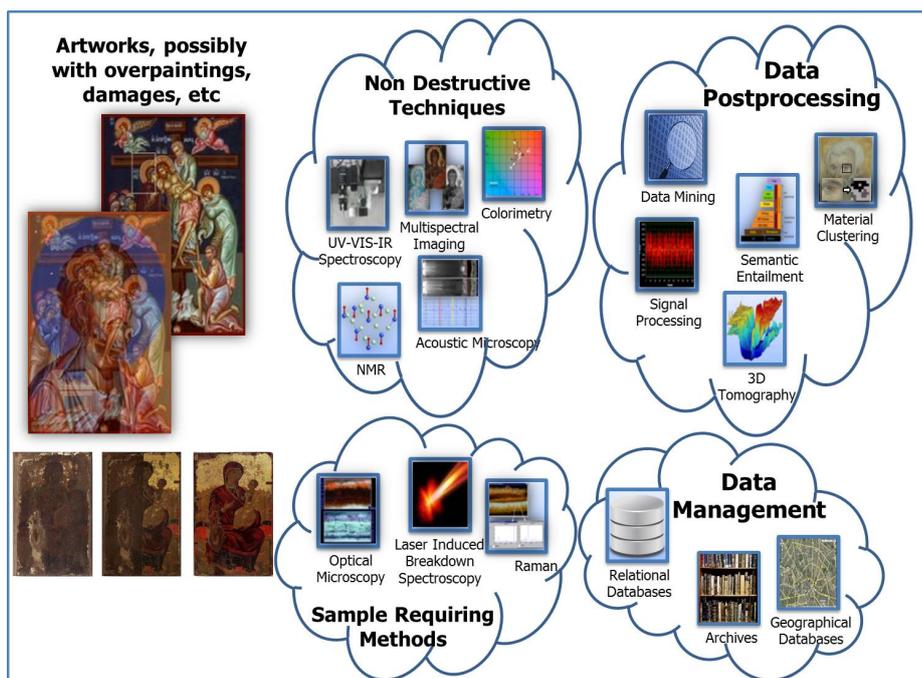

Fig. 1. Some of the techniques used in the analysis and characterization of artifacts.

2   Current Use of Semantics in Cultural Heritage

Cultural heritage is a field where semantic technologies have already been introduced, at least in a scientific or prototype level [4]. An important contribution in the domain is the CIDOC-CRM ontology [2], a formal ontology intended to facilitate the integration, mediation, and interchange of heterogeneous cultural heritage information. The CIDOC Conceptual Reference Model (CRM) represents the culmination of more than a decade of standards development work by the International Committee for Documentation (CIDOC) of the International Council of Museum (ICOM). Although CIDOC-CRM is now an ISO standard, there are only a few complete use cases published. The same applies for SKOS (Simple Knowledge Organization System), a W3C recommendation for sharing and linking knowledge organization systems via the Web.

Integrated tools for assisting the transformation and retrieval of cultural information have been developed in the past [9], but considerable R&D effort is still required, before delivering truly semantically empowered, friendly to use tools. Description Logics and formal semantics have also been employed [6] for real-life knowledge discovery and integration over distributed sources. They address a number of issues towards developing a robust inference platform, namely systematic accumulation of common concepts and inference rules; extending available ontologies with metaclasses; accumulation of factual and categorical knowledge; incorporation of fuzzy inference into the inference engine, and improvement of performance and scalability in the inference engine.

There are some notable efforts that take extended advantage of the available semantic technologies, like the semantic publishing of British Museum's collection[1], the "CulturaSampo"[2], a semantic portal of Finnish museums on the Semantic web, the DigiCULT[3],CASPAR[4] and CHARISMA[5] projects, as well as several other research funded projects. Nevertheless, the mainstream use of Semantic web technologies has not yet been achieved.

3   Needs, Shortcomings and Future Directions

In this section we discuss current shortcomings in the semantic management of cultural resources and future research directions that could benefit the cultural heritage field. We identify seven main open research problems that require solutions. We believe and hope that in the future most of these problems will have been addressed and resolved by the Semantic Web community.

Although we focus on the cultural heritage domain, we believe that similar shortcomings and solutions also apply in several other domains (e.g. in the medical research domain [1]).

