

Article

Open Access to Data about Silk Heritage: A Case Study in Digital Information Sustainability

Jorge Sebastián Lozano ^{1,*}, Ester Alba Pagán ^{1,*}, Eliseo Martínez Roig ¹, Mar Gaitán Salvatella ¹, Arabella León Muñoz ¹, Javier Sevilla Peris ², Pierre Vernus ³, Marie Puren ³, Luis Rei ^{4,5} and Dunja Mladenich ⁵

¹ Department of Art History, Universitat de València, Av. de Blasco Ibáñez, 28, 46010 València, Spain

² Institute of Robotics and Information and Communication Technologies (IRITIC), Universitat de València, 46980 Paterna, Spain

³ Laboratoire de Recherche Historique Rhône-Alpes (LARHRA), Université Lumière Lyon 2, 14, Avenue Berthelot, F-69363 Lyon, France

⁴ Jožef Stefan Institute, Jamova Cesta 39, 1000 Ljubljana, Slovenia

⁵ Jožef Stefan International Postgraduate School, Jamova Cesta 39, 1000 Ljubljana, Slovenia

* Correspondence: jorge.sebastian@uv.es (J.S.L.); esther.alba@uv.es (E.A.P.)

Abstract: This article builds on work conducted and lessons learned within SILKNOW, a research project that aimed at enhancing the preservation and digital dissemination of silk heritage. Taking the project and this heritage typology as a case study in the digital transformation of cultural heritage institutions, it illustrates specific challenges that these institutions must face and demonstrates a few innovative answers to meet those challenges. The methodology combines approaches typical of the humanities and others usual in ICT, being inductive regarding materials and methods (consisting of a detailed review of existing online repositories and research projects devoted to textile heritage) and descriptive for the results and discussion (which explain at length the development of some tools and resources that responded to the needs detected in the previous analysis). The article reports on the state of the art and recent developments in the field of textile heritage, the tools implemented to allow the semantic access and text analysis of descriptive records associated with silk fabrics, and the spatiotemporal visualization of that information. Finally, it argues that institutional policies, namely the creation and free dissemination of open data related to cultural heritage are just as important as technical developments, showing why any future effort in these areas should take data sustainability, both in its technical and in institutional aspects, into account, since it is the most responsible and reasonable approach in terms of efficient resource allocation.

Keywords: silk; cultural heritage; museums; open-access data; information sustainability



Citation: Sebastián Lozano, J.; Alba Pagán, E.; Martínez Roig, E.; Gaitán Salvatella, M.; León Muñoz, A.; Sevilla Peris, J.; Vernus, P.; Puren, M.; Rei, L.; Mladenich, D. Open Access to Data about Silk Heritage: A Case Study in Digital Information Sustainability. *Sustainability* **2023**, *15*, 14340. <https://doi.org/10.3390/su151914340>

Academic Editors: Mauro Francesco La Russa and Andrea Macchia

Received: 7 July 2023

Revised: 20 September 2023

Accepted: 25 September 2023

Published: 28 September 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction: From Silk Heritage to Digital, Public Humanities

This article is based on a research project about silk heritage, arguably a very specific field, but also a valuable case study of the challenges that cultural heritage institutions must face nowadays while pursuing their overall digital transformation. In that regard, the management of their collections' information constitutes a first step for most Cultural Heritage Institutions, also known as CHIs, or GLAMs: Galleries, Libraries, Archives, and Museums. Therefore, our first aim is to provide a detailed look into the pitfalls and opportunities involved in any approach to heritage information management, as exemplified by this case, digital information about silk textiles. Secondly, we argue that any sustainable approach towards that goal should aim at interoperability among repositories and as open access to heritage data as possible. These are not just technical or practical issues; in fact, they touch on deep institutional features within the humanities; mostly, the intellectual property associated with cultural objects and the roles of the professionals involved in their documentation and conservation. Finally, we intend to show the requirements and some

possible outcomes of the sectorial management of museum collections, as exemplified by some tools and resources created within the SILKNOW research project. In short, this article offers a case study about the heterogeneous and very demanding requirements posed by digital management of collections in cultural heritage institutions and demonstrates the answers provided for those needs in a fully developed research project, wherein considerations about the technical and institutional sustainability of digital resources played a key role.

Correspondingly, the article is structured in these main sections:

- “Materials and Methods” offers an in-depth analysis of the existing information resources about textile heritage, with a main focus on silk fabrics. It also provides an updated look at the sector-wide trend towards open access to cultural heritage information. This is motivated, among other reasons, by the growing understanding that the long-term sustainability of heritage data can be best guaranteed by efforts focused on interoperability, open access and decentralized management.
- “Results and Discussion” shows some tools and resources created within the SILKNOW research project, that aimed at providing sustainable answers to the limitations of those existing resources. More specifically, a multilingual thesaurus to standardize textile terminology across four European languages; an ontology based on CIDOC CRM, created to allow the mapping and interoperability of data about silk textiles, coming from different catalogs and very heterogeneous in nature; some cross-lingual models that can automatically infer and provide annotations to complete missing properties of cataloging records and a tool for the spatiotemporal visualization of information associated with records gathered from independent repositories.
- “Conclusions” argues that in addition to the technical requirements and possibilities inherent in digital repositories for cultural heritage, institutional and intellectual issues also have paramount importance. Open access to cultural heritage data must be claimed and facilitated by everyone involved before the promise of truly public humanities can be fully accomplished.

