SILKNOW

# D7.1. Testing report in a controlled scenario

| | |
|---|---|
| Deliverable number and name: | D7.1. Testing report in a controlled scenario |
| Due date: | 30/09/2020 |
| Work Package: | WP7 |
| Deliverable leader: | UNIPA |
| Authors: | Maurizio Vitella, Valeria Seidita, Georgia Lo Cicero, Dunja Mladenic, Besher Massri, Franz Rottensteiner, Raphael Troncy |
| Reviewers: | Marie Puren, Pierre Vernus |
| Approved by: | Jorge Sebastián |
| Dissemination level: | CO |
| Version: | V1.2 |

European Commission

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 769504

| Document revision history | | | |
|---|---|---|---|
| Version | Date | Contributor | Comments |
| 0.1 | 5/09/2020 | Valeria Seidita | Initial document with TOC |
| 0.2 | 10/09/2020 | Valeria Seidita and Georgia Lo Cicero | Description of the SILKNOW System |
| 0.5 | 25/09/2020 | Valeria Seidita and Georgia Lo Cicero | Validation method and results reported |
| 0.6 | 27/09/2020 | Maurizio Vitella, Valeria Seidita and Georgia Lo Cicero | Feedback and comments added |
| 0.7 | 28/09/2020 | Valeria Seidita, Georgia Lo Cicero, Dunja Mladenic | Section 2 revised by JSI |
| 0.8 | 29/09/2020 | Valeria Seidita, Georgia Lo Cicero, Franz Rottenstei-ner, Raphael Troncy | Sections 2 and 3 revised |
| 0.9 | 30/09/2002 | Valeria Seidita, Georgia Lo Cicero | First final version revised |
| 1.0 | 7/10/2020 | Valeria Seidita | Reviewers' comments met |
| 1.1 | 28/10/2020 | Jorge Sebastián, Valeria Seidita and Dunja Mladenic | Quality control and further edits |
| 1.2 | 29/10/2020 | Jorge Sebastián | Final version |

| List of acronyms | |
|---|---|
| **CIDOC - ICOM** | International Committee for Documentation - International Council of Museums |
| **CRM** | Conceptual Reference Model |
| **ICT** | Information and Communication Technologies |
| **V&V** | Validation and Verification |

**Table of contents**

The current deliverable reports the very first considerations on the validation activities performed in Task 7.1. It provides feedback from partners involved in the development of the text analytic module, the knowledge graph and the image analysis module. This feedback will be used for improvements and for performing further validations in Task 7.2. The report mainly describes how validation was carried out, and its results.

## 1. INTRODUCTION

In this report, we illustrate the work done within the first task of WP7 - Work Package 7, Performance testing and pilots evaluation. It has aimed at testing the performance of the SILKNOW system in a controlled scenario.

The SILKNOW project intends to deliver a web-based platform that provides texts, images, and printable models for improving knowledge of historical silk textiles. The system uses data from existing digital catalogs of textile collections, that are analyzed and processed to homogenize content, and retrieve semantic information from text in four languages (Spanish, English, French and Italian). These data are processed employing text analytics and image algorithms that have been implemented in specific software modules.

WP7 focusses on these software modules and, starting from an end-user's viewpoint, is mainly devoted to validating the system performance and its usability, in addition to evaluating the integration of SILKNOW results. Therefore, it includes a set of Verification and Validation activities (V&V) [1][2] on the whole system (Tasks 7.1, 7.2 and 7.6) and a set of activities devoted to assessing the usability of the software.

In Task 7.1, testing has been executed in a controlled scenario, by project partners themselves, mostly domain experts in the field of historical silk textiles. This task has been performed in close coordination with the development of the software modules in WP3 and WP4.

Nevertheless, WP7 in general and Task 7.1 in particular involve activities not strictly related to the development of software. Their main aim is to provide feedback on the software modules, from experts in the domain of silk and cultural heritage, to guarantee quality and alignment with the SILKNOW project objectives. Following this rationale, three fundamental software modules of the SILKNOW system have been considered in Task 7.1. Namely, they are the *text analytic module*, the *knowledge graph module* and the *image retrieval module*.

The text analytic module [5] is a tool for the textual analysis of data from museum collections. The result of this analysis is used to populate the ontology underpinning the whole SILKNOW system [6], which is based on CIDOC-CRM. New classes and properties have been added [6] to fully represent and describe the domain of silk historical textiles.

The knowledge graph module is a tool used for browsing the SILKNOW ontology. It is populated with text and images extracted from records, originally created for documenting the collections of museums. Such records have been harvested from publicly available data sources or supplied by data providers that collaborate with SILKNOW.

The image retrieval module (named MODULE II in the general workflow of the SILKNOW system [3] [4]) is a deep-learning based module aiming at predicting the properties of silk fabric by processing images. It involves a classification tool whose objective is to process data from images in order to allow semantic correlation among objects from different collections. More importantly, it also involves an image retrieval tool, as explained in full detail below.

These three modules are of fundamental importance for the development of the final system. They provide the semantic analysis of the input data to the system consistent with the SILKNOW ontology and allow the implementation of the intelligent part of the entire system.

Testing of these modules has been done during the development phase by the responsible partners (JSI, EURECOM and LUH) and will continue to be done before the final integration of the system, to be tested and validated during other WP7 tasks. In Task 7.1 it was considered essential for the development of the system to validate the three modules separately. Since SILKNOW is strongly interdisciplinary, experts in the field of fabrics and cultural heritage in general (domain experts from now on) have been strongly involved from the very beginning. Task 7.1 allowed domain experts to interact with developers and give them feedback on some central elements. As explained below, we focused on the semantic annotation of the text analysis module, the verification of the correct mapping within the ontology of the elements coming from the collections, the correctness of the value assigned to the "production location" field and the similarity of retrieved images.

This report is organized as follows: section 2 describes the functionalities and main features of all three software modules; section 3 describes the rationale underlying the modules validation; sections 4, 5 and 6 respectively show the method followed for each validation and the related results; section 7 reports comments by the domain experts; and finally section 8 draws some conclusions.

## 2. SILKNOW MODULES DESCRIPTION

### 2.1. **Text Analytic Module**

The Text Analytic Module in SILKNOW is a fundamental part of the final software development of the project. It is being carried out by JSI, within Work Package 3. It aims to support the development of an intelligent system for a better understanding and use of language related to silk heritage. For this purpose, it is based on the SILKNOW multilingual Thesaurus, that allows to analyze the content of the text within the objects' records, in the form of descriptive sentences or short paragraphs.

Also, the Text Analytic Module aims to implement semantic annotation, metadata prediction, and extraction of information from the textual descriptions in catalog records. These features, in turn, will support rich visualization and semantic search.

The semantic annotation tool has been developed based on machine learning techniques, extending an existing Wikifier service previously developed by JSI, which allows multilingual annotation of texts with Wikipedia concepts. The basic approach uses a Pagerank type method to identify a coherent set of relevant concepts considering the input document as a whole.

A text annotation tool has been developed for domain experts, accessible with credentials through the web service in http://relationextraction.com/ It allows the automatic extraction of information that is not considered in the existing metadata, and the training of the software.

## 2.2. **Knowledge Graph Module**

The SILKNOW Knowledge Graph integrates work done in WPs 2, 3, 4 and 5. It represents information extracted from records about silk textiles and other types of objects found in the participating museums or collections. Objects are described in line with the classes present in the SILKNOW Ontology. The Knowledge Graph allows users to navigate and search among objects in the database, and to retrieve information using complex queries.

At the time of the validation reported in this document, the list of museums from which data have been integrated are:

- Garin 1820
- Rhode Island School of Design Museum (RISD)
- Red Digital de Colecciones de Museos de España (CERES)
- Centre de Documentació i Museu Tèxtil de Terrassa (IMATEX database)
- Musée des Art Décoratifs
- Metropolitan Museum of Art
- Boston Museum of Fine Arts
- Musée des Tissus et des arts décoratifs
- Victoria and Albert Museum
- Museo Diocesano di Caccamo
- Museo Diocesano di Palermo
- Chiesa Madre di Termini Imerese
- Cattedrale di Palermo
- Duomo di Monreale

Navigating the Knowledge Graph is possible through a faceted browser available at http://data.silknow.org/fct/ (see Figure 1).

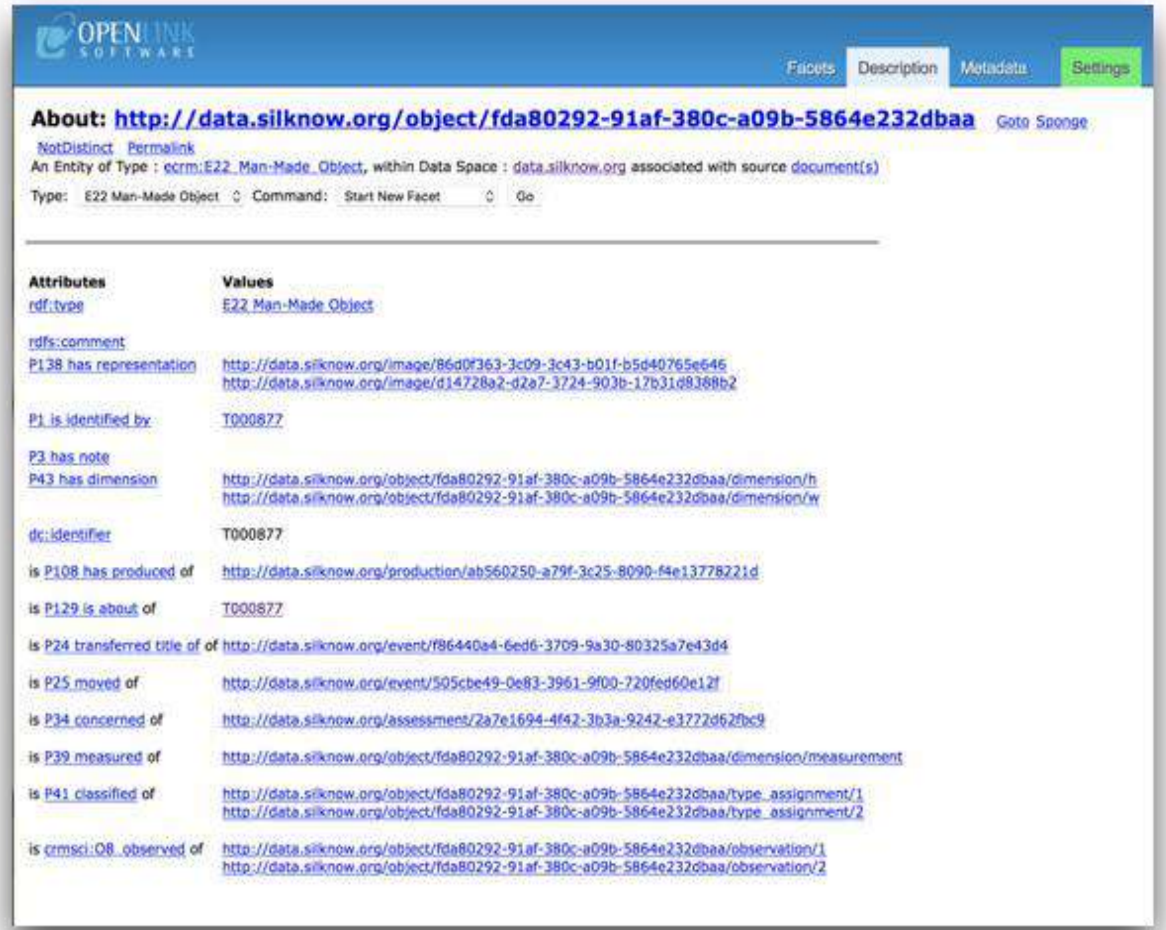



Figure 1. SILKNOW's faceted web browser

## 2.3. **Image Analysis Module**

This module is developed by LUH in WP4. It has two main components: image classification (the analysis of semantic properties of images from the image data files), and image retrieval (the discovery of images on the basis of the similarity of a given query image with images existing in the database). Both components use Convolutional Neural Networks (CNN), a supervised classification technique; they were described in D4.4 and 4.5, respectively.

