



D2.2

Refined digital cultural resource data & data structure

2017-02-27

Andreas Vlachidis (UCL); Antonis Bikakis (UCL); Mellisa Terras (UCL);
Daphne Kyriaki-Manessi (TEI-A); Evgenia Vasilakaki (TEI-A); Kalliopi Kontiza
(NG); Yannick Naudet (LIST)

This deliverable presents the CrossCult digital datasets for each project pilot. It contains the specification of the data structure, which will enable further configuration and extensions, and a sample set of cultural resources modelled with respect to the Upper-level ontology semantics. It consists of two documents: a report describing the data modelling exercise and a set of owl files containing the data resources.



www.crosscult.eu

| | |
|------------------------|---|
| Project acronym | CROSSCULT |
| Full title | CrossCult: Empowering reuse of digital cultural heritage in context-aware crosscuts of European history |
| Grant agreement number | 693150 |
| Funding scheme | Innovation Action (IA) |
| Work programme topic | H2020-REFLECTIVE-2014-2015/REFLECTIVE-6-2015: Innovation ecosystems of digital cultural assets |
| Project start date | 2016-03-01 |
| Project duration | 36 months |

| | |
|-------------------------------|--|
| Workpackage | WP2 |
| Deliverable lead organisation | University College London |
| Authors | Andreas Vlachidis (UCL); Antonis Bikakis (UCL); Mellisa Terras (UCL); Daphne Kyriaki-Manessi (TEI-A); Evgenia Vasilakaki (TEI-A); Kalliopi Kontiza (NG); Yannick Naudet (LIST) |
| Reviewers | Catherine Jones (UL), George Lepouras (UOP) |
| Version | 1 |
| Status | Review Complete |
| Dissemination level | PU: Public |
| Due date | M12 (2017/02/28) |
| Delivery date | 2017/02/27 |

Table of Contents

| | |
|---|----|
| 1. Introduction | 6 |
| 2. Data Modelling Method | 7 |
| 2.1. Data Modelling Stages | 7 |
| 3. Pilot Data Modelling | 11 |
| 3.1. Pilot 1 | 11 |
| 3.2. Pilot 2 | 13 |
| 3.3. Pilot 3 | 14 |
| 3.4. Pilot 4 | 15 |
| 4. CrossCult Keywords Classification | 16 |
| 4.1. The use of Keywords in CrossCult | 16 |
| 4.2. Methodology for Constructing the CrossCult Classification Scheme | 18 |
| 5. Summary and Future Work | 20 |
| Appendix | 21 |
| 5.1. Distinct Type of Physical Item | 21 |
| 5.2. Physical Item Has Current Location | 21 |
| 5.3. Physical Item Produced by Production Event | 22 |
| 5.4. Reflective Topics – Flat List | 23 |
| 5.5. CrossCult Classification Scheme | 24 |

List of Tables

| | |
|--|----|
| Table 1. Mappings between relational database tables and fields, and ontology classes and properties. | 10 |
| Table 2 Examples of Associative Relationships between Reflective Topics and CCCS Keywords | 17 |

List of Figures

| | |
|---|----|
| Figure 1: Data Modelling Stages from unstructured data input to final OWL output | 8 |
| Figure 2: Table structures and relationships used for storing the data of Pilot 3..... | 9 |
| Figure 3: Data model of the painting “Portrait of Girolamo Fracastoro”, National Gallery (London). | 12 |
| Figure 4: Data model of the Sanctuary of Apollo Maleatas, Epidauros Greece..... | 13 |
| Figure 5: Data model of museum exhibit 2279 (Archaeological Museum of Tripolis, Greece) ... | 14 |
| Figure 6: Data model of a point of interest (Luxembourg City)..... | 15 |
| Figure 7: Relationships between Keywords and other CrossCult entities..... | 16 |

1. Introduction

This report presents the results of a collaborative effort of the project partners, which took place in months M2-M12 of the project. It focuses on the production of semantically modelled digital cultural resources for the four project pilots, which has been based on the semantics of the CrossCult ontology, as described in deliverable D2.3. The deliverable aim was to integrate data from the four pilots, using the semantics and terms of the CrossCult ontology, so that they share the same level of granularity and structure. This task consisted 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 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 deliverable is organised as follows. Section 2 describes in detail our modelling methodology, the process we followed to convert the pilot data, which was initially available in diverse structures and formats into a coherent, semantically enriched resource. Section 3 provides an overview of the pilot data and presents four representative data modelling examples, one for each pilot. Section 4 describes the role of keywords in CrossCult and presents the CrossCult Classification Scheme, which organises the keywords in a hierarchy. Section 5 provides a summary of the results.

2. Data Modelling Method

Data modelling in the context of this deliverable refers to the specific process of applying the conceptual arrangements and definitions of the CrossCult Upper-level ontology to a range of disparate data resources. The proposed upper-level ontology, as discussed and presented in deliverable D2.3, is a CIDOC-CRM compliant ontological structure supplemented by ontological definitions from SKOS and FOAF, and extended with the project specific class Reflective Topic. It is a flexible and robust schema of standardised conceptual abstractions, capable of supporting the data modelling needs of cultural heritage data which enjoy different degrees of complexity. The data modelling exercise relies on a rigorous set of Upper-level ontology definitions in order to express a diverse range of cultural heritage data on the same level of semantics and with the same degree of granularity. It is a process of iterative data modelling which runs throughout the life cycle of the CrossCult project with each modelling cycle refining and incrementing the volume of data that contribute to the project. The current data modelling method deals with the tasks of data decoupling, cleansing and semantic enrichment and establishes the foundations for further refinement in future iterations.

The origin of the data as well their coverage and granularity vary significantly. Four distinct pilots contribute data to the CrossCult project covering a unique range of cultural heritage venues across Europe. From the large venue of National Gallery in London to the considerably smaller venue of the Archaeological Museum in Tripolis (Greece) and from the archaeological site of thermal springs in Montegrotto (Italy) to the historical points of interest in the cities of Luxembourg and Malta. The project ingests a wide range of diverse data associated to cultural heritage objects, events and subjects that span from Antiquity to Modern times. Such disparate data means there is a wide array of formats, technologies, management and classification approaches relevant to each data provider or source. Hence, the data modelling method caters for solutions that address issues relating to the diversity of content types, data formats, and level of data detail. The following paragraphs discuss the various stages of the method and reveal the design techniques and development choices for modelling a range of datasets from well-structured normalised databases to unstructured textual descriptions of museum exhibits.

2.1. Data Modelling Stages

The process of modelling the data of the four CrossCult pilots can be abstracted into three main stages. At the lowest level is the stage of selecting and curating the source data for each pilot, then comes the process of data cleansing, transformation and data mapping to the Upper-level ontology (assertions).

Table 1). Then at the top level we align the data to the final ontology to ensure its compliance. A more detailed view of the modelling stages is shown in Figure 1, which presents all the intermediate stages of the data modelling task.

Since this was the first iteration in data modelling, the pilots provided a manageable and representative sample of data. The size of the sample did not vary significantly across the four pilots, with each pilot contributing data for 20 to 30 unique items (i.e. exhibits, painting, monuments, points of interest, etc.). However, in the cases of Pilot 2 and 3, due to the size of the

original datasets, the sample was already covering over the 80% of the available items. Overall, the data modelling exercise delivered 80 uniquely identified items that are composed of 102 Physical Man Made Objects and 17 Physical Man Made Things. This translates to 3440 ontology (OWL) statements of named individual declaration and property assertion.

