



HAL
open science

Understanding and preserving the European Silk Heritage: producing accessible, interoperable and reusable Cultural Heritage data with the SILKNOW ontology

Marie Puren, Pierre Vernus

► To cite this version:

Marie Puren, Pierre Vernus. Understanding and preserving the European Silk Heritage: producing accessible, interoperable and reusable Cultural Heritage data with the SILKNOW ontology. Darial Annual Event 2019, May 2019, Varsovie, Poland. . hal-02443486

HAL Id: hal-02443486

<https://hal.science/hal-02443486>

Submitted on 17 Jan 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

Why SILKNOW ?

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)** (<http://silknow.org/>) aims to produce an intelligent computational system in order to improve our understanding of European silk heritage.

The SILKNOW platform

SILKNOW will offer a single platform give access to a wide variety of data describing silk-related objects to researchers, museums curators or general public with a single interface.

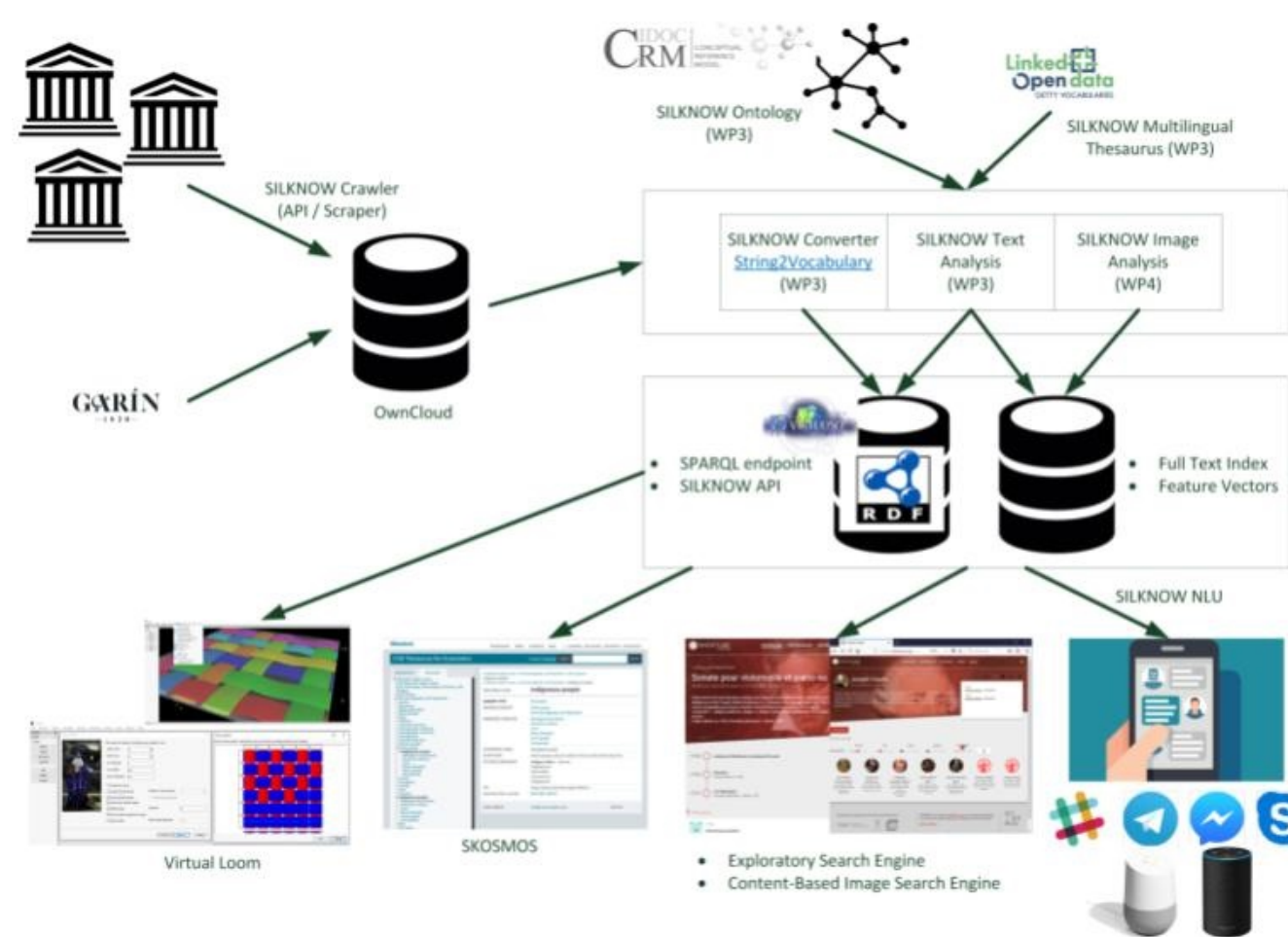


Fig. 1: The architecture of the SILKNOW platform

The SILKNOW crawler (<http://github.com/silknow/crawler/>) has collected datasets from websites or online databases of Cultural Heritage Institutions (CHIs) preserving silk-related artefacts : 30k records and 65k images ; 4 languages being used. The content of the fields of the museum records are normalized in JSON format. The JSON files stick with the data models of the source museums.

The SILKNOW middleware is composed of various software modules which goals are to semantify the original metadata records and to extract new semantic information either by analyzing the textual metadata records or the image itself. The semantic metadata are modeled using the RDF standard, a graph based data model, and all RDF graph is then loaded in a triple store.