In the classification component, the true values of the semantic properties to be predicted have to be known for evaluation, and this evaluation forms an intrinsic part of the software development by the ICT experts. Consequently, an additional evaluation by cultural heritage experts does not contribute much to the improvement of the method. This is different for the image retrieval module for reasons that will be given below. Thus, the consortium decided that the evaluation of the image analysis module by domain experts would focus on the image retrieval results.

The goal of the image retrieval module is to use visual information, i.e. an image, as a key to the database in order to find records that are similar to the given image in some sense. The image retrieval software and its description and documentation from deliverable D4.5. As described in that deliverable, the definition of similarity used to train the CNN is based on the similarity of semantic properties: two images in the training dataset are supposed to be more similar the more semantic properties associated with these images in the database match. The CNN essentially works as a feature extractor that computes a feature vector (called *descriptor*) for each image which has a small Euclidean distance from descriptors of similar images according to the definition of similarity just given, and a large distance from descriptors of dissimilar images. Up to five semantic properties (referred to as variables in D4.5) can be considered to define semantic similarity. They were chosen because they are the only ones in the SILKNOW data model considered to be related to the visual appearance of the fabric (cf. D4.1). The five variables are:

- Production timespan
- Production place
- Material
- Technique or procedure
- Subject depicted type

Each variable has been identified from the data model described in D2.1 and has a precise counterpart in a variable used in the knowledge graph.

The image retrieval software is implemented starting from the generation of training samples on the basis of records extracted from the knowledge graph (`Data Preparation`; cf. D4.5). This part is greatly dependent on the raw annotations for the individual classes in the knowledge graph, thus on domain experts' knowledge, and the way this knowledge is represented in the Thesaurus. Having generated the training data, the `Training` subsystem produces a pre-trained CNN model that can predict feature vectors from input images. This CNN is applied to all images in the database, and a spatial index of the corresponding descriptors is generated by the `Build Descriptor Index` subsystem, which forms the basis for image retrieval. The subsystem `Get Nearest Neighbours` takes an image as input, applies the CNN to compute its descriptor and then uses the spatial index of the descriptors of the images in the database to detect the k nearest neighbors in that descriptor space, i.e. the k records associated with images whose descriptors have the most similar descriptors (according to the Euclidean distance) to the query image.

For the purpose of numerical evaluation, the semantic properties of the neighbors thus retrieved from the database can also be predicted by a majority vote, but this is not the main purpose of the image retrieval module. For more details, cf. D4.5.

## 3. RATIONALE UNDERPINNING THE VALIDATION OF THE MODULES

The work done in Task 7.1, and described in this report, was not a "technical" validation activity. This means that it was not carried out by experts in the ICT sector but by experts in the field of cultural heritage, more specifically textile experts. It was therefore not an activity aimed at verifying the software developed. A technical validation had already been done by the partners who developed the three modules during their development phase, and is described in the deliverables D3.3 (text analysis) [5], D6.5 (knowledge graph) [7] and D4.5 (image retrieval) [8].

At the beginning of Task 7.1 the SILKNOW software was not yet fully integrated and the three modules were still in the development phase. Therefore, in agreement with the consortium partners, we decided to carry out a validation of the individual modules. The type of validation that would be most useful to improve the quality of the modules during their subsequent development was extensively discussed.

The rationale was to involve the domain experts, since they were expected to be the first end users of the resource. Additionally, they had been involved from the early stages of the project to draw up the scenarios, to build the thesaurus and then populate the ontology of the system. Now, their involvement was again fundamental to validate some choices that led to the software modules in their current form.

As regards the Text Analytic module, training machine learning models for classification requires labelled data. Domain experts have been required to identify relevant information about the artefacts that is not available in the provided metadata. These include the number of sections, foreground or background color of the material etc. (D3.3 for details on the results). In order to collect the labelled data as quickly as possible, we have adopted a machine learning technique called Active learning, where domain experts are required to provide labels for the most informative museum artefacts while the system is generating the model.

As regards the Image Retrieval module, domain experts identified five variables (see previous section) related to an expected conception of visual appearance of the fabric and to be predictable by image analysis methods.

As explained earlier in the description of the image retrieval module, the training of the software is dependent on how the annotations were made in the knowledge graph. Therefore, during the validation of the module, feedback from domain experts was necessary to validate the entire image retrieval chain. First, the experts of the domain had to determine if, starting from a sample image, the $k = 10$ images that the module considered to be the most similar ones in the database actually met a similarity criterion according to their expertise. This evaluation was used to find out whether for a given query image, the software produced at least some meaningful results, i.e. whether there were at least some retrieved images for a given query image that matched the experts' criteria.

The result of this validation should provide an indication whether the definition of similarity based on the semantic properties of images is sufficient to produce image retrieval results that are meaningful for the domain experts and whether the semantic properties chosen for that definition are appropriate. In this way, this evaluation will give hints for improving the image retrieval software, e.g. by increasing the emphasis of some variables over others or by including visual cues into the definition of similarity. We believe that the focal point is in the data preparation part of the classification module, because the results of a supervised method cannot be better than the quality of the training data, in this case extracted from the knowledge graph.

Therefore, the comments and responses provided by the experts will be used to validate those five variables chosen and the quality of the annotations of the variables in the knowledge graph, which may also give indications for improving the methodology. Namely, by reducing the influence of variables for which the annotations are very unreliable for the definition of similarity.

Both previous modules work upon the knowledge graph which is central for the SILKNOW system. Validation from the domain experts on the relationship between the knowledge graph, the thesaurus and the SILKNOW ontology is of great importance, since it greatly affects the other two modules and is the core of the SILKNOW system. Consequently, in this first validation task, we decided to check whether the records describing the museums' collections had been correctly represented in the knowledge graph.

# 4. TEXT ANALYTIC VALIDATION

## 4.1. **Objectives**

The objective of this validation is to map fragments of the texts describing the museum artifacts with specific labels capturing meta information about historical silk textiles. Labels chosen for this task have been: background color, foreground color, number of sections, and other numbers (used for yarns, dimensions, rapport...). In order to train the system, we involved domain experts from three project partners: University of Palermo, Garin and University of Valencia.

## 4.2. **Method**

To perform the experiments, JSI has developed a web service that provides textual descriptions of the museum artefacts and gives an interface for domain experts to label them. The initial experiments, defining a color annotation task and a number annotation task, were based on instances from the Victoria and Albert Museum (VAM), given the average high quality of its cataloguing records. In the first task, we asked the domain experts to annotate the selected parts of texts from the museum that contain a mention of some color. The possible annotations related to specific colors are background color, foreground color, border color or other. In the second task, we asked the domain experts to annotate the selected parts of texts from the museum that contain a mention of some number. The possible annotations related to specific numbers are yarns, dimension, section, rapport, and colors. Yarns represents the number of threads used to make the fabrics; Dimensions represents the measures inherent to the fabric; Section refers to the construction of the fabric and the numbers related to it; Rapport indicates the dimensions of the complete design and how many times it is repeated in width and height in a fabric. Each time a number does not refer to the previous labels the domain expert has to annotate it as "other".

We have selected these two tasks as relevant for the Virtual Loom developed in WP5, where the number of yarns, the number of colors and the background color can be used for the digital visualization of a textile.

The web service has been used by textile domain experts to train the machine to recognize specific textile terms, appearing in the records provided by museums. At the time of this validation, the records available for performing the annotation came from VAM, IMATEX and CERES. The records from VAM were grouped into 4 sections, according to the words that were to be analyzed.

- VAM Colors: The word underlined by the software inside the displayed record had to be a color. The categories from which the domain expert could choose were 'Background Color', 'Foreground Color', 'Border Color', 'Other'. Afterwards, the expert could indicate which category the color belonged to, by deducing it from the text on the item's record. When it was impossible to deduce from the text which category the indicated color belonged to, the domain expert clicked on ‘Skip’ to go to the next record. The ‘Other’ button was used if the color did not belong to any category (see Figure 3 for an overview).

- VAM n. Section: A number is indicated within the text of the textile item record. The domain expert could choose between ‘Yes’ or ‘No’ to indicate whether this number indicated a number related to the weave of the fabric. If the number was difficult to frame as belonging to the weave you could choose to click the ‘Skip’ button to go to the next record (see Figures 7 and 8).
- VAM numbers: A number is highlighted in the record. The domain expert had to indicate if the number referred to the number of garments of the yarn by clicking on ‘Yarn’, to dimensions by clicking on ‘Dimension’, if the number belonged to the weave by clicking on ‘Section’, if the number indicated the pattern ratio by clicking on ‘Rapport’, a color by clicking on ‘Color’, or something else by clicking on ‘Other’ (see Figure 2 and 10).
- VAM foreground colors: Analyzing the textile product record, the word indicating a color was highlighted. Subsequently, the domain expert had to indicate whether the color was in the foreground or not in the fabric's chromatic structure, choosing between ‘Yes’ or ‘No’; otherwise the ‘Skip’ button led to another record (see Figure 2).

At a later stage, record written in Spanish were added, from the IMATEX and CERES databases, again because of their average good quality. Subsequently, the domain experts used the same system to instruct the software to recognize whether the number underlined in the text belonged to the Yarn, Dimension, Section, Rapport or Color categories, or by highlighting a word indicating a color, whether it was a background color or foreground color.

### 4.3. **Results**

The http://relationextraction.com/ web service is well designed and easy to use but some problems were found that made the analysis of the selected words difficult. For example, often it was not possible to understand the context in which the word underlined by the platform was framed. This situation mainly happened when the platform was extrapolating too short a piece of text from the artefact record. It would be necessary to read the context of the extrapolated sentence to understand it better.

For instance, as shown in the first example (see Figure 2), sometimes it was hard to understand whether a color was a background or foreground one. A fragment of golden fabric, made with golden warp and silk weft, gives misleading indications. Indeed, if we say that the fabric is golden, it will look like the main color, but if gold is actually part of the warp, and also the weft is made of silks of different colors, gold is no longer the foreground color. Indeed, the presence of gold thread only marginally determines the dominant color of the fabric.

