Conceptual Modelling of the European Silk Heritage with the SILKNOW Data Model and Extension

Marie Puren[1][2] and Pierre Vernus[3][4]

Abstract

Silk is a particularly important material in European history. Its trade has facilitated the growth of new economic centres, stimulated the development of innovative weaving techniques, and allowed the creation of exceptional objects. While many institutions specialise in its conservation, this fragile Cultural Heritage is under threat. In this article, we explain the methodology used by the Horizon 2020 SILKNOW project to better promote these Cultural Heritage collections. We present the CIDOC CRM-based data model that we designed for the creation of a Knowledge Graph. We also present the methodology used to integrate, in the data model, the annotations generated by the analysis of texts and images. Finally, we present the SIlKNOW extension, designed to model the complex semantics describing the production process of silk fabrics.

Introduction

For the past ten years, the history of fashion has attracted a growing public. Museums are increasingly highlighting it in their exhibitions (Petrov 2019), in the hope of attracting new visitors and new patrons (Stewart and Marcketti 2012). However, despite the interest shown in these collections, the Textile Heritage remains under threat. This is particularly true of the Silk Heritage, whose fragility poses major conservation problems and whose protection requires substantial investment. The fragility of silk makes its conservation all the more difficult; in addition, there is the need to safeguard the weaving techniques used to produce these fabrics, in order to facilitate their restoration, but also their possible reproduction. Few materials are of such historical, cultural and artistic importance. Both tangible and intangible, this Cultural Heritage brings together objects of all kinds (costumes, furniture, accessories, etc.), craft skills and age-old traditions. Ancient silk fabrics are thus an integral part of the collections of major museums; but a significant part of the Silk Heritage is preserved by institutions that are small in size and generally lack the means to promote their collections (Portalés et al. 2018; Gaitán et al. 2019; Pagán et al. 2020).

Outside of the major events devoted to them, we also note the under-use of textile collections by visitors (Stewart and Marcketti 2012). The growth of the Internet and the

---

[1] EPITA, LRE, F-94270 Kremlin-Bicêtre, France

[2] Ecole nationale des Chartes, Centre Jean-Mabillon, F-75002 Paris, France

[3] LARHRA, F-69363 Lyon, France

[4] Université Lumière Lyon 2, F-69007 Lyon, France

new uses linked to the social Web are encouraging museums to take advantage of this new medium, both to attract new visitors and to enhance their collections. The digitization of Cultural Heritage - and its corollary, the exploitation of the digital data thus generated - has become one of the crucial issues in the promotion of Cultural Heritage (Estermann 2015). Institutions that preserve textile collections have clearly understood the importance of developing their online presence, particularly by providing access to their inventories, in order to broaden their audience and increase their influence. To enhance the value of these collections, which are poorly known and scattered among various institutions, one solution is to create a single access point via an online portal. The idea of aggregating and providing access to data from various sources is based on the assumption that this will make these collections better known and facilitate their re-use. In a context of economic crisis, Cultural Heritage institutions are seeking to make themselves more visible, but also to create new partnerships that will make it easier for them to develop large-scale valorisation projects (Suls 2017).

Portals are emerging, that allow dispersed Cultural Heritage collections to be explored simultaneously, such as the *Paris Musées* portal[5] in France. On the fashion side, the *Europeana Fashion* portal provides access to the collections of over thirty European institutions. However, while *Europeana Fashion* offers users a single point of access to explore these collections, it aggregates data from major European museums only (Suls 2017). This type of solution therefore leaves out small museums and does not allow more specialised collections, such as those devoted to silk, to be highlighted. In order to overcome this shortcoming, the H2020 SILKNOW project (*Silk Heritage in the Knowledge Society: from punched card to Big Data, Deep Learning and visual/tangible simulations*)[6] has developed an exploratory search engine that enables federated searching of Cultural Heritage collections dedicated to silk.

Using public Cultural Heritage data, SILKNOW has created a *Knowledge Graph* that compiles results from different sources and annotates them with new semantic information. The SILKNOW *Knowlledge Graph* represents the data contained in the catalogues of 21 Cultural Heritage institutions[7] , following a subset of the CIDOC Conceptual Reference Model or CIDOC-CRM. Collected online or provided directly by the institution concerned[8], the data is characteristically descriptive of objects incorporating silk in their manufacturing processes. These objects may be made entirely of silk, or they may incorporate other textile materials (natural fibres, metallic threads), or they may be made of composite materials (for example, wood in the case of furniture). The SILKNOW project focused on objects produced or consumed in Europe between the mid-15th to the middle of the 19th century.

In this paper, we present the different steps that allowed us to build the SILKNOW data model. We also describe the methodology we used to create a conceptual model extending the CIDOC CRM, and to integrate the annotations produced with machine learning methods.

---

[5] The portal can be accessed here: https://www.parismuseescollections.paris.fr/fr.
[6] The project website can be found here: https://silknow.eu/.
[7] A full list of these institutions can be found here: https://ada.silknow.org/fr/museums.
[8] Datasets have been made available to the project by *Garin1820*, the *Sicily Cultural Heritage* and the *Musée d'Art et d'Industrie de Saint-Etienne*.

In a first part, we discuss the contributions of Semantic Web technologies for the protection and enrichment of Silk Heritage. In a second part, we explain the approach we adopted to develop the SILKNOW data model. In the third part, we show how we used the Provenance Data Model or Prov DM to integrate the semantic annotations produced on these data thanks to the analysis of texts and images. We also propose to extend the CIDOC CRM with new classes and properties, so as to be able to express the complex semantics of the data describing the production process of silk fabrics.

Promoting Europe's Silk Heritage through the Semantic Web

Heterogeneous metadata

When users are interested in textile Cultural Heritage online, they are generally faced with two situations. The large generalist museums have the technical, financial and human resources to disseminate their data on the Internet, but they do not systematically highlight their textile collections on their websites. Conversely, small museums do not have the same resources as their better-endowed counterparts (Claerr and Westeel 2017) to promote specialised collections online, which can be particularly rich. This is particularly the case for Silk Heritage, as the example of the Rhône-Alpes region clearly illustrates (Foron-Dauphin and Cano 2016; Fournier et al. 2016). These institutions do describe their collections, but the digital data they produce is, most of the time, inaccessible online. Without a digital presence, these collections remain generally unknown, except to a few connoisseurs and specialists.

The records produced by all these institutions generally provide the same types of information: type of object, place and date of production, materials and techniques used, dimensions, and illustration(s). But the way in which this metadata is structured can vary greatly depending on the practices of its producers. These data come from institutions using various cataloguing standards and respecting their own description standards. Figure 1 illustrates this problem by comparing the list of metadata describing two objects (here, two damasks[9]) held respectively by the *Museu Tèxtil Terrassa*[10] and by the *Victoria & Albert Museum*[11] . The same types of information are used to describe these two objects, but they are expressed differently by the two museums. For example, the *Victoria & Albert Museum* indicates in the same field the material and technique used, whereas the *Museu Tèxtil Terrassa* clearly separates this information.

[9] For a definition of “damask”, please consult the SILKNOW thesaurus: https://skosmos.silknow.org/thesaurus/en/page/168.
[10] http://imatex.cdmt.es/
[11] https://www.vam.ac.uk/collections.