---

[1] http://collection.britishmuseum.org/
[2] http://www.kulttuurisampo.fi/
[3] http://www.digicult.info/
[4] http://www.casparpreserves.eu/
[5] http://charismaproject.eu/

The seven needs and their respective open research fields we identified are:

1. **The multidisciplinary nature of analytical data** available in the cultural heritage field requires advanced techniques for optimal data integration and knowledge reuse. Semantic web technologies play a crucial role in improving data integration, but this step should go beyond simple merging of different concepts and URIs in the same ontology. Instead, the objectives should be shifted towards truly conceptual integration, ontology matching and interlinking of semantically related entities. The integration process should also take into account that data have been created by different, possibly overlapping, methods and by users with different scientific backgrounds.

2. **The conceptual merging and true integration** of all this multidimensional information has the potential to uncover new knowledge about the interpretation, aesthetics and condition of artworks. In order for this leap of knowledge to take place, deep semantics are necessary. Semantic technologies that can use the multidisciplinary information to derive fuzzy inference rules and probabilistic description logics, as well as reasoning over dynamically evolving data are necessary. Bear in mind that we must be extremely careful when dealing with reasoning rules, as in an open world assumption tight rules can rarely be always true, since exceptions and subjectivity are most of the times present. For example, even though icons of Coptic style use a specific mixture of materials, there may be icons of different styles that are composed of this rare mixture of pigments that we are not currently aware of. For this reason, we must be able to handle uncertainty, define thresholds and confidence levels. Unfortunately, OWL and DL are not currently equipped with such operators, thus more advanced (and, as of yet, immature) techniques should be considered, such as uncertainty reasoning, representing and reasoning under uncertainty, etc.

3. **Personalized semantics** are necessary for the expression of subjective opinions and "subjective inference". It is a common case for an art object with a set of fixed attributes to have different aesthetical or varying interpretation meanings to people with different backgrounds and interests. For example, Byzantine icons may be appreciated quite differently by people of different ethnic and religious backgrounds; likewise, controversial artworks may be treated with mixed reviews and feelings, all of which may be valid and useful. Semantic inference, either using description logics, or semantic knowledge extraction should be able to derive the different semantics and personalize them accordingly.

4. **Ontology validation** is also a field where new research achievements can benefit cultural heritage. Although one may start with CIDOC-CRM as a base ontology, extending its model may be necessary, which can lead to a number of defects, like syntactic defects, unintended redundancy, modeling and conceptual defects, and semantic defects such as unsfatisfiability problems and inconsistencies.

5. Most institutions store their data in their local language. The Linked Open Data initiative encourages cultural heritage institutions to make their data

available to the public. Even if this practice becomes widely accepted, it will ensure that institutions will publish their data collections in languages other than they are available in. In order to make this information available to an international community, **multilingual knowledge representation, access and translation** are an impending need. Thus, technologies like localization and language technology processing, translation representation and natural multilingual data management for evolving linked data should be further advanced in the next years.

6. **Automatic and easy to use tools** are also necessary for the mainstream uptake of the Semantic web in the cultural heritage community. Cultural heritage institutions are encouraged to start processing their data with semantic technologies and publish them as Linked data but automatic and easy to use tools that can undertake this massive burden are still missing. Despite the fact that some semi-automatic tools already exist, further improvements are necessary if the Semantic web is to be fully embraced by the cultural heritage organizations.

7. Finally, museums, libraries and relevant foundations may turn to be very reluctant at sharing, or publishing information if copyright issues are not fully resolved and agreed upon. Thus, **ownership, permission of use, trust, and copyrights issues** should be addressed, resolved, and incorporated into semantic web technologies before initiatives like the Linked data movement are fully accepted by the administration of cultural heritage institutions.

4   The Future of Semantic Web for Cultural Heritage

We present the advantages of the adoption of Semantic web technologies in the cultural heritage field through a not so distant scenario. Imagine an archaeologist in 2022 who is performing an excavation in an ancient Macedonian tomb in Northern Greece. Her crew has already unearthed the dome of the tomb and they are now working towards revealing the rest of the structure. While she is updating the progress in the "SemExcavation4u" application using her smart link to the web (and to the world), the application queries a Linked open data repository, finds information of similar excavation sites, and inferences that given the materials found in another similar tomb, the ongoing operations, and the reported humidity of the soil, there is a considerable risk of collapse. Thus, the device issues a strong recommendation that all operations should halt and supports should be placed to hold the dome.