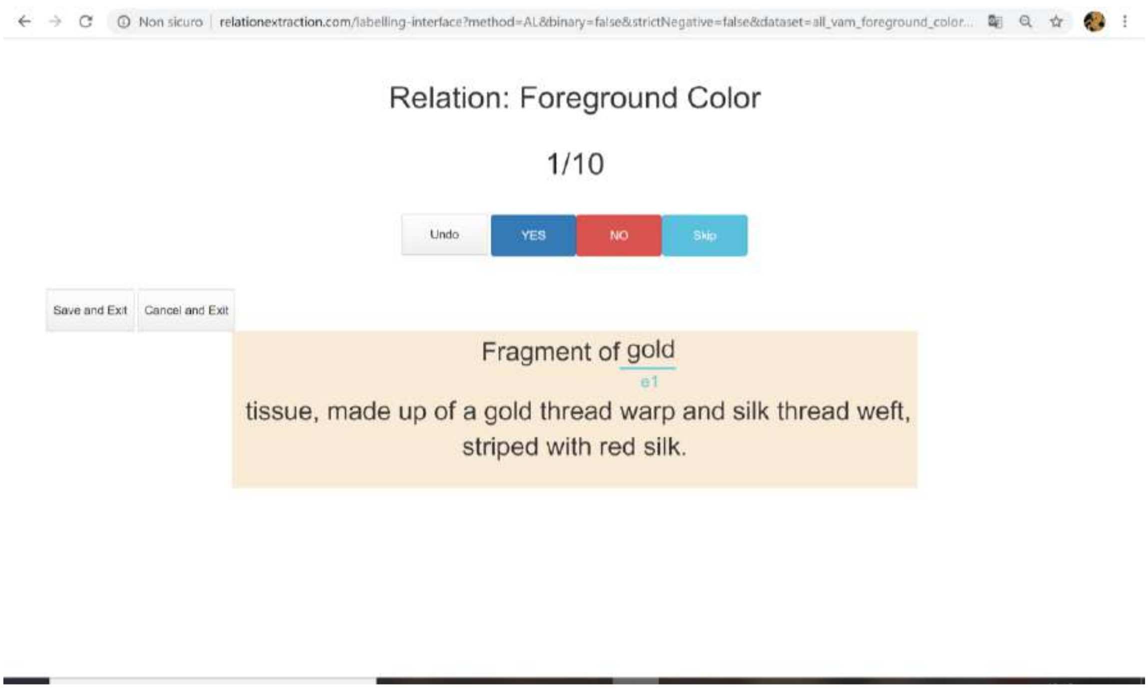



Figure 2. An example of a too short text.

As shown in Figure 3, sometimes we found descriptions of objects irrelevant to the silk and textile domain. The source database contained records of artefacts unrelated to textiles, such as a metal mourning ring. Therefore, it was difficult to understand whether the color highlighted by the web application was a foreground color, background color or border color. In this case, color refers to letters decorated with enamel, a technique that has nothing to do with fabric. However, the domain experts tried to do semantic annotation to train the system.

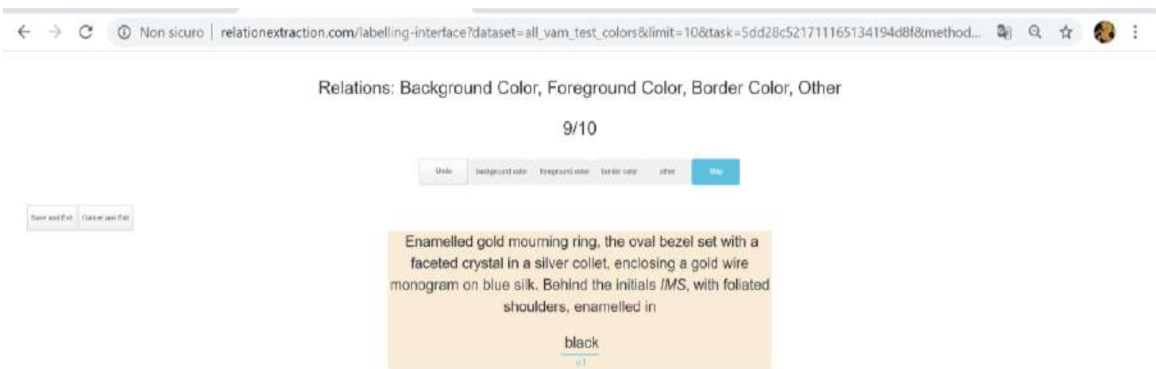



Figure 3. Unrelated record, since it represents a metal artifact.

Similar cases are shown in Figures 4 and 5. The term highlighted by the web application is a number that can be linked only to the label Other, or must be simply skipped, out of lack of context.

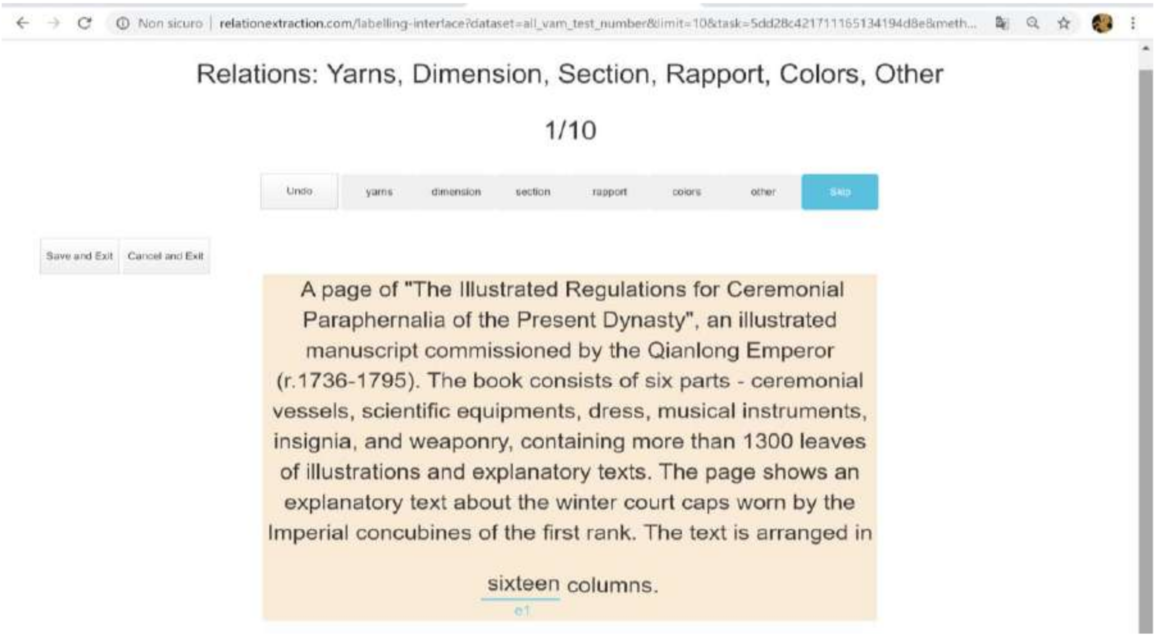



Figure 4. An example of text where the underlined number is describing an object that is not a fabric, but a printed page.

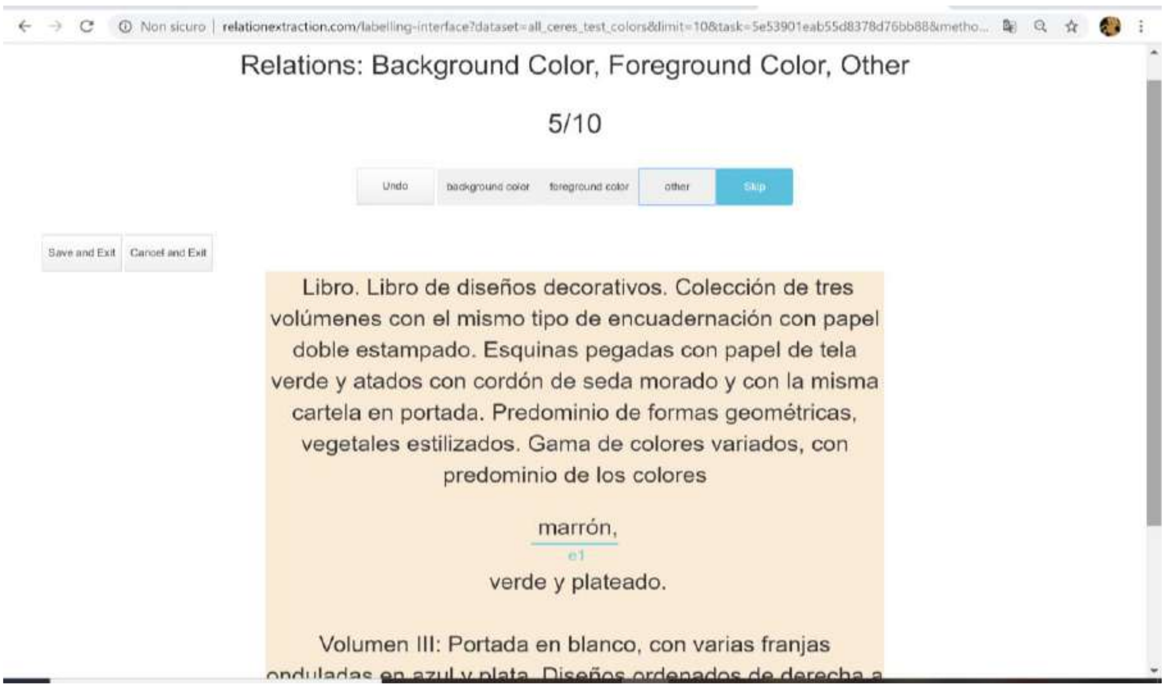



Figure 5. A color with an unclear label.

Beyond these practical problems, however, annotation by domain experts was developed in a useful and correct manner. Based on the initially annotated texts, we have formed two active learning models, one for each task, focusing on the most frequent label from the task: foreground color and the number of sections. The domain experts were then asked to annotate text excerpts for these two labels, answering yes/no.

We have represented texts with commonly used (Term Frequency Inverse Document Frequency) TFIDF word-vector representation and used a Support Vector Machine algorithm to train the models on English and Spanish texts. For English texts, the foreground color model was trained on 100 examples and tested on 200 examples. The model has achieved 65.5% classification accuracy and 78.4 F1 measure. For the number of sections, we have trained a model on 63 examples and tested on 139 examples, while the model for the number of sections has achieved 57.6% classification accuracy and 57.6 F1 measure.

For Spanish texts, the model was trained on almost 100 examples and tested on about 100 examples, achieving 75% classification accuracy and 47.0 F1 measure for the foreground color and 79% classification accuracy and 52.0 F1 measure for the number of sections. Analysing more in detail, we noticed that on Spanish texts precision was higher than recall, indicating that the model was correct when identifying positive examples in over 70% of cases but missed almost 40% of the positive examples.

# 5. KNOWLEDGE GRAPH VALIDATION

## 5.1. Objectives

This validation focused on the faceted browser (available at http://data.silknow.org/fct/) and followed two different rationales. The first validation activity aimed at verifying the coherence of what the knowledge graph presents, in comparison with the information stored in the records provided by the members of the consortium. The second one aimed at validating the correct interpretation of the Location attribute.