Figure 1. Metadata describing the techniques and materials used: examples from the *Museu Tèxtil Terrassa* and the *Victoria & Albert Museum*

In other words, the *Victoria & Albert Museum* and the *Museu Tèxtil Terrassa* do not use the same data model. This is because the catalogues are designed as separate data silos, which are not supposed to communicate with each other. If one wants to collect this metadata on a large scale, it will therefore be difficult to process it automatically due to the heterogeneity of its structure, but also due to the lack of structure of the metadata. Different types of information are often found in the same field (as shown in Figure 1), which makes it difficult to process the data. By nature, Cultural Heritage data are heterogeneous: they are produced by various institutions, they describe disparate objects - which does not make them easily accessible by search engines (Freire et al. 2018). To improve the discoverability of this data on the Internet, Semantic Web technologies offer the possibility of aggregating and providing access to metadata from different sources (Freire et al. 2019). It is in this context that the SILKNOW project has created an exploratory search engine that enables a federated search of 21 collections dedicated to the conservation of European Silk Heritage.

An ontology to build a Knowledge Graph

Since the emergence of the Semantic Web, we have seen new changes in the use of the Internet by museums, particularly the increasing openness of the data produced by these institutions (Marden et al. 2013; Estermann 2015). The Semantic Web implements linked data (Heath and Bizer 2011), with the aim of enabling machines to understand the semantics of information published on the Web, and thus to create new, even unexpected, links between this information. The Semantic Web thus opens up new perspectives for the visibility of Cultural Heritage on the Internet, by making Cultural Heritage data available to all - and not only in a museum context -; by aggregating, publishing and enriching this data on the Web to facilitate its reuse; by linking this data to each other and to other data, so as to constitute an information network; and finally by making proposals for enriching this data.

SILKNOW hypothesised that the use of Semantic Web technologies, and more specifically the creation of a Knowledge Graph based on a data model, could facilitate the integration, exploration and retrieval of Cultural Heritage data describing historical silk objects and fabrics. By containing this data, the SILKNOW Knowledge Graph is intended to create a single point of access to this data online. This approach has proven particularly successful for other types of Cultural Heritage data: for example, bibliographic data describing musical works for the DOREMUS project (Achichi et al. 2018), archival data for the ArchOnto project (Koch, Ribeiro, and Teixeira Lopes 2020), or historical data for the WarSampo project (Koho et al. 2021).

In concrete terms, data is downloaded or collected from partner institutions; and these files are converted to create the Knowledge Graph using the Resource Description Framework (RDF) as the data model. The graph, which contains all the metadata, is then loaded into a triple store. The converted data is then accessible online via an exploratory search engine (Schleider, Troncy, Ehrhart, et al. 2021; Schleider, Troncy, Gaitán, et al. 2021). The creation of the Knowledge Graph requires an RDF conversion, which is based on a manual mapping of the data. This mapping consists in putting in equivalence the data models describing similar objects (Christen 2012). When working with two data sources A and B, data mapping can consist of matching each field in database A with those in database B. In the SILKNOW project, we are using many different sources; we have chosen to use a pivotal data model that makes it easier to perform this mapping between all the collected datasets. We have thus created a subset of the CIDOC Conceptual Reference Model (CIDOC CRM) ontology, which is instantiated at the time of the RDF conversion (Schleider, Troncy, Ehrhart, et al. 2021; Schleider, Troncy, Gaitán, et al. 2021).

The SILKNOW data model

Expressing the semantics of Silk Heritage information with the CIDOC CRM

We chose CIDOC CRM (Le Boeuf et al. 2015) because it is a conceptual model specifically developed for Cultural Heritage by the International Committee for Documentation (CIDOC) of the International Council of Museum (ICOM). An ISO standard since 2006, renewed in 2014[12] , the CIDOC CRM is today a standard intended to ensure the interoperability of Cultural Heritage data. However, this standard is not set in stone; thus, at the time of writing this article, version 7.2.1 has been published[13] . However, we used version 6.2[14] - the most recent version at the time of the start of the project in 2018. CIDOC CRM is also a flexible and extensible model. If necessary, this model can be extended by creating new classes and properties to express new information, without changing the basic structure of the model. These features allow the creation of specialised extensions such as FRBRoo (Functional Requirements for Bibliographic Records), a conceptual model for bibliographic data[15], or CRMtex, a conceptual model for the study of ancient texts[16] . In all, there are ten compatible models on the official CIDOC CRM website[17].

---

[12] https://www.iso.org/standard/57832.html
[13] https://cidoc-crm.org/Version/version-7.2.1
[14] http://www.cidoc-crm.org/Version/version-6.2.3
[15] IFLA Study Group on the Functional Requirements for Bibliographic Records, *Functional Requirements for Bibliographic Records: Final Report*, IFLA, 2009. URL: https://repository.ifla.org/handle/123456789/811.

CIDOC CRM allows the underlying semantics of Cultural Heritage information to be expressed, and thus heterogeneous data to be modelled in a homogeneous way. If we take the descriptions of the two damask fabrics preserved by the *Victoria & Albert Museum* and the *Museu Tèxtil Terrassa* (Figure 1), we can express in a unique way with CIDOC CRM the information about the technique and the materials used during the weaving process. CIDOC CRM allows to model each of the descriptive fields with triplets Class (C) - Property (P) - Class (C), which can be represented as graphs as in Figure 2.

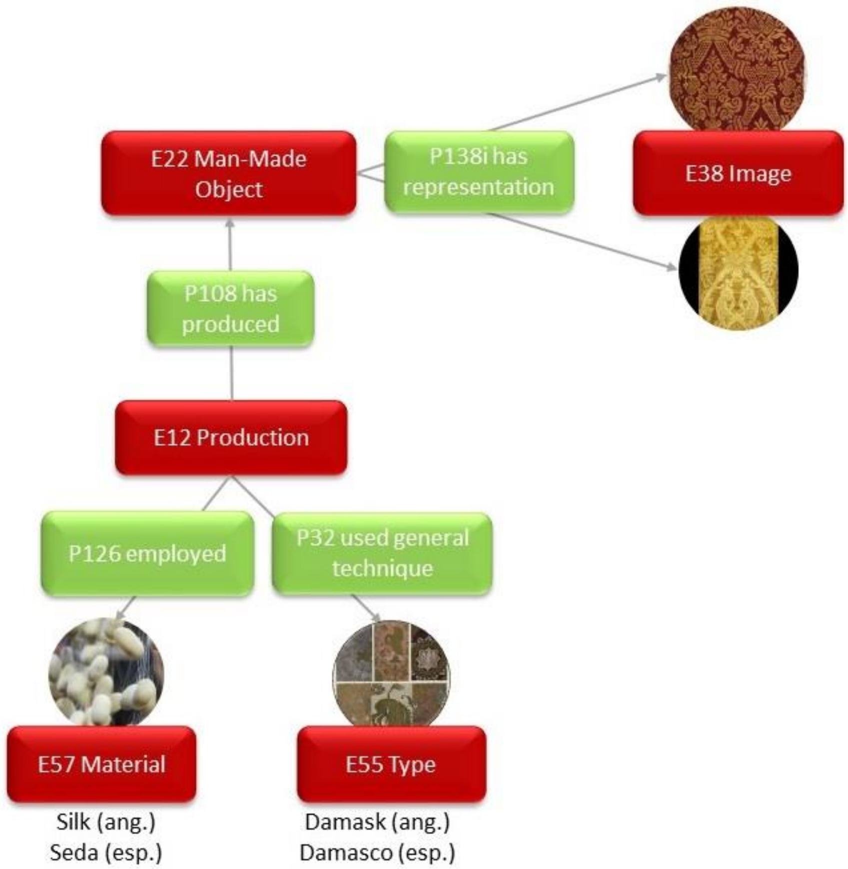

Figure 2. Graph in accordance with the CIDOC CRM conceptual model

In this figure, we can see that we have chosen to use the E22_Man Made Object class to represent the object described by the museum metadata. In version 6.2 of CIDOC CRM, this class is indeed used to model "physical objects purposely created by human activity" (Le Boeuf et al. 2015).