Research Background and Literature Review

The sustainability of digital information is not a new area of research and concern, although, for many cultural institutions, it may still come as a relative novelty. One of the earliest attempts to define it stated that it comprises not just the technical or structural issues associated with digital information, but also the related organizational, socio-technical, and economic infrastructure [1]. Later characterizations have stressed its dispersed, shifting nature, using the metaphor of an ecosystem that contains hardware devices, program files, data files, and also “the social elements which lead to the creation and use of digital artifacts” [2]. The increasing concern for sustainable development and its usual threefold structure, economic, environmental and social, has also impacted existing ideas about it, thus broadening even more the extra-technical considerations incorporated within the concept [3].

GLAMs have been very active, for centuries, in the generation of information about cultural heritage objects and environments. That information has been preserved, presented, disseminated and accessed in many ways, thus making them information specialists long before the discipline existed. Museums have joined the digital transformation probably later than many other scientific, entrepreneurial, or educational institutions; even within the same field, libraries have always been ahead of museums, as far as digital technologies are concerned, but by now they are closing that gap. They are in an extraordinary position to become key players in the scenario of digital content, within a datacentric society like ours, since they are producers and custodians of vast amounts of information, often highly curated, and potentially relevant for many areas of human experience.

Digital information about cultural heritage is very important for a number of reasons: easy access to information about precious and fragile objects, without compromising their physical safety; broad availability of that information, regardless of space and time

limitations; enhanced opportunities for resources' discovery and the establishment of connections between them; enriched combination of multimodal data (audiovisual, recorded, rendered. . .); advanced AI processing for annotation or search purposes, etc. It must face some challenges, too. Language and concepts in the humanities are usually complex and rarely univocal, often resisting quantification and discretization. Information standards do exist, but there are too many of them, and their heterogeneity usually results in separate siloes of information, that users must access separately, despite their shared traits.

In these regards, textile heritage is not different from other kinds of cultural heritage [4]. Its conservation requires its prior study and documentation. Through time, this highly specialized task has been conducted by individuals (researchers, scholars) and institutions (museums, collections, companies. . .). This has led to its increased respect, enjoyment and appreciation by society at large, although it suffers from a certain lack of attention, in comparison with other kinds of heritage. Not so many people pay attention to them; fewer institutions, in many cases of small or medium size and scarce resources, are devoted to it. It is a small field within cultural heritage, even while its objects and intangible resources are connected in various ways to most people. As a result of the high specialization that it requires, the efforts for its documentation show results of varying quality. An added problem is the inherent physical fragility of the textiles themselves.

In the last few decades, GLAMs have been transitioning towards digital tools and platforms, a process that is still ongoing. This general trend within heritage institutions and the humanities has also affected textile heritage. In some cases, large, national, well-funded museums have been able to carry out large digitization projects. For the rest, the path can be much more uneven. Some parts of their collections might have been cataloged (rarely in their entirety), in-house systems are developed and then discontinued, databases become obsolete as technology evolves. . . However, a fair amount of information in various digital forms is already available. Its growth is not just a matter of conservation for fragile objects, but also a push for emulation and collaboration between similar collections: once they become accessible to the public, the pressure on other museums to follow suit increases. For example, modern needlewomen, when they see vintage patterns online, may create their own products based on those patterns, thus contributing to the growth of these collections, as shown in [5].

All too often, these efforts suffer from the lack of a sustainable approach. Individuals, instead of teams, are in charge of them. Cataloging staff may have little or poor training in digital tools, and in all cases, funding and resources are scarce. The field is severely affected by an irregular, limited adoption of cataloging standards. Terminology tends to lack normalization, a problem replicated and aggravated within each national or local language, as well as by the diversity of technical or historiographical approaches to the subject [6]. The very technicalities that form the core of textile production (weaves, looms, patterns. . .) make it hard to deal with the topic, although the final products are very appealing to large sections of the population.

All these limitations make the digital transformation of small textile museums a very demanding issue. However, there is simply no alternative to the proper and sustainable management of information about their collections. The current situation seems positive, to the existence of some good information resources, as Section 2 presents in full detail. Nevertheless, these are siloed, self-enclosed databases, that require users to search across dozens of separate resources, in different languages, and follow very different standards. This approach is unsustainable, by definition.