During the first validation activity, we involved domain experts from the UNIPA. Data providers in connection with this partner had provided several records with information about objects and textiles:

- Museo Diocesano di Caccamo
- Museo Diocesano di Palermo
- Museo di Termini Imerese
- Cattedrale di Palermo
- Duomo di Monreale

During previous tasks of the project, the records from UNIPA were imported into the knowledge graph based on the Data Model Definition described in D2.3.

This validation activity objective is to verify whether all the attributes present in the UNIPA records have been correctly reported in the Knowledge Graph and are available through the web faceted browser.

During the second validation activity, we involved domain experts from UNIPA to validate data related to the production location, i.e., the place of origin where the textile was made. Experts had to verify if all data inserted in this field were correct, whether they effectively represent the place of origin and not the place where they are located now, or any previous location.

## 5.2. **Method**

The method employed for the first validation activity consists in comparing the original database records to its correspondence in the web faceted browser. Figure 6 shows an example of an original record and Figure 7 the corresponding record in the web faceted browser.

We validated 29 records about objects and information from the above said museums. From the analysis of how the record in the faceted browser is structured against how the data providers have created the record, we generalized the following object identifiers to check:

- Construction and Technique from the original record are combined in the “P32 used general technique” entity inside the “P108 has Produced” entity. Since information about construction and technique are two different rows in the original records, how many of them have been combined?
- Some fields in the original records (Historical Critical Information, Warp, Weft, Width and Description of pattern) are combined in “P3 has note” in the SILKNOW data model. How many of them have been combined in the faceted browser?
- “crmsci:08_observed” put together several rows of the original records. Does this class contain all the rows, all previous information, for each record?
- Availability of information about the dimensions of the fabric (or the object at large). Is that information available, in the original record and in the faceted browser, through the class “P34 has dimension (Y/N)”?

Domain experts created a table where the existence or the lack of the identifier is reported along with some comments where necessary.

As regards the second validation, we navigated the Knowledge graph starting from “Man-Made Object” (E22 with reference to the ontology) and the related “P108 has produced entity”. We considered objects mainly coming from two museums namely the one labelled with the identifier 95.71.XXX and GP00XXX but also from a few others.

For each object, we checked the production page (by clicking on “P108 has produced entity”) and the related “P8 took place on or within” entity.

| | |
|---|---|
| Time chronology | Prima metà del XVIII secolo (1745 – 1750) |
| Geography | Italia |
| Region production | Sicilia o Campania |
| Description | Parato composto da pianeta, stola e manipolo |
| Technique | Taffetas broccato à liage rèpris |
| Museum | (PA). |
| Language | Italiano |
| Dimensions | cm 108x68; cm 216x22; cm 100x20 |
| State of preservation | discreto; locali slegature delle trame supplementari |
| Width | non rilevabile |
| Pattern unit | cm 48.5x n.r.; numero dei campi: 1(?); tipo di campo: a ritorno. |
| Warp | fili/cm |
| Weft | anima in seta ritorta, 2 capi, S, colore giallo; III oro filato avvolto su anima in seta, 2 capi, S, |
| Construction | trame broccate, metalliche e serich, fissate al dritto in diagonale 3 lega 1, direzione S, da ¼ dei |
| Embroidery | assente |
| Description of the pattern | composizioni vegetali caratterizzate da cespi di piccole foglie verdi che si affiancano ad altre |
| Galloon | strutture interne, cm 2, con motivo a ventaglio lungo lo scollo. |
| Lining | in tela di lino di colore azzurro |
| Historical Critical Information | disegnativo, ripropone una tipologia una tipologia decorativa diffusa verso la metà del XVIII |
| Images (names of the images in the documen | SCHEDA88.jpg; SCHEDA88.1.jpg |
| Author of the technical analysis | R. Civiletto |
| Author of the Historical Critical Information | M. Vitella |

Figure 6. The table reporting an example of the UNIPA record named Caccamo 7.

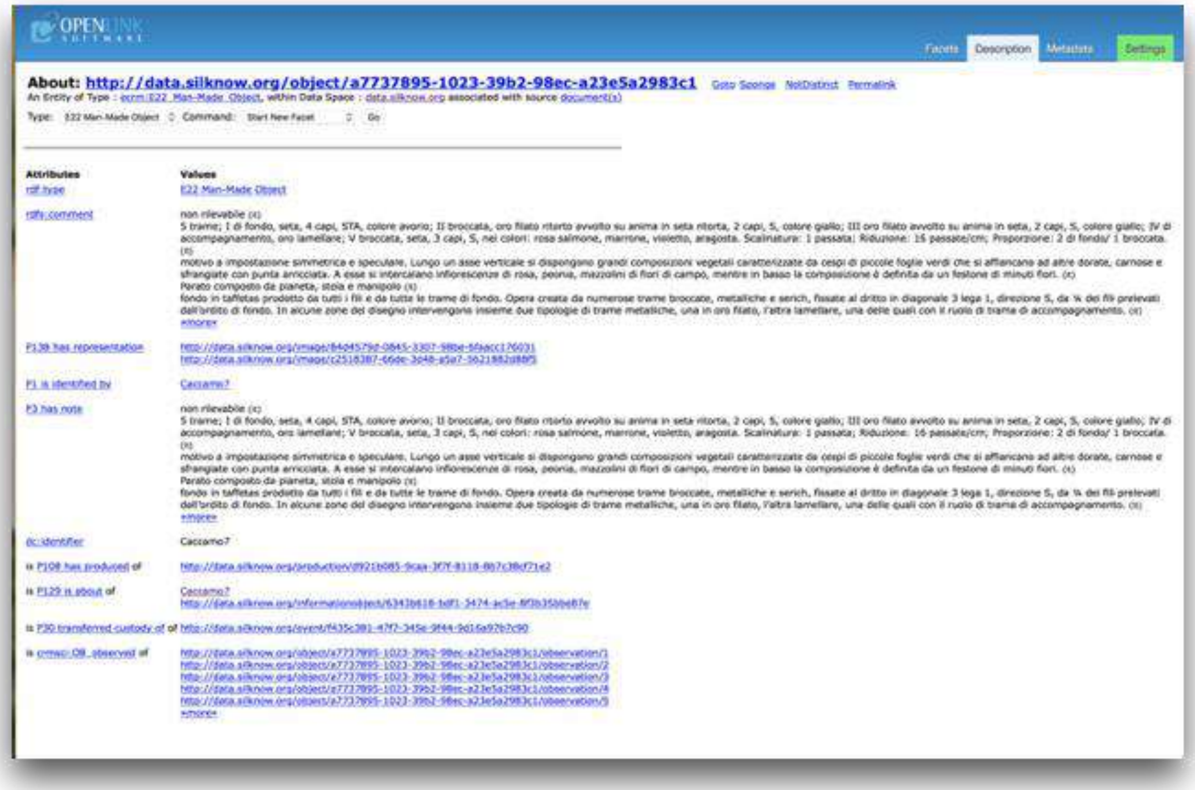



Figure 7. The web faceted browser for Caccamo 7 record.

## 5.3. Results

As regards the first validation, results are reported in Table 1 and can be summarized as follows:

- more than 50% of records in the faceted browserinformation about object dimensions in the field “P34 has dimension”; for instance, some records are related not to individual pieces but to sets of textiles (for instance Caccamo7 presents 3 pieces in a single record);
- in the 100% of records, construction and techniques have been adequately combined in “P32 used general technique”;
- 100% of historical critical information, warp, weft, width and description of a pattern has been adequately combined;
- 100% of the previous information are in the “crmsci:08_observed” entity;

All these remarks will be used to improve the search module that relies on the KG.

| | Validation#1 | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Obect identfier | Costruction and Tecnhnique are combined in the P32 used general technique | Historical critical Information, Warp, Weft, Widht, Description of pattern | In 08 observed each observation refers to one record og | PatternUnit, Lining, Galloon, Embroidery, State of preservation are lacking | P34 has dimension (Y/N) | | some typos | | |
| DiocesanoPA6 | y | y | y | y | y | | | | |
| Caccamo7 | y | y | y | y | N | 3 manufatti in uno stesso record | | | |
| CattedralePa2 | y | y | y | y | Dimensions only refer to Pianeta, dimensions of stola, manipolo, borsa | | | | |
| DiocesanoPA2 | y | y | y | y | y | | | | |
| Monreale1 | y | y | y | y | y | | | | |
| Caccamo6 | y | y | y | y | N | | | | |
| Caccamo1 | y | y | y | y | y | | | | |
| DiocesanoPa3 | y | y | y | y | y | | | | |
| DiocesanoPa8 | y | y | y | y | n | | | | |
| Terminiimerese4 | y | y | y | y | Dimensions only refer to one Pianeta, dimensions of the other three Pianetas,manipolo, velo di calice and | | | | |
| Caccamo8 | y | y | y | y | n | | | | |
| DiocesanoPA4 | y | y | y | y | y | | | | |
| Terminiimerese2 | y | y | y | y | Dimensions only refer to one Pianeta, dimensions of the other Pianetas, 3 manipolo, 1 velo di calice, 2 stolas, 1 | | | | |
| Caccamo4 | y | y | y | y | n | | | | |
| Terminiimerese6 | y | y | y | y | y | | | | |
| Caccamo2 | y | y | y | y | y | | | | |
| Caccamo3 | y | y | y | y | n | | | | |
| Caccamo5 | y | y | y | y | n | | | | |
| CattedralePa1 | y | y | y | y | Dimensions only refer to the Piaviale, dimensions of the Pianetas, manipolo, stola , borsa e palla lack | In P32used_genera_Technique the | | | |
| DiocesanoPA1 | y | y | y | y | y | | | | |
| DiocesanoPa7 | y | y | y | y | n | | | | |
| Monreale2 | y | y | y | y | y | | | | |
| CattedralePA3 | y | y | y | y | The dimension of one stola lacks | | | | |
| DiocesanoPA5 | y | y | y | y | n | | | | |
| Monreale3 | y | y | y | y | y | | : isolati | observatio 4 refuso >> | |
| Terminiimerese1 | y | y | y | y | only one dimension (wxh) instead of 4 | | | | |
| Terminiimerese5 | y | y | y | y | only one dimension (wxh) instead of 3 | | | | |
| Terminiimerese7 | y | y | y | y | y | | | | |
| Terminiimerese5 | y | y | y | y | y | | un e senza accento | | |

Table 1. Results of the first validation.

As regards the second validation, results are summarized as follows. Figure 8 shows that all records coming from Garin and owning the identifier T000877 present in the class “E2_production”, a value of “P8_took_place_on_or_within” equals to “chalet Garin”. This value cannot be associated with the production place, but with the current storage location.

For records mapped from RISD, Figure 9 shows that the description of the object does not allow the expert to understand whether the value of “P8_took_place_on” is the right place of origin. Several records, like the one shown in Figure 10 do not present the value of “P8_took_place_on”, so they cannot be evaluated.