Three main dataset types were provided by the pilots (see Figure 1) with varying structures: from the unstructured (word) or semi-structured (excel) to a structured data model (database). A large quantity of sample data that was provided by Pilot 2 and 3 was in the form of text files. These data described a range of museum exhibits and archaeological monuments in terms of their unique inventory number, associated descriptions and relevant keywords. Pilot 4 provided semi-structured data from spreadsheets, which arranged data in dedicated cells of labelled columns and distinct rows. However, some cells contained non-atomic values, which required further normalisation to indivisible structures. Pilot 1 provided a sample of structured data in the form of relational database tables, which were then mapped to the CrossCult ontology using automated methods as discussed below.

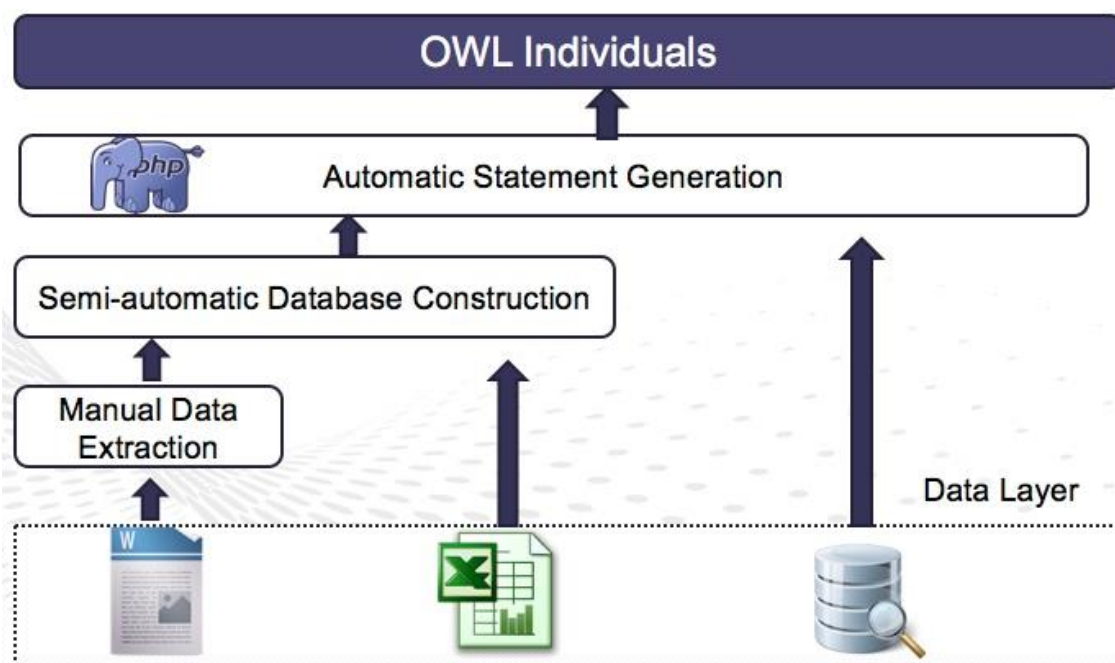


Figure 1: Data Modelling Stages from unstructured data input to final OWL output

2.1.1. Manual Data Extraction

The manual data extraction was dedicated to impose a data structure across a range of unstructured sample data available in text format. Such sample data contained a unique reference to a particular museum exhibit together with a rich description covering a wide range of facts and information about the exhibit, as shown below.

***2279:** Marble pediment tombstone with a representation of a family (enfance). The female figure bears a chiton and a cloak. The male figure and the boy bear a short chiton. On the architrave there is the inscription ANTIOXIC ΦΟΡΤΟΥΝΑΤΟΥ ΘΥΓΑΤΗΡ ΚΑΛΛΙΣΤΗ. Found in Herod Atticus villa in Loukou, Kynouria. Roman era work (middle Antonine era, 161 A.D - 180 A.D.). Dimensions: Height 1.60m, Width 0.82m. Location: Room 15, 1st floor (see Section 3.3)*

The volume of the data was not such to justify the development of a Natural Language Processing application for the automatic extraction of information from textual snippets. We manually extracted pieces of information from the text that could then be expressed as structured data and served by the upper-level ontology in the form of class instances. The manual extraction task focused on identifying entities of interest that would support the information retrieval needs of the various pilot scenarios. Thus, the task identified textual instances of relevant types (i.e. type of exhibit and related material), temporal and spatial information, dimensions, and other features of interest such as inscriptions or visual representations. The extracted pieces of information were stored as atomic values in intermediate formats, which were then utilised by the subsequent semi-automatic database construction stage, described below.

2.1.2. Semi-automatic Database Construction

The aim of the semi-automatic database construction was to populate a database with a set of structured data as a series of relational database tables. This stage received spreadsheets directly from the pilots and from the previous manual data extraction stage. With respect to the former input, we also applied some normalisation techniques for decoupling information from overloaded data entries and resolving such entries to their atomic values. The spreadsheet data were converted into comma separated values (CSV) and ready for populating a relational MySQL database.

The structure of the relational database that received the CSV files enabled the automatic generation of the final OWL statements. The relational database acted as a mediating layer between the semi-structured data files and the final OWL output feeding the routines of the Automatic Statements Generation stage with structured data. The database introduced a series of tables that stored the different types of CSV data such as temporal, spatial, dimension, features, and other information associated to the cultural heritage data. Figure 2 presents the database tables and their relationships to each other that form the database model for Pilot 3 (The Museum of Tripolis).

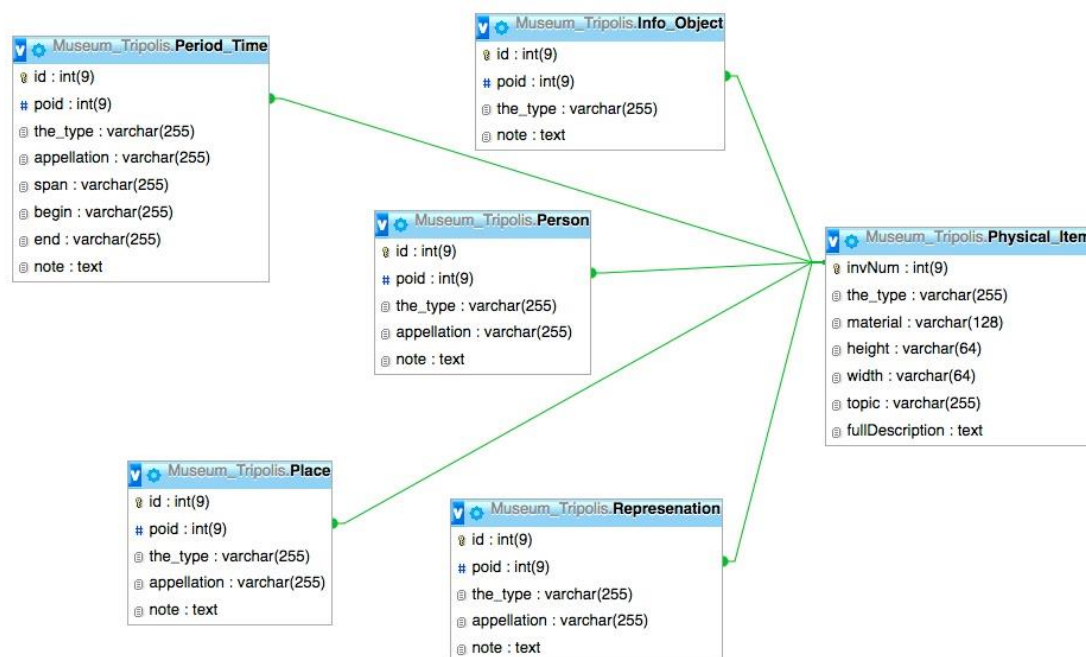


Figure 2: Table structures and relationships used for storing the data of Pilot 3

At the core of the structure is the *Physical_Item* database table. This table holds data directly related to the item such as type, material, dimensions, etc. Supplementary database tables such as *Place* and *Period_Time* hold spatial and temporal data, whereas other types of related to the Item information, such as visual depictions, inscriptions, context of use, etc. were stored in the tables *Info_Object* and *Representation*.

