Mingei

Deliverable 8.5 – Standardisation Report

Release Date	29.11.2019
Version	1.0
Dissemination Level	Public

Project Number	822336
Project Acronym	Mingei
Project Title	Representation and Preservation of Heritage Crafts

Deliverable Number	D8.5
Deliverable Title	Standardisation Report
Deliverable Type	Report
Dissemination Level	Public
Contractual Delivery Date	November 2019 (M12)
Actual Delivery Date	29.11.2019
Work Package	WP8 – INFORM & ENGAGE: Dissemination, Communication and Exploitation
Authors	Valentina Bartalesi, Daniele Metilli, Carlo Meghini, Xenophon Zabulis, Nikolaos Partarakis
Number of pages (incl. cover)	27

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 822336.
http://www.mingei-project.eu/

Executive summary

This deliverable describes the main standards for the Cultural Heritage sector, focusing in particular on knowledge representation (standard metadata schemas and standard ontologies), and on 3D formats for the representation of Cultural Heritage objects. The deliverable also reports the potential contribution of the project to standardisation activities in the domain of Cultural Heritage and Heritage Crafts. The updated versions of the deliverable will be submitted at M24 and M36.

The deliverable is structured as follows:

- **Section 1** provides an introduction to the context and purpose of this deliverable

- **Section 2** provides an overview of the three basic activities which are necessary for creating, using and sustaining digital heritage

- **Section 3** provides a summary of metadata schemes and standard ontologies of Semantic Web and Cultural Heritage

- **Section 4** makes an overview of the standards used in the Mingei Crafts Ontology

- **Section 5** provides an overview of 3D standard formats and the formats used by Mingei

- **Section 6** provides conclusions and future work

This deliverable is submitted in the context of T8.3. of Mingei. This is the first version of the deliverable. The next version of this deliverable will be submitted on M24.

Keywords

Cultural heritage, Craft heritage, Digital heritage, Standardisation, Semantic Web, Ontologies, Metadata schemas, 3D formats

Document History

Date	Version	Author/Editor	Affiliation	Comment
05.08.2019	0.1	Valentina Bartalesi	CNR	Initial draft
05.09.2019	0.2	Daniele Metilli	CNR	Updated Draft
05.10.2019	0.3	Carlo Meghini	CNR	Updated Draft
20.10.2019	0.4	Xenophon Zabulis	FORTH	Updated Draft, integration of 3D standardisation formats
10.11.2019	0.5	Nikolaos Partarakis	FORTH	Updated Draft, integration of Mingei 3D formats
25.11.2019	0.7	Valentina Bartalesi	CNR	Final version addressing review comments
27.11.2019	0.8	Margerita Antona	FORTH	QA review and final checks

Abbreviations

3DS	3D Studio
AAT	Art and Architecture Thesaurus
BVH	Biovision Hierarchy
CAD	Computer-aided design
CH	Cultural heritage
CIDOC	International Committee for Documentation
CiTO	Citation Typing Ontology
COLLADA	Collaborative Design Activity
CRM	Conceptual Reference Model
CrO	Crafts Ontology
CSG	Constructive solid geometry
DCMI	Dublin Core Metadata Initiative
EAD	Encoded Archival Description
EDM	Europeana Data Model
FBX	Filmbox
FOV	Field of view
FRBR	Functional Requirements for Bibliographic Records
FRBRoo	Functional Requirements for Bibliographic Records – Object Oriented
GLAM	Galleries, libraries, archives, museums
GPS	Global Positioning System
GPU	Graphics processing unit
HiCO	Historical Context Ontology
IFLA	International Federation of Library Associations and Institutions
IGES	Initial Graphics Exchange Specification
IRI	Internationalized Resource Identifier
ISO	International Organization for Standardization
LIDO	Lightweight Information Describing Objects
METS	Metadata Encoding and Transmission Standard
MTL	Material Template Library
NURBS	Non-Uniform Rational B-Spline
OAI-PMH	Open Archives Initiative Protocol for Metadata Harvesting
OWL	Web Ontology Language
PARTHENOS	Pooling Activities, Resources and Tools for Heritage E-research Networking, Optimization and Synergies
PEM	Parthenos Entities Model
PROV-O	Provenance Interchange Ontology

RDF	Resource Description Framework
RDFS	Resource Description Framework Schema
SKOS	Simple Knowledge Organization System
SIG	Special Interest Group
SPAR	Semantic Publishing and Referencing Ontologies
STEP	Standard for the Exchange of Product model data
STL	Stereolithography
UNESCO	United Nations Educational, Scientific and Cultural Organization
VRA	Visual Resources Association
VRML	Virtual Reality Modeling Language
W3C	World Wide Web Consortium
X3D	Extensible 3D Graphics
XML	Extensible Markup Language