[16] Martin Doerr, Francesca Murano and Achille Felicetti, *Definition of the CRMtex. An Extension of CIDOC CRM to Model Ancient Textual Entities*, Version 1.0, currently maintained by Francesca Murano and Achille Felicetti, June 2020. URL: https://cidoc-crm.org/crmtex/ModelVersion/version-1.0-0.
[17] For the full list: https://cidoc-crm.org/collaborations.

What classes and properties to represent this information?

As a conceptual model, CIDOC CRM is intended to model all kinds of Cultural Heritage data and should be able to cover most cases. CIDOC CRM thus provides a wide range of classes and properties. For example, there are 89 classes and 153 properties offered by version 6.2[18]. In general, therefore, we do not need to use all the classes and properties offered by CIDOC CRM.

In order to select the classes and properties needed to model the metadata collected by the project, it is first necessary to list the descriptive fields used to describe the textile objects. To do this, we analysed and compared a large number of records, based on the documentation used by Cultural Heritage institutions to create this metadata (notably (Grant, Nieuwenhuis, and Petersen 1995; Briatte 2012; « Europeana Data Model » s. d.)). This analysis allowed us to draw up a list of descriptive fields commonly used to describe textile objects, while eliminating descriptive fields that were not relevant to the project - namely, fields providing information on the administrative management of the object.

These descriptive fields were then classified into Information Groups, which form the Data Dictionary of the SILKNOW project. These information groups allowed us to identify the categories used by Cultural Heritage institutions to describe ancient textiles. We defined 22 Information Groups, from the most commonly used (e.g. "Object Title Information Group" or "Object Measure Information Group") to the most rarely encountered (e.g. "Object Missing Part Information Group" or "Exhibition Information Group"). These Information Groups then allowed us to rely on the Functional Overviews provided by the official CIDOC CRM documentation[19], and to more easily select the most appropriate classes and properties to represent these information groups. Functional Overviews classify CIDOC CRM classes and properties into information categories that can be easily aligned with the Information Groups in our data dictionary. For example, the Functional Overviews provide an "Object Title Information" category that corresponds to our "Object Title Information Group".

The SILKNOW data model consists of the classes and properties selected to represent the information groups we have previously defined. It is accessible and documented online (Puren and Vernus 2021b) thanks to OntoME (Beretta 2021), which is an ontology manager developed and maintained by the LARHRA laboratory. The SILKNOW data model is mainly based on the CIDOC CRM Core version 6.2 (71 properties and 21 properties). We also selected two classes and one property from an extension of CIDOC CRM, the Scientific Observation Model (CRMsci) version 1.2.3[20] (Kritsotaki et al. 2017). The CRMsci, which is a conceptual model designed to integrate the metadata produced by scientific observations, seemed to us to be very appropriate for modelling the historical and technical analyses formulated by experts in the field on ancient textiles. For example, the record describing a tapestry ("*lé de tenture*" in French) preserved in the collections of the Château de

[18] A list of these classes and properties can be found on the OntoMe website: https://ontome.net/namespace/1

[19] The functional overviews are available online: http://www.cidoc-crm.org/functional-units.

[20] The latest version 1.5 was released in January 2022.

Versailles[21] includes a History field ("*Historique*") and a Comment field ("*Commentaire*"). Using the CIDOC CRM and CRMsci models, we have expressed this information with the following triplets: *S4_Observation O8_observed E22_Man-Made Object; S4_Observation P2_has type E55_Type (Historique / Commentaire); S4_Observation P3_has note E62_String* (Figure 3). We chose to use the *S4_Observation* class provided by the CRMsci extension because it is defined as "the activity of gaining scientific knowledge about particular states of physical reality gained by empirical evidence, experiments and by measurements" (Kritsotaki et al. 2017). The observations made by the experts on ancient textiles thus seemed to us to fall under this type of scientific activity.

Figure 3. Use of the *S4* class *Observation*

Evaluating the ontology

The quality and robustness of this ontology is then assessed by a mapping process. This process consists of verifying that it is possible to represent all the metadata fields using the classes and properties selected in the ontology. The mapping is produced manually by domain experts, who then provide mapping rules for each of the collected datasets. Concretely, we chose two objects described extensively[22] in the catalogues of the selected institutions. Then, we interpreted each of the descriptive fields in the form of triplets Entity (E) - Relation (R) - Entity (E). To do this, we relied on the collected metadata files, converted into an intermediate JSON format. Figure 4 shows an extract corresponding to the metadata contained in Table 1.

[21] Available here: http://collections.chateauversailles.fr/#e7fe900c-4795-42f5-b5aa-26f2b3f649e3.

[22] That is, with the highest number of descriptive fields filled in the records.

```
"fields": [
    {
        "label": "title",
        "value": "lé de tenture"
    },
    {
        "label": "Désignation",
        "value": "lé de tenture"
    },
    {
        "label": "N° d'inventaire:",
        "value": "VMB 14527"
    },
    {
        "label": "Domaine",
        "value": "Textiles"
    },
    {
        "label": "Date de création",
        "value": "XVIIIe siècle"
```


Figure 4. Extract from a JSON metadata file

As shown in Table 1, we have therefore decomposed the whole metadata schema with Class (C) - Property (P) - Class (C) triples from the SILKNOW data model.

Field Value	**Value**	**Path**
Désignation	Lé de tenture	E22_Man-Made Object P102_has title E35_Title
N° d'inventaire	VMB 14527	E15_Identifier Assignment P140_assigned attribute to E22_Man-Made Object E15_Identifier Assignment P37_assigned E42_Identifier E15_Identifier Assignment P14_carried out by E40_Legal Body (Château de Versailles) E42_Identifier P2_has type E55_Type (N° d'inventaire)
Date de création	XVIIIe siècle	E12_Production P4_has time-span E52_Time-Span

		E52_Time-Span P78_is identified by E49_Time Appellation

Table 1. Table containing mapping rules

We then represented these triples as graphs as in Figure 5.