The SILKNOW project has aimed at providing answers to some of these challenges, as Section 3 in this article shows. This has been conducted thanks to digital tools and approaches, combined with scholarly expertise (from silk specialists, art historians and historians, and textile engineers. . .). The final goal was to provide methods and best practices for heritage institutions that want to take their textile collections into the information and knowledge age. It paid particular attention to small and medium-size institutions, often lacking the technical resources and staff to venture into cutting-edge ICT and research. In

this regard, the project offered paths toward the sustainability of their collections, and their data, as well as to the fulfillment of their mission and institutional motivations.

For instance, the aggregation of digital reproductions of heritage objects and metadata related to them will only be efficient, in the long run, if the principles of Linked Open Data and semantic web technologies are applied correctly. Indeed, never before has so much of the world's cultural heritage been available to us thanks to the numerous projects to digitize and open up museum catalogs, the reality is still far from the promise of linked data, particularly as implementation remains relatively complex [7]. What is more, while many institutions are trying to seize the opportunities offered by new technologies, the task of aggregating data on specific types of objects from multiple information systems, existing in different languages, located in several countries and presented in a variety of formats, has not been taken up commonly. Section 3.1 deals with the creation of a multilingual thesaurus, mentioned as a gap in previous literature [6] but not developed before, in order to help bridge and access information within records written in English, Italian, Spanish and French. Section 3.2 of the article shows how the project has innovatively used the CIDOC Conceptual Reference Model (CIDOC CRM), recognized by museums as an international standard to link to express the underlying semantics of cultural heritage documentation, to bring data from different sources together and thus derive new information [8,9].

The ability to automatically infer and add new information to the existing records can dramatically transform the operations of many CHIs and their professionals. Section 3.3 shows the path followed by the SILKNOW project in this regard, demonstrating that it is possible to infer properties of silk fabrics from text descriptions associated with the digitized objects. We show this by learning a text classification model on data coming from a single archive. Second, we show that transfer learning across archives is possible by training a model on data from one archive and evaluating it on data from a different archive. Third, we show that transfer learning across languages is also possible by training a cross-lingual model on data from an archive with English text descriptions and evaluating it on data from an archive with Spanish text descriptions. In terms of application, the method proposed can be used to fill in missing properties in the metadata fields of digitized cultural heritage artifacts. It can also be used to align this categorical metadata with a specific controlled vocabulary, with the ultimate goal of facilitating the discovery and exploration of cultural heritage in catalogs which include objects from multiple sources (e.g., multiple museums).

One last area for exploration incorporated in our work is spatiotemporal visualization (Section 3.4). Historical information is obviously and intimately related to chronology. Heritage objects also embody spatial features, the locations where they have been produced, traded, collected, exhibited, etc. Those two properties have been traditionally visualized with timelines, maps, graphs and their combinations [10,11]. Geographic Information Systems (GIS) have undergone a profound revolution thanks to the interactivity and multimedia features added by the digital tools, amounting to a full "geographical turn" in the humanities [12]. In this context, STMaps is another resource developed by the SILKNOW project that provides new functionalities for an unusual situation in the humanities: visualizing massive amounts of objects (or object records) according to their geographical and chronological features. Based on the knowledge graph and repository created by the project, it answers user queries not in the ordinary fashion of item lists or thumbnail images, but through a dynamic, spatiotemporal map. Thus, it is designed to help identify connections, parallels and patterns in ways unaffordable for traditional visualization techniques.

2. Materials and Methods

2.1. Access to Textile and Fashion Heritage: Some Approaches

Let us look now at the ways in which CHIs, individual researchers and very big digital players are paying close attention to textile heritage cataloging and dissemination. This detailed description of the state of the art in this particular sector will provide an

updated, fact-based approach to the myriad of circumstances and hurdles that many heritage institutions must face when they try to move towards digital transformation.

2.1.1. “Universal” Repositories

We Wear Culture, from Google Arts and Culture, illustrates one possible approach to the gathering and online dissemination of digital information about clothing and textile heritage. It is an aggregator of ad-hoc, usually highly curated content. In a way, it is the traditional answer, and a very successful one, if performed properly. The innovation here is bringing together content from almost 200 different institutions and providing it in a compelling way, prioritizing extraordinary photographs and audiovisual content over consistent and extensive documentation.

In all likelihood, most non-expert users will feel more attracted to this approach. A pre-selection has been conducted by each institution and approved by the Google team. However, in terms of discovery of new information, or of specific pieces (beyond the usual collection highlights), it is fairly limited. The searching functionality is very basic and irregular in results.

Sharing information across institutions in structured repositories is another approach, the one we are dealing with in the next two examples. It offers clear advantages and some prospects that are, at least, worth exploring.