2.1.3. Automatic (OWL) Statements Generation

The final stage of the method utilised the structured data of the relational database to digitally transform the data to ontology individuals. We first mapped associations between the relational database tables and fields to ontology classes and properties (see Table 1). We then employed a series of PHP routines driven by SQL queries for retrieving selected database records and declaring them as ontology individuals using OWL class and property assertions.

Table 1. Mappings between relational database tables and fields, and ontology classes and properties.

| Database | Ontology |
|---|---|
| Physical_Item | E22.Man-made Object (Pilot 1, Pilot 3) or E24.Man-made Thing (Pilot 2, Pilot 4) |
| Physical_Item.InvNum | E42.Identifier |
| Physical_Item.material | E57.Material |
| Physical_Item.height Physical_Item.width | E54.Dimension |
| Physical_Item.topic | E55.Type |
| Physical_Item.note | P3.has note (xsd:String) |
| Info_Object | P73.Information Object |
| Person | E39.Actor |
| Person.appellation | E82.Actor Appellation |
| Place | E53.Place |
| Place.appellation | E44.Place Appellation |
| Period_Time | E52.Time-Span |
| Period_Time.appellation | E49.Time Appellation |
| Period_Time.begin | P82a.begin_of_the_begin (xsd:dateTime) |
| Period_Time.end | P82b.end_of_the_end (xsd:dateTime) |

A series of bespoke PHP routines were executed in a web server environment to generate several hundred OWL statements that implemented the mappings. The routines cater for the automatic generation of statements with respect to individual(s) declaration, class assertion, object property assertion, and data property assertion [Appendix 5.1]. The routines also applied string cleansing techniques for the generation of URI friendly values whereas in many cases complex SQL Join statements were used for retrieving record relationships across the database tables [Appendix 5.2]. In addition, the routines whenever necessary introduced declarations of ontology classes that were not explicitly defined by the mapping exercise but were required for the construction of CIDOC-CRM compliant relationships. For example, temporal data (161 AD) associated to a museum exhibit (2279 Tombstone) via an implicit E12.Production event (PRD2279), which is not apparent in the relational database but is generated by the PHP routines in order for the final output to comply with the arrangements of the Upper-level ontology [Appendix 5.3]

3. Pilot Data Modelling

The following section discusses four separate data modelling examples that highlight the key modelling requirements of the different pilots. Each pilot exploits parts of the ontology to satisfy specific modelling needs as reflected by the user scenarios (Deliverable D2.1) and the range of the selected data. The examples reveal data modelling choices that are common across the four pilots as well as choices that reflect individual requirements for each pilot.

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. For example, common structures are used for associating physical items with properties, such as current location, identified by, consists of, has dimension, is referred to by, was produced by, etc. The following examples focus on particular data modelling challenges whereas the design solutions are applicable across the four pilots.

3.1. Pilot 1

Pilot 1 contributes data using a sample of approximately 20 paintings from the National Gallery (London) collection. The source data is in a normalised structure with consistent information available for all items in the dataset. The data includes information about each painting's artist, medium and support, dimensions, date of production, location in the gallery, and other data explicitly related to the painting. There is also an extensive use of various types that describe paintings in terms of their design techniques, styles and materials while a range of keywords is also available to support retrieval of paintings and their relation to different topics.

Figure 3 presents the modelling arrangements for the painting titled "Portrait of Girolamo Fracastoro". It is an oil painting on canvas painted by Titian around 1528, having dimensions 84 x 73.5 cm and being displayed in Room 2 of the National Gallery. The diagram draws particular attention to the fine distinction between support (i.e. canvas) and medium (i.e. drying oil) of the artwork. This distinction is a particular modelling requirement relevant to paintings and useful to Pilot 1. In addition, the model incorporates information about the painting's dimensions and the place and time it was created. The model caters for the assignment of precise time-span values. Other conceptual characteristics of the painting, such as the name of the person(s) portrayed (depicted) and the painting's relation to reflective topics and relevant subjects (i.e. medicine) are also incorporated in the model using precise semantics available from the Upper-level ontology. The diagram also displays a number of semantic enrichment links to standardised external resources, such as the DBpedia and the Arts and Architecture Thesaurus of Getty, which provide additional definitions to the participating entities.

3.2. Pilot 2

Pilot 2 contributes data from four separate archaeological sites located across Europe: Chaves (Portugal), Epidauros (Greece), Lugo (Spain), and Montegrotto (Italy). As a result, the sample of data varies significantly in terms of format, level of detail and structure. The data ranges from extended descriptions of monuments and objects of the archaeological sites to simple, almost telegraphic entries of objects and their associated subject keywords. The pilot data contains references to entities other than physical objects (i.e. museum exhibits or gallery paintings) including monuments, physical features, and activities such as, historical locations, important people visits to places. Therefore, pilot 2 presents some particular modelling requirements for accommodating data relevant to monuments, physical features, and people visits.

Figure 4 presents the modelling arrangements for the “Sanctuary of Apollo Maleatas” which is located in archaeological site of Epidauros (Greece). The model uses the *E24.Physical Man-made Thing* definition of CIDOC-CRM for capturing the semantics of the monument (sanctuary) instead of the *E22.Physical Man-made Object* class, which is employed in the model of Pilot 1. The two classes enjoy a parent class relationship hence, they carry supplementary semantics, with E22 being a specialisation of E24. The former class (E22) assumes that objects if not too heavy can be moved, however the pilot requires modelling man made things which are impossible to move, such as mineral water thermal pools. The diagram accommodates the semantics of an activity (visit) that took place in Epidauros by the Hadrian Emperor and establishes the relationship between a monument being located in place and a visit to the same place by an important historical person. It is a modelling requirement specific to the scenario needs of Pilot 2. Additional relationships are also facilitated by the model for enabling conceptual relationships between the monument, reflective topics and subject keywords. The diagram also includes some examples of semantic enrichment that connect instances with standardised definitions from external semantic web resources; for example, E21.Person (Hadrian Emperor) is the same entity as the DBpedia resource <http://dbpedia.org/resource/Hadrian>.