The SILKNOW ontology is an implementation of the SILKNOW model as based on the CIDOC Conceptual Reference Model (CRM) which provides definitions and a formal structure for describing the underlying semantic of the structure of documentation on cultural heritage. The SILKNOW ontology consists in a selection of CIDOC CRM classes and properties as well as new classes and properties that are useful to fully describe silk-related materials.

The SILKNOW converter is a software tool that aims to convert the museum metadata records from their proprietary formats into the common SILKNOW ontology model. **The String2Vocabulary** module (<https://github.com/DOREMUS-ANR/string2vocabulary>) is a generic component that complements the SILKNOW converter. Its purpose is to substitute specific literals with URIs from controlled vocabularies.

The SILKNOW ontology

The CRM is an event-centric model. It defines a limited set of objects with which it is possible to describe complex realities. More precisely, the CRM is a core ontology with more specialist extensions (for instance the FRBRoo, an ontology designed to represent the underlying semantic of bibliographic information). The SILKNOW ontology is based on this model.

The CIDOC CRM ontology consists of a limited set of classes. A CRM class is a category of items sharing common traits. It is identified by numbers preceded by the letter "E" and is named using noun phrases with title case. A CRM property defines a relationship of a specific kind between two classes. By convention, it is identified by a number preceded by the letter "P" and named using verbal phrases in lower case.

We evaluate the goodness of the ontology by providing mapping rules between the museum metadata records and the SILKNOW ontology. We observe that all fields can be represented using classes and properties from the SILKNOW ontology so far.