- Opportunities for discovery: large databases can provide “windows” of visibility for less-known pieces, many times kept in storage, that are less likely to attract the attention of the general public, but which can be interesting for other experts or targeted audiences. This is the main benefit and one that “only” requires cataloging data in digital formats, adapting them to existing standards, and sharing them through available repositories.
- Workflow optimization: information generated primarily for institutional, internal usage can to a certain extent be repurposed for later, external reuse, instead of incurring the costs of time-consuming, one-off curated content publishing.
- Multilingualism: joint efforts are helping to overcome linguistic barriers. Thanks to automated translation and, in specialized contexts, multilingual thesauri, it is possible to gather information in different languages, and not just the language employed by the user to interrogate the system. Museum catalogs tend to be rather specialized resources that use scholarly terminology. Therefore, it will always be better to count on multilingual controlled vocabularies and not just general automated translation.
- Opportunities for automatization of some tasks. Large bodies of information (for instance, objects covering an entire period or style) are hard to grasp in their entirety, even for experts. Artificial intelligence and big data might be ready to help us in some cases, where computers can take care of repetitive and cumbersome tasks. For instance, searching for previously unknown shared features, or for unexpected patterns within large numbers of objects and records, both in visual analysis and in textual analysis. Automated annotation might be a great help for catalogers, providing suggestions based on comparison with many other instances, but always ensuring that the AI-generated content is curated and supervised by domain experts.

Summing up, massive, shared repositories can help to go beyond the walls of individual institutions, of whatever size; but they should be particularly useful for smaller museums and collections, such as the ones scattered throughout Europe as memories of the essential roles that textile industries once played across the continent.

Europeana

Europeana is one of the largest experiments in open access to culture that the internet is making possible. It works with European archives, libraries and museums in order to share cultural heritage, providing access to millions of books, music, artworks and other content.

The important thing here is that Europeana brings together cataloging records and digital surrogates from literally thousands of institutions. This might seem like a simple

accumulative effort, but far from it, it is a technical feat of data harmonization and interoperability. A large part of it comes from national libraries, but museums and collections of material culture also have an important presence in the repository. It is decentralized in nature, working through national and thematic aggregators instead of a single data ingestion node. Very diverse institutions, regardless of size, share their contents and become data providers for Europeana.

It is very revealing that one of the first thematic clusters within it was devoted to fashion and costume (<https://pro.europeana.eu/project/europeana-fashion> (accessed on 12 September 2023)). Europeana Fashion currently gathers around one million records of cultural objects related to fashion, from catwalk photographs to drawings from the great designers of couture brands. It was born as a research project that later became a network of fashion-related institutions and an aggregator for this kind of content in Europeana.

Wikimedia Commons

A different model is that of Wikimedia Commons, the file repository that hosts public domain and freely licensed media content for the various projects of the Wikimedia Foundation. Wikipedia is just one (and the most used) among them.

This approach is very different from the previous ones. Wikimedia Commons can be used to find images (or multimedia) of cultural heritage, but not as a direct provider. Instead, search engines increasingly rely on Commons as the first option in image searches about historical cultural objects. Since, most often, images uploaded to Wikimedia have little or no limitations to their reuse, they get downloaded and copied in ever-increasing numbers. Any cultural institution should ask itself what to do, in this regard: whether to fight a long, uphill battle to obtain its website as the top search result for objects kept by them; or to “join the enemy” and simply make sure that the image shown in Wikimedia is one provided by them, properly referenced and linked to the owning institution.

We have outlined just two global repositories that contain data about textiles and fashion, among many other subjects, of course. This short overview simply aims to make clear that digital platforms offer various models for the dissemination of cultural heritage data—particularly about fashion and textiles. Those platforms evolve over time, and in this regard as in others, Europeana seems the most stable option for any institution within Europe that wants to share their collection information beyond their own digital resources.

In any case, it does not have to be an either/or dichotomy. All three approaches can be useful for the same institution, and even for the same user.

2.1.2. National Databases

A few countries have followed an approach that aims at a central, state-wide repository for material heritage, or at least some of its varieties. These catalogs require a substantial effort to establish cataloging guidelines, use common schemes for the records (i.e., data models) and employ controlled vocabularies. Some kind of shared software, within a decentralized system, or an online platform for a centralized one, are common features for these databases, too. These coordination efforts pay off in the longer run, enabling users to access data from dozens or hundreds of museums through a single gateway, instead of having to search for them at each individual institution. They also provide greater financial efficiency, sharing one information system and software application among many museums, instead of having each one of them make the expense to develop or acquire their own solution. On the downside, it is also worth noting that a single record scheme may not always do justice to the many kinds of data about heterogeneous museum objects, such as Baroque altarpieces, folk crafts, traditional African masks, filmed records of performance art, or textiles. The advantages of easy, homogeneous access must balance the losses in information specificity and detail.

- Joconde is the classical model in this regard. This database, created and maintained by the French Ministry of Culture, as of 31 August 2023, gathers 667,466 records from more than 250 museums having received the legal status of “Musée de France” [13].

Its records are also shared through other platforms, notably, on POP, the Plateforme Ouverte du Patrimoine (<https://pop.culture.gouv.fr/> (accessed on 12 September 2023)). By late 2020, some 21,000 of them were related to textiles and costumes [14].