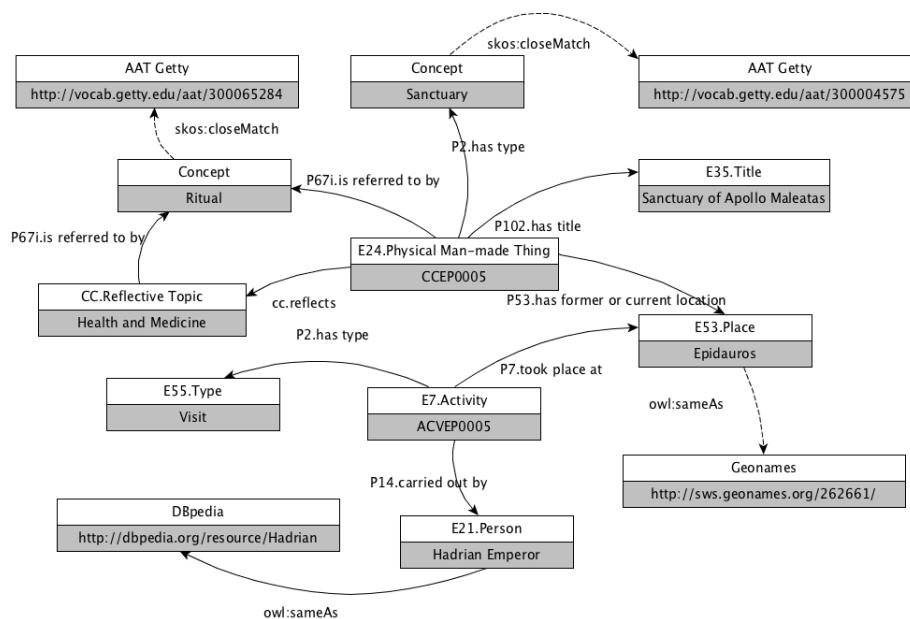


Figure 4: Data model of the Sanctuary of Apollo Maleatas, Epidauros Greece

3.3. Pilot 3

Pilot 3 contributes data from a sample of approximately 25 museum exhibits on display at the Archaeological Museum of Tripolis (Greece). The provided data contains rich descriptions including information about the temporal, geometrical, spatial and contextual characteristics of the exhibits. The descriptions do not vary significantly in terms of size and level of detail albeit some descriptions are a little longer than others. The modelling requirements for Pilot 3 have some parallels with the Pilot 1 requirements in terms of accommodating semantics of temporal, spatial and contextual definition. In particular, exhibits from the museum of Tripolis carry inscriptions and features that can be modelled in a similar way to the depictions in a painting. The Upper-level ontology, being a CIDOC-CRM compliant, contains a rich set of properties and classes for accommodating such requirements. The pilot presents some specific requirements with respect to the modelling of the provenance of exhibits. The provenance information of the exhibit is accommodated by an *E5.Event* of type excavation that took place in Kynouria (Greece).

Figure 5 presents the modelling arrangements for the museum exhibit 2279, which is a tombstone of the Middle Antonine era with an approximate production date between 161-180 AD. The diagram captures the semantics of the tombstone with respect to its static characteristics of dimension, date of production, material, and location. In addition, the model accommodates relationships to conceptual characteristics that describe the artefact in terms of its reflective topic and subject keywords. The inscription of the tombstone is modelled with precise semantics available from the upper-level ontology where the specialised property *P128.carries*, enables the relationship between the actual artefact and the carried inscription to be fully expressed. It is a different semantic relationship than the *P62.depicts* that is used for connecting an artefact with a depicted visual item. This fine distinction between depiction and carried inscription demonstrates the flexibility and breadth of the ontology to deal with precise semantics when required.

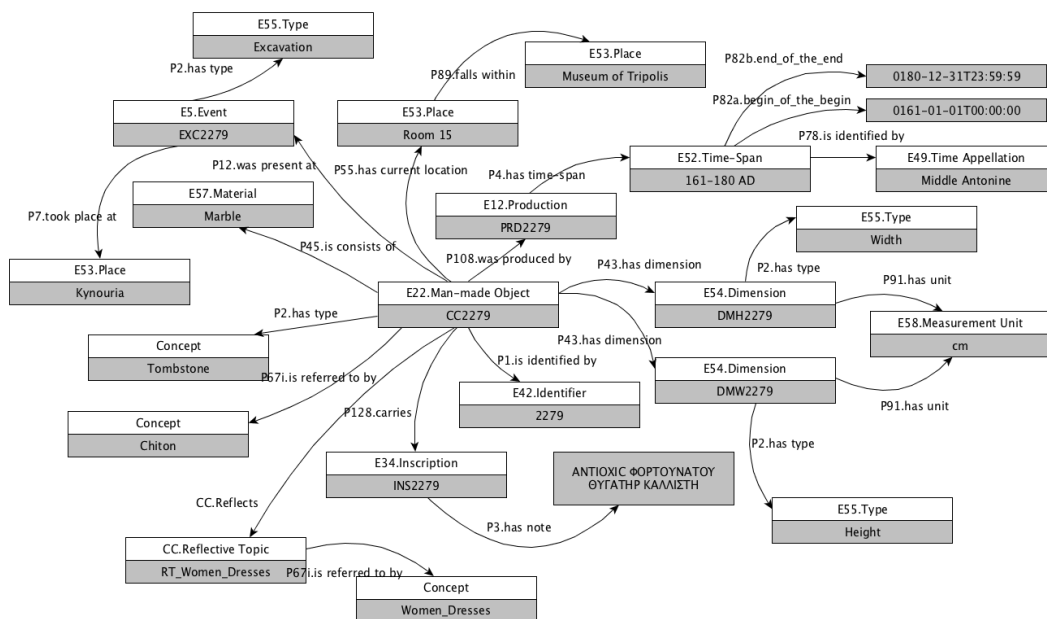


Figure 5: Data model of museum exhibit 2279 (Archaeological Museum of Tripolis, Greece)

3.4. Pilot 4

Pilot 4 contributes data from a sample of several Points of Interest (POI) located in the cities of Luxembourg and Malta. The data focuses on the relationship of POIs with specific reflective narratives and multimedia that drive the narratives and navigate the users of the application towards the location of POIs. The scenarios of Pilot 4 take place in a city, hence the venue is a city space in the context of a treasure hunt activity rather than a visit to a controlled environment of a museum, archaeological site, or an art gallery. Pilot 4 data focuses less on the static attributes of the POIs that contain spatial, geometric and temporal information and more on data that drives the reflective narrative and engagement with the POIs. This is a modelling requirement which is not only relevant to Pilot 4 but also applicable to the other pilots that also engage with different modes of reflection. The model draws some parallels with the Pilot 2 requirements in terms of accommodating semantics with respect to immovable man made things, hence the class *E24.Physical Man-made Thing* is used instead of *E22.Man-made Object* for capturing the semantics of POIs.

Figure 6 presents the modelling arrangements for the POI “Mir wëlle bleiwen wat mir sinn” (we want to remain what we are) which is a national motto of Luxembourg displayed in the balcony on the Façade of the Monkey's Bar on the Rue de la Loge (Luxembourg). The diagram presents the modelling arrangements with respect to the precise location of the POI in terms of its spatial coordinates. This is an arrangement particularly relevant to Pilot 4 due to the outdoor location of POIs. In addition, the model accommodates navigational information modelled as *E73.Information Object*, which is composed of the specialised visual and audio components, *E38.Image* and *E33.Linguistic Object*, respectively. The reflective narrative requirements of the pilot are addressed by the *E89.Propositional Object* class which is composed by a range of elements including *E35.Title*, *E38.Image* and body note whereas, reflective questions that relate to the narrative are also modelled as components of the same *E89.Propositional Object*.