Table of Contents

Executive summary 2

Keywords 2

1. Cultural Heritage 6

2. Digital Heritage 9

3. Semantic Web and Cultural Heritage 11

3.1 Standard Metadata Schemas 12

3.2 Standard Ontologies 13

3.3 Semantic Web Standard Languages 17

4. Standards in the Mingei Ontology 18

5. 3D Standard Formats for Cultural Heritage 19

1. Introduction

Generally speaking, Cultural Heritage is the legacy of physical artefacts and intangible attributes of a group or society that are inherited from past generations. The precise extent of what constitutes "heritage" is not fixed, being the result of a selection process by society [1].

According to the United Nations Educational, Scientific and Cultural Organization (UNESCO)[1], the term Cultural Heritage (CH) refers to two main categories of heritage: tangible cultural heritage and intangible cultural heritage. The first category is composed of: (i) movable cultural heritage, e.g. paintings, sculptures, manuscripts; (ii) immovable cultural heritage, e.g. monuments, archaeological sites; (iii) underwater cultural heritage, i.e. shipwrecks, underwater ruins and cities. UNESCO[2] defines intangible cultural heritage as "the practices, representations, expressions, knowledge, skills – as well as the instruments, objects, artefacts and cultural spaces associated therewith – that communities, groups and, in some cases, individuals recognize as part of their cultural heritage". For example, intangible cultural heritage may include oral traditions, performing arts, social practices, and crafts.

Cultural heritage institutions – galleries, libraries, archives, museums (GLAM) – play a fundamental role in preserving and providing access to cultural heritage collections, and to fulfill these goals it is crucial that their collections are properly digitised.

Digitisation allows CH collections to enter the digital economy, providing significant opportunities for improving public access and reuse of their assets [2]. Tools and applications based on CH collections can be applied in several fields (e.g. tourism, education, media), thereby increasing the value that CH collections provide to the public. Furthermore, digital tools and techniques (e.g. 3D scanning) are crucial for allowing the preservation and restoration of CH objects.

The Digital Agenda for Europe[3] promotes the digitisation of CH collections and their economic exploitation, calling for the use of innovative technologies to improve public access to them. Furthermore, the Directive 2013/37/EU[4] mandates that CH objects from GLAM institutions shall be re-usable for both commercial and non-commercial aims, promoting the achievement of this goal through the use of open standards and machine-readable open formats.

The Commission Recommendation on the digitisation and online accessibility of cultural material and digital preservation (2011/711/EU) asks Member States to increase the online accessibility of CH through mutual cooperation and partnerships with the private sector. Additionally, the Member

[1] http://www.unesco.org/new/en/culture/themes/illicit-trafficking-of-cultural-property/unesco-database-of-national-cultural-heritage-laws/frequently-asked-questions/definition-of-the-cultural-heritage/

[2] https://ich.unesco.org/en/convention

[3] https://s3platform.jrc.ec.europa.eu/digitisation-of-cultural-heritage

[4] https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32013L0037

States are encouraged to support large-scale digitisation through the implementation of appropriate legal frameworks and licensing mechanisms.

2. Digital Heritage

There are three basic activities which are necessary for creating, using and sustaining digital heritage: digitisation, access and preservation [3].

2.1. Digitisation

Digitisation is the process of transforming information into a digital format, in which the information is organized into bits. Through digitisation, an analogue CH object is converted into digital form. If the CH object does not have an analogue original, this step is replaced by the creation of the digital object itself.

Digitising information means turning it into a form that can be easily read by a computer.[5] Different techniques are used for different types of object – e.g. text, audio, video – and they consist of specialized hardware, software, protocols and standards, policies and procedures.

2.2. Access

Providing access to digital CH collections requires the creation of efficient and intuitive resource discovery tools. This access has to be guaranteed for different categories of users, with different backgrounds and affiliations.

2.3. Preservation

Preservation of digital CH objects is the process of guaranteeing that these objects will be kept accessible and usable in the future. In this context, the creation of metadata schemas is a key activity in order to allow tools to discover and access digital objects in a collection.