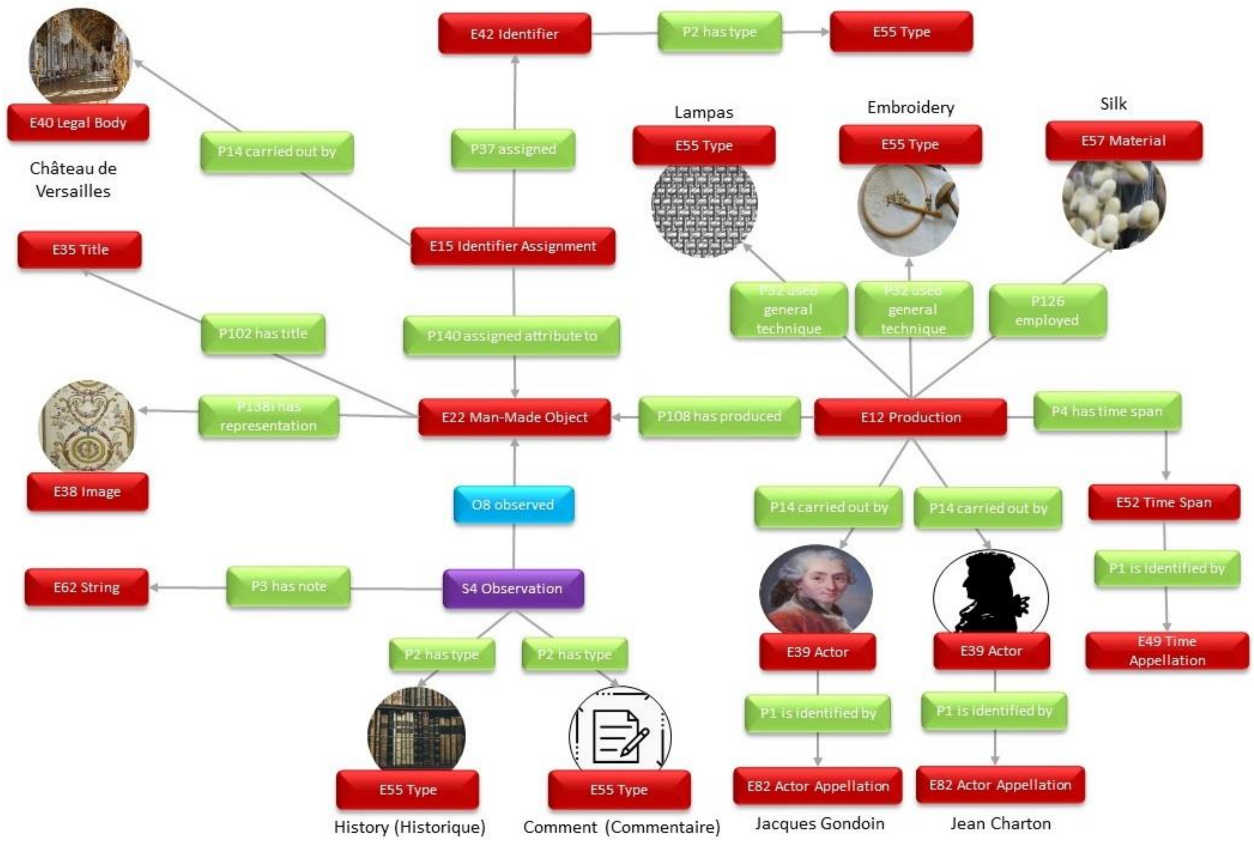

Figure 5. Triplets (extracts) used to model the metadata of the record shown in Figure 3

These graphs allow us to visualise the triplets and to check their consistency more easily. On Figure 5, we visualize the triplets modelling the information on the production conditions of the tapestry conserved at the Château de Versailles, namely:

- the production date: *E12_Production P4_has time-span E52_Time-Span*,
- the producer/creator: *E12_Production P14_carried out by E39_Actor*,
- the materials used: *E12_Production P126_employed E57_Material*,
- techniques used: *E12_Production P32_use general technique E55_Type*.

After providing these mapping *rules*, we found that all the metadata fields could be expressed using the SILKNOW data model. The JSON files were then converted to create the Knowledge Graph using an RDF converter. The graph, which contains the metadata describing 36,210 objects illustrated by 74,527 images, is finally loaded into a triple store, while the images are uploaded separately to a media server. All converted data is accessible via a SPARQL endpoint based on Virtuoso 7[23]. This data is also available via a faceted browser[24] . This browser can also be used via a graphical user interface which provides access to the SILKNOW multilingual thesaurus dedicated to the vocabulary of historical silk textiles[25] (see "The SILKNOW extension"). All the data collected is accessible with the exploratory search engine ADASilk (*Advanced Data Analysis for Silk Heritage*[26]) (Schleider, Troncy, Ehrhart, et al. 2021). Figure 6 shows the record describing the tapestry preserved in Versailles, as it appears in ADASilk[27] .

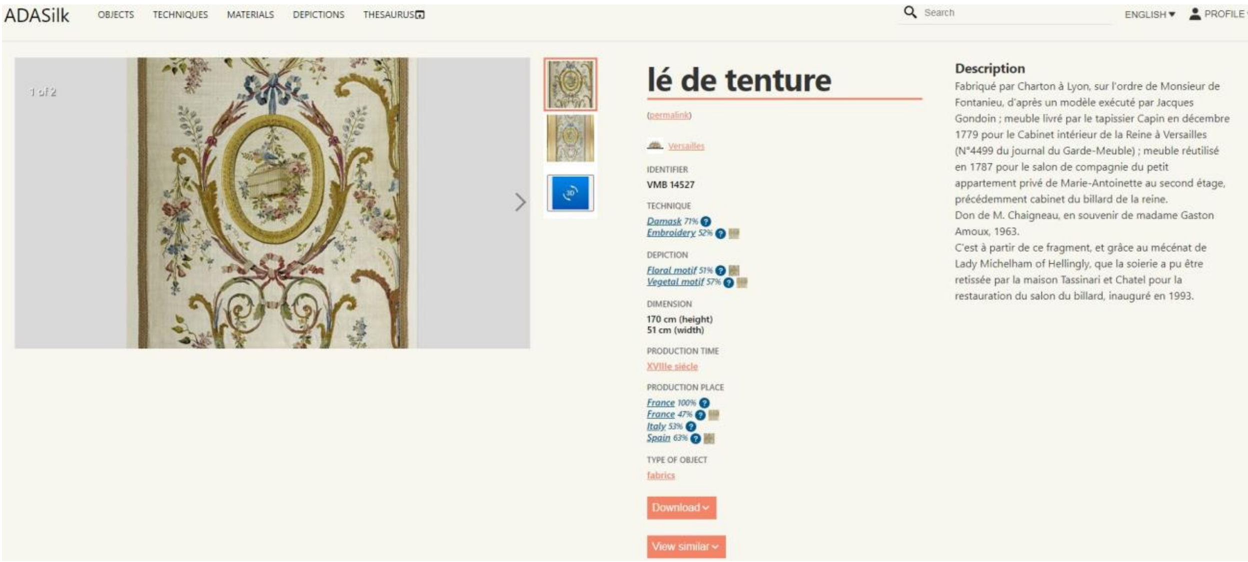

Figure 6. Example of a record in ADASilk

Integrating annotations into the data model

Prov DM: Provenance Data Model

SILKNOW has developed text (Massri and Mladenić 2020; Schleider and Troncy 2021) and image (M. Dorozynski, Clermont, and Rottensteiner 2019; Clermont et al. 2020) analysis methods that infer new properties about objects from the metadata describing them, so as to enrich the existing metadata. If we examine the record shown in Figure 6, we find the following information in the "Technique" field: Damask 71% and Embroidery 52% and in the "Depiction" field: Floral motif 51% and Vegetal motif 57% (see Figure 7).

---

[23] https://data.silknow.org/sparql.
[24] https://data.silknow.org/fct/.
[25] https://skosmos.silknow.org/thesaurus/en/.
[26] https://ada.silknow.org/fr.
[27] To consult it directly online: https://ada.silknow.org/object/2b347a9f-2c51-30f9-9e8d-ca9396a3228e.

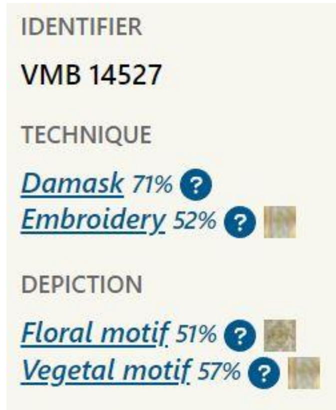

Figure 7. Fabric predictions and associated confidence scores: example

This new information was obtained using convolutional neural networks trained on the data (texts and images) contained in the Knowledge Graph. Based on textual descriptions of silk objects or on images representing them, text (Rei 2021) or image (Mareike Dorozynski and Rottensteiner 2021) classification models are able to predict the values of certain properties: the period of production, the place of production, the technique and the material for the text classification model; the previous properties, plus the patterns represented for the image classification model. Each of these predictions is associated with a confidence score expressed as a percentage. In the previous example, the text classification model predicted the "damask" value with a confidence score of 71% for the "Technique" property; the image classification model predicted the floral motif value with a confidence score of 51% for the "Depiction" property.

We have modelled the integration of the new data produced by these analyses, with the aim of making a clear distinction between these predictions and the original data. We also want to provide ADASilk users with information on the degree of reliability of the results of these analyses. To model these annotations, we have chosen to use the conceptual data model Provenance Data Model (Prov DM), recommended by the W3C (Belhajjame et al. 2013). Figure 8 shows the entities and relationships from Prov DM that we used in graph form[28] .

