The 11th International Symposium on Virtual Reality, Archaeology and Cultural Heritage VAST (2010)
A. Artusi, M. Joly-Parvex, G. Lucet, A. Ribes, and D. Pitzalis (Editors)

LIDO and CRM$_{dig}$ from a 3D Cultural Heritage Documentation Perspective

D. Pitzalis[1,2], F. Niccolucci[2], M. Theodoriou[3] and M. Doerr[3]

[1]Centre de Recherche et de Restauration des Musées de France, Paris,France
[2]Science and Technology for Archaeology Research Center, The Cyprus Institute, Nicosia, Cyprus
[3]Institute of Computer Science, FORTH-ICS, Greece

Abstract

The most important characteristic of Digital Libraries is their flexibility in exposing content. Typically a DL provides a search interface which allows resources to be found. These resources can be local or remote, depending on how the data are organised within the DL and on how these data are made available for harvesting from/to other DLs. This kind of communication is possible because the structures of different DLs are expressed in formal specifications. In particular, especially in Cultural Heritage where we need to describe an extremely heterogeneous environment, some metadata standards are emerging and mappings are proposed to allow metadata exchange and enrichment. The CIDOC-CRM is an ontology designed to mediate contents in the area of tangible cultural heritage and it is ISO 21127 : 2006 *standard. In particular an extension of the CIDOC-CRM, known as CRM$_{dig}$, enables to document information about data provenance and digital objects in a very precise way. LIDO is a rich metadata schema suitable for handling museum-related data, still under development but very promising. In this paper we propose an update of the CIDOC-CRM to LIDO mapping and using a case study we will compare how CIDOC-CRM$_{dig}$ and LIDO handle the digital information of an object.*

Categories and Subject Descriptors (according to ACM CCS): D.2.12 [Software Engineering]: Interoperability—Data mapping, H.3.7 [Information Storage and Retrieval]: Digital Libraries—, I.3.m [Computer Graphics]: Miscellaneous—

1. Introduction

The choice of different metadata systems to be used in a digital library framework depends on a number of different factors: the nature of data, their intended use, and the interests and research methodology of the relevant community of use. The quest for ease of use and simplicity, accompanied by a limited need of manipulating the digital content, pushes a preference for schemas as simple as possible, which may then be perceived as inappropriate when the scope of the repository extends to cover other domains and other research goals. This has been the case of DC-based metadata schemas, suitable for managing literary works but not capable to deliver all the richness of content required by tangible cultural heritage, for example museum content. On the other hand, overarching schemas such as CIDOC-CRM have been labelled as too complex in cases where a flat structure, with only a small set of elements, satisfies the needs of the related users' community. The pacific, and fruitful, co-existence of digital objects pertaining to different culture domains is then assured by mapping the relevant metadata schemas to each other, the first step of interoperability.

In this paper we will consider a rich metadata schema, LIDO (Light Information Describing Objects), proposed to handle museum-related content in the framework of Europeana. Besides being a self-sufficient schema to be possibly used in the museum framework, LIDO is proposed by the European project ATHENA as the standard for digital content aggregators. A two-step process is envisaged: mapping individual repository schemas to LIDO and mapping (once for all) the latter to the current Europeana schema.

According to its proponents, LIDO [LID09] is a a metadata schema suitable for harvesting museum data developed by an international consortium [LID10] and adopted by the EU ATHENA [ATH09] project. LIDO is based on previ-

ous museum schemas such as CDWALite [CDW09], museumsdat [MUS09] and SPECTRUM [SPE09], and strongly relying on the CIDOC-CRM [CDG*09] reference model. From the museum schemas, LIDO derives flexibility, ease of use for museum personnel and coverage of most of the needs arising in a museum environment. Being CIDOC-CRM compliant, LIDO adopts the event-oriented approach and guarantees a high level of interoperability. LIDO has not been conceived as another collection management system, but as an harvesting schema for the delivery of metadata.

A LIDO record is conceptually organised in 7 areas called Wrappers: Object Identification, where the physical Object is identified; Object Classification, including information about its type; Relation, with the relations of the Object with other objects and its subject; Events, describing events in which the Object took part; Rights; Record, carrying the record information; and Resource, containing information about the Object's digital representation.

LIDO is being adopted by ATHENA as a common metadata schema for the aggregation and provision of digital content to Europeana. The schemas of individual repositories are mapped to LIDO, and, via the mapping of the latter to the Europeana metadata structure (currently ESE, Europeana Semantic Elements), ingestion of digital content is eventually made possible.

LIDO is still work in progress: version 0.8 has been released with release notes and is available in the above-mentioned ATHENA web site, while v0.9 is experimentally being adopted in the test ingestion phase.

