Using LIDO to handle 3D cultural heritage documentation data provenance

Denis Pitzalis, Franco Niccolucci, Matthieu Cord

Using LIDO to handle 3D Cultural Heritage Documentation Data Provenance

Denis Pitzalis
The Cyprus Institute
STARC
Nicosia, Cyprus
d.pitzalis@cyi.ac.cy

Franco Niccolucci
The Cyprus Insitute
STARC
Nicosia, Cyprus
f.niccolucci@cyi.ac.cy

Matthieu Cord
UPMC-Sorbonne Universities
LIP6
Paris, France
matthieu.cord@lip6.fr

Abstract

It is important for Digital Libraries (DL) to be flexible in exposing their content. Typically a DL provides a search/browse interface which allows resources to be found and a service to make the data available for harvesting from/to other DLs. This kind of communication is possible because the structures of different DLs are expressed following formal specifications. In particular 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. CIDOC-CRM is an ontology designed to mediate contents in the area of tangible cultural heritage and was published as ISO $21127 : 2006$ *standard. Lately an extension of CIDOC-CRM, known as CRM*$_{dig}$*, enables to document information about data provenance and digital surrogates in a very precise way. Another metadata schema suitable for handling museum-related data is LIDO. In this paper we propose a case study where we show how CIDOC-CRM*$_{dig}$ *and LIDO handle the digital information of an object and specially the data provenance.*

1. Introduction

The choice of the metadata schema to be adopted in a digital library depends on a number of different factors: the nature of data, their intended use, and the interests and research methodology of the relevant community. Simplicity 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 or other research goals. This has been the case of metadata schemas based on Dublin Core, suitable for managing websites 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. 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) [12], 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 [3] as a 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 Data Model schema (EDM) [9].

According to its proponents, LIDO is a metadata schema suitable for harvesting museum data developed by an international consortium [13] and adopted by the EU ATHENA project. LIDO is based on previous museum schemas such as CDWALite [6], museumsdat [14] and SPECTRUM [21], and strongly relies on the CIDOC-CRM [7] 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 a harvesting schema for the delivery of metadata. The current version of LIDO is 1.0.

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.

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 [11] and an update has been published in [19]. These documents sketch the correspondence between the two schemas. Recently, in order to capture provenance information of digital objects, an extension of CIDOC-CRM, named CRM$_{dig}$, has been developed [22] in the framework of the CASPAR [5] first and 3D-COFORM [1] later EU projects [10] . 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 is important to enable this feature for LIDO as well.

Such an assessment, together with the complete mapping, provided separately [16] is of paramount importance to foster the adoption of the LIDO metadata schema in the cultural heritage community. So far, as already mentioned, metadata schemas used in digital libraries were considered by heritage professionals as not being rich enough to convey the information necessary for current heritage research and practice. Others were perceived as too complex for practical use. LIDO is at a time lightweight and rich enough. Demonstrating the capability of managing provenance information and compliance to CIDOC-CRM through a mapping is therefore a significant step towards standardization. Moreover, a theoretical compliance and a formal mapping are not convincing enough: that is why we chose the approach of demonstrating it in a significant example.

The goal of the present paper is to assess the capability of LIDO to deal with 3D cultural objects, possibly proposing improvements in this direction. What is presented here is not just an example, it is exemplary. Hopefully, heritage professionals may follow this approach and apply it to different collection of objects in a much easier way than just referring to a formal description of the mapping.

2 A case-study scenario

The examples used in the paper refer to the following case-study scenario: the "Cylinder Seal of Ibni-Sharrum" (Figure 1). The original artwork is exposed at the Louvre Museum in Paris, France, more precisely in the department of Near Eastern Antiquities. The story of this art piece is interesting and rich enough to give it the title of "masterpiece of glyptic art". Engraved by Ibni-Sharrum probably during the Agade period, under the reign of Sharkali-Sharri (c. 2217-2193 BC) it depicts two buffaloes that have just slaked their thirst in the stream of water spurting from two vases held by two naked kneeling heroes. In particular, in 2008 a 3D model has been acquired using a multi technique 3D acquisition[17]. We can use a hypothetic database, based on the information stored at C2RMF (Centre de Recherche et de Restauration des Musées de France) where