[28] A description of the panel preserved in the *Musée des Arts Décoratifs* can be found at the following address: https://ada.silknow.org/fr/object/99528c86-aac9-3231-a9d9-b84f6e4756fd.

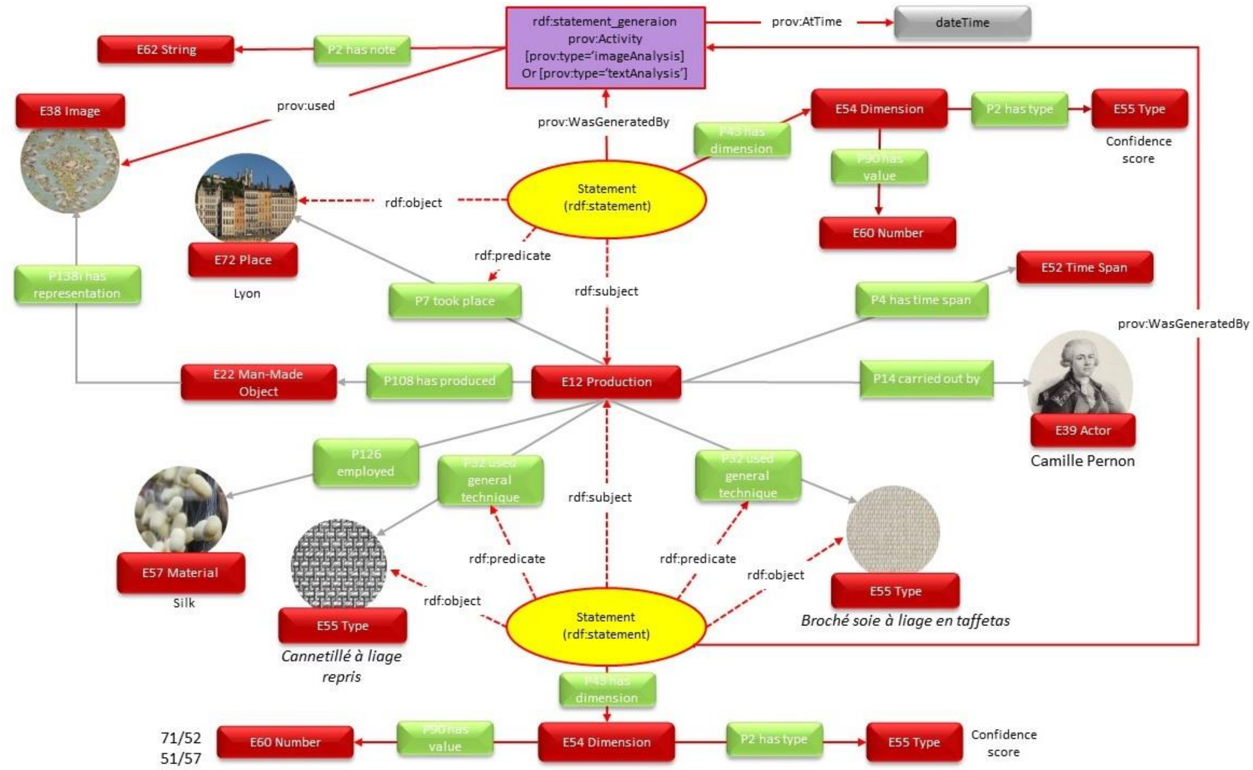

Figure 8. Integration of image analysis data with ProvDM

Image or text analysis is represented as a *Prov:Activity* which can be qualified by a type (image analysis or text analysis). Depending on the case, this *Prov:Activity* takes an *E38_Image* (image analysis) or a text or *E62_String* (text analysis) as input (*prov:used*), and produces two declarations as output (*prov:WasGeneratedBy* properties). Each of these statements has an *E54_Dimension*. The date of the analysis can be specified (*prov:AtTime*). If necessary, the analysis module can be specified with a *prov:Agent* class (of type *Software Agent*) and documented (*E31_Document*).

The SILKNOW extension

During our examination of the metadata describing ancient silk fabrics, we noticed that the technical and historical analyses written in free text were particularly rich. These sometimes very long blocks of text provide information about the weaving process, such as which weave(s)[29] , or which type(s) of yarns were used. However, this semantic richness is difficult to capture in the CIDOC CRM. The SILKNOW data model proposes to use the two triplets *S4_Observation O8_observed S4_Observation* and *S4_Observation P3_has note E62_String* to model the data provided by the metadata fields corresponding to these analyses (cf. Figure 3). The function of the *P3_has note* property is to be "a container for all informal descriptions about an object that have not been expressed in terms of CRM

---

[29] The term "weave" refers to the way in which the weft and warp threads are interwoven. This is called satin weave, for example. For more information on this term, please consult the SILKNOW thesaurus: https://skosmos.silknow.org/thesaurus/en/page/318.

constructs" (Le Boeuf et al. 2015). It is used to model the way in which an object is characterised (e.g. its appearance, structure, etc.), but it does not allow to express "in a structured form, everything that can be said about an item". This is one of the consequences of the CIDOC CRM formalism which does not have the function of capturing everything that can be said about an object.

These blocks of free text provide particularly precise information on the techniques used to produce the textiles described; often they also provide detailed iconographic descriptions. The textual analysis carried out on these texts has thus made it possible to extract technical information (weaves and weaving techniques) and other information relating to the iconographic description (motifs and styles), as illustrated in Figure 9. These different types of information may be of interest to the general public looking for patterns, to textile specialists working on the history of techniques, or to designers and creative industries wishing to draw inspiration from a historical style. Giving access to this information via a search engine would also provide a richer user experience by allowing the user to make more refined queries on the data.

When the data is integrated into the Knowledge Graph, some information (techniques and patterns) is annotated using the SILKNOW thesaurus (for more information on this thesaurus, see (Léon et al. 2019)). This semantic enrichment is achieved in the following way: if the value of a character string for a technique or pattern can be associated with one of the character strings proposed by the thesaurus, this value is replaced by a unique identifier. Modelling these concepts by means of a formal language would make it possible to integrate these annotations into the Knowledge Graph, to give access to this new data via the exploratory search engine, and to create links between these concepts and the thesaurus more easily.