Due to the increasing importance of LIDO for the documentation of cultural heritage, a mapping of CIDOC-CRM v5.0.1 to LIDO v0.7 has been undertaken and a concise representation of the mapping is available through the CIDOC-CRM web site [KD10]. This document sketches the correspondence between the two schemas. As it refers to LIDO v0.7, although it covers most of the LIDO elements, it needs updating. Recently, in order to capture provenance information of digital objects, an extension of CIDOC-CRM, named CRM_{dig}, has been developed [TTD*10] in the framework of the CASPAR [CAS09] first and 3D-COFORM later EU projects [3D-09]. Such information is paramount when dealing with digital replicas of cultural objects, in order to guarantee the transparency of the relation between the digital replica and the real physical original, therefore it seems important to enable this feature for LIDO as well.

The goal of the present paper is to update the LIDO to CIDOC-CRM mapping to include the most recent version of both schemas; to extend the mapping to CRM_{dig}; and to assess the capability of LIDO to deal with 3D cultural objects, possibly proposing improvements in this direction.

2. A case-study scenario

The examples used in the paper refer to the following case-study scenario: the"Mona Lisa" painting. The original painting is exposed at the Louvre Museum in Paris, France, more precisely in the newly restored "Salle des État" with the Venetian Paintings. The life of the painting itself is very interesting and rich. Painted by the Italian genius "Leonardo da Vinci" between $1503-1506$ it depicts the portrait of Lisa Gherardini, wife of Francesco del Giocondo. In particular, in 2004 a 3D model has been acquired using a laser scanner developed by CNRC (National Research Council Canada), based on a 3 laser technology at 3 different colour wavelength [BGM*07].

We can describe an hypothetic database, based on the information stored at C2RMF (Centre de Recherche et de Restauration des Musées de France) where every artwork corresponds, in a relation $1:1$, to a record structured like:

struct oeuvre {

string *oeuvre_recordId* = **REC1**;
string *oeuvre_title* = **(fr) La Joconde, (en) Mona Lisa, (it) La Gioconda**;
string *oeuvre_title_alternative* = **(fr) Portrait de Lisa Gherardini** $(1479-1550ca.)$**, épouse de Francesco del Giocondo, (en) Portrait de Lisa Gherardini** $(1479-1550ca.)$**, wife of Francesco del Giocondo, (it) Ritratto di Lisa Gherardini** $(1479-1550ca.)$**, sposa di Francesco del Giocondo**;
struct oeuvre_artist {
string *oeuvre_artist_name* = **(fr) Léonard de Vinci, (en) Leonardo da Vinci, (it) Leonardo da Vinci**;
string *oeuvre_artist_nationality* = **(fr) Italien, (en) Italian, (it) Italiano**;
string *oeuvre_artist_date_born* = $15/04/1452$;
string *oeuvre_artist_date_death* = $02/05/1519$;
string *oeuvre_artist_school* = **(fr) Italienne, (en) Italian, (it) Italiana**;
}
struct oeuvre_owner {
string *string oeuvre_owner_place* = **(fr) France, Paris, Musée du Louvre, (en) France, Paris, The Louvre Museum, (it) Francia, Parigi, Museo del Louvre**;
string *oeuvre_owner_inventoryId* = **INV 779**;
string *oeuvre_owner_collection* = **(fr) peinture, (en) painting, (it) pittura**;
}
string *oeuvre_category* = **(fr) peinture, (en) painting, (it) pittura**;
time *oeuvre_date_creation_begin* = 1502;
time *oeuvre_date_creation_end* = 1506;
string *oeuvre_material* = **(fr) peuplier, (en) poplar, (it) pioppo**;
string *oeuvre_technique* = **(fr) peinture á l'huile,(en) oil on wood, (it) olio su tavola**;
struct ouvre_size {

string *oeuvre_size_width* = 530*mm*;
string *oeuvre_size_height* = 770*mm*;
string *oeuvre_size_depth* = 30*mm*;
}
string *oeuvre_mainBiblio* = **catalogue sommaire illustré des peintures du musée du Louvre t.II Italie, Espagne, Allemagne, Grande-Bretagne et divers. Paris RMN 1981 p.192**;
time *oeuvre_dataEntry* = 01/01/2010;
string *oeuvre_ownerEntry* = **The Mapper**;
url *oeuvre_thumb* = **http://www.louvre.fr/...**
}

}

For every artwork we can have multiple digital resources, with the relation $1:N$ where 1 is the artwork and N the number of digital resources. For clarity we will illustrate just the information concerning the 3D acquisitions event.

struct film {

string *film_recordId* = **DIG2**;
string *film_oeuvreId* = **REC1**;
string *film_technique* = **laser scanning**;
string *film_mime* = **ply**;
string *film_device* = **camera CNRC**;
time *film_date* = 29/10/2004;
string *film_author* = **CNRC team**;
string *film_rigtht* = **Centre de Recherche et de Restauration des Musées de France**;
string *film_view* = **whole**;
string *film_size* = **700M vertex**;
string *film_path* = **/path/to/image.ply**;

}

According to [Doe00], [BSM*05] and [PLP*06] we can represent the oeuvre record in CIDOC-CRM as:

Struct oeuvre
E84.Information_Carrier "*Mona Lisa*" →
P70B.is_documented_in → E31.Document "*our database*"
oeuvre_recordId *REC1*
E84.Information_Carrier "*Mona Lisa*" →
P48F.has_preferred_identifier → E42.Identifier "*REC1*"
oeuvre_title *Mona Lisa*
E84.Information_Carrier "*Mona Lisa*" → P102F.has_title
→ E35.Title "*Mona Lisa*"
E35.Title "*Mona Lisa*" → P2F.has_type → E55.Type
"*Main Title*"
→ P72F.has_language → E56.Language "*En*"
→ P73F.has_translation → E35.Title "*La Joconde*"
E35.Title "*La Joconde*" → P72F.has_language →
E56.Language "*Fr*"
→ P73F.has_translation → E35.Title "*La Gioconda*"
E35.Title "*La Gioconda*" → P72F.has_language →
E56.Language "*It*"
oeuvre_title_alternative *Portrait de Lisa Gherardini . . .*
E84.Information_Carrier "*Mona Lisa*" → P102F.has_title
→ E35.Title "*Portrait de Lisa Gherardini . . .* "
E35.Title "*Portrait de Lisa Gherardini . . .* " →
P2F.has_type → E55.Type "*Alternative Title*"
→ P72F.has_language → E56.Language "*En*"
→ P73F.has_translation → E35.Title "*Portrait de Lisa Gherardini . . .* "
E35.Title "*Portrait de Lisa Gherardini . . .* " →
P72F.has_language → E56.Language "*Fr*"
→ P73F.has_translation → E35.Title "*Ritratto di Lisa Gherardini . . .* "
E35.Title "*Ritratto di Lisa Gherardini . . .* " →
P72F.has_language → E56.Language "*It*"
struct oeuvre_artist
E84.Information_Carrier "*Mona Lisa*" →
P108B.was_produced_by → E12.Production "*The Painting of Mona Lisa*"
E12.Production "*The Painting of Mona Lisa*" →
P14B.carried_out_by → E21.Person "*Leonardo da Vinci*"
E21.Person "*Leonardo da Vinci*" → P14.1B.in_the_role_of
→ E55.Type "*Artist*"
oeuvre_artist_name *Leonardo da Vinci*
E21.Person "*Leonardo da Vinci*" → P131B.is_identified_by
→ E82.Actor_Appellation "*Leonardo da Vinci*"
oeuvre_artist_nationality *Italian*
E21.Person "*Leonardo da Vinci*" →
P107B.is_current_or_former_member_of → E74.Group
"*Italian Nationality*"
E74.Group "*Italian Nationality*" → P71B.is_listed_in →
E32.Authority_Document "*Nationality*"
→ P102F.has_title → E35.Title "*Italian*"
oeuvre_artist_date_born 15/04/1452
E21.Person "*Leonardo da Vinci*" → P98B.was_born →
E67.Birth "*the birth of Leonardo*"
E67.Birth "*the birth of Leonardo*" → P4F.has_time-span →
E52.Time-span → P81.ongoing_throughout →
E61.Time_Primitive "15/04/1452"
oeuvre_artist_date_death 02/05/1519
E21.Person "*Leonardo da Vinci*" → P100B.died_in →
E69.Death "*the death of Leonardo*"
E69.Birth "*the death of Leonardo*" → P4F.has_time-span
→ E52.Time-span → P81.ongoing_throughout →
E61.Time_Primitive "02/05/1519"
oeuvre_artist_school *Italian*
E21.Person "*Leonardo da Vinci*" →
P107B.is_current_or_former_member_of → E74.Group
"*Italian School*"
E74.Group "*Italian School*" → P71B.is_listed_in →
E32.Authority_Document "*School*"
hspace*1cm → P102F.has_title → E35.Title "*Italian*"
struct **oeuvre_owner** *The Louvre Museum*
E84.Information_Carrier "*Mona Lisa*"
→P52F.has_current_owner → E40.Legal_Body "*The Louvre Museum*"
oeuvre_owner_place *France, Paris, The Louvre Palace*