- CERES, the Red Digital de Colecciones de Museos de España, offers a similar framework. It is built on an information system named Domus, developed by the Spanish Ministry of Culture and currently used by 195 museums throughout Spain, both public and private. While the system was originally built for the internal management of the collections, sharing the catalog records through the Ministry's centralized repository is a permanent feature of the software. This repository is then made public online through the CERES website. It also relies on a set of common controlled vocabularies and cataloging rules. It covers large parts of Spanish heritage kept in museums, but it cannot be said to be fully comprehensive, either. Some regions have developed their own, independent systems. Even among museums contributing to CERES, the quantity and depth of their records on the platform can vary widely. Despite such shortcomings, it offers a tremendous amount of information and serves as an outstanding example of the feasibility and advantages of centralized repositories. In its current version, it offers more than 341,000 records from 118 museums (<http://ceres.mcu.es/> (accessed on 12 September 2023)). It is particularly useful for small and medium institutions: among them, many specializing in textile heritage. They can benefit greatly from shared resources like Domus and CERES, as they usually lack the funding, human resources and expertise to embark on large digitization campaigns on their own.
- A partly similar approach lies at the basis of BeWeB, the census of heritage owned by Catholic dioceses and institutions from Italy (<https://beweb.chiesacattolica.it/> (accessed on 12 September 2023)). While organized by a private institution, the Italian Bishops' Conference (CEI—Ufficio Nazionale per i beni culturali ecclesiastici e l'edilizia di culto), offers coverage even larger than the ones just mentioned. It contains records on more than 10 million objects, including archival documents and books, with historical and artistic objects exceeding 4 million [15]. Inevitably, the quality and standardization of all these records is quite a challenge and often offers ample room for improvement. The resource itself, however, is staggering in its ambition and reach and offers a good example for private owners of art historical heritage.
- The last two instances show, on the other hand, some of the limitations of the centralized model. Even when controlled vocabularies are available, cataloguers do not always follow them consistently. In CERES, identical pieces may be cataloged as either "Textiles" or "Tejidos" (or as any of their many subtypes), which makes systematic recovery quite unpredictable sometimes. Semantic web technologies can help to overcome these problems, but only to a certain extent. On the other hand, these repositories bring together records prepared over long periods of time (decades, sometimes), in widely different institutions, about very heterogeneous records, by catalogers with varying levels of expertise and dedication to the task. The resulting records are also dissimilar in quality, depth and scientific validity. In any case, the main advantages of these large repositories are, again, the new opportunities they provide for discovery into the less visible parts of our heritage, for the cross-reference of objects between institutions or across disciplines, for quantitative analysis and innovative visualizations.

2.1.3. Major Museums

In many developed countries, large museums of national -and sometimes even global, encyclopedic-reach are guardians of massive historical holdings. Some of them are owned by the State, usually originating from royal -and national, later-collections. This is the situation in many European countries. In other cases, equivalent institutions have been formed during the 20th century, through acquisitions and bequests. Some of them are more focused on fine arts than on decorative arts, but their textile holdings can be impressive, in

any case. Others are specialized in decorative arts, traditional crafts or modern design, but they remain large-scale institutions with dazzling collections.

For a number of institutional reasons, these large museums tend to offer their catalogs independently, not as part of national repositories like the ones in the previous section. This gives them more visibility and reinforces their exceptionality. Moreover, their leading position often means that they count on the human and technological resources required to have a strong online presence, including comprehensive, in-depth cataloging information on most of their holdings. Some of them are also champions of the open-access movement among museums.

We may group in this category museums such as:

- The Victoria and Albert Museum in London was created in the mid-19th century with a focus on the applied arts and science. Cataloging records on part of their collections are available on their website, numbering more than 1.2 million. Their holdings of textiles and fashion are truly impressive: their online presence almost reaches 80,000 pieces, not including embroidery and fashion items. They are also accessible through an API, an uncommon but forward-looking feature for museums (<https://developers.vam.ac.uk/> (accessed on 12 September 2023)). Similar institutions in other European countries are the Musée des Arts Décoratifs in Paris and the Museum für angewandte Kunst in Vienna.
- The Musée des Tissus in Lyon (<https://www.museedestissus.fr/> (accessed on 12 September 2023)), until recently known as the Musée des Tissus et des Arts Décoratifs. Widely considered the best European collection of historical silks, it is also exceptionally rich in fashion and other textiles, with 2.5 million pieces in total. Online access to such a vast collection is quite limited, however. As a private establishment, dependent on both public and private funding, it has suffered serious institutional crises in the last years that seem to have been overcome by now.
- The Metropolitan Museum of Art, in New York City, is an encyclopedic art museum, and a public/private partnership. It houses world-class collections of many kinds of fine and decorative arts. In the last decades, its Costume Institute has gained huge visibility for its temporary exhibitions and celebrity-oriented events, such as the annual Met Gala. Less known but equally important are the textile collections. It is a global leader in the field of open access to collections, providing full information and high-resolution images on some 400,000 pieces, from the total 1.5 million objects it holds (<https://www.metmuseum.org/about-the-met/policies-and-documents/open-access> (accessed on 12 September 2023)). Some 40,000 catalog records on textiles are freely available on its website.
- The Smithsonian Institution is a system of research centers, libraries and museums, part of the US federal administration. One of its 19 museums is the Cooper Hewitt Design Museum, located in New York City. While it may seem focused on modern design, it houses impressive historical collections, too, with textiles among them. It is also a champion of open access, as part of the Institution's general policy, featuring an API (<https://edan.si.edu/openaccess/apidocs/> (accessed on 12 September 2023)) and an image repository shared across all of the Smithsonian's collections (<https://www.si.edu/openaccess> (accessed on 12 September 2023)).

