

An ontological model for digital preservation

Panos Constantopoulos^{1,2} and Vicky Dritsou¹

¹Information Systems and Databases Laboratory, Athens University of Economics
and Business

²Digital Curation Unit, R.C. Athena

Email: {panosc | vdritsou}@aueb.gr

Abstract

The long-term preservation of digital resources requires assigning appropriate metadata. Several recommendations exist for preservation metadata reflecting the purposes and major preoccupations of the organizations that proposed each one. Digital resources are either digital surrogates of non-digital objects, or original digital objects. In either case we may assume that, if a digital resource is selected for preservation, it can be considered as a cultural object that belongs to some collection, has a certain value and must be appropriately preserved, used and documented. In other words, although the digital nature of the resource induces specific requirements in terms of preservation policies and techniques, basic commonalities can be traced with other cultural objects at the functional level. Based on this premise, the work reported here aims at harmonizing the information structures supporting digital preservation with those for documenting non-digital cultural objects. The approach taken is to develop a conceptual model for preservation metadata that complies with a standard ontology for cultural documentation, namely CIDOC CRM.

The perils faced by digital resources are related either to physical causes or to technological evolution. Correspondingly a number of preservation strategies have been developed and a preservation life cycle, independent of preservation strategy, has been identified. Furthermore, several proposals for preservation metadata have been made, five of which, appearing to be the most interesting and influential, namely the Dublin Core Metadata Element Set, the Open Archival Information System (OAIS), the Curl Exemplars Digital Archives (CEDARS) model, the Pittsburgh Project and the National Library of Australia (NLA) proposal, we analyze in order to understand their differences and to draw upon for our own model.

Our model is derived from the CIDOC CRM taking into account the above metadata sets. It is intended to capture the kinds of events occurring in the digital preservation life cycle, the kinds of objects involved and the relationships among those. Thus it could be considered as an application ontology for digital preservation derived from CIDOC CRM. It includes a minimal set of concepts appropriately inter-related and specialized. The resulting model yields a metadata set that displays significant overlap with the selected pre-existing ones. Its merit lies in the inference capability stemming from the explicit semantic structure, as well as in the integration with the domain of cultural documentation.