The Association for Library Collections and Technical Services[6] Preservation and Reformatting Section[7] of the American Library Association defines digital preservation as the combination of "policies, strategies and actions that ensure access to digital content over time". Therefore, digital preservation refers to the problem of retaining the meaning of a digital object unaltered for an evolving designed community [4]. The general steps [5] required for digital preservation are: (i) the original input is the physical storage (on some form of long-term storage media) of the sequence of bits which encodes the digital object in some format; (ii) by reading these bits from the storage media, the user obtains a sequence of bit values representing the digital object; (iii) by rendering

[5] https://www.collinsdictionary.com/dictionary/english/digitize

[6] http://www.ala.org/alcts/

[7] http://www.ala.org/alcts/mgrps/pars

these bits, the user obtains a representation of the object; (iv) by interpreting the representation, the user figures out its meaning. The goal of preservation is to allow the fulfillment of this whole process, including rendering and understanding the digital CH object, at any time in the future.

3. Semantic Web and Cultural Heritage

Digital Cultural Heritage collections represent very large amounts of information, resulting from the digitisation of CH collections with the goal of improving public access to them.

In the last decades, the World Wide Web has impacted the distribution of information in a significant way. The first approaches to bringing CH collections on the Web have focused on presentation of the contents of the collection, e.g. through Web sites and online access to their databases. But as the amount of available information has increased, there has also been a demand for “targeted global search, comparative studies, data transfer, and data migration between heterogeneous sources of cultural contents” [6]. The development of this kind of functionalities requires new approaches that would benefit greatly from standard ways to represent the semantic level of knowledge about CH objects.

The Semantic Web [7] is an extension of the World Wide Web through standards developed by the World Wide Web Consortium (W3C). According to the W3C, "The Semantic Web provides a common framework that allows data to be shared and reused across application, enterprise, and community boundaries" [7]. The Semantic Web employs formal methods for representing knowledge.

Formal methods are useful to make sense of the vast amounts of information that is published on the Web. It is quite challenging to handle CH objects through formal methods, because they are heterogeneous and also because information about the past is intrinsically incomplete [8]. Efforts towards the interoperability of knowledge representations about CH objects have led to the development of standard ontologies, such as the CIDOC CRM[8]. Where a shared semantic structure did not seem feasible, common metadata schemas have been created, the most relevant example being the Dublin Core[9] schema. At the same time, standard terminologies have been developed, such as the Library of Congress Subject Headings[10] and the Art and Architecture Thesaurus[11]. In the CH field, there now exist hundreds of data structures, metadata structures, and terminology systems, including dozens of standards. In the following, we report the most relevant standards in the context of the Mingei project.

[8] http://www.cidoc-crm.org/
[9] https://www.dublincore.org/specifications/dublin-core/dcmi-terms/
[10] http://id.loc.gov/authorities/subjects.html
[11] https://www.getty.edu/research/tools/vocabularies/aat/

3.1. Standard Metadata Schemas

Standard metadata schemas, such as the Dublin Core, represent "a common denominator by far too small to fulfill advanced requirements" [8], but can be useful as a sort of "metadata pidgin" [9] that enables shared understanding of common concepts.

3.1.1. Dublin Core

The Dublin Core schema is a small vocabulary developed by the Dublin Core Metadata Initiative (DCMI)[12]. The aim of the Dublin Core is to describe digital (video, images, web pages, etc.) and physical (books, CDs, artworks, etc.) resources. The metadata included in the vocabulary can be used for different goals, e.g. to combine different vocabularies of standard metadata or to allow the interoperability of metadata vocabularies in the Semantic Web architectures.

3.1.2. SKOS

The Simple Knowledge Organization System (SKOS)[13] is a data model for sharing and linking knowledge organization systems via the Web. Knowledge organization systems (e.g. thesauri, taxonomies) often have similar structures, and are used in similar applications. The SKOS allows representing this similarity in order to share data and technology among different applications. Furthermore, the SKOS data model provides a standard migration path for using existing knowledge organization systems into Semantic Web applications.

3.1.3. Soggettario Nazionale

Nuovo soggettario[14] is a subject indexing tool for several kinds of information resources. It is edited by the National Central Library of Florence, in compliance with the International Federation of Library Associations and Institutions (IFLA) recommendations. This tool can be used for indexing resources from general and specialized libraries as well as from museums, multimedia libraries, archives and documentation centres. The Nuovo soggettario system has been used by the Italian National Bibliography since 2007.

