5th International Congress on

"Science and Technology for the Safeguard of Cultural Heritage in the Mediterranean Basin"

Istanbul, Turkey

22 – 25 November 2011

EDITING METADATA TO SUPPORT THE ACQUISITION, CONTENT ANALYSIS, STORAGE AND RETRIEVAL OF ANCIENT DOCUMENTS[*]

Debole Franca[1], Salerno Emanuele[1], Savino Pasquale[1], Tonazzini Anna[1]

[1]National Research Council of Italy, ISTI-CNR, Pisa, Italy, name.surname@isti.cnr.it

Preservation, accessibility, readability improvement and content analysis of ancient documents are nowadays urgent needs, to prevent loosing much of our past memory. Indeed, the damages that they have suffered over time still cause a progressive decay, and the fragility of rare and important historical documents prevents their direct access by scholars and historians. On the other hand, low contrast and possible artifacts, make their reading difficult. Finally, interesting features are often barely detectable in the original documents, whereas they could represent the most important information from a historical and cultural point of view.

Multispectral and multisensory imaging and digital image processing techniques can be used to enhance readability and seek out new information. Digital processing may require many different techniques to be applied. The results of these elaborations may then be subjected to other processes for document interpretation, such as OCR. Elaborations of this kind are, or could become, standard practice in libraries and archives, thus concurring to enrich the documentation and improve the access and usability of cultural heritage objects. A major challenge is thus the creation of structured digital libraries, where, for each stored document, the plurality of its representations is suitably managed. These include all the acquisition channels available and all the subsequent processed versions, along with the corresponding parameters. This rich description of acquisitions and image processing results should support the archival and the retrieval based on the features of the images and on the processing they have undergone. At the same time, the availability of traditional descriptive metadata should support content based search, such as those usually done in a Digital Library.

In this paper, we propose a metadata schema model to support such a combination of classical and new ways of describing a document and its analysis process, and we illustrate a Metadata Editor (ME) that supports the creation and editing of the metadata. This enables the results achieved to be better evaluated, and an expert knowledge database to be created, to extract cross-document observations and conclusions about the artifact formation, its ageing, etc. This is important information for art historians and art restorers as well. The proposed metadata schema satisfies three broad requirements. On the one hand, it enables the representation and description of a cultural heritage object (an ancient document in this specific case), to support its retrieval and access. On the other hand, it supports the description of the complete acquisition process, and of all the processing performed on the digital representation of the object. Finally, it is fully compliant to existing standard metadata schemata, so that interoperability of existing archives and reuse of existing resources is guaranteed. The metadata schema is composed of different interrelated entities describing the physical object – i.e. the actual physical work of art – its digital representation – i.e. how the object is represented in digital form – and its digital elaboration – i.e. all processing activities performed on the digital elaboration in order to improve its quality and readability. All these entities can be linked to entities describing the authors of the cultural objects and the organizations which are responsible for their preservation.

The paper also describes a metadata editor, a tool providing an easy way to create and edit the metadata records. The metadata editor allows users to create, search and edit metadata records. It also supports the use of controlled vocabularies for the different metadata elements.