A few months later, a restorer is working on a golden crown found in the tomb. The crown is heavily damaged, so he is not sure whether is of Macedonian or Persian style. He accesses the object's report, which contains spectroscopic and μRaman analysis data that identify the object's materials. He performs a query and the inference machine answers that there is a 99% chance the crown is of Persian style. The reasoner further explains that its decision was based on the orpiment mixture found in the object, which used to be extracted from the mines of Persia and at the creation time of the crown it was common among Persian blacksmiths.

Whether scenarios like the above will become possible and the Semantic web will grow in the next 10 years as an integral part of the cultural heritage domain, or will remain just another promising buzzword used in research projects and scientific publications, depends on the ability of the Semantic web community to deliver the necessary advances in technology, as well as the willingness of the cultural heritage community to embrace the necessary changes in their data management processes.

Advances in semantic expression, interoperability and integration, reasoning and validation, multilingualism and copyright issues, provided by powerful, (semi-)automatic, and easy to use tools may lead to this change in cultural data management and help the core of the community embrace semantic technologies. If this is the case, it will happen step-by-step, with the first step being the recruitment of data scientists with semantic background in the cultural heritage organizations, a step already taken by most of the large institutions.

Aristotle University of Thessaloniki, with its 42 Departments and its long and strong commitment to archaeological research, conservation and restoration, has embarked on a large scale, multidisciplinary project to record, characterize, and cross-reference the cultural heritage in Northern Greece. Semantics will play a tremendous role in this monumental effort.

References

1. Agorastos, T., Koutkias, V., Falelakis, M., Lekka, I., Mikos, T., Delopoulos, A., Mitkas, P.A., Tantsis, A., Weyers, S., Coorevits, P., Kaufmann, A.M., Kurzeja, R., Maglaveras, N.: Cancer Informatics (8), 31–44
2. Doerr, M.: The cidoc conceptual reference module: an ontological approach to semantic interoperability of metadata. AI Mag. 24(3), 75–92 (2003)
3. Karagiannis, G., Alexiadis, D., Damtsios, A., Sergiadis, G., Salpistis, C.: Diffuse reflectance spectroscopic mapping imaging applied to art objects materials determination from 200 up to 5000 nm. Rev. of Scient. Instruments 81(11), 113104 (2010)
4. Karagiannis, G., Vavliakis, K., Sotiropoulou, S., Damtsios, A., Alexiadis, D., Salpistis, C.: Using signal processing and semantic web technologies to analyze byzantine iconography. IEEE Intelligent Systems 24(3), 73–81 (2009)
5. Karagiannis, G.T., Alexiadis, D.S., Damtsios, A., Sergiadis, G.D., Salpistis, C.: Three-dimensional nondestructive "sampling" of art objects using acoustic microscopy and time-frequency analysis. IEEE T. Instrumentation and Measurement 60(9), 3082–3109 (2011)
6. Lin, C.H., Hong, J.S., Doerr, M.: Issues in an inference platform for generating deductive knowledge: a case study in cultural heritage digital libraries using the cidoc crm. Int. Journal on Digital Libraries 8(2), 115–132 (Apr 2008)
7. Salvadó, N., Butí, S., Tobin, M.J., Pantos, E., Prag, A.J.N.W., Pradell, T.: Advantages of the use of sr-ft-ir microspectroscopy: applications to cultural heritage. Analytical Chemistry 77(11), 3444–3451 (2005), pMID: 15924374
8. Van Grieken, R., Corporation., E.: Non-destructive Micro Analysis of Cultural Heritage Materials. Elsevier,, Burlington : (2005)
9. Vavliakis, K.N., Symeonidis, A.L., Karagiannis, G.T., Mitkas, P.A.: An integrated framework for enhancing the semantic transformation, editing and querying of relational databases. Expert Systems with Applications 38(4), 3844 – 3856 (2011)