Multilingual Issues and Access to Digital Cultural Heritage in Greece
National report in the framework of the MINERVA PLUS project

Dimitrios Koutsomitropoulos, cand. Dr., M.Sc., Georgia Solomou, cand. M.Sc.,
Paraskevi Fragouli, cand. M.Sc.
Computer & Informatics Engineers
High Performance Information Systems Laboratory, University of Patras
{kotsomit, solomou, fragouli}@hpclab.ceid.upatras.gr

According to the 2001 survey of the National Statistics Agency the population of Greece consists of about 10.934.087 inhabitants. Of them 99% are Greek and the other 1% percent is divided between about 5 major people groups that posses other citizenships as well. There are no language or national minorities currently recognized in Greece. The only officially recognized minority is the religious minority of Greek Muslims in western Thrace.

Greek is the official written/spoken language and the vast majority of the population speaks Greek. However, some very small language groups speak other languages and dialects as well, like Romanika, Vlachika or Turkish.

In the framework of the MINERVA PLUS project we conducted a survey regarding multilingualism and use of thesauri in Greek cultural web sites. For this survey we evaluated 62 cultural web sites in terms of their multiple-language availability, search facilities and use of multilingual indexing and cataloguing structures such as vocabularies and thesauri. The methodology followed was the evaluation of these web sites from an evaluation team in combination with information requests from people responsible for the sites, when necessary.

Indeed, there is a long recognized need for multilingualism, especially in cultural web sites and digital collections. Most important reason for this are the benefits ripped from the wider dissemination and promotion possibilities that Greek cultural and educational content achieves through multilingualism. Multilingual support also appears to highly enhance marketing and diffusion of the Greek tourist product.

This situation is also reflected in the fact that almost all of the Governmental web sites (ministries, national agencies etc) are at least bilingual (Greek-English) while the vast majority of cultural portals and cultural institutions sites are also multilingual.

Nevertheless, the process of "going multilingual" encounters a set of difficulties mainly focused:

a) on *Conceptual level*, because it seems quite difficult to achieve consensus on standardization and translation of domain terms and vocabularies, especially about Greek cultural heritage and

b) on *Technical level,* because efficient technical means still have to be found that ease and streamline concurrent authoring and presentation not only of the site interface but also the content itself and its description (metadata) in multiple languages.

In accordance to our survey, most of the web-sites and on-line collections in Greece are indeed multilingual (in fact almost every site supports at least Greek and English) but there are currently only a few sites that support thematic vocabularies or even thesauri. Some of them employ "free text indexing" by submitting queries to an underlying data base system. Most often than not, they also offer structured navigation services among their content. These "structure" however commonly reflects no standardized taxonomy or just a monolingual one. In fact, only recently a major initiative has started in order to digitize cultural collections all over Greece and make them available through the Web possibly utilizing controlled vocabularies (main guideline is to use the CIDOC-CRM ontology for describing artefacts. Nevertheless no concrete results are yet available).

Survey Results

For our survey we evaluated 62 web sites, held by several cultural institutions in Greece. More than a half of them (54.8%) present the profile and collections of several well-known Greek museums, whereas the rest are grouped among other type of institutions. In particular, 25.8% of them belong to libraries, 12.9% are cultural sites and only 4% are characterized as web sites for digital archives.

As far as multilingualism is concerned, it is encouraging that a significant majority of our test group (59.7%) seem to present themselves in both English and Greek language, so they can be characterized as fully bilingual sites. On the whole, most web sites (67.7%) are available at least in English, whereas some of the rest give many signs of work in progress, by describing some of their resources in English as well. Museums appear to be more interested in working towards this direction. Unfortunately, three or more languages, other than Greek and English, are rarely seen. In particular, only two cultural institutions (3.2%) have their sites available in three languages, another one (1.6%) uses five, and there was just one library with content available in a total of seven different languages.

An important criterion we took in to account during the evaluation was whether the sites provide *actual access* to digital cultural content or not (MINERVA PLUS WP3 Meeting in Budapest, Feb. 2005). By this we mean, that the site should provide access to tangible digital resources, like photos, video etc and not just to textual information or metadata. Therefore, among the evaluated cultural web sites, a relative small percentage (51.6%), provide access to its digital collections. The majority of these sites belong to museums, which present usually photographs of their most significant and representative exhibits. There were also found cultural sites, mainly held by libraries, that provide access to a rich collection of digitized documents and other material.