E40.Legal_Body "*The Louvre Museum*" → P74F.has_current_or_former_residence → E53.Place "*Louvre*"
E53.Place "*Louvre*" → P87B.is_identified_by → E44.Place_Appellation "*France, Paris, The Louvre Palace*"
oeuvre_owner_inventoryId *INV 779*
E40.Legal_Body "*The Louvre Museum*" → P14B.performed → E15.Identifier_Assignment "*Assign Inventory Number*" → P37B.assigned → E42.Identifier "*INV 779*"
E42.Identifier "*INV 779*" → P2F.has_type → E55.Type "*Inventory number*"
oeuvre_owner_collection *Italian Paintings*
E84.Information_Carrier "*Mona Lisa*" → P46B.forms_part_of → E78.Collection "*Italian Paintings*"
E78.Collection "*Italian Paintings*" → P109F.has_current_or_former_curator → E40.Legal_Body "*The Louvre Museum*"
oeuvre_category "*Painting*"
E84.Information_Carrier "*Mona Lisa*" → P2F.has_type → E55.Type "*Painting*"
E55.Type "*Painting*" → P2F.has_type → E55.Type "*Category*"
struct oeuvre_date 1502 – 1506
E12.Production "*The Painting of Mona Lisa*" → P4F.has_time_span → E52.Time-Span * → P82F.at_some_time_within → E61.Time_Primitive "1502 – 1506"
oeuvre_material *Poplar*
E12.Production "*The Painting of Mona Lisa*" → P126F.employed → E57.Material "*Poplar*"
E57.Material "*Poplar*" → P71B.is_listed_in → E32.Authority_Document "*Materials*"
oeuvre_technique *Oil on wood*
E12.Production "*The Painting of Mona Lisa*" → P32F.used_general_technique → E55.Type "*Oil on wood*"
E55.Type "*Oil on wood*" → P71B.is_listed_in → E32.Authority_Document "*Techniques*"
oeuvre_size_width *530 mm*
E84.Information_Carrier "*Mona Lisa*" → P43F.has_dimension → E54.Dimension "*Mona Lisa's Width*"
E54.Dimension "*Mona Lisa's Width*" → P2.has_type → E55.Type "*Width*"
hspace*1cm → P90.has_value → E60.Number "*530*"
hspace*1cm → P91.has_unit → E58.Measurement_Unit "*mm*"
oeuvre_mainBiblio *Catalogue sommaire...*
E84.Information_Carrier "*Mona Lisa*" → P70B.is_documented_in → E31.Document "*book*"
E31.Document "*book*" → P2F.has_type → E55.Type "*Catalogue*"
hspace*1cm → P1B.is_identified_by → E75.Conceptual_Object_Appellation "*Catalogue sommaire...*"
oeuvre_dataEntry 01/01/2010
E42.Identifier "*REC1*" → P37B.was_assigned_by → E15.Identifier_Assignment "*Recording Mona Lisa*"
E15.Identifier_Assignment "*Recording Mona Lisa*" → P4F.has_time-span → E52.Time-Span * → P82F.at_some_time_within → E61.Time_Primitive "01/01/2010"
oeuvre_ownerEntry *The mapper*
E15.Identifier_Assignment "*Recording Mona Lisa*" → P14F.carried_out_by → E21.Person "*The mapper*"

The schema can be *synthesised* in Figure 1. The film structure will be presented after the introduction to the CRM$_{dig}$ here below.

3. An overview of CRM$_{dig}$

The issue of provenance of digital artefacts is gaining increasing importance as digital technologies acquire an important role in cultural heritage research and practice. Both culture and science assign a key position to the uninterrupted chain linking the original and the processed outcome. Provenance in science means experiment repeatability and verifiability, in culture means being original and not counterfeited.

If some steps of a documented "chain of custody" are missing for an artefact, they must be laboriously reconstructed, reverse engineered, based on the analysis of the features of the object. When culture and technology are intermixed, as it happens for 3D replicas of cultural objects, both motivations determine the necessity of ascertaining the provenance of digital objects. In this case the hiatus between reality (the real thing) and virtuality (the digital replica) is the most delicate step, because the methodology needs to swap from the tangible to the intangible (digital). Continuity of the "chain of custody" relies thus on documentation, which is in charge of keeping track of the (accepted) alterations and of providing evidence that no other undocumented modification took place. A similar care must be paid when a digital object is processed, for example to "clean" a 3D model or to simplify its structure, with a purely "soft" process. For this reason an extension of the London Charter [LON09] to safeguard provenance information of cultural digital objects has been proposed, and an extension of CIDOC-CRM, called CRM$_{dig}$, has been defined to document provenance metadata [TTD*10].

To monitor all relevant parameters of digital provenance it is assumed that a suitable interactive Workflow Monitoring Tool is available and that machine action is completely determined by the specification of the machine and its input parameters therefore there is no need to further decompose it in the provenance record. Digitisation will operate on a finite set of physical objects and will produce digital output for each of them and ultimately generate a 3D model. The modelling approach is event centric and follows a hierarchical workflow structure.