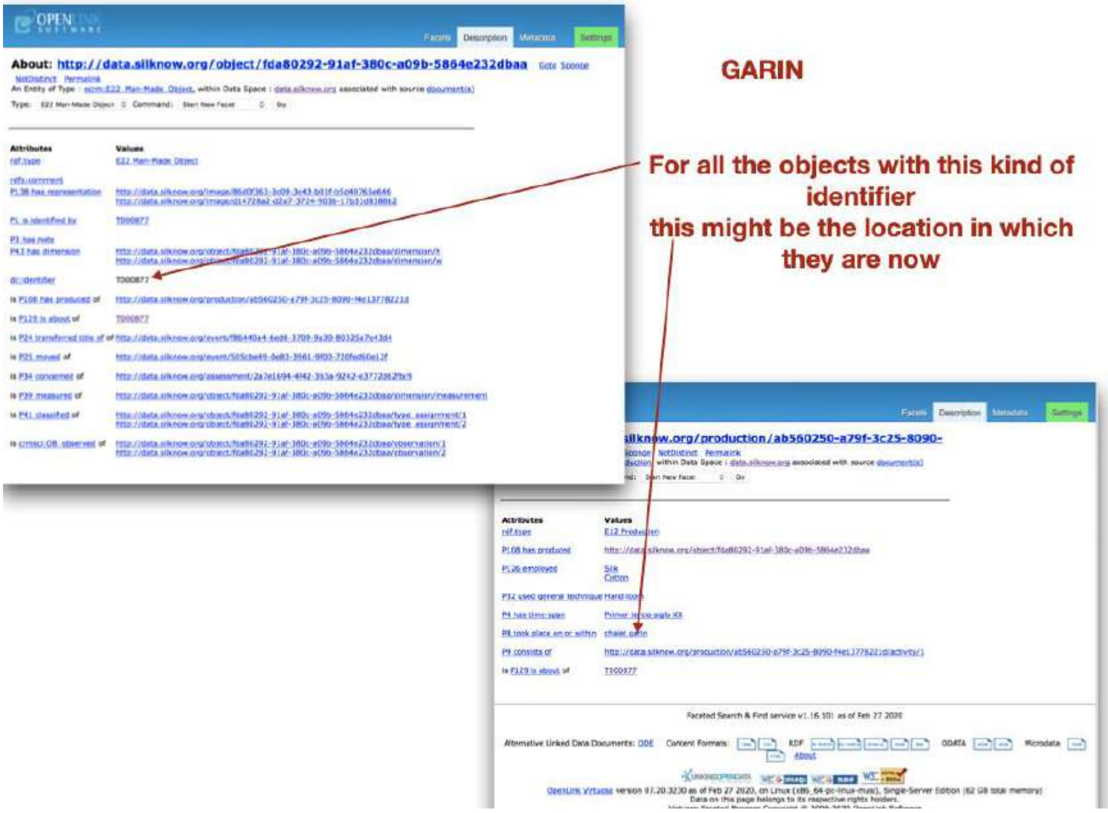



Figure 8. First example – records from GARIN.

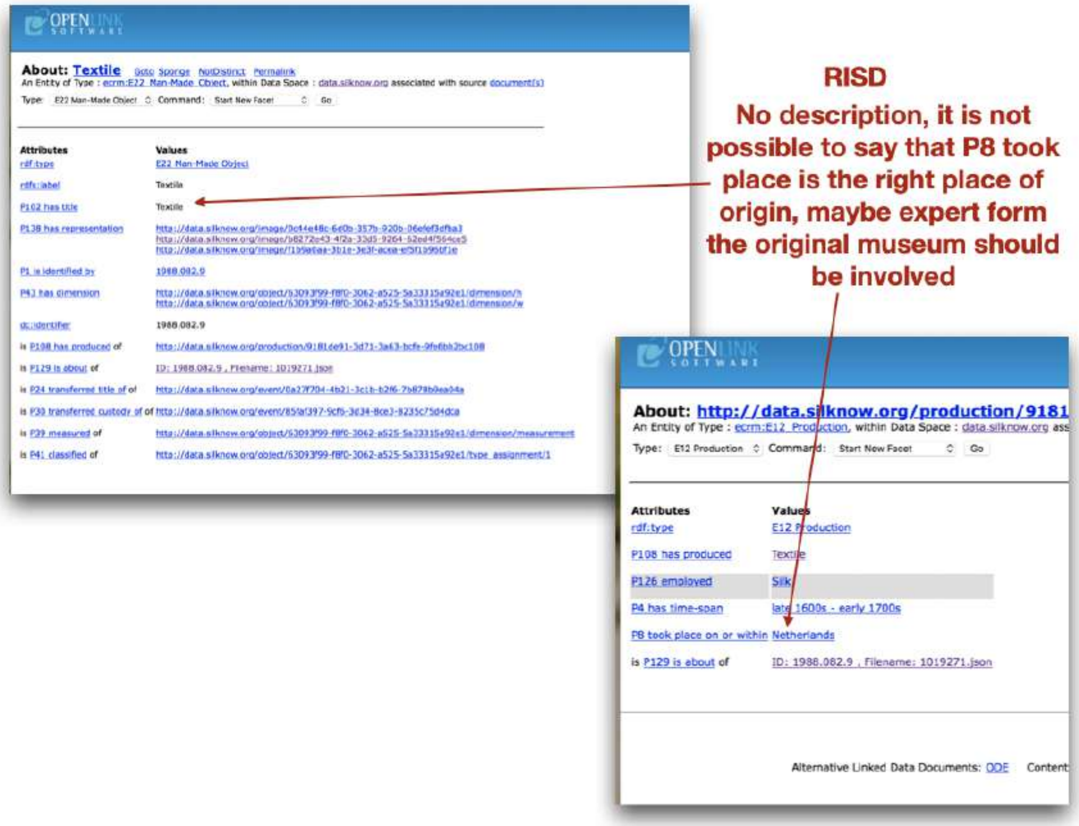



Figure 9. Second example - records from RISD

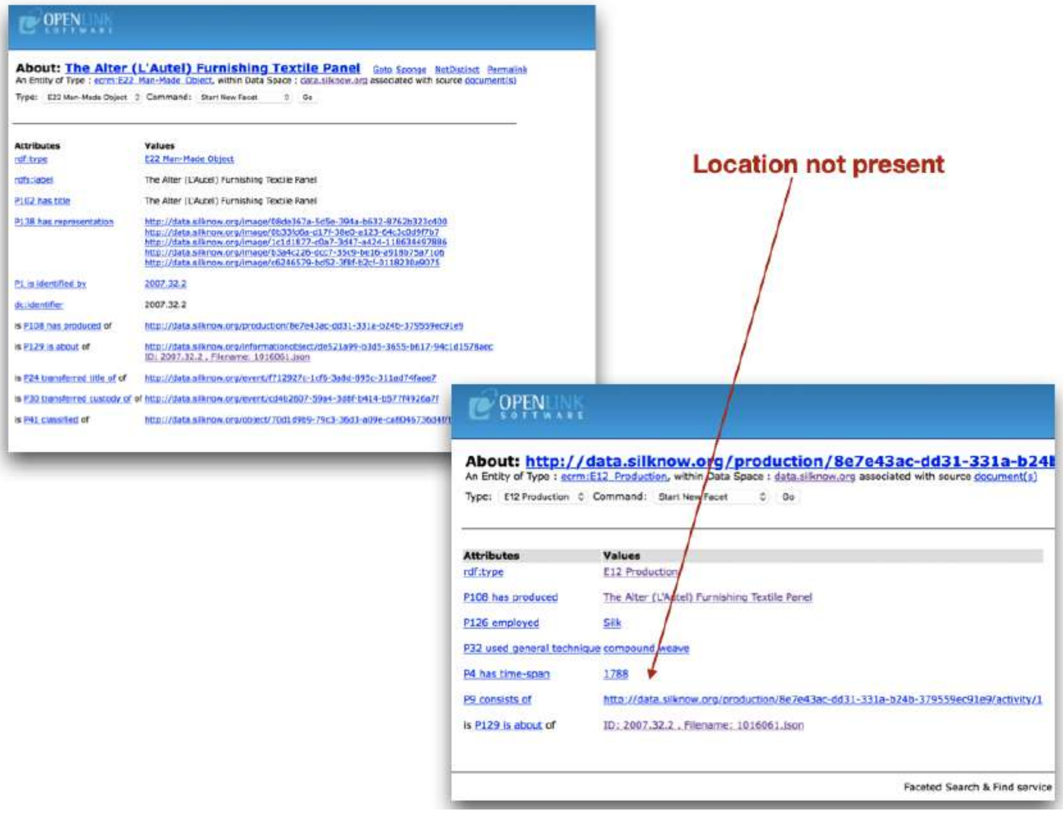



Figure 10. An example of record where location has not been reported.