Figure 1. Cylinder Seal of Ibni-Sharrum, Musée du Louvre, AO 22303

every artwork corresponds, in a relation $1 : 1$, to a record "oeuvre" structured like:

string *oeuvre_recordId* = **REC1**;
string *oeuvre_title* = **(fr) Sceau Cylindre de Ibni-Sharrum, (en) Cylinder Seal of Ibni-Sharrum**;
struct oeuvre_artist
string *oeuvre_artist_name* = **(fr) Ibni-Sharrum, (en) Ibni-Sharrum**;
string *oeuvre_artist_nationality* = **(fr) Mesopotamia, (en) Mesopotamia**;
string *oeuvre_artist_school* = **(fr) Regne de Sharkali-Sharri , (en) Reign of Sharkali-Sharri**;
struct oeuvre_owner
string *string oeuvre_owner_place* = **(fr) France, Paris, Musée du Louvre, (en) France, Paris, The Louvre Museum**;
string *oeuvre_owner_inventoryId* = **AO 22303**;
string *oeuvre_owner_collection* = **(fr) Antiquités Orientales, (en) Near Eastern Antiquities**;
string *oeuvre_category* = **(fr) sculpture, (en) sculpture**;
time *oeuvre_date_creation_begin* = **2217 BC**;
time *oeuvre_date_creation_end* = **2193 BC**;
string *oeuvre_material* = **(fr) Serpentine, (en) Serpentine**;
string *oeuvre_technique* = **(fr) gravure, (en) engrave**;
struct ouvre_size
string *oeuvre_size_diameter* = **26 mm**;
string *oeuvre_size_height* = **39 mm**;
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. In the special case of a 3D model we can have multiple digital sources.

struct film
string *film_recordId* = **DIG2**;
string *film_oeuvreId* = **REC1**;
string *film_technique* = **photogrammetry**;
string *film_mime* = **dae**;
string *film_device* = **Nikon D2X**;
time *film_date* = **29/07/2008**;
string *film_author* = **John Doe**;
string *film_rigtht* = **C2RMF**;
string *film_view* = **whole**;
string *film_size* = **500K vertex**;
string *film_path* = **/path/to/model.dae**;

According to [8], [4] and [18] we can represent the "oeuvre" record in CIDOC-CRM in a similar way than in [19]. The complete mapping is available at the CIDOC-CRM website.

The schema can be synthesised in Figure 2. 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. Provenance in science means experiment repeatability and verifiability.

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 surrogate) is the most delicate step, because the methodology needs to swap from the tangible to the intangible (digital). 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 CIDOC-CRM, called CRM$_{dig}$, has been defined to document provenance metadata [22].

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 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*

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 superclasses. Thus the common properties that could be inherited between super and sub events can be grouped with four main questions about: who, where, when and what. For further details on the above events, see [20].

Digitisation Process
D2.Digitization_Process "*3D Scanning of the Cylinder*" → L11F.had_output → D9.Data_Object "*Cylinder Seal Model*"
D2.Digitization_Process "*3D Scanning of the Cylinder*" → L1.digitized → E84.Information_Carrier "*Cylinder Seal*"
Struct film
D13.Digital_Information_Carrier "*3D of Cylinder Seal*" → P70B.is_documented_in → E31.Document "*our database*"
film_recordId *DIG2*
D13.Digital_Information_Carrier "*3D of Cylinder Seal*" → P48F.has_preferred_identifier → E42.Identifier "*DIG2*"
film_oeuvreId *REC1*
D13.Digital_Information_Carrier "*3D of Cylinder Seal*" → L19F.stores → D9.Data_Object "*Cylinder Seal Model*"
E84.Information_Carrier "*Cylinder Seal*" → P48F.has_preferred_identifier → E42.Identifier "*REC1*"
film_technique *photogrammetry*
D2.Digitization_Process "*3D Scanning of the Cylinder*"→ P2F.has_type → E55.Type "*photogrammetry*"
film_mime *dae*
D9.Data_Object "*Cylinder Seal Model*" P2F.has_type → E55.Type "*mimetype:dae*"
film_device *Nikon D2X*
D2.Digitization_Process "*3D Scanning of the Cylinder*"→ L12F.happened_on_device → D8.Digital_Device "*Nikon D2X*" → P2F.has_type → E55.Type "*photogrammetry*"
film_date 29/07/2008
D2.Digitization_Process "*3D Scanning of the Cylinder*"→ L31.has_starting_datetime → E61.Time_Primitive "29/07/2008"
film_author John Doe
D2.Digitization_Process "*3D Scanning of the Cylinder*" → L30.has_operator → E21.Person "*John Doe*"
film_right *Centre de Recherche et de Restauration des Musées de France*