2.1.4. Other Projects

Research projects are providing yet another model for the study and dissemination of textiles by means of online repositories. Sometimes they are linked to individual institutions, while in some other cases, one of their goals is precisely to bring together resources from separate organizations, adding to the challenge of interoperability of shared information. Some have been born as aids for a personal research project, later outgrowing that framework in order to incorporate data from other researchers. Others are born from a private grant or thanks to competitive funding received from national or international research agencies. The sustainability of these projects is open to question, since the information

obtained and the expertise gathered during the years of their design, implementation and publication, are seldom kept in use for a long time after the funding disappears. However, proper institutional support, one that incorporates future maintenance and when necessary, technical support for migration and scalability, can help to overcome this problem. It is also essential to adopt an open framework for the information, allowing present or future semantic linking between information resources.

- SilkMemory is, according to its own website <https://silkmemory.ch/> (accessed on 12 September 2023), a “web portal [that] provides access to the archive database of the Lucerne School of Art and Design with digitized text and image sources about the silk industry of the Canton of Zurich”. Born after the commercial demise of the once-thriving Swiss silk industry, it was funded by the Zurich cantonal government and went online in 2018. It provides a thoughtful answer to a danger that is common to many European countries: the dispersal and loss of the valuable archival and material heritage generated by those industries, most of which have gone out of business during the last decades. It offers a database of fabrics, books and images kept in those archives, together with a selection of some personal or institutional stories obtained from the same archival fund.
- ART-CHERIE (Achieving and Retrieving Creativity Through European Fashion Cultural Heritage Inspiration—<https://www.artcherie.eu/> (accessed on 12 September 2023)) was a project funded by the Erasmus programme of the European Commission, lasting from December 2016 to May 2019. It brought together partners from Belgium, Greece, Italy and the United Kingdom, from a quite broad scope, including an interesting connection to the training of fashion designers. Among other outputs, it aimed at providing a Digital Database or “Catalogue and Digitisation of Museo Prato Exhibits and Collection”. However, it is not openly accessible.
- María Judith Feliciano, an independent scholar specializing in Medieval Iberian textiles, is the principal investigator of the “Medieval Islamic Textiles in Iberia and the Mediterranean” research project (<https://maxvanberchem.org/en/scientific-activities/projects/art-history/16-histoire-de-l-art/160-medieval-islamic-textiles-in-iberia-and-the-mediterranean-2> (accessed on 12 September 2023)). Funded by the Fondation Max van Berchem in 2016–17, it was a crossroads for a number of research projects from other scholars in the same field, like Ana Cabrera, Laura Rodríguez Peinado and Therese Martin. It reportedly aimed at producing a website and database to make available the results of the research carried out, but these have only been disseminated through articles and essays.
- Tetiana Brovarets is one case of an independent scholar working outside the usual funding schemes and doing valuable self-supported research. For example, she has published a database where textiles with embroidered verbal texts are collected: mostly *rushnyks*, Ukrainian towels from the late 19th and early 20th centuries (<https://volkovicher.com/> (accessed on 12 September 2023)). Thanks to this database, it is possible to study different combinations of one and the same images and inscriptions on textiles, as shown in [16].
- IMATEX is the online database offering information about the collection of the Centre de Documentació i Museu Tèxtil de Terrassa (<http://imatex.cdm.tcat/> (accessed on 12 September 2023)). Created in 1996, it was originally built as a gateway for designers searching for inspiration in CDMT’s historical collection and later transformed into a generic online information resource, open to everyone [17]. It is extremely rich in content, including costumes, accessories, designs, paraments, sample books, a library and an outstanding collection of more than 9000 textiles. Available in Catalan, Spanish and English, initially it was made possible by the European Regional Development Fund, and by the CDMT’s own budget afterward.
- The MINGEI project aims to explore the possibilities of representing and making accessible both tangible and intangible aspects of craft as cultural heritage (<https://www.mingei-project.eu/> (accessed on 12 September 2023)). One of the crafts under study is

silk weaving, led by one of the project partners, Haus der Seidenkultur in Krefeld. It is a Horizon 2020 project, led by FORTH, and it is being carried out between 2019 and 2022. The project does not directly intend to build a database, but rather a repository of innovative storytelling models, including interactive Augmented Reality and Mixed Reality. It does have a strong emphasis on developing content description tools that comply with existing semantic web standards, such as CIDOC-CRM.