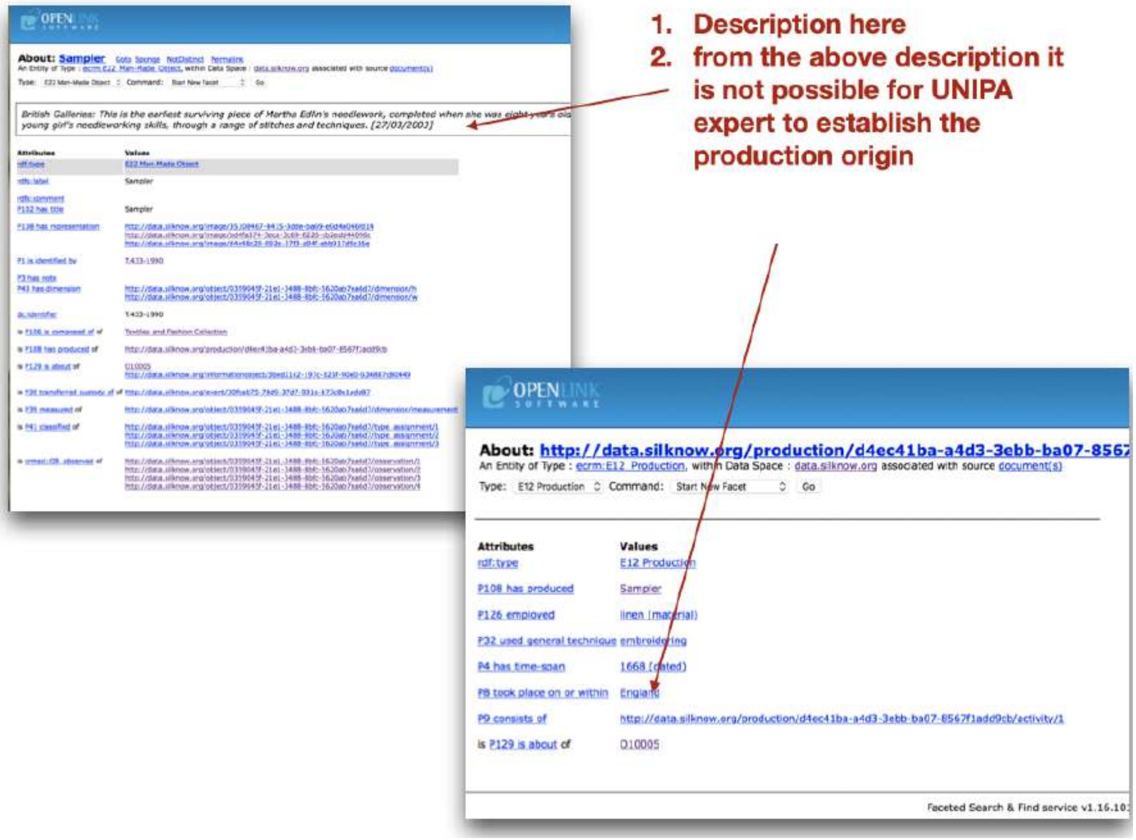



Figure 11. An example of a too short description

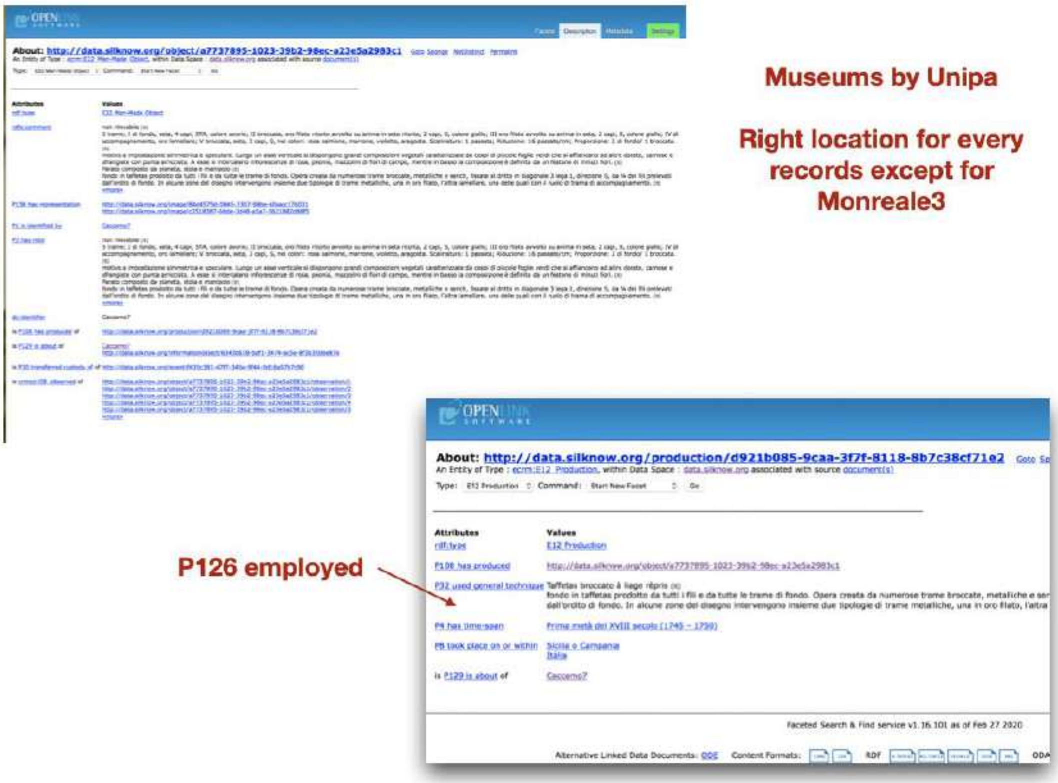



Figure 12. An example of record from which the right location may be inferred.

Some records (see Figure 11) present a description of the objects that do not allow the expert to establish if the value in “P8_took_place_on” is the place of origin. More details in the description are needed. Most records from Italian museums (see Figure 12) present the right location with the exception of the ones from Monreale.

## 6. IMAGE RETRIEVAL VALIDATION

### 6.1. **Objectives**

As mentioned above (see section 2.3), the evaluation of the image retrieval module has focused on trying to establish whether the image retrieval algorithm gives results, regarding the similarity of two images, that are meaningful for the experts.

LUH provided metrics and a technical evaluation of the algorithm based on the application of the image retrieval method as a basis for classifying the images (deliverable D4.5). Thus, in Task 7.1, there is no further need for a numerical evaluation of the method’s potential to support a classification. In contrast, it is more important to establish the quality of the results using the expert and common users’ habits.

## 6.2. **Method**

The method employed was adopted to determine image similarity between two textiles, from a domain expert point of view. As an initial, working hypothesis, we decided to use the following concepts or properties of the textiles:

- Pattern - this refers to the decorative motifs, for instance birds, stripes, flowers, garlands… The meaning of this concept regards the potential desire of the user to find fabrics whose images present similarities in this formal or iconographic aspect.
- Color - it is a direct visual feature of an image, from a very basic point of view, maybe the most common similarity parameter perceivable for every person. Besides, it has great historical relevance, in most cases.
- Appearance - using this term domain experts refer to the overall aspect of the fabric represented in the image. It also includes disposition of the decorative elements, outline and shape, geometric form, etc. This is also a characteristic commonly recognized by any average user.

These concepts have no direct counterpart with the variables from the knowledge graph that were used by the algorithm. They would be related to "subject depicted type", but for reasons given below this variable was not considered in the training process. Anyway, we decided to perform this first validation employing elements closer to the most frequent users, the textile expert domain and the cultural heritage professional. For instance, the identification of Pattern and Color had already been proposed in the end users' profiles and functional requirements outlined in task 2.4 (cf. D2.4).

To carry out the evaluation, LUH has provided the domain experts with two sets of image retrieval results related to two different CNNs and testing environments:

- imatex - training the CNN with all but 100 randomly picked images just from the IMATEX database;
- all_museums - training the CNN with all but 100 randomly picked images just from all museum datasets in the knowledge graph (garin, imatex, joconde, mad, mfa, risd - status 10th of April 2020).

Thus, we used two different CNNs, one where the records are relatively homogeneous in terms of the annotations and the appearance of the fabrics in the images, because they were all from the same dataset (imatex), and one with a more heterogeneous origin (all museums). In both cases, the annotations for the semantic variables "production place", "production timespan", "material" and "technique or procedure" were used to define similarity. The variable "subject depicted type" was not used to train the CNNs, since at the time when the results were produced, the annotations for that variable in the knowledge graph were still very heterogeneous.

For each of these two datasets (i.e., imatex and all museums), the results of the image search were delivered in a folder named "similar_image_folder". This folder contains one subfolder for each of the 100 query images of the corresponding dataset. The folder name is identical to the SILKNOW knowledge graph object URI of the record corresponding to the query image. The results of image retrieval are stored in a subfolder called "10_most_similar_images".

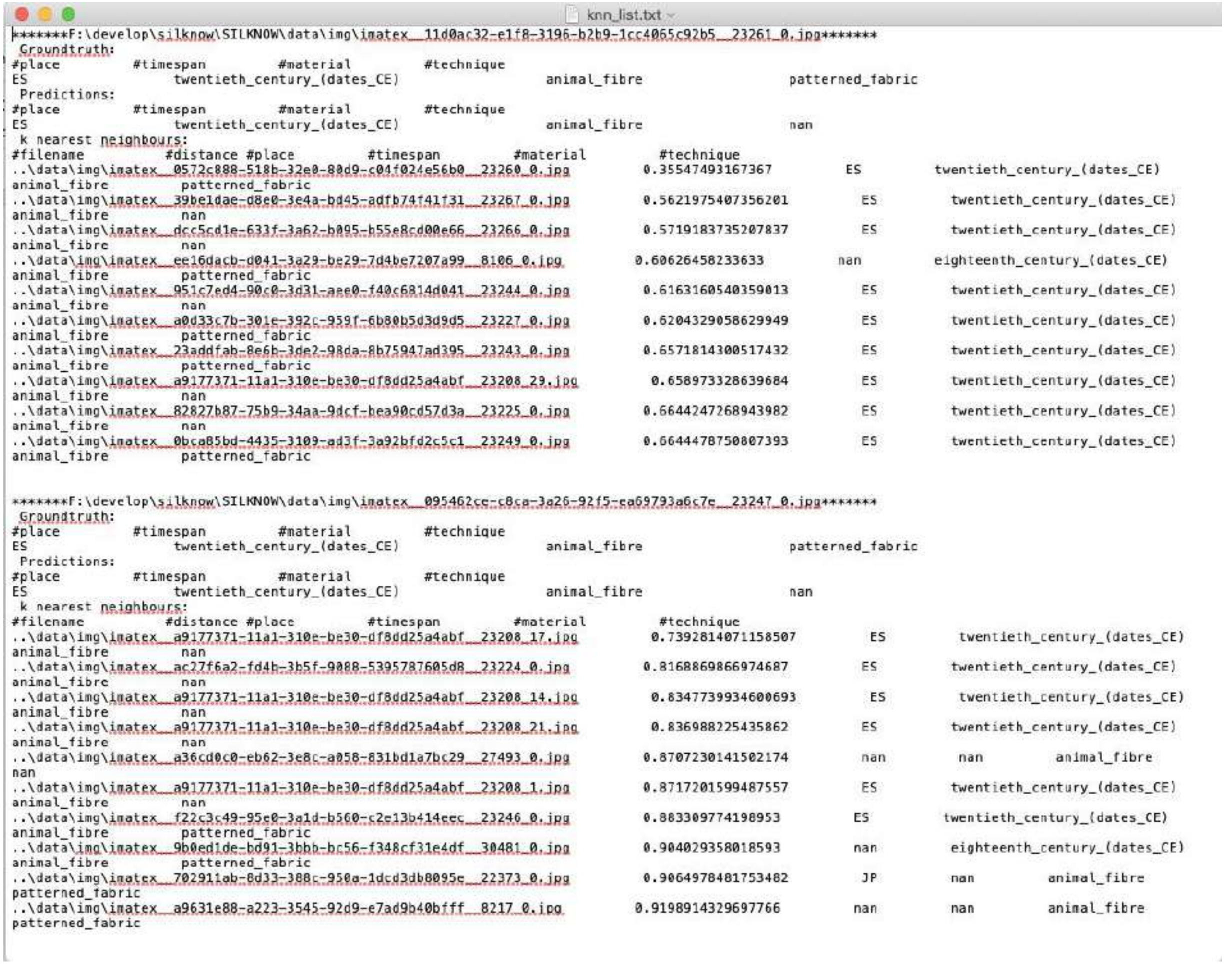
It contains the 10 images considered to be most similar to the query images delivered by the image retrieval software. The 10 images can be ordered by the matching score, which is the beginning of each image name. Further, each image name contains the museum where it is from, the object URI and the actual image name. Thus, each image file name follows the scheme: "matchingscore_museum__URI_name.jpg".

In addition, the information about the annotations (both reference and predicted values) is provided in a file "10_nearest_neighbours_list.txt" for every query image in the same folder as that image. This information is listed in a compact form for all query images in a file "knn_list.txt" for every dataset (see Figure 18) .

The final objective of the evaluation is to know whether the retrieved images are related to the search image from an expert's viewpoint. In case some images may not match the search image, the experts have been requested to indicate:

- Which images exactly do not match (image name)?
- What is the reason for that decision?
- What are the criteria for validating the retrieval results?