The main data acquisition process is an event referred to

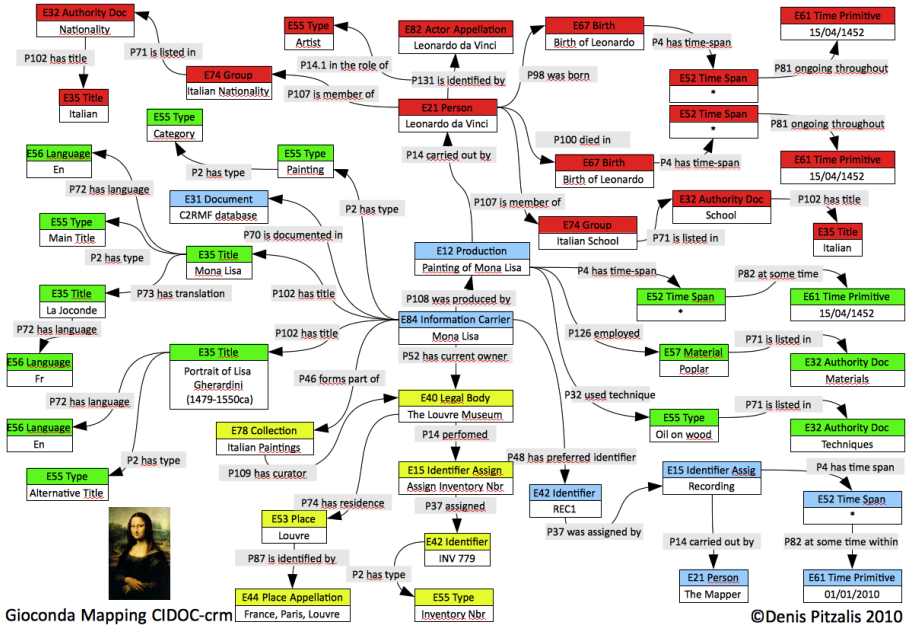

Figure 1: *CIDOC-CRM mapping of the "Mona Lisa" record*

the Data Acquisition Event, a super-event comprising of sub-events that describe the details of the process. The Data Acquisition Event includes generic, set-up information about the acquisition process that is valid for all sub-events unless it is overwritten. The **Data Acquisition Event** can exist on its own without sub-events and is identified either by a UUID or by a URI of the form: *http://"responsible organisation's URI":digitisation:"set of objects ID":date*

The **Data Acquisition Event** is further classified by a specific type, according to the acquisition method that it is based on: "Photogrammetric", "RTI Acquisition", "In-hand scanning", and so on, from a pre-defined controlled vocabulary. It consists of a set of sub-events, one per object, named Object Acquisition Event. There are four types of digital objects that participate in an acquisition process: the primary data objects; digital objects that contain calibration information (colour charts, grey scale charts, photo of a ruler etc.); digital objects that document the acquisition planning and setup; and finally digital objects that document information about the physical object. The above objects are associated with three generic events, the Capturing Event that captures a physical object's digital representation, the Calibration Event that captures calibration information; and the Digital Documentation Event that captures the acquisition planning and setup and/or the physical object's documentation.

The **Object Acquisition Event** in some cases has to be developed in steps of one or more sequential procedures. Thus it is the container for either sequential or independent capturing sub-events, using correspondingly the types of Sequence Event and Capturing Event.

The **Capturing Event** represents the capturing unit using one capturing device. The Sequence Event contains one or more sub-events of the Capturing Event type, appropriately ordered.

The **Calibration Event** is used to describe the calibration process during object acquisition.

The **Digital Documentation Event** is used to describe a specific capturing event that is used for documentation of the Data Acquisition Event.

Each of the events used in our model has its own properties (links to other classes) according to the class it belongs to and also complies with the class hierarchy concepts which means that it inherits properties from its super-classes. Thus the common properties that could be inherited between super and sub events can be grouped with four main questions about:

who : the persons or organisations playing role in the event;
where : the place the event was done;
when : the time the event was done;
what : the things involved in the event.

For further details on the above events, see [RI09].

Digitisation Process
D2.Digitization_Process "*3D Scanning of Mona Lisa*" → L11F.had_output → D9.Data_Object "*Mona Lisa Model*"
D2.Digitization_Process "*3D Scanning of Mona Lisa*" → L1.digitized → E84.Information_Carrier "*Mona Lisa*"
Struct film
D13.Digital_Information_Carrier "*3D of Mona Lisa*" → P70B.is_documented_in → E31.Document "*our database*"
film_recordId *DIG2*
D13.Digital_Information_Carrier "*3D of Mona Lisa*" → P48F.has_preferred_identifier → E42.Identifier "*DIG2*"
film_oeuvreId *REC1*
D13.Digital_Information_Carrier "*3D of Mona Lisa*" → L19F.stores → D9.Data_Object "*Mona Lisa Model*"
E84.Information_Carrier "*Mona Lisa*" → P48F.has_preferred_identifier → E42.Identifier "*REC1*"
film_technique *laser scanning*
D2.Digitization_Process "*3D Scanning of Mona Lisa*"→ P2F.has_type → E55.Type "*laser scanning*"
film_mime *ply*
D9.Data_Object "*Mona Lisa Model*" P2F.has_type → E55.Type "*mimetype:ply*"
film_device *camera CNRC*
D2.Digitization_Process "*3D Scanning of Mona Lisa*"→ L12F.happened_on_device → D8.Digital_Device "*CNRC camera*" → P2F.has_type → E55.Type "*laser scanner*"
film_date 29/10/2004
D2.Digitization_Process "*3D Scanning of Mona Lisa*"→ L31.has_starting_datetime → E61.Time_Primitive "29/10/2004"
film_author CNRC Team
D2.Digitization_Process "*3D Scanning of Mona Lisa*" → L30.has_operator → E21.Person "*CNRC Team*"
film_rigtht *Centre de Recherche et de Restauration des MusÃl'es de France*
D9.Data_Object "*Mona Lisa Model*" → P105F.right_held_by → E39.Actor "*Centre de Recherche et de Restauration des Musées de France*"
film_view *whole*
D9.Data_Object "*Mona Lisa Model*" → P2F.has_type → E55.Type "*whole*"
film_size *700M vertex*
D9.Data_Object "*Mona Lisa Model*" → P90F.has_value → E60.Number "*700M*"
→ P91F.has_unit → E58.Measurement_Unit "*vertex*"
film_path */PATH/TO/OBJECT*
D13.Digital_Information_Carrier "*3D of Mona Lisa*" → P48F.has_preferred_identifier → E42.Identifier "*/PATH/TO/OBJECT.ply*" → P2F.has_type → E55.Type "*path*"