3.1.4. OAI-PMH

---

[12] https://www.dublincore.org/specifications/dublin-core/dcmi-terms/

[13] https://www.w3.org/TR/skos-reference/

[14] https://thes.bncf.firenze.sbn.it/index_eng.html

Open Archives Initiative Protocol for Metadata Harvesting[15] (OAI-PMH) was established in 2002 (Version 2.0 and the document was last updated in 2015) as a protocol for harvesting metadata descriptions of records in an archive in order to allow the construction of services using metadata from many archives. The aim of OAI-PMH is to facilitate broad access to digital resources for eScholarship, eLearning, and eScience. An implementation for the OAI-PMH must support the representation of metadata in Dublin Core.

3.1.5. AAT Thesaurus

AAT[16] (Art and Architecture Thesaurus) is a vocabulary used for managing information in the art and architecture domains. AAT includes over 90,000 words and phrases that describe "art, architecture, decorative arts, material culture, and archival materials". Among others, he AAT can be used for describing objects, materials, activities, agents, periods.

3.2. Standard Ontologies

As reported in [10], "there is a set of rich conceptual models or core ontologies of relationships for the digital world that are completely integrated and cover, in a complementary way, a vast spectrum of key conceptualizations for memory institutions and the management of digital content. Such core ontologies of relationships are fundamental to schema integration and play a vital role in practical knowledge management completely different to the role played by specialist terminologies".

Complexity of CH objects requires constant extension of domain ontologies. Therefore common standard ontologies listed below are under permanent development.

3.2.1. CIDOC CRM

The CIDOC Conceptual Reference Model (CRM)[17] is a core ontology that aims at enabling information exchange and integration between heterogeneous sources of CH information, archives and libraries. It provides semantic definitions and clarifications needed to transform heterogeneous information sources into a coherent global resource.

The CIDOC CRM promotes a shared understanding of cultural heritage knowledge by providing a common and extensible semantic framework for integrating cultural heritage information. CIDOC CRM is intended as a common language for domain experts and developers to identify the

15 https://www.openarchives.org/pmh/

16 https://www.getty.edu/research/tools/vocabularies/aat/

17 http://www.cidoc-crm.org/

requirements that the information systems have to satisfy and to define a good practice of conceptual modelling.

The CIDOC CRM is the result of over 20 years of work, originally done by the CIDOC Documentation Standards Working Group and then by the CIDOC CRM Special Interest Group (SIG). Since December 2006, CIDOC CRM is an official ISO standard. The standard was renewed in 2014 and can be found at ISO 21127:2014. The latest version of the ontology 6.2.7 which was released on October 2019 consists of 99 classes and 191 properties.

3.2.2. FRBRoo

FRBRoo[18] is a formal ontology that aims at capturing and representing the underlying semantics of bibliographic information and capturing the integration, mediation, and interchange of bibliographic and museum information.

FRBRoo is a joint effort of the CIDOC Conceptual Reference Model and Functional Requirements for Bibliographic Records international working groups.

The original FRBR model was designed as an entity-relationship model by a study group of the International Federation of Library Associations and Institutions (IFLA) and was published in 1998.

In summer 2006, a first draft of FRBRoo was completed. It is a logically formal model interpreting conceptualizations expressed in FRBRer and formulated as an extension of the CIDOC CRM, where any conflicts occurring in the harmonization process have been resolved.

3.2.3. Europeana Data Model

The Europeana Data Model[19] (EDM) aims at structuring the data that Europeana will be ingesting, managing and publishing. The EDM is a major improvement on the Europeana Semantic Elements (ESE), the basic original data model of Europeana. Each of the different heritage sectors represented in Europeana uses different data standards, and ESE reduced these to the lowest common denominator. On the contrary, EDM goes beyond the data standard used by the different heritage sectors represented in Europeana, i.e. the museums, archives, audiovisual collections and libraries. Indeed, EDM implements an open, cross-domain Semantic Web-based framework able to represent the range and richness of particular standards such as LIDO for museums, EAD for archives or METS for digital libraries.

3.2.4. Parthenos Entities Model

[18] http://www.cidoc-crm.org/frbroo/home-0
[19] https://pro.europeana.eu/resources/standardization-tools/edm-documentation

The PARTHENOS Entities Model (PEM) [11] aims at capturing and representing the knowledge generation process: which actors and which services are involved in the knowledge creation, resource curation, and management chain. The PEM is formalized by using CIDOC CRM and its extension CRMdig [12]. The former is able to capture the knowledge of cultural heritage objects, while the latter to describe the provenance of information and the digitization process. The current version (version 2.1) defines a total of 33 classes and 37 properties extending CIDOC CRM and CRMdig entities. The PE Model has been.