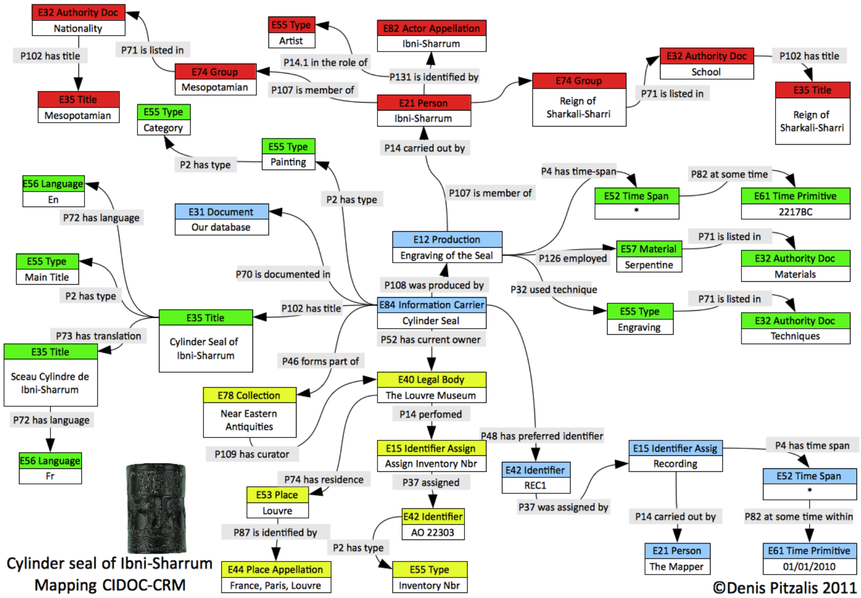

Figure 2. CIDOC-CRM mapping of the "Cylinder Seal of Ibni-Sharrum" record

D9.Data_Object "*Cylinder Seal Model*" → P105F.right_held_by → E39.Actor "*Centre de Recherche et de Restauration des Musées de France*"
film_view *whole*
D9.Data_Object "*Cylinder Seal Model*" → P2F.has_type → E55.Type "*whole*"
film_size *500K vertex*
D9.Data_Object "*Cylinder Seal Model*" → P90F.has_value → E60.Number "*500K*"
→ P91F.has_unit → E58.Measurement_Unit "*vertex*"
film_path */PATH/TO/OBJECT*
D13.Digital_Information_Carrier "*3D of Cylinder Seal*" → P48F.has_preferred_identifier → E42.Identifier "*/PATH/TO/OBJECT.dae*" → P2F.has_type → E55.Type "*path*"

As is, this mapping just describes the final result of our acquisition process but does not take into 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 [22] we can extend our system to cover other information on the digitisation event itself including the provenance information (Figure 3)

Although in this schema, for the sake of simplicity, we are not taking into account sub-Events *P9.consist_of*: "Calibration Event", "Documentation Event" and "Object Acquisition Event", it is clear that CIDOC-CRM, with the extension CRM_{dig}, provides a powerful and flexible infrastructure to document information about data provenance in a very precise way especially in cases like the one used as example, where the process acquisition requires complex steps.

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 [11] we will present now the mapping between LIDO v1.0 and CIDOC-CRM v5.0.2 using as example the record of the Ibni-Sharrum's cylinder seal described before. As it is not in the scope of this article to propose a complete mapping we will demonstrate the mechanism via 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, both event centric. 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 definitions 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 [15].

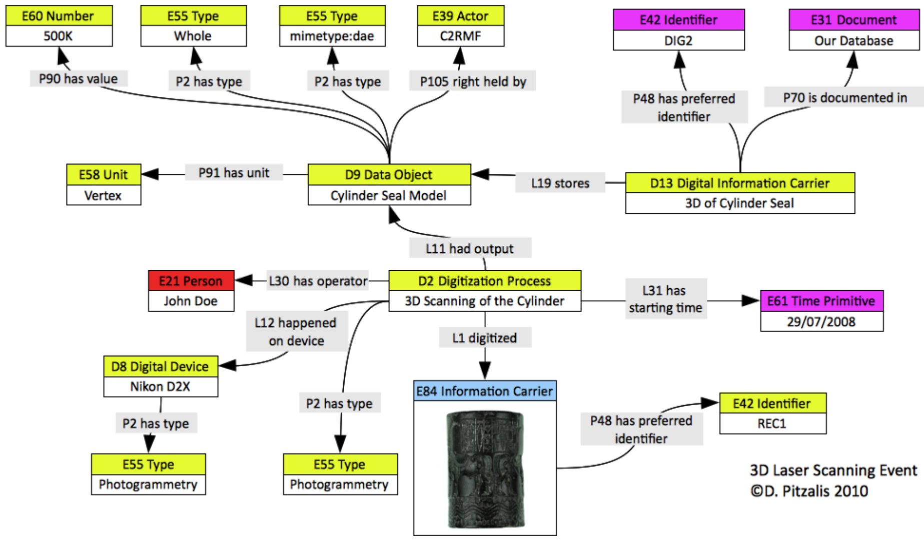

Figure 3. CRM$_{dig}$ representation of a photogrammetry acquisition

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 describes 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.