- The PARVENUE project was recently funded by the German Federal Ministry for Education (<https://www.parvenue-projekt.de/> (accessed on 12 September 2023)). Led by art historians from the Heinrich-Heine-Universität in Düsseldorf, it is built on a tight collaboration with the Deutsche Textilmuseum in Krefeld (<https://www.deutschestextilmuseum.de/> (accessed on 12 September 2023)), one of the European capitals of the silk industry. One of the areas of the project is built on preliminary cataloging of the collection of 30,000 fabrics and costumes in the museum, not yet available online.
- Another recent initiative is the Restaging Fashion project, based in the Lipperheidesche Kostümbibliothek—Sammlung Modebild in Berlin, and the Fachhochschule Potsdam (<https://uclab.fh-potsdam.de/projects/restaging-fashion/> (accessed on 12 September 2023)). Active between 2020 and 2023, and also funded by the German Federal Ministry for Education, it purports to build an online catalog of costumes, prints and drawings, held in different institutions, adding 3D visualizations of some of these objects.
- Finally, some authors of this article have been involved in three projects focused on the dissemination and sustainability of heritage via digital tools and platforms. One of them is, as already mentioned, SILKNOW, a Horizon 2020 project active between 2018 and 2021, that among other things has built ADASilk, a repository of some 40,000 records about silk-related objects from different museums and collections (<https://ada.silknow.org/> (accessed on 12 September 2023)). The joint team of art historians and computer scientists in Universitat de València has also been working on SeMap (<https://www.uv.es/semmap/> (accessed on 12 September 2023)), a project funded by Fundación BBVA and built on the data from Spanish museums made available by CERES, the web portal of the Red Digital de Colecciones de Museos de España presented above. Finally, some members of the same team are currently working on the ClioViz project (<https://www.uv.es/clioviz> (accessed on 12 September 2023)), funded by the Spanish government (2022–2025). It explores advanced techniques for the visualization of historical information, thanks to the collaboration of the already-mentioned CERES national repository.

2.2. Sustainability of Heritage Information: Toward Open Access

Before speaking about information sustainability, it may seem worthwhile to speak about why museums or heritage institutions should be interested in this topic. Access to heritage and culture should be at the center of that conversation. It could well start two centuries ago, with the birth of large national museums across Europe, but we will not go that far.

Let us just remind now that access to culture is recognized as a human right in article 27 of the Universal Declaration of Human Rights, and that access to information about cultural objects is one of the best ways to protect and preserve them, for today's society, but also for future generations, so that both they and we can enjoy those objects, share our experiences around them, and learn about the people that created or used them.

The advent of the internet (or even more generally, the Information Age and the Network Society, following Castells' terminology [18]) has made this discussion even more poignant for cultural heritage institutions, traditionally charged with those tasks of preserving the memory of the past while making it accessible and understandable for current citizens.

The practical implementations of all these general principles vary greatly, of course, depending on several circumstances:

- Ownership: public institutions (meaning state-owned), or collections owned by private organizations or individuals.
- Funding models: be they fee-based, paid through taxes, established as non-profits but aiming at sustainability, and many different mixed approaches.
- Intellectual property rights (moral and economic ones) connected to the works, or their associated documentation have varying consequences on the display, dissemination and reuse of all that information.
- Information tools: catalogs and inventories kept only for professional and scholarly (internal) use; sometimes slowly and partially published over decades or centuries; sometimes disseminated through exhibition catalogs or research journals.
- Digital availability: varying degrees of transition to digital tools and repositories for all that information, a true wealth of data kept by heritage institutions.

One of the many aspects in this discussion has been the adoption of open-access policies within GLAMs, or CH institutions. The definition of “open” in open access is itself a contested topic, but for our purposes let us agree that “open refers to a policy or practice that allows reuse and redistribution of materials for any purpose, including commercial” [19,20].

Examples of important museums that have made all or large parts of their holdings freely available on the web are well known: the Rijksmuseum, the Metropolitan, the Getty, museums belonging to universities such as Yale or Harvard... But is this policy only available to well-funded institutions?

Incidentally, some examples indicate that a temporary closing for renovations or other reasons was an important push to take those decisions, making their collections available online while various reasons made physical access to them impossible. To our knowledge, COVID lockdowns did not have this exact kind of consequence, a big effort for the massive digitization of collections, among other museums. For good reasons, of course: lockdowns were not planned, and many other issues were more pressing for museums during that difficult time. Adaptation to changing environments and visitor behaviors, however, has now become much more evident as a good reason to invest in digitization and open access to museum collections.