As is, this mapping just describes the final result of our acquisition process but does not take in account any information about the creation of the digital surrogate itself. This means that the experiment is not repeatable and we don't know how we obtained the final model. Following the model suggested in [TTD*10] we can extend our system to cover other information on the digitisation event itself including the provenance information (Figure 2)

Although in this schema, for the sake of simplicity, we are not taking in account sub-Events *P9.consist_of*: "Calibration Event", "Documentation Event" and "Object Acquisition Event", it is clear that CIDOC, and in special CRM$_{dig}$, provide us with a powerful and flexible infrastructure to document information about data provenance in a very precise way.

4. Mapping LIDO to CIDOC-CRM: an update

In general, LIDO elements contain descriptive information in the familiar scheme path → label-content. To map such an approach on CIDOC-CRM, an equivalent triple must be identified. Based on the mapping proposed by the FORTH CIDOC team [KD10] we will present now the mapping between LIDO v0.8 and CIDOC-CRM v5.0.2 using as example the record of the "Mona Lisa" painting described before. As there is no space for the complete mapping and it is not the scope of this article to propose one, a complete mapping will be made available for download at the CIDOC-CRM website; instead we will show some relevant examples.

According to the Lido Data Structure we show now part of the mapping based on our dataset. As mentioned before both structures, CIDOC-CRM and LIDO, are data transfer mechanism and not metadata format. As a metadata format LIDO is just not aimed at covering all collection management needs, but at delivering metadata to online services. It is important to understand the difference between the two definition above: a data transfer mechanism offers a mediation between alternative representations; a metadata format offers a set of rules and recommendation about how to describe the content for a kind of object [NIS04].

In this specific case the mapping presented here is not intended to be reversible and does not have a relation of equivalence. Instead such a mapping describe how to transform data exposed in one structure into an equivalent description with the same meaning in both LIDO and CIDOC. We will end up with a "good" inspiration of a mapping that can be used as starting point for your specific dataset.

Basic information about the object are in the **Object Identification Wrapper**. The title, or the name of the object, is a mandatory field that correspond to *oeuvre_title* in our example.

LIDO [Obj Id]> titleWrap> titleSet> appellationValue:preferred "*Mona Lisa*"
in CIDOC-CRM corresponds to:
E84.Information_Carrier "*Mona Lisa*" → P102F.has_title → E35.Title "*Mona Lisa*".

If there is more than one title in LIDO v0.8 we can repeat

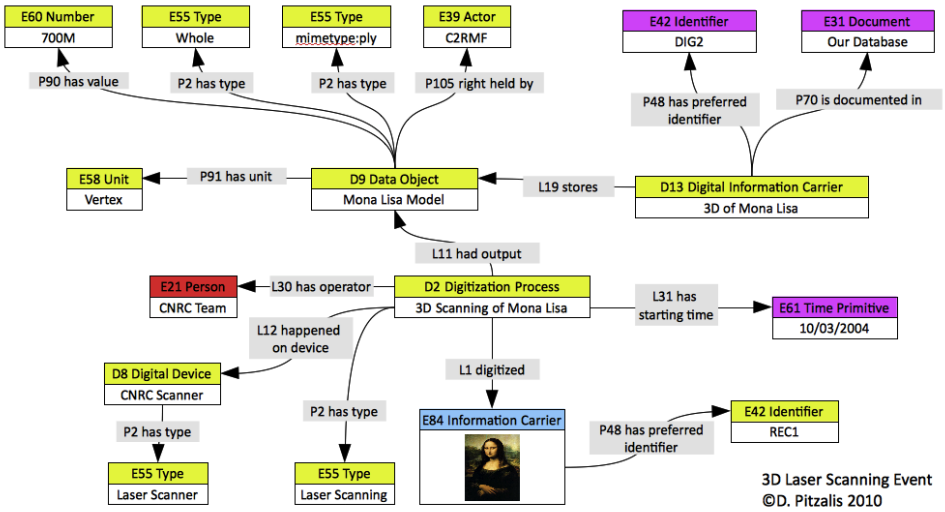

Figure 2: *CRM_{dig} representation of a laser scanning acquisition*

the Title Set element has many time as we need and "preferred" or "alternate" can be specified in the pref attribute of the appellationValue element. We can use the sourceAppellation element to identify the alternative title's source.