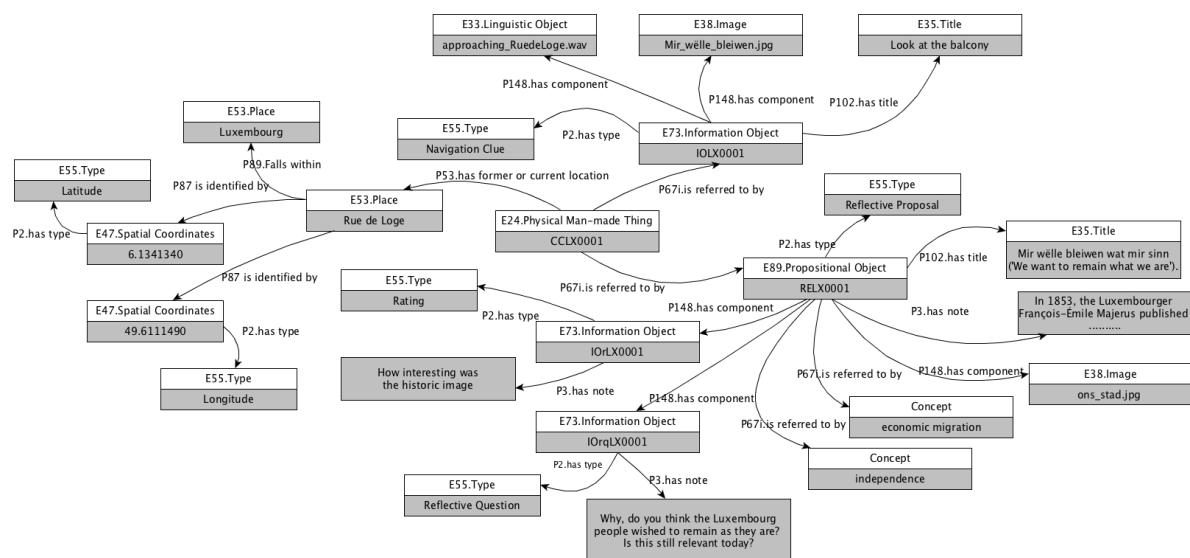


Figure 6: Data model of a point of interest (Luxembourg City)

4. CrossCult Keywords Classification

4.1. The use of Keywords in CrossCult

The *CrossCult Classification Scheme* (CCCS) is a terminological structure, supplementary to the ontology, which accommodates the keyword requirements of the project through a controlled vocabulary of concepts. The role of CCCS is not to classify objects according to their characteristics, which is handled by the ontology, but to provide a supplementary layer of terminology (as subjects, types etc.) that can be useful during retrieval.

Another distinct role of the classification system is to complement the flat list of Reflective Topics [Appendix 6.4] with vocabulary terms. The Reflective Topics (RT) are predetermined topics aimed at driving the reflection and reinterpretation qualities of the CrossCult application. Keywords from the CCCS can be linked with the RT in order to provide to the topics additional vocabulary and conceptual meaning. The task of associating RT with keywords is ongoing. Table 2 presents an example of relationships between RT and CCCS keywords that have been currently identified and applied to the pilot data models.

In CCCS, the Keywords are organised and defined in a hierarchical order whilst, where possible, are linked to external semantic definitions from the Arts and Architecture Thesaurus of Getty (AAT), the EUROVOC and the Library of Congress Subject Authorities (LC) vocabulary. Keywords from CCCS can be used both as “types” (instances of the E55 Type class), a form for further specialisation of the existing class hierarchy used in CIDOC-CRM, and as “propositional objects” (instances of E89) to describe the subjects related to a CrossCult entity. Figure 7 depicts how such uses can be supported by the CrossCult ontology.

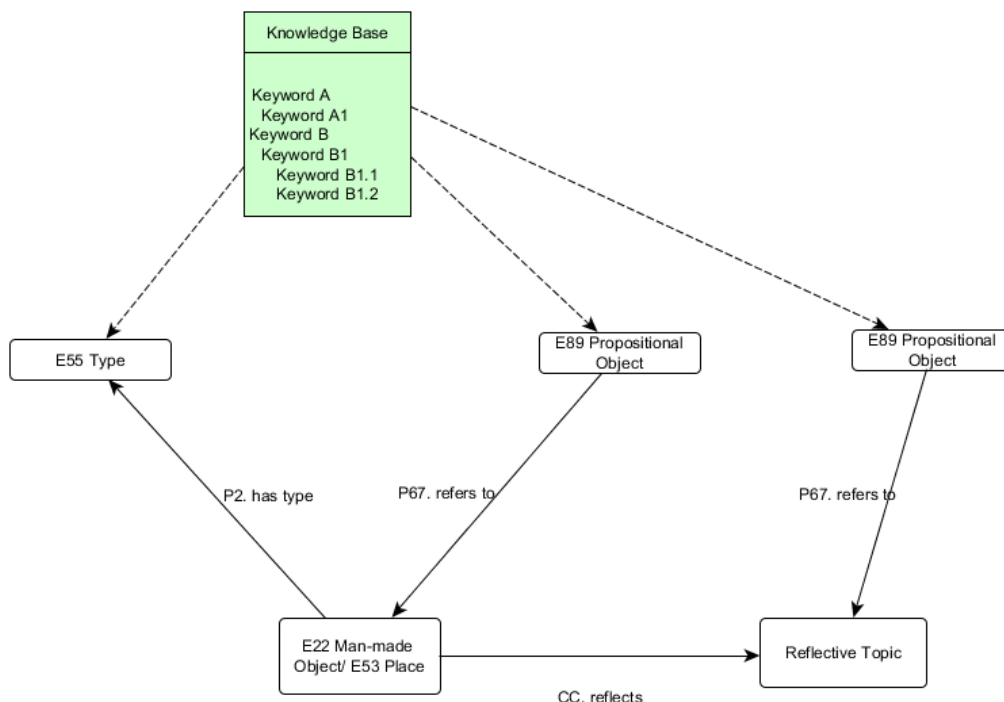


Figure 7: Relationships between Keywords and other CrossCult entities.

Table 2 Examples of Associative Relationships between Reflective Topics and CCCS Keywords

| Reflective Topic | Keywords |
|---|---|
| Human Senses and optical effects | Senses Sight |
| Objects as symbols in Altarpieces of the collection | Religious Work Iconography Altar Pieces |
| Costume_and_fashion_in_Paintings | Clothing |
| Thermal_Springs_and_Health | Healing Medical Treatment Healing Practice Mineral Water |
| Social Status | Mother Status of Women Under authority |
| Thermal Bath Buildings | Hyperthermal Roman baths Roman bath spaces Mineral water Thermal bath buildings Pool buildings |
| Religion and Rituals | Births Votive offerings Deities Ritual events People in religion |
| Religion and Thermal Springs | Shrines Ex voto shops Ablutions Ritual Events Sacred springs Deities Nymphae Apollo Asclepius |
| The Role of women in society | Daily life of women Social Status |
| Medicine and healing in art | Illness Medical treatment Health |
| Human senses and audible performances | Senses Hearing |
| Colours and Pigments through the ages | Trade routes Pigment Colorant material Materials |
| Life in the Netherlands in 17 th century | Daily life |

4.2. Methodology for Constructing the CrossCult Classification Scheme

The development of CCCS relies to a large extent on the reuse of terms from standard and widely known controlled vocabularies. The reuse of standardised resources ensures the validity of the CCCS structure and the consistency in the use of its terms. The classification scheme incorporates terms from the following vocabularies.