In order to answer these questions, the experts arranged a table with six columns: the name of the test image, the names of the ten most similar images, pattern, colour, appearance and the column for indicating whether the 10 most similar images match the test image (see Figure 13).



```
knn_list.txt
*******F:\develop\silknow\SILKNOW\data\img\imatex__11d0ac32-e1f8-3196-b2b9-1cc4065c92b5__23261_0.jpg*******
 Groundtruth:
#place        #timespan         #material         #technique
ES               twentieth_century_(dates_CE)                   animal_fibre                      patterned_fabric
 Predictions:
#place        #timespan         #material         #technique
ES               twentieth_century_(dates_CE)                   animal_fibre                      nan
 k nearest neighbours:
#filename            #distance #place         #timespan         #material          #technique
..\data\img\imatex__0572c888-518b-32e0-80d9-c04f024e56b0__23260_0.jpg          0.35547493167367          ES        twentieth_century_(dates_CE)
animal_fibre          patterned_fabric
..\data\img\imatex__39be1dae-d8e0-3e4a-bd45-adfb74f41f31__23267_0.jpg          0.5621975407356201          ES        twentieth_century_(dates_CE)
animal_fibre          nan
..\data\img\imatex__dcc5cd1e-633f-3a62-b095-b55e8cd00e66__23266_0.jpg          0.5719183735207837          ES        twentieth_century_(dates_CE)
animal_fibre          nan
..\data\img\imatex__ee16dacb-d041-3a29-be29-7d4be7207a99__8106_0.jpg          0.60626458233633          nan        eighteenth_century_(dates_CE)
animal_fibre          patterned_fabric
..\data\img\imatex__951c7ed4-90c0-3d31-aee0-f40c6814d041__23244_0.jpg          0.6163160540359013          ES        twentieth_century_(dates_CE)
animal_fibre          nan
..\data\img\imatex__a0d33c7b-301e-392c-959f-6b80b5d3d9d5__23227_0.jpg          0.6204329058629949          ES        twentieth_century_(dates_CE)
animal_fibre          patterned_fabric
..\data\img\imatex__23addfab-8e6b-3de2-98da-8b75947ad395__23243_0.jpg          0.6571814300517432          ES        twentieth_century_(dates_CE)
animal_fibre          patterned_fabric
..\data\img\imatex__a9177371-11a1-310e-be30-df8dd25a4abf__23208_29.jpg          0.658973328639684          ES        twentieth_century_(dates_CE)
animal_fibre          nan
..\data\img\imatex__82827b87-75b9-34aa-9dcf-bea90cd57d3a__23225_0.jpg          0.6644247268943982          ES        twentieth_century_(dates_CE)
animal_fibre          nan
..\data\img\imatex__0bca85bd-4435-3109-ad3f-3a92bfd2c5c1__23249_0.jpg          0.6644478750807393          ES        twentieth_century_(dates_CE)
animal_fibre          patterned_fabric

*******F:\develop\silknow\SILKNOW\data\img\imatex__095462ce-c8ca-3a26-92f5-ea69793a6c7e__23247_0.jpg*******
 Groundtruth:
#place        #timespan         #material         #technique
ES               twentieth_century_(dates_CE)                   animal_fibre                      patterned_fabric
 Predictions:
#place        #timespan         #material         #technique
ES               twentieth_century_(dates_CE)                   animal_fibre                      nan
 k nearest neighbours:
#filename            #distance #place         #timespan         #material          #technique
..\data\img\imatex__a9177371-11a1-310e-be30-df8dd25a4abf__23208_17.jpg          0.7392814071158507          ES        twentieth_century_(dates_CE)
animal_fibre          nan
..\data\img\imatex__ac27f6a2-fd4b-3b5f-9088-5395787605d8__23224_0.jpg          0.8168869866974687          ES        twentieth_century_(dates_CE)
animal_fibre          nan
..\data\img\imatex__a9177371-11a1-310e-be30-df8dd25a4abf__23208_14.jpg          0.8347739934600693          ES        twentieth_century_(dates_CE)
animal_fibre          nan
..\data\img\imatex__a9177371-11a1-310e-be30-df8dd25a4abf__23208_21.jpg          0.836988225435862          ES        twentieth_century_(dates_CE)
animal_fibre          nan
..\data\img\imatex__a36cd0c0-eb62-3e8c-a058-831bd1a7bc29__27493_0.jpg          0.8707230141502174          nan        nan        animal_fibre
nan
..\data\img\imatex__a9177371-11a1-310e-be30-df8dd25a4abf__23208_1.jpg          0.8717201599487557          ES        twentieth_century_(dates_CE)
animal_fibre          nan
..\data\img\imatex__f22c3c49-95e0-3a1d-b560-c2e13b414eec__23246_0.jpg          0.883309774198953          ES        twentieth_century_(dates_CE)
animal_fibre          patterned_fabric
..\data\img\imatex__9b0ed1de-bd91-3bbb-bc56-f348cf31e4df__30481_0.jpg          0.904029358018593          nan        eighteenth_century_(dates_CE)
animal_fibre          patterned_fabric
..\data\img\imatex__702911ab-8d33-388c-950a-1dcd3db8095e__22373_0.jpg          0.9064978481753482          JP        nan        animal_fibre
patterned_fabric
..\data\img\imatex__a9631e88-a223-3545-92d9-e7ad9b40bfff__8217_0.jpg          0.9198914329697766          nan        nan        animal_fibre
patterned_fabric
```

Figure 13. An excerpt of knn_list file for Imatex.

The criterion used was: if at least two properties among pattern, color and appearance match between two images, then they may be considered similar.

### 6.3. **Results**

The results of the validation performed by domain experts, although performed on a limited number of images, helped to give a general overview of the system.

The analysis revealed that in some cases the retrieved “10_most_similar_images” were too similar to each other. Indeed, the system had collected images of the same piece of fabric viewed from different angles or images of very similar fabrics. In other cases, however, the retrieved images fulfilled only one or none of the requirements considered relevant by the domain experts (i.e. pattern, color or general appearance.)

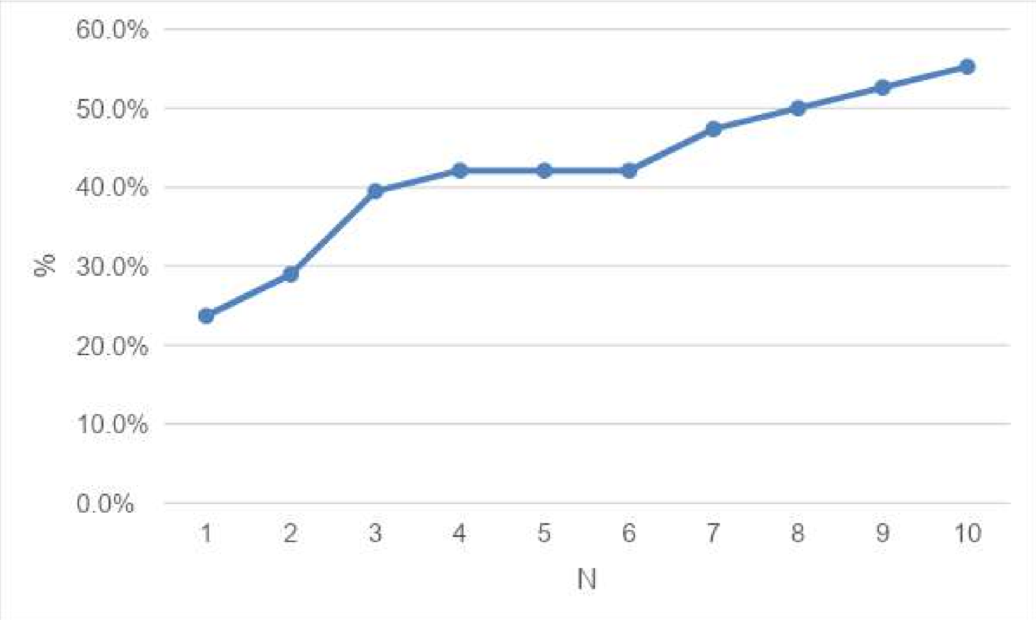



Figure 14. Percentage of query images for which the image retrieval software delivered at least one similar image. Basis: domain expert evaluation based on 38 query images from the IMATEX dataset.

More precisely, a statistical analysis revealed that in 55.7% of the inspected cases there was at least one meaningful matching record among the 10 most similar images retrieved from in the database. Figure 14 shows the percentage of query images for which there was at least one similar image (according to the domain experts' criteria) among the first N retrieved images. The figure also shows that only in 23% of the cases, the most similar image according to the software was a match. There was no case in which all retrieved images matched the domain experts' criteria; the maximum amount of similar images retrieved for a specific query image was 7 (1 case only). Whereas these results are far from perfect, they indicate that a database search via an image only can deliver at least one meaningful result in more than half of the cases.

Figure 15 shows an example of a test image and Figure 16, 17 and 18 belong to the 10_most_similar_images retrieved by the module. They respectively refer to lines 13,15 and 16 in the table of results shown in Figure 19. As can be seen, the first image has been considered similar since all three criteria match. The second image matches only for one criterion whereas the third image does not present any criteria of similarity from the point of view of the domain experts.

Figure 15. imatex__050efea2-8089-339b-9dae-e5c4ac5fa3a6__8332_0.jpg – Test image

Figure 16. 0.7838_imatex__5aed1e6b-90fd-3ee9-a168-bb678538c74a__7474_0.jpg – Image completely matching.

Figure 17. 0.7953_imatex__aee306f6-eceb-34bc-8464-faf66816ea69__20175_0.jpg – Image partially matching.