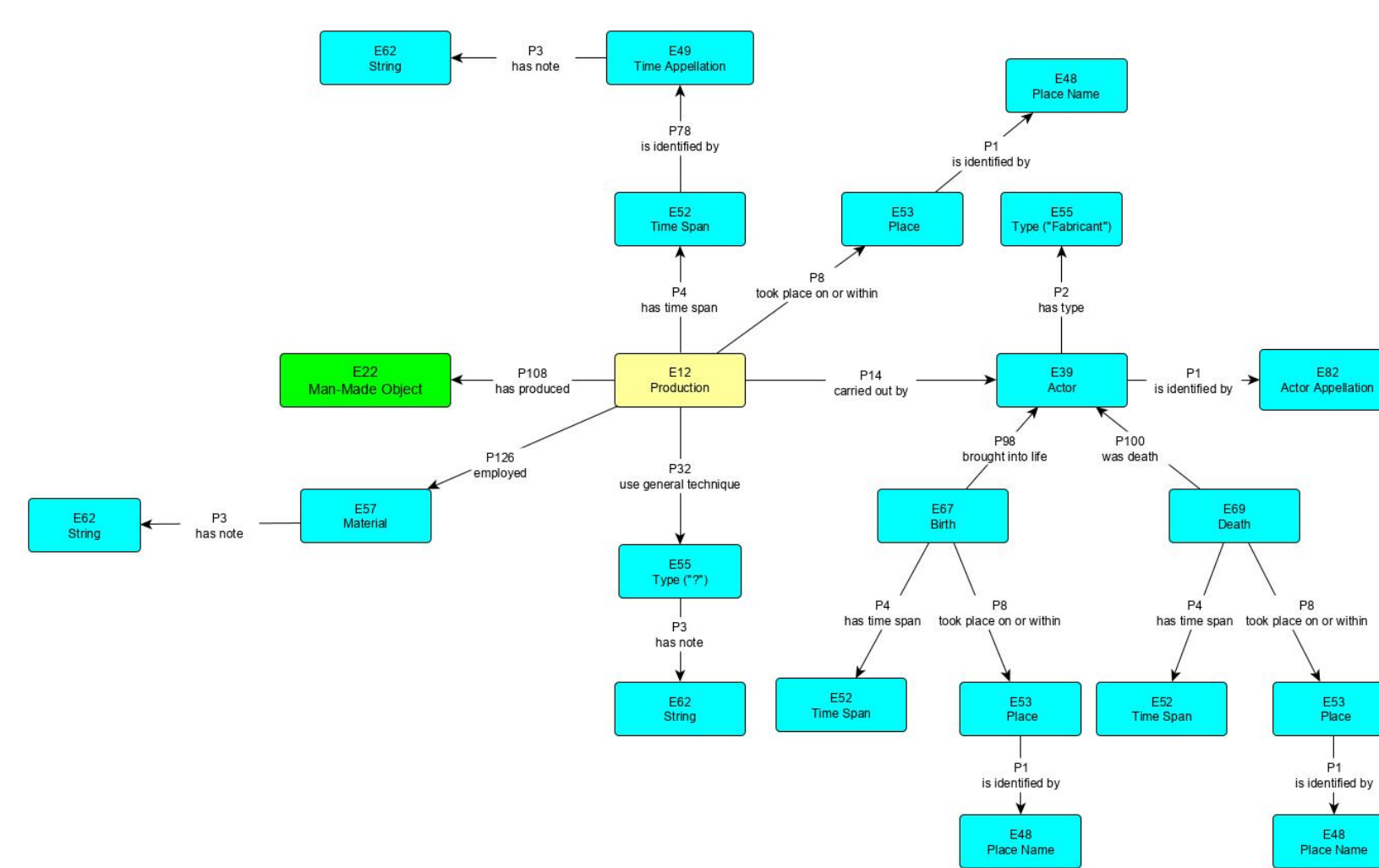


Fig. 2: RDF graph generated using the SILKNOW ontology for one representative record (Musée des Tissus, Lyon)

Improving the SILKNOW ontology

To provide the mapping tables and the associated RDF graphs, we have selected two representative records for each CHI. But this manual selection can be biased in some ways: by example, SILKNOW has crawled 6299 records from the Victoria Albert Museum database. The data describing silk-related artefacts are in fact various and very rich, and this first mapping can not accurately reflect this richness.