- Terms referring to arts and cultural heritage elements (e.g. art collections, painters, iconography, etc.) have been drawn from AAT (<http://www.getty.edu/research/tools/vocabularies/aat/>). Due to the nature of the project, most of the keywords have been mapped to terms of this vocabulary.
- Terms related to policies, social issues (e.g. migration), politics and the environment have been drawn from EUROVOC (<http://eurovoc.europa.eu>), an EU Thesaurus that covers extensively the particular subjects.
- More specific terms that do not fall within any of the above controlled vocabularies have been verified against the LC Subject Authorities (<http://authorities.loc.gov/>).

To ensure the comprehensiveness of CCCS and to maintain the project specific focus of the terminology, the contributing terms are derived from the scenario descriptions of the four pilots and the descriptions of relevant cultural heritage objects, including their meaning, symbolism, materials, cultural context and construction techniques.

The general methodology for constructing CCCS consists of the following steps:

- S1. We identify the terms used in the four pilots (as explained above) and discard duplicate terms.
- S2. We verify the terms against the authority vocabularies. For example, the term <collectors> used in Pilot 1 is available from AAT (<http://vocab.getty.edu/aat/300025234>), the term <emigration> of Pilot 4 available from EUROVOC (<http://eurovoc.europa.eu/724>), and the term <women in society> used in Pilot 2 and 3 can be found in LC Authorities (<https://lcn.loc.gov/n84736267>).
- S3. For each term that doesn't directly match any of the terms in the authority vocabularies, we examine the best possible match. For example, the term <Medical treatment> is used in pilots 1 and 2 and is a reference term in both EUROVOC and LC with the indication to use the valid term <therapeutics>. In this case, the term <therapeutics> is incorporated in the structure as the preferred term.
- S4. We place the term in the CCCS hierarchy, taking into account the structures of the external vocabularies. This process is described in more detail below.
- S5. We add further inter-term relationships, following the guidelines of the authority vocabularies. For example, the term <Healing> verified in LC is related to the term <traditional medicine> is defined in AAT.

The hierarchy of the current version of the CCCS (i.e. the “broader term” and “narrower term” (NT) relationships among the CCCS terms) is consistent with the structures of AAT and EUROVOC, which comply with the ISO 25964 standards

(<http://www.niso.org/schemas/iso25964/>) for thesaurus construction. We use the structures of the two vocabularies to construct CCCS, and specifically to:

- *establish broader and narrower term relationships between the CCCS terms*, following the AAT and EUROVOC hierarchies.

For example, for the term <pigment> used in Pilot 1, we established the following relationships, which are consistent with the AAT hierarchy.

<pigment> (<http://vocab.getty.edu/aat/300013109>)
 has BT <colorant (material)> (<http://vocab.getty.edu/aat/300013026>)
 has NT <blue pigment> (<http://vocab.getty.edu/aat/300013182>)

For a project-specific term, we first select the relevant facet from AAT/EUROVOC and then select the appropriate category, under which this term can be added. For example, the term <women in society> is not part of the authority records of AAT or EUROVOC but exists in LC Authorities. We can add the term under the broader EUROVOC category <social status>, therefore creating an original CrossCult structure within the Classification Scheme, i.e.

<social status> (<http://eurovoc.europa.eu/4277>)
 has NT: <women in society> (<https://lccn.loc.gov/n84736267>)

- *place at similar hierarchical level concepts of the same specificity*. This was necessary as AAT and EUROVOC are fully developed vocabularies accommodating thousands of terms and are much broader in scope than CCCS. For example, the terms <costume and fashion> have as narrower terms at the same level <costume accessories> and <hair styles>

Costume (mode of fashion) (<http://vocab.getty.edu/aat/300178802>)
 has NT Costume accessories (<http://vocab.getty.edu/aat/300209273>)
 has NT Hair styles (<http://vocab.getty.edu/aat/300262903>)

This is instead of having the full-fledged hierarchy of AAT, as it appears in <http://vocab.getty.edu/aat/300178802>.

- *create hierarchical relationships between project-specific terms*. For example, the term <appearance> (<http://www.crosscult.eu/KB#appearance>) used in Pilots 1, 2 and 3 does not appear in any of the controlled vocabularies. We added the term in the CCCS hierarchy under <culture related concepts> (<http://vocab.getty.edu/aat/300073689>) as it refers to a person's appearance that is perceived in various cultures.
- *to complete CCCS with concepts that were not originally identified in the four pilots*. For example, the term <jewelry and accessory components> (<http://vocab.getty.edu/aat/300387426>) accommodates a variety of object descriptions and it was considered as a useful addition for all pilots, albeit it was not originally identified in any of the pilots.

The current version of the CCCS is presented in Appendix 6.5.

5. Summary and Future Work

This deliverable describes the current status of the digital cultural resource data and data structure, which has been modelled following the semantic definitions of the upper-level ontology that are described in deliverable D2.3. The document discusses the details of the data modeling method revealing its stages and tasks in the process of homogenising disparate datasets under a common ontological layer. In addition, it describes the role of the CrossCult Classification Scheme as a project specific controlled vocabulary for the purposes of keyword assignment, and presents the development method of the scheme. The deliverable also includes four data modelling examples, which reveal modelling choices that are common across the four pilots and particular choices that reflect specific modelling needs of each pilot.

The data modelling exercise delivered a representative example of pilot data with respect to the semantics of the Upper-level ontology. The next steps of data modelling will expand to the full scale of the available data. The number of the contributing items is not expected to increase significantly for some of the pilots, which have already submitted data from over 80% of their available artefacts. The main focus of the next stages will be to augment 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. We will also further investigate the scope and structure of Reflective Topics and their relation to keywords, narratives and other reflection proposals. The CrossCult digital cultural resource data and data structure, and the CrossCult Classification Scheme will be further refined and their semantic enrichment will continue to expand until M25. Their final version will be described in the refined version of this deliverable at M25.

Appendix

Appendices 6.1-6.3 present sample PHP routines that automate the generation of OWL statements. Appendix 6.4 presents the current version of the Crosscult Classification Scheme.