oeuvre_title_alternative "*Portrait of Lisa Gherardini*"
LIDO [Obj Id]>titleWrap>titleSet>appellationValue:alternate "*Portrait of Lisa Gherardini*"
>titleWrap>titleSet>sourceAppellation "*The Louvre museum*"
CIDOC E84.Information_Carrier "*Mona Lisa*" → P102.has_title → E35.Title "*Portrait of Lisa Gherardini*"
E35.Title → P2F.has_type → E55.Type "*Alternative Title*"

Information about the record are stored into the **Record Wrapper**.

oeuvre_recordId "*REC1*"
LIDO >RecordID "*REC1*"
CIDOC E84.Information_Carrier "*Mona Lisa*" → P48F.has_preferred_identifier → E42.Identifier "*REC1*"
oeuvre_dataEntry 01/01/2010
LIDO >RecordInfoSet>recordMetadataDate "01/01/2010"
CIDOC E42.Identifier "*REC1*" → . . . → P82F.at_some_time_within → E61.Time_Primitive "01/01/2010"

5. Mapping Lido 0.8 Resource to CRM_{dig}

LIDO resource wrapper need to be handled with special care. In the last [KD10] mapping it has no been considered because out of scope within the CRM structure. Nowadays with the introduction of CRM_{dig} we are able to propose a mapping for the two structures.

LIDO>ResourceWrap>linkResource
CIDOC D1.Digital_Object → P48.has_preferred_identifier → E42.Identifier → P2.has_type → E55.Type "Web resource"
LIDO>ResourceWrap>resourceID
CIDOC D1.Digital_Object → P48.has_preferred_identifier → E42.Identifier
LIDO>ResourceWrap>resourceRelType
CIDOC D1.Digital_Object → P2.has_type → E55.Type → P2.has_type → E55.Type "Resource Relationship"
LIDO>ResourceWrap>resourceType
CIDOC D1.Digital_Object → P2.has_type → E55.Type → P2.has_type → E55.Type "Resource"
LIDO>ResourceWrap>rightsResource
CIDOC D1.Digital_Object → P104.is_subject_to → E30.Right → P75B.is_possessed_by → E39.Actor
LIDO>ResourceWrap>resourceViewDescription
CIDOC D1.Digital_Object → P3.has_note → E62.String
LIDO>ResourceWrap>resourceViewType
CIDOC D1.Digital_Object → P3.has_note → E62.String
LIDO>ResourceWrap>resourceViewSubjectTerm
CIDOC D1.Digital_Object → P2.has_type → E55.Type → P2.has_type → E55.Type
LIDO>ResourceWrap>resourceViewDate
CIDOC D1.Digital_Object → L11B.was_output_of → D7.Digital_Machine_Event → L31F.has_starting_date-time → E61 Time Primitive
LIDO>ResourceWrap>resourceViewDate
CIDOC D1.Digital_Object → L11B.was_output_of → D7.Digital_Machine_Event → L32.has_ending_date-time

LIDO>ResourceWrap>resourceSource
CIDOC D1.Digital_Object → P70B.is_documented_in → E31.Document → P67.refers_to → E39.Actor

6. LIDO v0.9: what has changed

As mentioned at the beginning LIDO is a young schema under constant development. One of the biggest changes we can appreciate in the new version of LIDO available in the ATHENA website, the LIDO v0.9, involve the presence of a "thumbnail" and a "master" digital replica, possibly accompanied by other "resources" into the "resource wrapper".

It is clear that these additions are necessitated by compliance of the ATHENA repository with the Europeana requirements and they are not, at this moment, part of LIDO. If for a standard photo (the simplest two dimensional representation of reality) the latter appear as an inelegant "quick and dirty" way of handling the outcomes of a search to a digital library by an user, they are meaningless for any other digital media object, be it multispectral 2-dimensional imagery, or a 3D model.

CIDOC-CRM already offer the instruments to document such complex cases i.e. [KVS*09] and [RPSL08] and CRM$_{dig}$ is a good reply to provenance requirement. It is therefore hoped that in future releases of Europeana metadata schemas a different solution than the one suggested by LIDO is adopted for such media, . If not, a number of difficult decisions will need to be taken by 3D modellers: which "thumbnail" resource (in 2D) is the best representation of the 3D model? and what is the "master resource", raw data as acquired or the final, clean model? Furthermore, usually a 3D scan generates a number of files (for example, rotating the object to scan it completely), which one is the "master"? And even in 2D, and in the simple example of a coin, which face produces the "master"?

Many other similar questions, all without a reasonable answer, raised by the above-mentioned requirements show that this is not the correct solution to the "digital" problem.

7. Conclusions and future work

This paper shows that LIDO can manage provenance information in a way that complies with CRM$_{dig}$, and perhaps suggests some small changes to better satisfy the necessity of documenting provenance as provided by the latter. Such possibility is a substantial step forward, in the direction of guaranteeing the reliability of digital objects as substitutes of physical ones for several tasks, for research and documentation as well as for communication. As more and more cultural institutions, we hope, will adopt LIDO as their own metadata schema, or will map their schemas to LIDO, this implies a wider and more confident use of digital objects in Cultural Heritage applications and by heritage professionals.