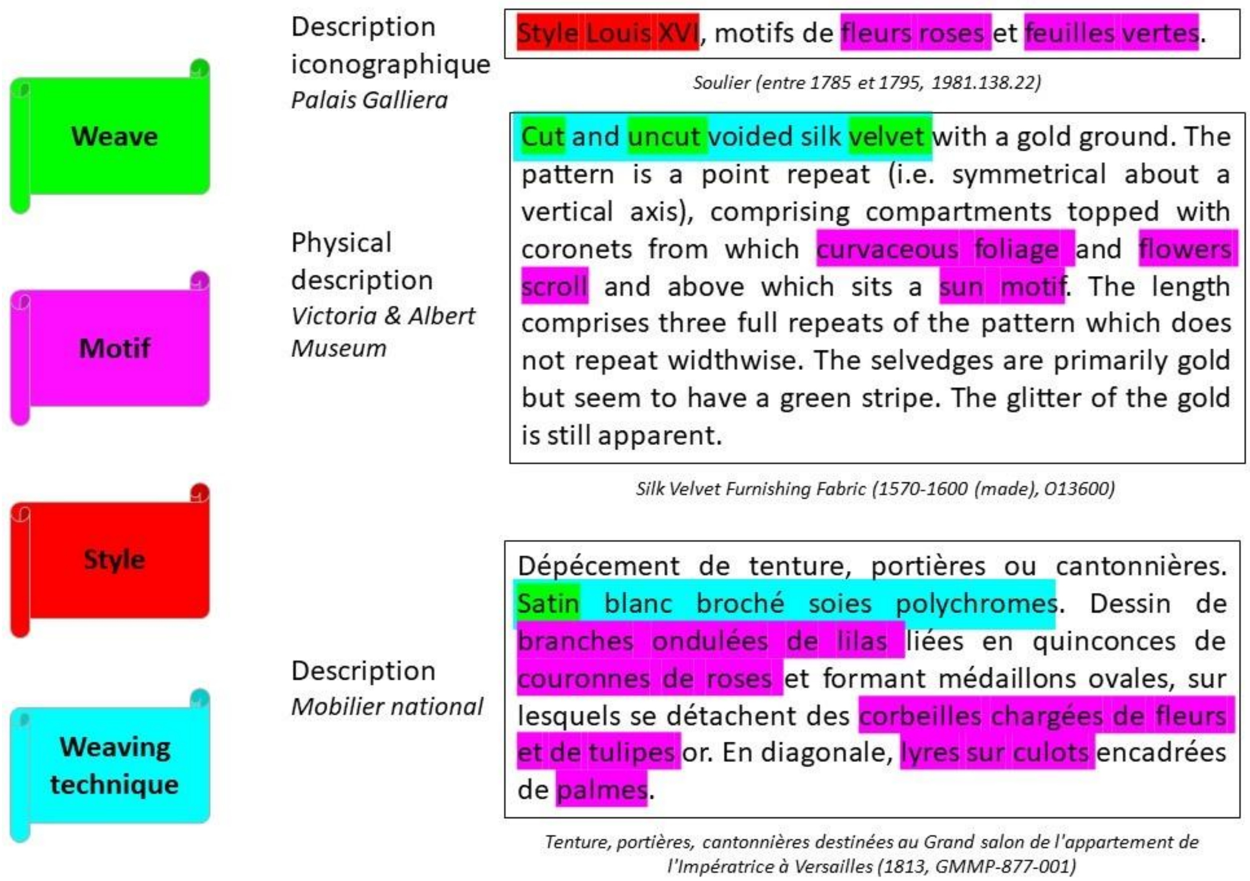

Figure 9. Types of information extracted from the descriptions: examples from the catalogues of the *Palais Galliera*, the *Victoria & Albert Museum* and the *Mobilier National*

The inherent flexibility of CIDOC CRM allows for the creation of new classes and properties to express new kinds of information. The aim is not to capture everything that can be said about an object, but to express some of the things that have been said about it - in this case, information about the process that led to the production of a silk fabric. Based on the results of the text analysis and the resulting data annotation, we created a CIDOC CRM-compatible model to formally express the process of creating and producing a silk fabric. This specialised extension, accessible via OntoME, offers 23 classes and 12 properties (Puren and Vernus 2021a). To create this extension, we adopted a bottom-up approach, relying primarily on the examination of collected data. We also worked with domain experts to verify the validity of these new classes and properties.

Among these 23 classes and properties is class *T1_Weaving*, which is a subclass of *E12_Production*. *T1_Weaving* refers to the activity of interlacing at right angles the warp threads (the threads along the length of the fabric, which are stretched on the loom) and the weft threads (the perpendicular threads that run alternately between the warp threads). This activity is usually carried out using a loom[30] . The product of the weaving is a fabric,

[30] The "weaving" entry can be found in the SILKNOW thesaurus: https://skosmos.silknow.org/thesaurus/en/page/526.

represented by the class *T7_Fabric*, itself a subclass of *E22_Man-Made Object*. These two entities can be linked with the CRM properties *P108_has produced* (*T1_Weaving P108_has produced_T7 Fabric*) and *P31_has modified* (*T1_Weaving P31_has modified T7_Fabric*). The production of this fabric is carried out following a technical procedure(s) (*T25_Weaving Technique*). The weaving process always includes the use of at least one specific weave (*T21_Weave*) (*T32_Weave Type*). It then produces a fabric that is often decorated with patterns (*T18_Motif*) of various kinds (*T34_Motif Type*), sometimes characteristic of a style (*T11_Style*) identified by experts in the field (*T13_Style Assignment*). In Figure 10, we express this information with these new classes and associated properties, for a set of tapestries for the Grand salon of the apartment of the Empress[31] in Versailles[32] kept in the collection of the French *Mobilier national* (see the last description on Figure 9).

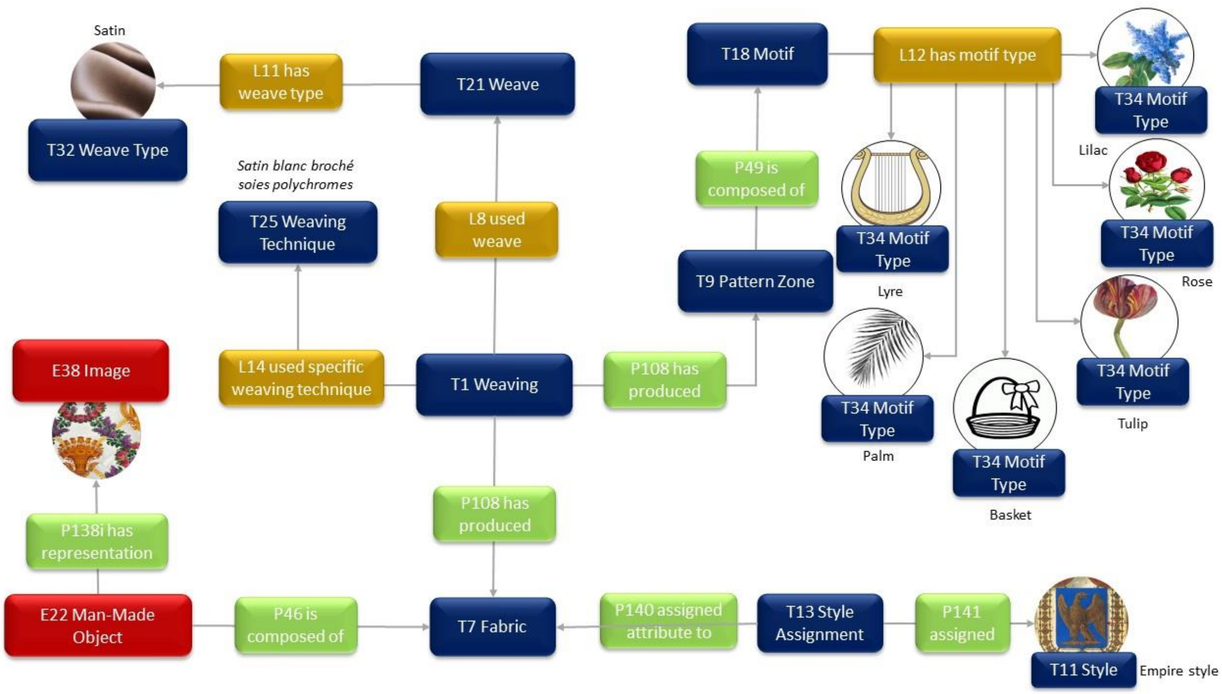

Figure 10. CIDOC CRM and SILKNOW extension compliant graph

As mentioned above, these new classes and properties facilitate the integration of semantic enrichments into the Knowledge Graph and allow the end user to query these annotations, but also to have access to additional contextualisation elements. The *T32_Weave Type* class, for example, is used to express the different weaves used during the weaving process and defined in the SILKNOW thesaurus. As shown in Figure 11, the information about the weaves used is modelled using the triplet *T21_Weave L11_has weave type T32_Weave Type*.