5.1. Distinct Type of Physical Item

```
//Select Physical Item Identifier
$sql = "SELECT invNum,fullDescription FROM Physical_Item";
$result = mysqli_query($conn, $sql);
if (mysqli_num_rows($result) > 0) {
    // output data of each row
    while($row = mysqli_fetch_assoc($result)) {
        if (!empty($row["invNum"])) {
            // Declaration E42_Identifier Individual
            echo "<Declaration>". "<br>";
            echo "<NamedIndividual IRI=\"#\".$row[\"invNum\"]}\"/>". "<br>";
            echo "</Declaration>". "<br>";
            // Declaration E22_Man Made Object Individual
            echo "<Declaration>". "<br>";
            echo "<NamedIndividual IRI=\"#CC\".$row[\"invNum\"]}\"/>". "<br>";
            echo "</Declaration>". "<br>";
            // ClassAssertion E42_Identifier
            echo "<ClassAssertion>". "<br>";
            echo "<Class abbreviatedIRI=\"ecrm:E42_Identifier\"/>". "<br>";
            echo "<NamedIndividual IRI=\"#\".$row[\"invNum\"]}\"/>". "<br>";
            echo "</ClassAssertion>". "<br>";
            // ClassAssertion E22_Man Made Object
            echo "<ClassAssertion>". "<br>";
            echo "<Class abbreviatedIRI=\"ecrm:E22_Man-Made_Object\"/>". "<br>";
            echo "<NamedIndividual IRI=\"#CC\".$row[\"invNum\"]}\"/>". "<br>";
            echo "</ClassAssertion>". "<br>";
            // ObjectProperty E22_Man Made Object Pl_is_identified_by E42_Identifier
            echo "<ObjectPropertyAssertion>". "<br>";
            echo "<ObjectProperty abbreviatedIRI=\"ecrm:Pl_is_identified_by\"/>".
"<br>";
            echo "<NamedIndividual IRI=\"#CC\".$row[\"invNum\"]}\"/>". "<br>";
            echo "<NamedIndividual IRI=\"#\".$row[\"invNum\"]}\"/>". "<br>";
            echo "</ObjectPropertyAssertion>". "<br>";
            // DataProperty Physical Item P3_has_note
            echo "<DataPropertyAssertion>". "<br>";
            echo "<DataProperty abbreviatedIRI=\"ecrm:P3_has_note\"/>". "<br>";
            echo "<NamedIndividual IRI=\"#CC\".$row[\"invNum\"]}\"/>". "<br>";
            echo "<Literal"
datatypeIRI=\"http://www.w3.org/2001/XMLSchema#string\"/>\".$row[\"fullDescription"
]\"/>\"/>". "<br>";
            echo "</DataPropertyAssertion>". "<br>";
        }
    }
} else {
    echo "0 results";
}
```

5.2. Physical Item Has Current Location

```
//Select Physical Item Has Current Location
```

```

$sql = "SELECT Physical_Item.invNum, Place.poid, Place.the_type, Place.appellation
FROM Physical_Item INNER JOIN Place ON Physical_Item.invNum=Place.poid WHERE
Place.the_type = 'curLoc'";
$result = mysqli_query($conn, $sql);
if (mysqli_num_rows($result) > 0) {
    while($row = mysqli_fetch_assoc($result)) {
        if (!empty($row["appellation"])) {
            $appellation = htmlentities($row["appellation"]);
            $appellationRefined = str_replace(" ", "_", $appellation);
            // ObjectProperty E22_Man Made Object P55_has_current_location
E53_Place
            echo "<ObjectPropertyAssertion>". "<br>";
            echo "<ObjectProperty
abbreviatedIRI=\"ecrm:P55_has_current_location\"/>". "<br>";
            echo "<NamedIndividual IRI=\"#CC\".$row[\"invNum\"]\"/>". "<br>";
            echo "<NamedIndividual IRI=\"#\".$appellationRefined\"/>".
"<br>";
            echo "</ObjectPropertyAssertion>". "<br>";
        }
    }
} else {
    echo "0 Physical Item Has Current Location results";
}

```

5.3. Physical Item Produced by Production Event

```

//Select Physical Item is Produced by P12 Production Event
$sql = "SELECT Distinct poid FROM Period_Time";
$result = mysqli_query($conn, $sql);
if (mysqli_num_rows($result) > 0) {
    while($row = mysqli_fetch_assoc($result)) {
        if (!empty($row["poid"])) {
            $appellation = htmlentities($row["poid"]);
            $appellationRefined = str_replace(" ", "_", $appellation);
            // Declaration E12_Production_Event
            echo "<Declaration>". "<br>";
            echo "<NamedIndividual IRI=\"#PRD\".$row[\"poid\"]\"/>". "<br>";
            echo "</Declaration>". "<br>";
            // ClassAssertion E12_Production_Event
            echo "<ClassAssertion>". "<br>";
            echo "<Class abbreviatedIRI=\"ecrm:E12_Production\"/>". "<br>";
            echo "<NamedIndividual IRI=\"#PRD\".$row[\"poid\"]\"/>". "<br>";
            echo "</ClassAssertion>". "<br>";
            // ObjectProperty E22_Man Made Object P108i_was_produced_by
E12_Production_Event
            echo "<ObjectPropertyAssertion>". "<br>";
            echo "<ObjectProperty
abbreviatedIRI=\"ecrm:P108i_was_produced_by\"/>". "<br>";
            echo "<NamedIndividual IRI=\"#CC\".$row[\"poid\"]\"/>". "<br>";
            echo "<NamedIndividual IRI=\"#PRD\".$row[\"poid\"]\"/>". "<br>";
            echo "</ObjectPropertyAssertion>". "<br>";
        }
    }
} else {
    echo "0 Physical Item is Produced by P12 Production Event";
}
//Select P12 Production Event has Time Span
$sql = "SELECT Distinct poid, span FROM Period_Time";

```

```

$result = mysqli_query($conn, $sql);
if (mysqli_num_rows($result) > 0) {
    while($row = mysqli_fetch_assoc($result)) {
        if (!empty($row["poid"])) {
            $appellation = htmlentities($row["poid"]);
            $appellationRefined = str_replace(" ", "_", $appellation);
            $span = htmlentities($row["span"]);
            $spanRefined = str_replace(" ", "_", $span);
            // ObjectProperty E22_Man Made Object P108i_was_produced_by
            E12_Production_Event
                echo "&lt;ObjectPropertyAssertion&gt;". "<br>";
                echo "&lt;ObjectProperty abbreviatedIRI=\"ecrm:P4_has_time-
span\"/&gt;". "<br>";
                echo "&lt;NamedIndividual IRI=\"#PRD\".$row[\"poid\"]\"/&gt;". "<br>";
                echo "&lt;NamedIndividual IRI=\"#\".$spanRefined\"/&gt;". "<br>";
                echo "&lt;/ObjectPropertyAssertion&gt;". "<br>";
        }
    }
} else {
    echo "0 P12 Production Event has Time Span results";
}

```

5.4. Reflective Topics – Flat List

Pilot 1

- Medicine and Healing in Art
- Colours and Pigments through the Ages: colour change
- Costume and fashion in Paintings
- Human Senses and audible performances
- Human Senses and optical effects
- Life in the Netherlands in 17th century
- Medicine and Healing in Art
- Objects as symbols in Altarpieces of the collection
- Spanish Masterpieces in the Collection
- The Role of Women in Society

Pilot 2

- Thermal Springs and Health
- Religion and Thermal Springs
- Pilgrimage Routes
- Social Status of Pilgrims
- Thermal Bath Buildings
- Wellness routes
- Daily Life
- Leisure

Pilot 3

- Appearance (dresses, hairs...)
- Immortality
- Religion and Rituals
- Names and Myths and Animals
- Social Status
- Education
- Daily Life

Pilot 4

- Migration and the industrial revolution in Luxembourg
- General migration trends in Luxembourg
- Reflection on Migration on Malta: a focus on Emigration in the 20th Century
- Impact of Migration on cultural representation and cultural literacy in Luxembourg
- Impact of Migration on cultural representation and cultural literacy in Malta
- The story of the Knights Hospitaliers (Malta), symbolism, language and community
- WW1 and its impact on migration in Luxembourg (linked to Uni of Lux WW1 exhibition.)