The paper suggests that a metadata harvesting schema, developed by and for the use of the museum community, should further evolve in order to be suitable for information related to multispectral imagery, 3D models and other advanced imaging technologies.

Thus the paper aims at progressing towards a mutual understanding of professional communities involved in the cultural heritage field. Otherwise, they may risk to work separately, developing their own documentation standards that may not be fully interoperable although based on the same reference models.

8. Acknowledgement

The authors gratefully acknowledges support through the EC project 3D-COFORM. 3D-COFORM is a co-funded Large Scale Integrating Project and has received funding from the European Community's Seventh Framework Programme (FP7/2007-2013) under grant agreement n°231809.

We also want to thank Regine Stein, Co-Chair of the Data Harvesting and Interchange Working Group of CIDOC and Gordon McKenna, IDM at Collection Trust for their support and their useful comments.

References

[3D-09] 3D-COFORM: Tools and expertise for 3D collection formation. http://www.3d-coform.eu, may 2009. 2

[ATH09] Athena access to cultural heritage networks across europe. http://www.athenaeurope.org, may 2009. 1

[BGM*07] BORGEAT L., GODIN G., MASSICOTTE P., POIRIER G., BLAIS F., BERALDIN J.-A.: Visualizing and analyzing the mona lisa. *Computer Graphics and Applications, IEEE 27*, 6 (nov.-dec. 2007), 60 –68. 2

[BSM*05] BOEUF P. L., SINCLAIR P., MARTINEZ K., LEWIS P., AITKEN G., LAHANIER C.: Using an ontology for interoperability and browsing of museum, library and archive information. In *International Council of Museums 14th Triennial Meeting* (2005), ICOM COmmittee for Conservation. 3

[CAS09] Caspar digital preservation user community. http://www.casparpreserves.eu/, may 2009. 2

[CDG*09] CROFTS N., DOERR M., GILL T., STEAD S., STIFF M.: *Definition of the CIDOC Conceptual Reference Model*. Tech. rep., ICS-FORTH, Mar 2009. 2

[CDW09] Categories for the Description of Works of Art - Research at the Getty. http://www.getty.edu/research/conducting_research/standards/cdwa/, may 2009. 2

[Doe00] DOERR M.: *Mapping of the Dublin Core Metadata Element Set to the CIDOC CRM*. Tech. Rep. 274, ICS-FORTH, July 2000. 3

[KD10] KUTRAKI M., DOERR M.: *Mapping LIDO v0.7 to CIDOC-CRM v5.0.1*. Tech. rep., ICS-FORTH, Mar 2010. 2, 6, 7

[KVS*09] KARAGIANNIS G., VAVLIAKIS K., SOTIROPOULOU S., DAMTSIOS A., ALEXIADIS D., SALPISTIS C.: Using signal processing and semantic web technologies to analyze byzantine iconography. *IEEE Intelligent Systems 24*, 3 (2009), 73–81. 8

[LID09] LIDO v0.8 - Lightweight Information Describing Objects. http://www.athenaeurope.org/getFile.php?id=535, may 2009. 1

[LID10] CIDOC Data Harvesting and Interchange Working Group. http://www.lido-schema.org, july 2010. 1

[LON09] London charter for the use of 3d visualization in the research and communication of cultural heritage. http://www.londoncharter.org/, Feb 2009. 4

[MUS09] Harvesting Format for Providing Core Data from Museum Holdings. http://www.museumdat.org/, may 2009. 2

[NIS04] NISO: *Understanding metadata*. National Information Standards Organization, 2004. ISBN 1-880124-62-9. 6

[PLP*06] PITZALIS D., LAHANIER C., PILLAY R., AITKEN G., RUSSELL A., SMITH D. A., SINCLAIR P. A. S., ADDIS M. J., LOWE R., HAFEEZ S., LEWIS P. H., MARTINEZ K., M. C. SCHRAEFEL: Semantically exposing existing knowledge repositories: a case study in cultural heritage. In *The first international conference on Semantics And digital Media Technology (SAMT 2006)* (2006). 3

[RI09] 3D-COFORM Integrated Repository Architecture and Design Specifications. http://www.3d-coform.eu/index.php/downloads/cat_view/34-public-deliverables/37-repository-infrastructure, may 2009. 5

[RPSL08] RIBES A., PILLAY R., SCHMITT F., LAHANIER C.: Studying that smile. *IEEE SIGNAL PROCESSING MAGAZINE 25*, 4 (JUL 2008), 14–26. 8

[SPE09] Collections Trust - SPECTRUM. http://www.collectionstrust.org.uk/spectrum, may 2009. 2

[TTD*10] THEODORIDOU M., TZITZIKAS Y., DOERR M., MARKETAKIS Y., MELESSANAKIS V.: Modeling and Querying Provenance by Extending CIDOC CRM. *Distributed and Parallel Databases* (2010). 2, 4, 6