



Executive Summary

D1.2 – Data Management Plan

The CrossCult project participates in the **Open Research Data Pilot** (OpenAIRE, <https://www.openaire.eu/opendatapilot>), which aims to make the research data generated by selected Horizon 2020 projects accessible with as few restrictions as possible, while at the same time protecting sensitive data from inappropriate access. This applies to (i) the data and metadata needed to validate results in scientific publications and (ii) other curated and/or raw data and metadata that may be required for validation purposes or with reuse value. Examples include anonymised user profiles, contextual information, statistics, results of experiments, measurements, observations resulting from fieldwork, survey results, interview recordings and images.

Deliverable 2.1 currently contains the initial version of the **Data Management Plan** (DMP) of CrossCult, describing how the research data collected or generated will be handled during and after the project. This will be revised at mid-term (August 2017) and at the end of the project (February 2019).

The DMP contains two types of datasets, namely general and pilot-specific ones:

- **General datasets** refer to the ontologies and vocabularies that provide common grounds for the knowledge management, reasoning, adaptation and personalisation procedures of the CrossCult platform. These include: (i) an upper-level ontology for cultural heritage, (ii) a venue ontology to formally represent the venues involved in CrossCult experiences, their infrastructure and exhibition characteristics, (iii) a user ontology that provides common foundations for different user modelling approaches, and (iv) classifications of terms and vocabularies defining common terminology, reflective topics, keywords, types, etc.
- **Pilot-specific datasets** include resources needed for the deployment and operation of the pilot applications, plus the capture and processing of experiment results. The four pilots have datasets with their own ontology extensions and literals, related to their topics of interest and reflection objectives. Other common datasets (realised with different structures and formats) include collections of user profiles, repositories of multimedia contents (predominantly, pictures and videos) and logs of user interactions and observations. Other disparate datasets are the collections of pre-defined routes and tours within the National Gallery from pilot 1, the game scenarios for pilot 2, the reflective narratives for pilot 3 or the geo-located object discovery sequences for pilot 4.

In June 2016, the Steering Committee made the decision to use Zenodo (<http://zenodo.org/>) as the main repository, plus a mirror managed by TEI-A (<http://hypatia.teiath.gr/>). The latter is linked to Google Scholar and to Europeana, which ensures extra dissemination at very little cost. Furthermore, it was decided to use Mendeley (<http://www.mendeley.com>) as a tool for reference management for the consortium. By March 2017, the Zenodo repository contains a first version of the general datasets (ontologies, terms and vocabularies).