Figure 18. 0.8099_imatex__7a8fb350-298d-3398-bb3f-051f494cb4bd__7816_0.jpg - Image not matching.

| | A | B | C | D | E | F |
|---|---|---|---|---|---|---|
| 1 | | | | | | |
| 2 | Test image | Retrieved images | Pattern | Colour | Appearance | Macthes |
| 3 | *imatex__9b200f0f-cca7-30c4-a296-38c4c4ad7a7d__8796_0.jpg* | 0.0602_garin__16de931b-eb9f-35eb-8dd8-408dccf72d...T000085_REV.jpg | | | | N |
| 4 | | 0.0637_garin__2fd133b8-43dc-38d4-9e06-060e6bb3e2...T000194_ANV.jpg | | | | N |
| 5 | | 0.0638_garin__56d19edb-608e-3456-bc8e-fe78903712...T000535_REV.jpg | | X | | N |
| 6 | | 0.0649_garin__dd94b714-22c7-39b2-a8de-eda16ba9e...T000488_ANV.jpg | | | | N |
| 7 | | 0.0649_risd__e9d0a81d-359d-3c8b-8129-e735b4251eef__1072951_0.jpg | | | | N |
| 8 | | 0.0650_garin__2d794629-5772-38f9-9717-015a913b02...T000589_ANV.jpg | | | | N |
| 9 | | 0.0651_garin__b708c72c-5d31-32f9-9767-0de95e5c43f...T000106_REV.jpg | | X | | N |
| 10 | | 0.0652_garin__182b8218-342e-367c-b285-de12c346fee...T000152_ANV.jpg | | | | N |
| 11 | | 0.0652_risd__12466010-6489-3686-8ccc-156c30aeb5a5__1118376_0.jpg | | | | N |
| 12 | | 0.0657_garin__1371396f-e5d2-30c8-8003-bcf335168aa...T001104_REV.jpg | | X | | N |
| 13 | *imatex__050efea2-8089-339b-9dae-e5c4a6__8332_0c5fa3a* | 0.7838_imatex__5aed1e6b-90fd-3ee9-a168-bb678538c74a__7474_0 | X | X | X | Y |
| 14 | | 0.7860_imatex__77fcc9d8-6ae9-3b11-b60f-daeb0a9cd6d4__8211_0 | X | X | X | Y |
| 15 | | 0.7953_imatex__aee306f6-eceb-34bc-8464-faf66816ea69__20175_0 | | | X | N |
| 16 | | 0.8099_imatex__7a8fb350-298d-3398-bb3f-051f494cb4bd__7816_0 | | | | N |
| 17 | | 0.8099_imatex__73fb7896-71fd-3b82-b2f7-0a9eb06a45b1__7848_0 | | | | N |
| 18 | | 0.8212_imatex__5ce75ea6-8770-3846-bd22-7c35f1d69d5c__20209_0 | | | X | N |
| 19 | | 0.8258_imatex__dc78cac2-45b8-3dfc-aa93-9caff68708ee__29564_0 | | | X | N |
| 20 | | 0.8295_imatex__ed269b01-9421-39c9-89ac-bb29d51663db__20872_2 | | | X | N |
| 21 | | 0.8350_imatex__ed269b01-9421-39c9-89ac-bb29d51663db__20872_1 | | | X | N |
| 22 | | 0.8363_imatex__25523484-5bc6-392b-894e-6ba29bf9f532__20144_0 | | | | N |
| 23 | imatex__51bcd373-2326-3209-94d3-04b415fa925e__3431_0 | 0.5114_imatex__014377fb-b1ea-35b3-89cf-e4109cff4d6c__27947_0 | | X | | N |
| 24 | | 0.5153_imatex__34ee1479-f9ce-3d06-a635-3b7527ca5bd7__4437_0 | | X | X | Y |
| 25 | | 0.5201_imatex__cbba34fd-d95e-3923-a2cb-3aeba78a1a2d__4179_0 | | | | N |
| 26 | | 0.5246_imatex__405b9ae9-8735-3a9b-bdff-36a5aabc3e92__28036_0 | | | X | N |
| 27 | | 0.5287_imatex__ddfef515-4a69-3361-951a-7e0fa53e0e6d__33170_0 | | | X | N |
| 28 | | 0.5398_imatex__b244ff42-287a-31df-9a9f-92e3e0503d0c__28030_0 | | | X | N |
| 29 | | 0.5404_imatex__faeb16f6-d759-30b0-90be-b4128a41e01b__3784_0 | | | | N |
| 30 | | 0.5411_imatex__1e72f2fb-5563-393d-8fab-7228f2ce58d5__5085_0 | | X | | N |
| 31 | | 0.5426_imatex__7237f361-8bee-3f9a-9f24-ba6e1818c406__24401_0 | | | | N |
| 32 | | 0.5469_imatex__872bced7-f7b6-31ce-a93e-9029537ab129__29551_0 | | X | X | N |

Figure 19. An excerpt of the table reporting the results about image retrieval.

The algorithm is based on the assumption that the more semantic properties match the more the images are similar. However, the evaluation criteria of the domain experts did not take into account the semantic properties used to train the algorithm. Given this fact, it does not come as a surprise that these very first results are not perfect yet, with about 45% of query images for which none of the retrieved images was considered to be similar by the domain experts. As a consequence, the partners are refining their working definition of similarity for training the image retrieval module. This discussion has led domain experts to reconsider the relevance of certain properties for similarity (e.g. the subject depicted is more important than production place or time). It has also led to the definition of some rules by the domain experts that are being incorporated in new developments. Furthermore, some more visual features derived from an analysis of the images in color space will also be integrated.

In the nearest future these results will be used to refine the image retrieval algorithm. This will mainly be done by considering additional terms in the loss function for training the networks beyond those used already in D4.5, with the main goal to make the machine converge to a concept of similarity closer to the one presumably used by an end user.

## 7. COMMENTS BY THE DOMAIN EXPERTS

In this section, we report all considerations made by the domain expert performing the validation. It is worth noting that the experts own skills in the domain of Cultural Heritage (textiles and so on) so their comments do not include ICT technical issues. These comments may be used, eventually, from ICT partners for improving all the software modules.

As regards the validation of the Text Analytic Module, domain experts have been actively involved in the training of the software through the use of the web service http://relationextraction.com/.

Regarding the validation of the Image retrieval software, the images submitted to the domain experts were grouped according to 5 semantic variables that undoubtedly represent key properties of the textile product. They are: the historical period of production (timespan); the Place of production (place); the Material (material); the Technique or procedure (technique); the Type of subject depicted (depiction). However, the depiction was not used in the training process (cf. section 6.2).

The images representing the silk fabric selected by the software had one or more of these characteristics in common, which should have led to a certain visual similarity. Unfortunately, the analysis by the domain experts revealed that the presence of one or more of these variables is not sufficient to select similar images in many cases. For instance, as regards the Timespan, it is a well-known fact that in any historical period fabrics were produced with different techniques and weaves, as well as decorative modules with significantly varying subjects: flowers, birds, geometric shapes of differing ratios, etc.

The same can be done for the place of production, which is often not even mentioned in the catalog records, because it is difficult to determine with certainty. When it is present, however, it cannot be an indication of a univocal formal feature, because products woven in the same place could be produced with different formal characteristics. The production material is an important feature in the artistic field in general, but the research area of the project is focused on silk fabrics, so it is quite obvious that the images selected by the software represent silk artefacts. On the other hand, the production technique is a good discriminating factor for similar fabrics; however, these can have widely dissimilar decorative motifs.

Furthermore, the recognition of a textile weft by the domain expert from a photo is something that a domain expert can only do if the image allows a very close view of the fabric. In the field of textile experts, in fact, we use a lenticular of the size of one square centimeter, with a magnification of 10x, that allows to “read” the construction of the weft and the warp.

Finally, another feature taken into consideration was the depicted subject (depiction). This characteristic opens up an interesting insight into the perceptive approach that every expert in the domain, as a human being, assumes in front of the patterns present on a fabric. Features listed above are extrapolated from the catalog records incorporated into the knowledge graph, records produced by different individuals, who will have a subjective perception and way of expressing themselves. Each cataloguer will have his or her own style in the production of the records, even though they are frequently conducted according to a guide, such as the grid provided by the CIETA (Centre International d’Étude des Textiles Anciens) cataloging methodology [9][10]. It is almost never possible to give a univocal description of a textile design. The simplest example is to describe the fabrics called 'bizarre', produced in the early eighteenth century or even the damasks of the sixteenth century. They are fabrics loaded with decorative elements such as flowers, leaves or imaginative and abstract elements that each cataloguer can interpret differently.

This already creates interpretative problems that, transferred into an algorithm, determine the lack of a real affinity between fabrics that just because they share an element, such as a flower or a geometric shape, would be similar. One problem in this regard was the variable patterned_fabric, because it could not represent a significant element for discernment, as most fabrics have a decorative pattern. The quality of the parameters chosen for the selection of the images is therefore underlined, but these must exist at the same time in order to have a greater possibility that the selected images are of similar artefacts and in any case remain insufficient to give a certainty of similarity. Variables such as color, the ratio of the design, or the weaving technique, make the appearance of one fabric different from another and are important for human perception.

Domain experts were also asked to evaluate the accuracy of the class Production place, as reported in the records. It was also asked whether the Production place coincided with the place of conservation of the artefacts. These two classes are very important to create a geographical map of textile production.

The class Production place is not present in all the records, as it is very difficult for cataloguers to identify it with certainty. As a matter of fact, in textile products some information such as the techniques used or the nature of the materials can indicate the possible origin in a territory, but they cannot define it with certainty. This is because the same techniques were replicated in different territories with imported materials. Every artefact that met the public's taste was replicated throughout Europe by the most expert weavers, or the craftsmen who were able to make it were requested.

Moreover, just by checking the class Production place, it was found that the records in the various databases collected were filled in differently, because they are filled in by different people with different methods. In fact, some information about the artifacts usually lies in the body of the text of the record -often, in a plain text, general description field- and not in the classes.

## 8. CONCLUSION AND FUTURE WORK

In this report we have described the work done in Task 7.1 on the validation of three important software modules underlying the SILKNOW system: text analysis, image retrieval and knowledge graph. The first two employ machine learning techniques for allowing the system to learn and predict specific features of the object coming from museum collections. The knowledge graph offers the possibility to search for useful information linked to ontology.

Providing a first validation of the functionalities of these modules was a very important step for guaranteeing the quality of the whole system that will be integrated. Validation performed in Task 7.1 was a first step towards a more complete validation of the modules and the software that will be performed in Task 7.2.

We obtained results in terms of improving semantic annotation by means of the domain experts' knowledge, fixing some mapping problems in the knowledge graph and laying the foundations for improving image retrieval algorithms. For instance, it has to be noted that there is no search engine (such as, e.g., Google) that will deliver only matching records for a query; in general, a user might already be content to obtain just a few meaningful matches. At least one similar image (out of 10) was retrieved for the majority of search images, but the number of query images with more than two matches is very low and there were also 45% of the images without any match.

These results will be strongly used during Task 7.2, with additional feedback from ICT experts on the comments reported in section 7 for carrying out a new iteration with the domain experts and ensuring the convergence towards a high-quality integrated system. In so doing, one of the most important and urgent issues is the incorporation of the depicted subject (depiction).

## REFERENCES


[1] *1012-2016 IEEE Standard for System, Software and Hardware Verification and Validation.* 2016. doi:10.1109/IEEESTD.2017.8055462
[2] I. Sommerville. *Software Engineering 10*. Harlow: Pearson Education Limited, 2016.
[3] As presented in the project's Description of the Action, part B, pages 15-16.
[4] SILKNOW Deliverable 3.3 - Design and implementation of text analytic module
[5] SILKNOW Deliverable 3.2 - Design of the SILKNOW Ontology and the Ontology
[6] SILKNOW Deliverable 4.4 – Design and implementation of deep learning based image classification system.
[7] SILKNOW Deliverable 4.5 – – Design and implementation of image retrieval method based on Siamese CNN
[8] SILKNOW Deliverable 6.5 - Integrated System
[9] https://cieta.fr/
[10] As explained, for example, in D. Devoti, *Premessa alla Mostra in Tessuti antichi nelle chiese di Arona*, 1981, Torino.

Universitat de València

Università degli Studi di Palermo

Leibniz Universität Hannover

MONKEYFAB

Institut "Jožef Stefan" Ljubljana, Slovenija

Instituto Cervantes

cnrs

EURECOM

GARÍN 1820