

Producing accessible, interoperable and reusable Cultural Heritage data with the SILKNOW ontology to preserve the European Silk Heritage

Silk played an important role in European history, mostly along the Western Silk Road's network of production and market centres. Silk, however, has become a seriously endangered heritage. Although many European specialized museums are devoted to its preservation, they usually lack size and resources to establish networks or connections with other collections. The H2020 SILKNOW project (Silk heritage in the Knowledge Society: from punched card to Big Data, Deep Learning and visual/tangible simulations)¹ aims to produce an intelligent computational system in order to improve our understanding of European silk heritage.

The short paper will present:

- The SILKNOW platform that will form a coherently integrated system to give access to a wide variety of data describing silk-related objects to researchers, museums curators or general public with a single interface ;
- The definition of the SILKNOW ontology, based on the CIDOC Conceptual Reference Model (CIDOC-CRM) and the specification of the entities to be represented by the ontology and the existing relationships between these entities ;
- The process of elaboration of new CRM classes and properties to describe silk textiles' data (and more generally textiles' data).

SILKNOW has crawled datasets from websites or online databases of Cultural Heritage Institutions (CHIs) preserving silk-related artefacts - such as the Musée des Tissus de Lyon (via the Joconde² database), the Victoria and Albert Museum³ or the Museos estatales del MEC⁴. The SILKNOW crawler is made in Node.js and its source code is available at: <https://github.com/SILKNOW/crawler/>. We have then analyzed the structure of the records on silk-related items from these different institutions.

In order to give access to these various datasets via a unique point of entry, it is necessary to harmonize them by designing and implementing a unique and complete data model, well adapted to Cultural Heritage data describing textile-related artefacts and more precisely silk-related artefacts. This data model is based on the CIDOC Conceptual Reference Model (CIDOC-CRM) which provides definitions and a formal structure for describing the underlying semantic of the structure of documentation on Cultural Heritage.

¹ <http://silknow.org/>

² <http://www2.culture.gouv.fr/documentation/joconde/fr/pres.htm>

³ <https://www.vam.ac.uk/>

⁴ <http://ceres.mcu.es/pages/SimpleSearch?index=true>

The CRM is an object-oriented model independent from any technical implementation framework. It defines a limited set of objects with which it is possible to describe complex realities. More precisely, the CRM is a core ontology – that is to say a formal representation of knowledge – with more specialist extensions (for instance the FRBRoo, an ontology designed to represent the underlying semantic of bibliographic information). It is an empirical ontology, elaborated from the analysis of the data produced by the cultural heritage experts. Moreover, the CRM data model is flexible and extensible. In other words, given that it is based on a double hierarchy of classes and properties, if needed, it is possible to add new subclasses and sub-properties, in order to express more specific relationships and properties, without modifying the basic structure of the model.

There is yet no CRM extension for dealing with the production of textile artefacts; something similar to FRBRoo, for the creation, production and expression process in literature and the performing arts. Therefore, some of the free-text fields, especially the complementary fields on the technical and material information, are still in need of a more thorough reflection. The more complex modeling of the semantics included in data about the creative and productive process of silk textiles requires elaborating new classes and properties. SILKNOW takes these digital silk textiles' data to analyze and processed them with advanced text analytics. A text analytic module is currently designed and developed, in order to analyse the text content from the data collection. The ontology is used to structure the analysed information and map this information to populate the SILKNOW ontology. The text semantic meaning is based on English and will be translated from/to the other languages (French, Italian, and Spanish) in order to be processed. Thus, when text content is analysed, many different natural language processing techniques are applied to splitting sentences, tokenization, and entity extraction. The result of these methods is used to enrich the SILKNOW ontology by employing matching algorithms to determine the correct corresponding semantic concept of a named entity.

The SILKNOW ontology will thus allow the project to elaborate new CRM classes and properties well adapted to describe silk textiles data. These results will be freely available, and will help to describe more precisely silk-related objects, and improve the way we analyze and understand these Cultural Heritage items. They can also be used as a basis to elaborate new CRM classes and properties for textiles' data and not only silk textiles' data. Moreover using the SILKNOW ontology will also allow to a wide variety of users to have access to an endangered heritage, and encourage new research on this heritage.