In this example from the Museo Diocesano di Palermo data collection, the field "Trama" ("Weft") is represented using the classes **S4_Observation** with the property **O8_Observation**, and **E55_Type** with the property **P2_has type** from the SILKNOW ontology :

"Trama : 1 trama, di fondo, 3 capi, s, colore cremisi.
Scalinatura: 4 trame. Riduzione: 25 trame al cm"

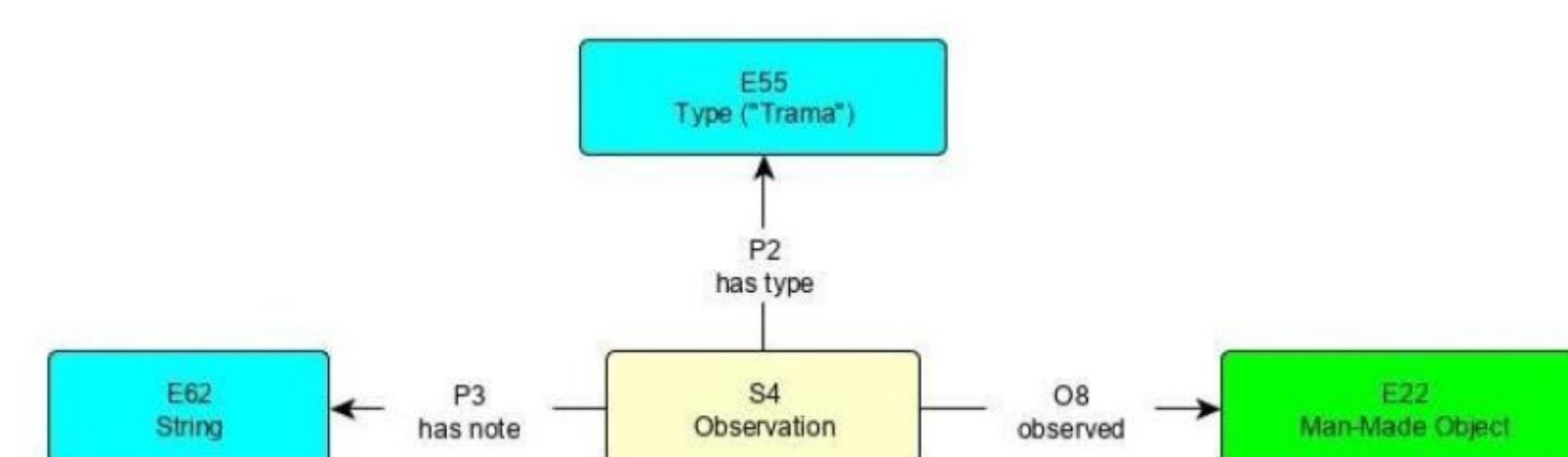


Fig. 3: RDF graph generated using the SILKNOW ontology for the property "Trama" (Museo Diocesano di Palermo, Palermo)

These mapping rules and the associated RDF graphs are clearly not satisfying. But a more thorough analysis of the free-text fields describing the creation and the production process of these silk artefacts allows us to refine this first mapping.

As shown by the RDF graph below, the CRM data model is flexible enough to offer classes and properties allowing to describe how a fabric is woven. (Some technical terms are still in French and will soon be translated by the team.)