5.5. CrossCult Classification Scheme**Build_environment**

Agoras
 Aqueducts
 Pipes
 Castelli_aquae
 Cisterns
 Fountains
 Gymnasiums
 Health_facilities
 Sanatoriums
 Houses
 Guest_houses
 Mansions
 Rural_houses
 Amphitheatres_[built_work]
 Bridges_[built_works]
 Buildings_[structures]
 Banquet_Halls
 Temples_[buildings]
 Cave_temples
 Wells_[structures]
 Villae
 Frescos
 Churches_[buildings]
 Basilicae
 Circuses_[Roman_arenas]
 Libraries_[buildings]
 Monasteries_[build_complexes]
 Sewers_[drainage_structures]
 Industrial_buildings
 Steel_meels
 Millaria

- Palaestrae
- Public_baths
 - Bath_accessories
 - Ceramics
 - Glass_vessels
 - Roman_bath_spaces
 - Steam_baths
 - Roman_baths
 - Thermal_bath_buildings
 - Pool_buildings
 - Mineral_water_pools
 - Private_pools
 - Public_pools
 - Sacred_pools
- Stadiums
- Theatres_buildings
 - Arena_theatres
- Demographics**
 - Labor_mobility
 - Migrant_workers
 - Cross_border_workers
- Economic_development**
 - Industrialization
- Education**
 - Literature
 - Music
 - Sciences
- Food**
 - Bread
 - Eggs
 - Fish
 - Fruits
 - Grains
 - Honey
 - Legumes
 - Meat
 - Game
 - Milk
 - Oil
 - Olives
 - Vegetables
 - Wine
- Furnishings_and_equipment**
 - Clothing
 - Buttons
 - Costume_accessories
 - Combs
 - Crinales
 - Foot_wear
 - Jewelry
 - Bracelets
 - Earrings
 - Gold_jewelry
 - Necklaces
 - Rings
 - Mirrors
 - Ointments
 - Perfumes
 - Drinking_vessels
 - Drinking_glasses
 - Furniture
 - Hair_styles
 - Vessels
 - Amphorae
 - Bowls
 - Dishes
 - Pans
 - Vases

Health

- Hygiene
- Illness
- Medical_treatment
 - Healing_practices
 - Healing_plants
 - Mud_bathing
 - Showering
 - Water_bathing
 - Water_drinking

Biological_concepts

- Senses
 - Hearing
 - Sight
 - Smell
 - Taste
 - Touch

Concepts_in_the_arts

- Attribution_qualifiers
 - Circle_of
 - Follower_of
 - Pupil_of
 - Studio_of
 - Workshop_of
- Colour_theory
- Iconography

Events

- Wars

Materials

- Colorant_[material]
 - Pigment
 - Azurite
 - Blue_pigment
 - Prussian_blue
 - Smalt
 - Ultramarine_blue

Migration

- Emigration
- Immigration
- Internal_migration
 - Rural_urban_migration
 - Urban_rural_migration

Objects

- Visual_works
 - Portraits
 - Self_portraits
 - Religious_works
 - Altar_pieces

People_in_the_arts

- Collectors

Political_concepts

- Government_policy
- Public_policy

Business_[commercial_function]

- Banking
- Trade_[function]
- Manufacturing
 - Ex_voto_shops

Culture_[related_concepts]

- Appearance
- Daily_life
 - Daily_life_of_children
 - Playing
 - Toys
 - Daily_life_of_men
 - Daily_life_of_women
 - Food_preparation
 - Raising_children

- Weaving
- Dowry
- Expeditions
- Language_[general_communication]
- Identity
 - Group_identity
- Immortality
- Islam
- Leisure
 - Bathing
 - Festivals
 - Games
 - Board_games
 - Dices
 - Reading
 - Sports
 - Theater
- Mortality
- Name_giving
- Nudity
- Philanthropy
- Western_culture
- Painting_[image_making]**
 - Mural_painting_[image_making]
- Religions_[belief_systems_and_cultures]**
 - Rituals_[events]
 - Ablutions
 - Births
 - Deaths
 - Healing
 - Votive_offerings
 - Coins
 - Inscriptions
 - Plaques
 - Marriage
 - Rites_of_passage
 - Funerals
 - Initiations
 - Last_rites
 - Purification
 - Holy_water
 - Sacrifices
 - Spondae
 - Weddings
 - Sacraments
 - Anointing_of_the_sick
 - Baptism
 - Confirmation
 - Eucharist
 - Holy_orders
 - Matrimony
 - Reconciliation
 - Mythical_or_legendary_beings
 - Creatures
 - People_in_religion
 - Deities
 - Apollo
 - Asclepius
 - Hygieia
 - Indigenous_deities
 - Nymphae
 - Nyrius
 - Monks
 - Nuns
 - Priestesses
 - Priests
 - Worshippers
- Pilgrimage

- Sacred_animals
 - Sacred_dogs
 - Sacred_horses
 - Sacred_snakes
- Sacred_objects
- Sacred_sites
 - Oracles
 - Sacred_wells
 - Sanctuaries
- Sacred_space
 - Cult_place
 - Shrines
 - Nymphaea
- Sacred_symbols
- Inhabited_places**
 - Cities
 - Villages
- Painting_techniques**
 - Acrylic_painting_[technique]
 - Mineral_painting_[technique]
 - Oil_painting_[technique]
 - Watercolour_painting_[technique]
 - Aquarelle_[technique]
 - Guache_[technique]
- Social_status**
 - Men_in_power
 - Emperors
 - Eparchs
 - Heads_of_states
 - Kings
 - Mayors
 - Rank
 - Clergy
 - Craftsmen
 - Free_citizens
 - Freed_people
 - Aristocracy_[social_class]
 - Industrialists
 - Knights
 - Magistrates
 - Merchants
 - Slaves
 - Soldiers
 - Status_of_Women
 - Autonomous
 - Courtesan_hetera
 - Under_authority
 - Daughter
 - Men's_ideal
 - Mother
 - Wife
- Styles-periods_and_Cultures_by_region**
 - Early_Western_World
 - Ancient_European_styles_and_periods
 - Mediterranean_[Early_Western_World]
 - Near_Eastern_[Early_Western_World]
 - European
 - Ancient_European
 - Early_Nederlandish
 - European_regions
 - Hellenic
 - Hispanic
 - Holy_roman_imperial
 - Early_Christian_Byzantine_styles_[cultures_and_periods]
 - Hun_[culture]
 - Irish_traveller_[culture]
 - Latin_[Latin_Europe]
 - Renaissance_[baroque_styles_and_periods]

- Rom_[culture]
- Walloon_[culture_or_style]
- Hutsul
- Medieval_styles_and_periods
- Modern_European_styles_and_movements
- Indo-European
- International_post_1945_styles_and_movements
- Hip-hop
- Punk_[international_movement]
- Minimal
- Postmodern
- Transportation_spaces**
 - Railroad_stations
 - Roads
 - Roman_roads
 - Routes
 - Pilgrimage_routes
 - Trade_routes
 - Transportation_vehicles
 - Busses
 - Carriages
 - Cars
 - Motorbikes
 - On_foot_transportation
 - Ships
 - Trains
 - Water_transport
 - Fluvial_routes
 - Maritime_routes
 - Ports
 - River_channels
- Travel**
 - Travel_stops
 - Travellers
- Visual_works**
 - Caricatures
- Water**
 - Drinking_water
 - Fresh_water
 - Caves
 - Lakes
 - Mineral_water
 - Alcaline
 - Bicarbonates
 - Bromic
 - Calcium
 - Chloride
 - Iodic
 - Radioactive
 - Saline
 - Sodium
 - Sparkling_water
 - Sulphate
 - Rivers
 - Springs
 - Sacred_springs
 - Thermal_springs
 - Cold
 - Hyperthermal
 - Hypothermal
 - Mesothermal
 - Wetlands
 - Salt_water
 - Oceans
 - Seas