3.2.5. CiTO Ontology

CiTO[20] [13] is an ontology for representing the types of citation, both factual and rhetorical. A bibliographic citation is intended as a reference in a citation work given to another publication (e.g. a journal article, a conference paper, or a web page)..

The ontology identifies three different types of citations: (i) direct, e.g. the list of the references of a scientific paper, (ii) indirect (e.g. a more recent version of a scientific paper written by the same authors on the same research topic), or implicit (e.g. artistic quotations or parodies).

3.2.6. VRA Core

VRA Core[21] is a standard ontology for representing visual culture, developed by the Library of Congress[22] (LC) and the Visual Resources Association[23]. The ontology includes several categories of metadata elements that describe different aspects of the works of visual culture (e.g. subject, material, techniques, authorship). VRA Core is employed by GLAM institutions for the representation of objects from several different domains, e.g. art and archaeology.

3.2.7. Historical Context Ontology (HiCO)

HiCO is an ontology for describing historical context of cultural heritage objects.

Historical context involves aspects explicitly described in an object of interest, e.g. the description of an event in a document, but also aspects that are implicitly described, e.g. a citation of art styles in a paint. Explicit and implicit aspects are useful elements in order to deeply understand the content of the object of interest.

---

[20] https://sparontologies.github.io/cito/current/cito.html
[21] http://www.loc.gov/standards/vracore/
[22] https://www.loc.gov/
[23] https://vraweb.org/

HiCO, using referenced models like SPAR ontologies [14] and a set of properties from PROV-O ontology[24], aims to describe these issues.

3.2.8. OWL Time

OWL Time[25] is "an ontology of temporal concepts aiming to describe the temporal properties of resources in the world or described in Web pages". The OWL Time vocabulary allows the representation of time instants and time intervals, and the expression of relations among them. In addition, it allows the description of duration and time position (based on XSD datatypes). OWL Time supports the standard Gregorian calendar and also other temporal reference systems**.**

3.2.9. PROV Ontology

The PROV Ontology[26] (PROV-O) is a W3C recommendation that provides a vocabulary to represent provenance information. Provenance refers to the entities, activities, and agents involved in producing a specific object, and representing this knowledge allows reconstructing the process of its creation. The ontology supports different contexts and it can be specialized to represent more specific provenance information about different domains.

[24] https://www.w3.org/TR/prov-o/
[25] https://www.w3.org/TR/owl-time/
[26] https://www.w3.org/TR/prov-o/

3.3. Semantic Web Standard Languages

Ontologies are commonly encoded using ontology languages, i.e. formal languages that allow encoding knowledge about specific domains and are usually based on first order logic or description logics. These languages usually allow reasoning on the knowledge through the use of specialised software. In 2004, the W3C announced its support for two Semantic Web technology standards, RDF (Resource Description Framework) and OWL (Ontology Web Language). RDF [15, 16, 17] is a language used to represent knowledge about Web resources in a machine-readable way. In RDF, each resource is identified through an IRI. An RDF statement is a triple in the form of subject-predicate-object. A collection of statements representing a "slice of reality" forms an RDF graph. RDF Schema (RDFS) is an extension of RDF that includes a vocabulary allowing the modeling of classes, i.e. groups of resources, and the definition of class hierarchies and property hierarchies.

OWL (Web Ontology Language) [18, 19, 20] is an extension of RDF/S, based on description logics. OWL is more expressive than RDF/S, allowing the representation of axioms about the knowledge, e.g. relations between classes, cardinality, and characteristics of properties. OWL 2, the newest version of the language, supports three subprofiles, each tailored to a specific kind of application: OWL 2 EL, OWL RL, and OWL 2 QL [21].

4. Standards in the Mingei Ontology

The Craft Ontology (CrO for short) is an application ontology [22] developed by the Mingei project. CrO is obtained by integrating the following existing ontologies:

- the CIDOC CRM, a top ontology and an ISO standard forming the conceptual backbone of the CrO;
- the FRBRoo, a domain ontology for bibliographic records, resulting from the harmonization of FRBR with CRM;
- OWL Time, a domain ontology recommended by W3C for the representation of time;
- the Narrative Ontology [23], a domain ontology focused on the representation of narratives.

Of these four ontologies, three are standard vocabularies: the CIDOC CRM is an ISO standard, FRBRoo and OWL Time are standard *de facto.*

CrO also uses the standard languages of the Semantic Web for modelling knowledge, in particular:

- the Resource Description Framework (RDF) for basic knowledge representation;
- OWL 2 DL for ontology modelling;
- Simple Knowledge Organization System (SKOS) for expressing terminologies;
- XML Schema[27] for datatypes;
- the RDF Content ontology[28] for text modelling.