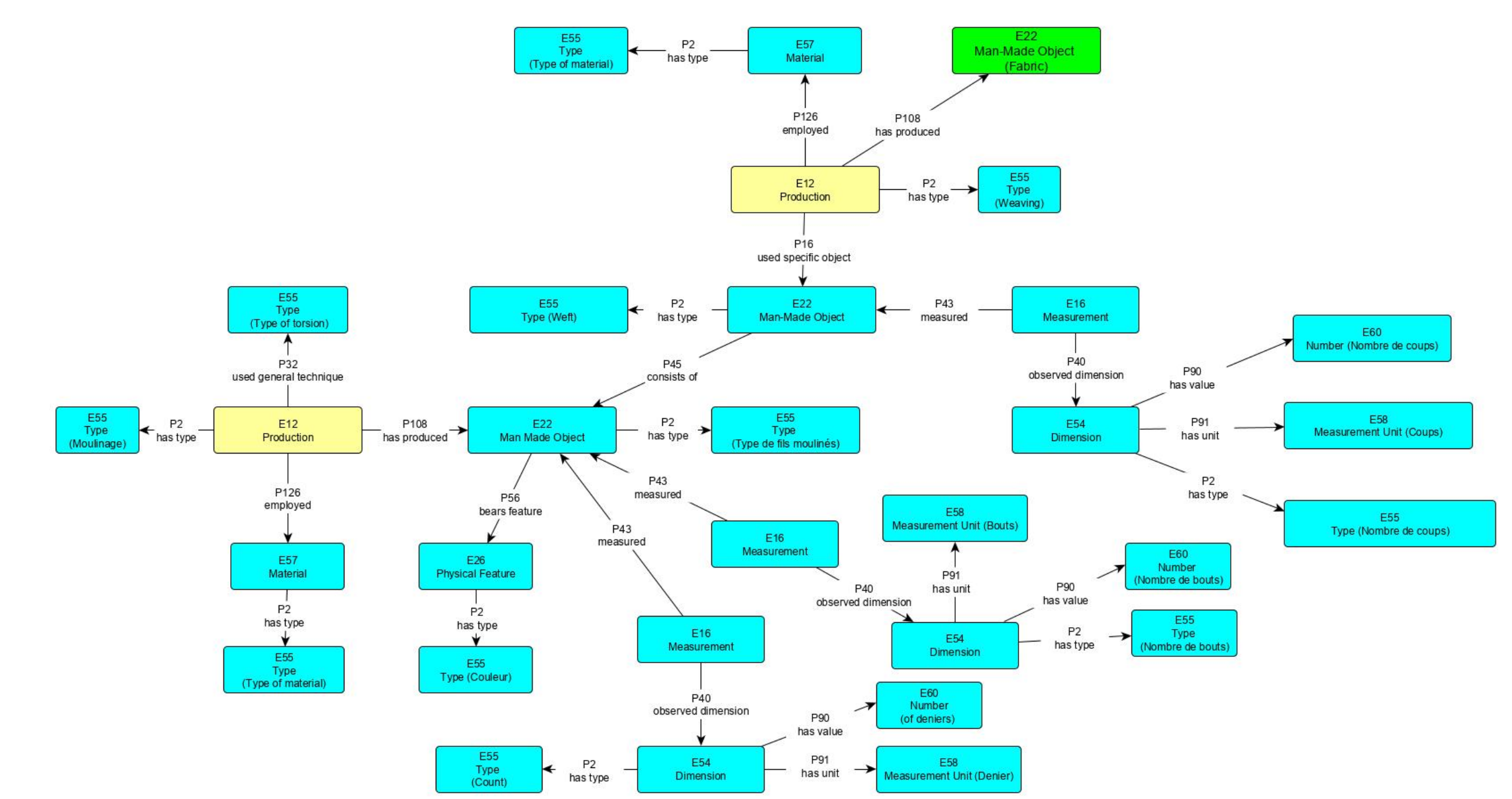


Fig. 4: RDF graph generated using the SILKNOW ontology to describe the weft of a silk woven fabric

Warp and weft are the weaving or embroidery terms for the directionality of the threads that make up a loomed or woven fabric. Warp threads are the threads that run along the length of the yardage (up-and-down, vertically) and parallel to the selvage (horizontal axis). Weft threads are the threads that run from selvage to selvage (side-to-side, horizontally).

Further developments

Automatic text analysis. It is necessary to thoroughly and systematically analyze the content of the free-text fields generally used by the CHIs to describe the artefacts they preserve and to precisely evaluate the goodness of the SILKNOW ontology. 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.

New CRM classes and properties? 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. For now, it does not seem that we need to create new CRM classes and properties to describe silk textile's data, and more broadly textile's data. But it is still an open question : for example, it could be useful to compare the results of the SILKNOW projects with the results of research projects working on digital data describing Cultural Heritage artefacts.