[31] Marie-Louise I, Duchess of Parma, and second wife of Napoleon I.
[32] Tenture, portières, cantonnières destinées au Grand salon de l'appartement de l'Impératrice à Versailles (1813, Bissardon, Cousin & Bony), Collections du *Mobilier national*, URL: https://collection.mobiliernational.culture.gouv.fr/objet/GMMP-877-001.

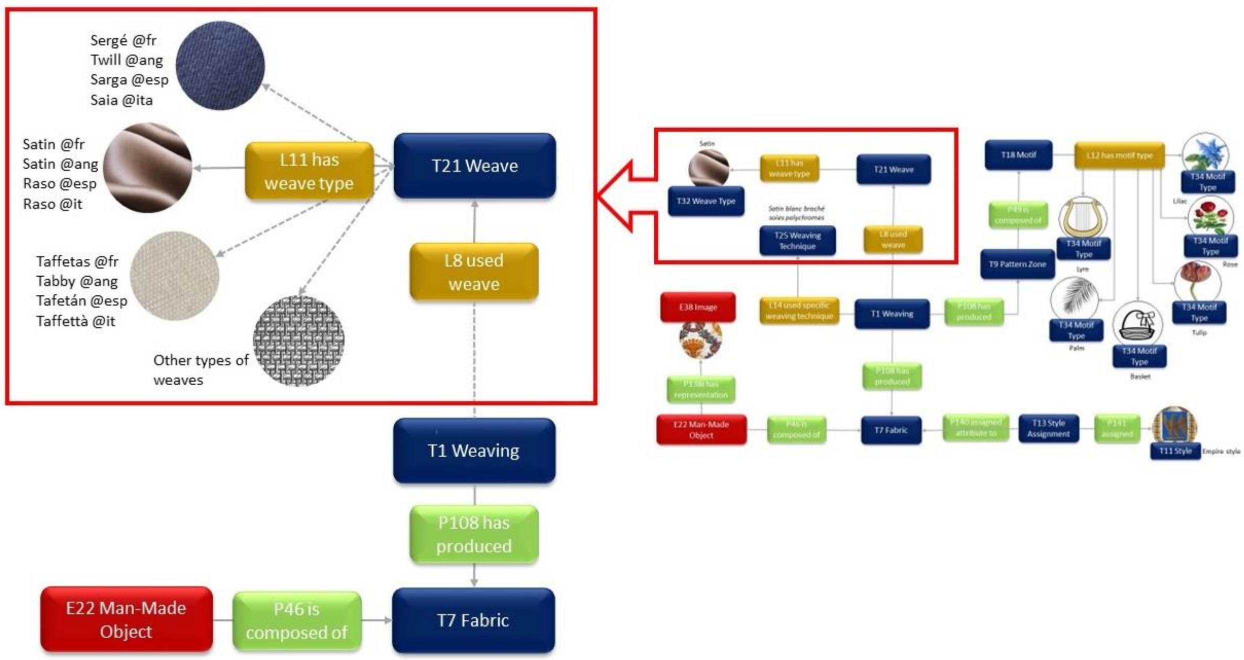

Figure 11. Using the *T32_Weave Type* class

We have done the same for the information on patterns by creating the class *T34_Pattern Type*. This class should also make it possible to create links with the definitions available in the thesaurus. For example, the user could consult the definition of the Palme motif[33] , represented on the tapestries held in the collections of the *Mobilier national*, without forgetting the bibliographical references associated with it.

Conclusion

Silk Heritage is of crucial importance for understanding European history and culture. But some conservation institutions lack the means to promote it and make it accessible to the general public. More generally, Silk Heritage suffers from a lack of visibility, especially on the Internet, which has become the preferred tool for searching and discovering new information. Semantic Web technologies, however, offer the possibility of disseminating Cultural Heritage data more easily on the Web by improving its discoverability, but also of enriching this data by linking it to other information.

The SILKNOW project wished to exploit these potentialities to enhance the visibility of the European Silk Heritage, by developing the exploratory search engine ADASilk dedicated to these specialised collections. ADASilk is based on a Knowledge Graph, whose creation is based on the SILKOW ontology - a subset of the CIDOC CRM and CRMsci ontology. The richness of the metadata contained in the Knowledge Graph has also led us to develop the SILKNOW extension, a conceptual model for the description of ancient silk textiles. We hope that this very specialised extension is only a first step in the development of a domain

[33] The "palm" concept is defined in the thesaurus: https://skosmos.silknow.org/thesaurus/en/page/763.

ontology for textile Cultural Heritage. We also wish that our work, and the methodology we have adopted, will be of benefit to other Cultural Heritage projects on the Web.

A project like SILKNOW shows the interest of semantic technologies for Cultural Heritage protection. The same strategy has been used in other Cultural eritage projects and can be considered as a "classical" approach. However, it is necessary to question its ease of use and application. Indeed, such projects require significant investments, and therefore remain inaccessible to small institutions, which would be the first to benefit from them. Having demonstrated the value of these approaches, it is therefore necessary to consider the development of tools to access these technologies, and thus make them accessible to the greatest number of people.

References

Achichi, Manel, Pasquale Lisena, Konstantin Todorov, Raphaël Troncy, and Jean Delahousse. 2018. « DOREMUS: A Graph of Linked Musical Works ». In *ISWC: International Semantic Web Conference*, LNCS:3-19. The Semantic Web – Part II. Monterey, CA, United States: Springer. https://doi.org/10.1007/978-3-030-00668-6_1.

Belhajjame, Khalid, Reza B'Far, James Cheney, Sam Coppens, Stephen Cresswell, Yolanda Gil, Paul Groth, et al. 2013. *PROV-DM: The PROV Data Model*. Edited by Luc Moreau and Paolo Missier. https://www.w3.org/TR/prov-dm/.

Beretta, Francesco. 2021. « A challenge for historical research: making data FAIR using a collaborative ontology management environment (OntoME) ». *Semantic Web – Interoperability, Usability, Applicability* 12 (2): 279-94. https://doi.org/10.3233/SW-200416.

Briatte, Katell. 2012. *HADOC. Modèle harmonisé pour la production des données culturelles*. Ministère de la Culture et de la Communication.

Christen, Peter. 2012. *Data Matching*. Berlin, Heidelberg: Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-31164-2.

Claerr, Thierry, and Isabelle Westeel, ed. 2017. *Numériser et mettre en ligne*. *Numériser et mettre en ligne*. La Boîte à outils. Villeurbanne: Presses de l'enssib. http://books.openedition.org/pressesenssib/414.

Clermont, D., Mareike Dorozynski, Dennis Wittich, and Franz Rottensteiner. 2020. « Assessing the Semantic Similarity of Images of Silk Fabrics Using Convolutional Neural Networks ». *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences ; 5,2* 5 (2): 641-48. https://doi.org/10.15488/10876.

Dorozynski, M., D. Clermont, and F. Rottensteiner. 2019. « Multi-task deep learning with incomplete training samples for the image-based prediction of variables describing silk fabrics ». In *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, IV-2-W6:47-54. Copernicus GmbH. https://doi.org/10.5194/isprs-annals-IV-2-W6-47-2019.

Dorozynski, Mareike, and Franz Rottensteiner. 2021. *SILKNOW Image Classification - final version WP6 - Multi-Task Classification*. https://doi.org/10.5281/zenodo.5091813.

Estermann, Beat. 2015. « Open data et crowdsourcing : un état des lieux du point de vue des musées ». *La Lettre de l'OCIM. Musées, Patrimoine et Culture scientifiques et techniques*, nº 162 (novembre): 41-46. https://doi.org/10.4000/ocim.1597.

« Europeana Data Model ». s. d. Europeana Pro. Accessed on 4 October 2022. https://pro.europeana.eu/page/edm-documentation.

