

Ontology-based Integration of Cultural Heritage Metadata

Prof. Christos Papatheodorou
Dept. of Archives and Library Sciences - Ionian University
72 Ioannou Theotoki str., 49100 Corfu, Greece
email: papatheodor@ionio.gr

Abstract. Managing heterogeneous data is a challenge for cultural heritage institutions, archives, libraries, and museums which usually develop collections with heterogeneous types of material, described by different metadata schemas. For example, the Library of Congress, USA, provides EAD metadata for the archives description, MARC 21 records for the description of a wide variety of material, such as books and photographs, Text Encoding Initiative (TEI) for documenting the text of digital reproductions in the American Memory Collection, etc. The wide use of a number of cultural heritage metadata schemas imposes the development of interoperability techniques that facilitate unified access to cultural resources. One of the widely implemented techniques is the Ontology-Based Integration. Ontologies provide formal specifications of a domain's concepts and their interrelations and act as a mediated schema between heterogeneous sources.

This presentation describes an ontology-based metadata integration architecture. Its components are the mediator, which is based on the CIDOC CRM, the local sources, whose schemas are XML-based metadata schemas, and the mappings between the local sources and the mediator. Four integration scenarios are proposed on this architecture. Moreover a mapping language is demonstrated, called Mapping Description Language (MDL), to define the mappings of the metadata schemas of the local sources to the CIDOC CRM ontology as rules. An application of MDL usage to define the mapping from Encoded Archival Description (EAD) to the CIDOC CRM ontology is exhibited. Finally an algorithm for the transformation of EAD metadata to CIDOC CRM, as well as a query transformation algorithm from XPATH to CIDOC CRM are demonstrated.