



Executive Summary

D2.2. Refined digital cultural resource data & data structure

The D2.2 presents the results of a collaborative effort of the project partners, which took place in months M2-M12 of the CrossCult project. It focuses on the production of semantically harmonized digital cultural resources based on the semantics of the CrossCult Upper-level ontology. Data modelling in the context of this deliverable refers to the specific process of applying the conceptual arrangements and definitions of the CrossCult (CC) Upper-level ontology to a range of disparate cultural heritage data. The CC ontology is a CIDOC-CRM compliant structure supplemented by ontological definitions from SKOS and FOAF, and extended with the project specific class Reflective Topic.

The task consists of identifying relevant data resources from the four pilots; cleaning up and formatting the data so that it aligns with the appropriate structure; modelling the data as OWL individuals and statements of the CrossCult (CC) ontology; enriching the data by linking it to external resources. The data we modelled is representative of the form and quality of data that will underpin the four pilots. To undertake this modelling we used a sample of data from each pilot in order to represent the diversity within each of the datasets.

The CC ontology ingests a wide range of diverse data associated to cultural heritage objects, events and subjects that span from Antiquity to Modern times. Integration of disparate data is inherited to a wide range of formats, technologies, management and classification approaches relevant to each data provider or source. Hence, the data modelling method addresses issues relating to the diversity of content types, data formats, and level of data detail. The modelling process abstracted into three main stages: i) selecting and curating the source data for each pilot, ii) data cleansing, transformation and data mapping to the Upper-level ontology and ii) automatic data assignment to CC ontology ensuring compliance with the model. The data sample of the four pilots describes museum exhibits, gallery items, archaeological sites and points of interest in terms of their unique identifier, associated descriptions, multimedia elements, and relevant keywords.

The adoption of the specialised classes E22 Physical Man Made Object and E24 Physical Man Made Thing, is a leading modelling choice that underpins most of the common requirements across the four pilots. The range of artefacts, paintings, museum exhibits, monuments, and points of interest that contribute to the four pilots are modelled as instances of the aforementioned classes. This choice enables the use of common semantics across the pilots for modelling spatial, temporal, geometrical, and other associative relationships. In addition, the CrossCult Classification Scheme (CCCS) supplements to the CC ontology, with

an additional layer of semantics through a controlled vocabulary of concepts. The terminological structure accommodates the keyword requirements of the project providing vocabulary for subjects and types.

The data modelling exercise delivered a representative example of pilot data with respect to the semantics of the Upper-level ontology. It managed to harmonise diverse data under a common semantic layer enriching their structure and enabling inference and retrieval. The next stages will focus on augmenting the data with media content and narratives that enhance their reflection and re-interpretation qualities and to semantically enrich the data with links to standardised semantic web resources.