Foron-Dauphin, Nathalie, and under the direction of Delphine Cano. 2016. « Le patrimoine des Soieries Bonnet (Jujurieux, Ain) : conserver et valoriser in situ un ensemble unique de l'industrie textile ». *In Situ. Revue des patrimoines*, nº 29 (juin). https://doi.org/10.4000/insitu.13890.

Fournier, Marianne, Clarisse Despierres, Cladie Mercandelli, and Emmanuelle Bernard. 2016. « Conservation et présentation des lieux et collections de patrimoine textile : l'exemple de la fédération « Transmission du savoir-faire textile en Rhône-Alpes » ». *In Situ. Revue des patrimoines*, nº 29 (juin). https://doi.org/10.4000/insitu.13463.

Freire, Nuno, Enno Meijers, René Voorburg, and Antoine Isaac. 2018. « Aggregation of Cultural Heritage Datasets through the Web of Data ». *Procedia Computer Science*, Proceedings of the 14th International Conference on Semantic Systems 10th – 13th of September 2018 Vienna, Austria, 137 (janvier): 120-26. https://doi.org/10.1016/j.procs.2018.09.012.

Freire, Nuno, René Voorburg, Roland Cornelissen, Sjors de Valk, Enno Meijers, and Antoine Isaac. 2019. « Aggregation of Linked Data in the Cultural Heritage Domain: A Case Study in the Europeana Network ». *Information* 10 (8): 252. https://doi.org/10.3390/info10080252.

Gaitán, Mar, Ester Alba, Arabella León, Manolo Pérez, Javier Sevilla, and Cristina Portalés. 2019. « Towards the Preservation and Dissemination of Historical Silk Weaving Techniques in the Digital Era ». *Heritage* 2 (3): 1892-1911. https://doi.org/10.3390/heritage2030115.

Grant, Alice, Joséphine Nieuwenhuis, and Tony Petersen. 1995. *International guidelines for museum object information : the CIDOC information categories.* ICOM International Committee for Documentation. https://icom.museum/en/ressource/international-guidelines-for-museum-object-information-the-cidoc-information-categories/.

Heath, Tom, and Christian Bizer. 2011. *Linked Data: Evolving the Web into a Global Data Space*. Synthesis Lectures on Data, Semantics, and Knowledge. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-031-79432-2.

Koch, Inês, Cristina Ribeiro, and Carla Teixeira Lopes. 2020. « ArchOnto, a CIDOC-CRM-Based Linked Data Model for the Portuguese Archives ». In *Digital Libraries for Open Knowledge: 24th International Conference on Theory and Practice of Digital Libraries, TPDL 2020, Lyon, France, August 25–27, 2020, Proceedings*, 133-46. Berlin, Heidelberg: Springer-Verlag. https://doi.org/10.1007/978-3-030-54956-5_10.

Koho, Mikko, Esko Ikkala, Petri Leskinen, Minna Tamper, Jouni Tuominen, and Eero Hyvönen. 2021. « WarSampo Knowledge Graph: Finland in the Second World War as Linked Open Data ». *Semantic Web* 12 (2): 265-78. https://doi.org/10.3233/SW-200392.

Kritsotaki, Gerald Hiebel, Theodoridou, Martin Doerr, and Yannis Rousakis. 2017. *Definition of the CRMsci. An Extension of CIDOC-CRM to support scientific observation. Version 1.2.3*. https://www.cidoc-crm.org/crmsci/ModelVersion/version-1.2.3.

Le Boeuf, Patrick, Martin Doerr, Christian Emil Ore, and Stephen Stead. 2015. *Definition of the CIDOC Conceptual Reference Model. Version 6.2.* https://cidoc-crm.org/.

Léon, A., M. Gaitán, J. Sebastián, E. Alba Pagán, and I. Insa. 2019. « SILKNOW. Designing a Thesaurus about Historical Silk for Small and Medium-Sized Textile Museums ». In *Science and Digital Technology for Cultural Heritage*, édité par Pilar Ortiz Calderón, Francisco Pinto Puerto, Philip Verhagen, and Andrés J. Prieto, 1re éd., 187-90. CRC Press. https://doi.org/10.1201/9780429345470-34.

Marden, J., C. Li-Madeo, N. Whysel, and J. Edelstein. 2013. « Linked open data for cultural heritage: evolution of an information technology. » *roceedings of the 31st ACM international conference on Design of communication*, 107-12. http://dx.doi.org/10.1145/2507065.2507103.

Massri, M Besher, and Dunja Mladenić. 2020. « Extracting Structured Metadata from Multilingual Textual Descriptions in the Domain of Silk Heritage ». *Information Society*, 4.

Pagán, Ester Alba, María del Mar Gaitán Salvatella, María Dolores Pitarch, Arabella León Muñoz, María del Mar Moya Toledo, José Marin Ruiz, Maurizio Vitella, et al. 2020. « From Silk to Digital Technologies: A Gateway to New Opportunities for Creative Industries, Traditional Crafts and Designers. The SILKNOW Case ». *Sustainability* 12 (19): 8279. https://doi.org/10.3390/su12198279.

Petrov, Julia. 2019. *Fashion, History, Museums: Inventing the Display of Dress*. London: Bloomsbury Publishing Plc. https://doi.org/10.5040/9781350049024.

Portalés, Cristina, Jorge Sebastián, Ester Alba, Javier Sevilla, Mar Gaitán, Paz Ruiz, and Marcos Fernández. 2018. « Interactive Tools for the Preservation, Dissemination and Study of Silk Heritage—An Introduction to the SILKNOW Project ». *Multimodal Technologies and Interaction* 2 (2): 28. https://doi.org/10.3390/mti2020028.

Puren, Marie, and Pierre Vernus. 2021a. « SIKNOW 0.1 ». *OntoME*. 2021. https://ontome.net/namespace/36.

———. 2021b. « Silknow generic ». *OntoME*. 2021. https://ontome.net/profile/7.

Rei, Luis. 2021. *SILKNOW Text Classifier Models* . https://doi.org/10.5281/zenodo.5070696.

Schleider, Thomas, and Raphael Troncy. 2021. « Zero-Shot Information Extraction to Enhance a Knowledge Graph Describing Silk Textiles ». In *Proceedings of the 5th Joint SIGHUM Workshop on Computational Linguistics for Cultural Heritage, Social Sciences,* *Humanities and Literature*, 138-46. Punta Cana, Dominican Republic (online): Association for Computational Linguistics. https://doi.org/10.18653/v1/2021.latechclfl-1.16.

Schleider, Thomas, Raphael Troncy, Thibault Ehrhart, Mareike Dorozynski, Franz Rottensteiner, Jorge Sebastián Lozano, and Georgia Lo Cicero. 2021. « Searching Silk Fabrics by Images Leveraging on Knowledge Graph and Domain Expert Rules ». In *Proceedings of the 3rd Workshop on Structuring and Understanding of Multimedia heritAge Contents*, 41-49. SUMAC'21. New York, NY, USA: Association for Computing Machinery. https://doi.org/10.1145/3475720.3484445.

Schleider, Thomas, Raphaël Troncy, Mar Gaitán, Dunja Mladenic, Avgustin Kastelic, Besher Massri, Arabella León, et al. 2021. « The SILKNOW Knowledge Graph ». *Semantic Web – Interoperability, Usability, Applicability.* https://www.semantic-web-journal.net/content/silknow-knowledge-graph-0.

Stewart, Tekara Shay, and Sara B. Marcketti. 2012. « Textiles, dress, and fashion museum website development: strategies and practices ». *Museum Management and Curatorship* 27 (5): 523-38. https://doi.org/10.1080/09647775.2012.738137.

Suls, Dieter. 2017. « Europeana Fashion: Past, Present and Future ». *Art Libraries Journal* 42 (3): 123-29. https://doi.org/10.1017/alj.2017.18.