[27] https://www.w3.org/TR/xmlschema-2/
[28] https://www.w3.org/TR/Content-in-RDF10/

5. 3D Standard Formats for Cultural Heritage

A 3D file format is used for storing information about 3D models. Popular 3D formats such as STL, OBJ, FBX, and COLLADA are widely used in a plethora of applications.

5.1. Content

A 3D file format stores information about 3D reconstructed models. The encoded information regard mainly the *geometry* and *appearance of the modelled object. Some formats include scene* information and *animations*.

Model geometry regards the representation of the 3D shape of the reconstructed objects. In most cases, this representation is restricted to the surface of the objects, as it is the visible part of the object.

Appearance regards the way the object is perceived by the human visual system. This is determined by the illumination of the scene and the reflectance properties of the object. These properties are, in turn, determined by the absorption spectrum of the material(s) an object is made of.

The colour percept is determined by several factors, including cognitive processes (i.e., colour constancy). Common 3D models simulate these factors and represent appearance by colours, textures, and type of material.

The scene of a 3D model includes the location, orientation, diffusion and spectrum of light sources.

Formats that include animation determine how a 3D model moves and/or deforms during a time interval.

5.2. Formats

There are more than a hundred 3D file formats, some of which are associated with specific software (i.e., Blender, Unity, and AutoCAD).

Most 3D file formats store only the 3D model geometry and appearance. Though conversion is possible between such formats it is not always straightforward to so do, for proprietary formats.

Neutral or *open source* formats were invented as intermediate formats for converting between two proprietary formats.

Two such examples are STL (with a .STL extension) and COLLADA (with a .DAE extension).

Most 3D modelling software supports reading and writing popular neutral formats. In addition, most software applications support reading and writing to a most popular of proprietary formats.

3D file format	Type
STL	Neutral
OBJ	ASCII variant is neutral, binary variant is proprietary
FBX	Proprietary
COLLADA	Neutral
3DS	Proprietary
IGES	Neutral
STEP	Neutral
VRML/X3D	Neutral

5.3. Geometry

Typically representation of object geometry is considered as the fundamental content of a 3D file.

In order to encode the surface geometry, the following three methods are available, each of them with the corresponding strengths and weaknesses:

1. approximate mesh
2. precise mesh
3. constructive solid geometry (CSG)

5.3.1. Approximate mesh

In the approximate mesh encoding, the 3D surface of the model is represented by a mesh of polygons. Triangles are the simplest and most widely used shape. The vertices of the triangles and the direction of the outward normal vector to the triangles are stored.

The process of covering a surface with non-overlapping geometric shapes is known as “tessellation” and these file formats are also called “tessellated formats”.

Triangles are the simplest polygon that defines a planar piece of surface. In this way, they approximate the smooth geometry of the surface. The smaller the triangles the more faithful the approximation is. On the other hand, the smaller triangles, the larger the file size of the model.

Tessellated formats are used when the final product has not to have a very high resolution, e.g. 3D printing.

5.3.2. Precise mesh

In cases where the tessellated encoding is not fine enough, such as scientific simulation, the precise mesh encoding is employed.

Precise file formats compensate for resolution lost in tessellation by a generative approach that employees a parametric (analytic) representation of surfaces.

Typically, *Non-Uniform Rational B-Spline* patches (or NURBS) are employed to introduce "knots", which are control points and a set of parameters. The reconstructed surface is computed by smoothly interpolation over the control points.

Precise meshes can be created from approximate meshes. They are mainly utilized to generate surfaces that are smooth, at any scale of observation. On the other hand, the surface is computed on-line, thereby requiring significant computational cost and inapplicable in interactive uses of the model.

5.3.3. Constructive solid geometry

In constructive solid geometry, 3D shapes are constructed by adding or subtracting some primitive shapes (e.g. spheres). This format is oriented in the CAD of models rather than the digitization of existing objects.

5.4. Appearance

The encoding of appearance is of profound importance in the digitization of artefacts, as it is relevant to the material they are made of and, furthermore, the artistic dimensions that they may contain.

Appearance represents (or simulates) surface properties as material type, texture, colour (absorption spectrum) which determine how model looks like when it is visualised.

Information about appearance is encoded in two different ways.