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

LIDO [Obj Id]> titleWrap> titleSet> appellationValue:preferred "*Cylinder Seal of Ibni-Sharrum*"

CIDOC-CRM E84.Information_Carrier "*Cylinder Seal*" → P102F.has_title → E35.Title "*Cylinder Seal of Ibni-Sharrum*".

If there is more than one title in LIDO v1.0 we can repeat the Title Set element as 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. Information about the record itself are stored into the **Record Wrapper**.

oeuvre_recordId "*REC1*"
LIDO >RecordID "*REC1*"
CIDOC E84.Information_Carrier "*Cylinder Seal*" → 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 1.0 Resource to CRM$_{dig}$

Almost in the same way we can map the LIDO resource wrapper although it need to be handled with special care. In the last [11] mapping it has not been considered for being out of scope within the CIDOC-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 Conclusions and future work

This paper shows that LIDO can manage provenance information in a way that complies with CRM_{dig}. Such possibility added to the use of event centric ontologies is a substantial step forward in the direction of guaranteeing the reliability of digital surrogates in use 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. Already some living repositories are using LIDO and are available online for consultation.

Acknowledgement

The authors gratefully acknowledge 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.

References

[1] 3D-COFORM: Tools and expertise for 3D collection formation, may 2009. http://www.3d-coform.eu.

[2] A. Artusi, M. Joly, G. Lucet, D. Pitzalis, and A. Ribes, editors. *VAST10: The 11th International Symposium on Virtual Reality, Archaeology and Intelligent Cultural Heritage*, Paris, France, 2010. Eurographics Association.

[3] Athena access to cultural heritage networks across europe, jan 2011. http://www.athenaeurope.org.

[4] P. L. Boeuf, P. Sinclair, K. Martinez, P. Lewis, G. Aitken, and C. Lahanier. Using an ontology for interoperability and browsing of museum, library and archive information. In *International Council of Museums 14th Triennial Meeting*. ICOM COmmittee for Conservation, 2005.

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

[6] Categories for the Description of Works of Art - Research at the Getty, may 2009. http://www.getty.edu/.

[7] N. Crofts, M. Doerr, T. Gill, S. Stead, and M. Stiff. Definition of the CIDOC Conceptual Reference Model. Technical report, ICS-FORTH, Mar 2009.

[8] M. Doerr. Mapping of the Dublin Core Metadata Element Set to the CIDOC CRM. Technical Report 274, ICS-FORTH, July 2000.

[9] M. Doerr, S. Gradmann, S. Hennicke, A. Isaac, C. Meghini, and H. Van de Sompel. The europeana data model (edm). In *World Library and Information Congress: 76th IFLA General Conference and Assembly*, 2010.

[10] M. Doerr, K. Tzompanaki, M. Theodoridou, C. Georgis, A. Axaridou, and S. Havemann. A Repository for 3D Model Production and Interpretation in Culture and Beyond. In Artusi et al. [2], pages 97–104.

[11] M. Kutraki and M. Doerr. Mapping LIDO v0.7 to CIDOC-CRM v5.0.1. Technical report, ICS-FORTH, Mar 2010.

[12] LIDO - Lightweight Information Describing Objects, jan 2011. http://www.lido-schema.org.

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

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

[15] NISO. *Understanding metadata*. National Information Standards Organization, 2004. http://www.niso.org/.

[16] D. Pitzalis. Mapping LIDO v1.0 to CIDOC-CRM v5.0.2. Technical report, The Cyprus Institute - STARC, Mar 2010.

[17] D. Pitzalis, P. Cignoni, M. Menu, and G. Aitken. 3d enhanced model from multiple data sources for the analysis of the cylinder seal of ibni-sharrum. In M. Ashley, S. Hermon, A. Proenca, and K. Rodriguez-Echavarria, editors, *The 9th International Symposium on VAST International Symposium on Virtual Reality, Archaeology and Cultural Heritage*, pages 79 – 84, Braga (Portugal), 2008. Eurographics.

[18] D. Pitzalis, C. Lahanier, R. Pillay, G. Aitken, A. Russell, D. A. Smith, P. A. S. Sinclair, M. J. Addis, R. Lowe, S. Hafeez, P. H. Lewis, K. Martinez, and 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.

[19] D. Pitzalis, F. Niccolucci, M. Theodoriou, and M. Doerr. LIDO and CRMdig from a 3D Cultural Heritage Documentation Perspective. In Artusi et al. [2], pages 87–95.

[20] 3D-COFORM Integrated Repository Architecture and Design Specifications, may 2009. http://www.3d-coform.eu/.

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

[22] M. Theodoridou, Y. Tzitzikas, M. Doerr, Y. Marketakis, and V. Melessanakis. Modeling and Querying Provenance by Extending CIDOC CRM. *Distributed and Parallel Databases*, 2010.