Examples and Good Practices

According to our survey only a small percentage of the evaluation group employs a thesaurus or taxonomy for thematic indexing. Tsafou and Hatzimari (2001) report that libraries make very little to no use of thesauri and identify for this the following main reasons:

- Limited strength of their collections.
- They use software developed abroad that does not support non-Latin alphabets in the construction of the information database.
- There exists no national coordinating institution that would undertake the development of suitable information processing tools.

To the above, one may add the lack of development or standardized translation of appropriate thesauri for the cultural domain.

Many sites (43.5%) provide a controlled vocabulary, either mono- or multilingual, as a means of describing and searching available resources. However these vocabularies are mostly proprietary, suitable for the documentation needs of each particular site and not standardized in any way. On the other hand, an interesting example (although monolingual) of standardization efforts on terminologies and descriptive formalisms is the translation in Greek of the **International Standard for Archival Description (ISAD)** by the Society of Greek Archivists (http://www.eae.org.gr). An example of use of this standard is the on-line collection of the Hellenic Literary and Historical Archive (http://www.elia.org.gr).

Out of the sites that provide some means of structuring information (ranging from vocabulary to thesaurus) a 63.3% maintain bilingual versions of these structures which become available on the basis of the interface language selected by the user. Usage examples of proprietary bilingual vocabularies for the cultural domain are **Myriobiblos**, the Digital Library of the Church of Greece (http://www.myriobiblos.gr), the on-line collection of the **Benaki Musem** (www.benaki.gr) and the **Museum of Cycladic Art** (www.cycladic-m.gr).

As mentioned earlier, our survey has confirmed that only few of the sites (a 9.7% of the evaluation group) employ a thesaurus per se for describing their collections. Even fewer support and maintain a multilingual one (3%). This kind of thesaurus is based almost solely on translations of well-known international standards and classification systems. Out of the multilingual versions of standardized taxonomies and thesauri currently in use by Greek cultural digital collections and sites the most prominent seem to be:

- **LCSH** (http://www.loc.gov/cds/lcsh.html): The Library of Congress Subject Headings (LCSH) is a thesaurus from which the subject indices of documents (books, articles etc) are selected. It is an accumulation of the headings established at the US Library of Congress since 1898. It currently contains over 220.000 terms and its organization is based on the ISO-2788 standard. There exist custom translated versions of LCSH which are used by the majority of Greek libraries that provide access to their items information on-line. Concurrent multilingual use of LCSH is not always the case; however some bilingual examples include the **Library of the Technological Educational Institute of Thessaloniki**

(http://www.lib.teithe.gr) which employs LCSH 27th version and the on-line catalogue of the **Eugenides Foundation** (http://www.eugenfound.edu.gr) which demonstrates clear thesauri relationships like USE, UF, BT, NT and RT.

- **SEARS:** The Sears List of Subject Headings originally developed by Minnie Earl Sears in 1928 is an alternative to LCSH for small libraries. It has lower complexity than LCSH, as its headings tend to be shorter and its subdivisions fewer. Multilingual use of SEARS in Greece is rare. The **Library of the Technological Educational Institute of Lamia** (http://www.lib.teilam.gr) employs a bilingual version of SEARS for the thematic indexing of its items.

- **NLG-LCSH**: The National library of Greece (NLG, http://www.nlg.gr) has used LCSH as the basis for developing a customized translation in Greek. NLG maintains this adjusted translation and makes it available to other interested libraries and institutions which then adjust it further, according to their needs. Use of NLG-LCSH in languages other than Greek is unclear. The **Public Central Library of Serres** (http://www.serrelib.gr/) supports a monolingual NLG-LCSH version blended with SEARS headings.

Even though support for multilingual thematic indexing appears to be limited in cultural electronic archives, still a twofold momentum towards overcoming this situation can be recognized: Cultural institutions and organizations show demonstrable awareness and willingness to render their collections accessible and discoverable by a growing number of non-native speakers. On the one hand, there are on-going standardization efforts that try to offer some choice and guidance for multilingual description of digital cultural resources rather than enforce their suggestions. On the other hand, institutions seem steadily to adapt to the multilingual challenge as an always growing number of them enable multilingual access to their collections. As promising as it be, the situation can only benefit from a tighter coordination at national and international level.