5.4.1. Texture mapping

In texture mapping, every point in the encoded 3D surface is mapped to a 2D image, called the texture map of the model. When visualizing the 3D model, every surface point is assigned a coordinate in the texture map. Typically, in the file format, only the vertices of the mesh are associated with location in the texture map. Points in between are usually coloured online. When using a mesh of triangles, each triangle of the model is mapped to a triangle in the texture map. Barycentric coordinates or affine/perspective transformation are employed to map the triangle texture of the 3D face of the model. Usually, this operation is implemented in hardware (in the GPU) to accelerate the process.

The 2D texture map is, in most cases, stored separately in another file that the geometry content.

5.4.2. Face attributes

Another way for encoding appearance is the association of faces and/or vertices with colour, texture and material attributes. The main advantage of corresponding formats, is that they include a *specular component* encoding the reflectance behaviour of the surface due to light sources or reflections from other surfaces. Moreover, the include a *transmissive component* that encodes the colour aberration as well we the intensity and spectral transformation that radiation (light) undergoes when passing through the material that the object is made of. Finally, transparent and semi-transparent materials change the direction of light passing through them, a behaviour that is an *index of refraction* property, associated with the material type.

It ought to be noted that face attributes do not provide a faithful emulation of light behaviour. They rather simulate the visual result employing various techniques from the domain of Graphics to achieve a realistic rendering.

In order to encode appearance information using face attributes context-related knowledge is required, such as the type of material. Typically, such knowledge is required to be provided by an external source (i.e., the curator), as its inference from visual data can be either difficult or an ill-posed problem.

5.5. Scene information

Encoding information about the scene is supported by some 3D file formats. The scene describes the layout of the 3D model in terms of light sources, GPS location. Some formats also determine the properties of the virtual camera (pose, FOV, resolution) to be used for visualization of the model.

The encoding of the light source depends on the nature of the light source. In the simplest case of a point source, its location, colour, and intensity are stored.

In Mingei, we do not encode camera and light attributes, as they are controlled by end users or the specific visualization application.

5.6. Animation

Some 3D file formats have the capability to store animations of a 3D model. This is very useful in game designing or movie making where animations are used heavily.

5.6.1. Skeletal animation

The most widely used format of encoding the motion of a 3D model is “skeletal animation”. In skeletal animation, each model is associated with an underlying skeleton. The skeleton is a hierarchy of virtual “bones”. The motion of parent bones affects child bones, similarly to the human body, where a movement of the shoulder joint affects the position of the wrist and fingers.

The bones are typically rigid. However most formats allows non-rigid motion and encode scale and shear, besides rotation. Bones are connected by “joints”, which are associated with constraints in the possible rotations of a bone in relation to its parent. Given a rigid skeletal hierarchy, its posture can be encoded only by the angles of the joints and a skeletal rotation.

5.6.2. Techniques of animation

There are several ways of creating and storing animations of skeletal structures. The most important techniques are forward kinematics, inverse kinematics, and key-frames[1].

5.7. Mingei Formats

The two most widely used formats for 3D digitisations are STL and OBJ. Comprehensive reviews of the numerous types of 3D file formats can be found in the EduTechWiki of the University of Geneva on 3D file formats and on Wikipedia on 3D graphics file formats.

STL (STereoLithography) is important in the domains of 3D printing, rapid prototyping, and computer-aided manufacturing. The corresponding file extension is .STL. STL encodes the surface geometry of a 3D model approximately using a triangular mesh. STL ignores appearance and scene information. It is one of the simplest and leanest 3D file formats.

OBJ is a widely used file format in 3D graphics, associated with the file extension OBJ. The OBJ file format supports both approximate and precise encoding of surface geometry. For approximate encoding, the surface mesh is not restricted to triangular facets, but polygons can be used. For precise encoding, it uses smooth curves and surfaces such as NURBS. The OBJ format can encode colour and texture information. This information is stored in a separate file with the extension .MTL (Material Template Library).

BVH is an ASCII file that contains motion capture data for three-dimensional characters. The BVH file format was originally developed by Biovision, a motion capture services company. The BVH format is widely used format, though it lacks a full definition of the basis pose (this format has only translational offsets of children segments from their parent, no rotational offset is defined). In Mingei, we use BVH to encode the hierarchy of different models.

6. Conclusions and future work

In this deliverable we have presented an overview of the three basic activities which are necessary for creating, using and sustaining digital heritage. A summary of metadata schemas, standard ontologies for representing Cultural Heritage collections, and of the standard languages of the Semantic Web used to implement these ontologies has been reported. Furthermore, we have provided an overview of the standards (i.e. ontologies and languages) applied in the Mingei Crafts Ontology as well as of 3D standard formats used for producing 3D models in the Mingei project.

As future work, this deliverable will be updated to report all further developments of the CrO Ontology. We will also keep in contact with all Mingei partners to identify other standards used by them for the development of the project. Furthermore, we will contact standardisation bodies to identify specific technologies developed in the project that may be incorporated into existing or new standards. Finaly, a detailed description of how the Mingei protocol is aligned with existing standards in the domain will be provided in order to ensure that end users of the protocol will gain the maximum out of exisisting standardised formats.

The updated versions of the deliverable will be submitted at M24 and M36.

7. References

[1] Logan, W. S. Closing Pandora's Box: Human Rights Conundrums in Cultural Heritage. In Silverman, Helaine., Ruggles, D. Fairchild. (eds.). Cultural heritage and human rights. New York, Springer, 2007.

[2] Varbova, V. & Zhechkov, R. A policy brief from the policy learning platform on environment and resource efficiency. Interreg Europe – European Regional Development Fund, 2018.

[3] Sotirova, K. et al. Digitization of cultural heritage–standards, institutions, initiatives. Access to digital cultural heritage: Innovative applications of automated metadata generation, 2012.

[4] Gladney, H. Preserving Digital Information. Springer-Verlag, 2007.

[5] Flouris, G. and Meghini, C. Some preliminary ideas towards a theory of digital preservation. In Proceedings of the 1st international workshop on digital libraries foundations, 2007.

[6] Doerr, M. The CIDOC CRM, an ontological approach to schema heterogeneity. Dagstuhl Seminar Proceedings. Schloss Dagstuhl-Leibniz-Zentrum für Informatik, 2005.

[7] Berners-Lee, T. The Semantic Web (PDF). Scientific American, 2001.

[8] Doerr, M. The CIDOC conceptual reference module: an ontological approach to semantic interoperability of metadata. AI magazine 24.3, 2003.

[9] Baker, T. A grammar of Dublin Core. D-lib magazine 6.10, 2000.

[10] Doerr, M. and Stead, S. Harmonized models for the Digital World CIDOC CRM, FRBROO, CRMDig and Europeana EDM. In Tutorial. 15th Int. Conf. on Theory and Practice of Digital Libraries, TPDL, Berlin, Germany, 2001.

[11] Bruseker, G., Doerr, M. and Theodoridou, M. Report on the Common Semantic Framework, 2017.

[12] Doerr, M. and Theodoridou, M. CRMdig an extension of CIDOC-CRM to support provenance metadata, 2014.

[13] Shotton, D. CiTO, the Citation Typing Ontology. Journal of Biomedical Semantics, vol. 1, no Suppl 1, 2010.

[14] Peroni, S., Shotton, D. The SPAR Ontologies. In: Vrandečić D. et al. (eds) The Semantic Web – ISWC 2018. ISWC 2018. Lecture Notes in Computer Science, vol 11137. Springer, Cham, 2018.

[15] Beckett, D. The design and implementation of the redland rdf application framework. Computer Networks, 39(5):577–588, 2002.

[16] Brickley, D., Ramanathan V. G. and McBride, B. RDF Schema 1.1. W3C recommendation, 2014.

[17] McBride, B. The Resource Description Framework (RDF) and its vocabulary description language RDFS. In Handbook on ontologies, pages 51–65. Springer, 2004.

[18] Antoniou, G. and Van Harmelen, F. Web Ontology Language: OWL. In Handbook on ontologies, p. 67–92. Springer, 2004.

[19] Dean, M., Schreiber, G., Bechhofer, S., Van Harmelen, F., Hendler, J., Horrocks, I., McGuinness, D. L., Patel-Schneider, P.F. and Stein, A. OWL Web Ontology Language reference. W3C Recommendation, 2004.

[20] Patel-Schneider, P.F., Hayes, P., Horrocks, I., et al. OWL Web Ontology Language semantics and abstract syntax. W3C recommendation, 2004.

[21] Motik, B., et al. OWL 2 web ontology language profiles. W3C recommendation, 2009.

[22] Guarino, N. (ed.). Formal ontology in information systems: Proceedings of the first international conference (FOIS'98), June 6-8, Trento, Italy (Vol. 46). IOS Press, 1998.

[23] Bartalesi, V., Meghini, C. and Metilli, D. A conceptualisation of narratives and its expression in the CRM. International Journal of Metadata, Semantics and Ontologies, 12(1), 35-46, 2017.

[24] Lundgren M. A comparison of 3D file formats. Bachelor thesis performed in Information Coding. LiTH-ISY-EX-ET–11/0384–SE Linköping, 2011.