This discussion is by no means recent [21]. Anyway, the important fact here is that some small and medium institutions have also adopted ambitious schemes of digitization, open access and (importantly) interoperability for the information about their holdings. For instance, OPENGLAM, an international network of heritage professionals, did some important work preparing a Declaration on Open Access to Cultural Heritage (<https://openglam.org/> (accessed on 12 September 2023)).

We are addressing here two of the benefits that this approach involves. Our argument is that, once information is digitized (both images and catalog records) and incorporated into a structured data repository, sharing it across institutions is really at hand, in most cases. Secondly, instead of only expecting users to find our website among millions of other websites, an additional path to follow is to aggregate our data, our cataloging records, into larger repositories.

There are many other reasons for shared repositories: ensuring the practice of better cataloging and information management strategies, opportunities for increased visibility for small and medium-size institutions (that usually do not have such a big public profile among the general audience), better chances of being indexed and made visible by generic search engines, guaranteeing a permanent exercise of citizens’ rights of access to culture in spite of failing institutional abilities during any given crisis, sustainability of digital resources against data obsolescence and in order to ensure the appropriate usage of the required investment.

Some challenges do appear quickly, too. Searching in a repository that contains dozens of millions of records seems daunting (but not much worse than any ordinary internet search). Many will say that more information does not equal more knowledge, invoking the all-too-present danger of infocipation. With data available in such huge amounts, doubts

about information quality, relevance and homogeneity are perfectly understandable. Will non-expert users be able to find useful data in such a deep ocean? How will the very demanding efforts required to generate good-quality datasets be professionally and socially recognized [3]?

Regarding textiles, if we are dealing with fabrics, fashions, and in general, with material objects of an inherently fragile material condition, digital preservation and access seem the only forward-looking option, as the following examples will show.

3. Results and Discussion: Silk Heritage Resources and Tools from the SILKNOW Project

All this background shaped the planning and work carried out within the SILKNOW research project. It provided a valuable case study of the requirements and implications of managing vast amounts of digital information about cultural heritage, specifically silk heritage. In this section, firstly we will present the SILKNOW thesaurus and ontology, two of the resources that had to be developed in order to make accessible, through a single interface, more than 30,000 museum records about silk textiles, created across dozens of museums, in four languages, and according to very different record structures. The thesaurus and the ontology provided the foundations to standardize and unify all that data through the ADAsilk exploratory search engine.

These interoperability efforts were also the basis for some advanced tools that explored possibilities for the exploitation and visualization of the information contained within the records describing the silk fabrics. Two of them will be presented here: textual analysis devised towards automated annotation, as well as the spatiotemporal visualization of data contained in the records. In this manner, digital artifacts such as technical descriptions of textiles can provide patterns and suggest descriptions of objects that lack human-made characterizations, something that can help to advance the existing knowledge [2].

These two resources and two tools offer useful insights into the demands and results deriving from building a large repository of cultural heritage information. All of them faced a very serious challenge: dealing with information written in four languages, often using specialized textile vocabulary, with enormous heterogeneity in data quality and granularity. Thus, they exemplify novel approaches to that challenge, for different functionalities and through innovative procedures. Aside from this common experiment, each one of them explores various aspects of semantic data in cultural heritage. Connecting objects or records across museum collections and scholarly disciplines is the great promise of data made interoperable and linked: “setting up the means by which the result of any search on a given theme, subject or person, would produce and display a map of its cultural history across disciplines and across media” [22].

3.1. The SILKNOW Thesaurus

Smeets [23] affirms that since traditional, intergenerational ways of transmission are losing ground worldwide, additional ones are being and must be explored. Most heritage contains both tangible and intangible elements, whose proper safeguarding requires careful documentation of the link between them, i.e., terminology. While cultural heritage institutions strive to use controlled vocabularies based on their own collections [24], some efforts have been conducted to standardize all available vocabularies, regardless of important differences among them, being produced by different professionals, by different nationalities or even different disciplines, etc. [25]. As recently as 2016, a proposal was made at an ICOM General Conference to develop a textile thesaurus based on merging and enlarging the existing vocabularies [6].

For instance, the CIETA Vocabularies have been and still are the most common terminological resource within the field of historical textiles. CIETA, the Centre International d'Étude des Textiles Anciens, based in Lyon, offers a hub for researchers, collections and institutions. Its vocabularies (<https://cieta.fr/cieta-vocabulaire/> (accessed on 12 September 2023)), available in all the major European languages, have been continuously expanded

and translated, with the latest versions being made fully available online only very recently, for the first time.

In fact, in most museums, textile vocabularies are often based on their own collections [24], such as The Textile Museum Thesaurus from the Textile Museum in Washington D.C. [26], developed with the aim of improving their collection catalog. In Europe, CIETA offers excellent vocabularies but focuses mainly on weaving techniques. The Europeana Fashion Vocabulary [27] focuses not only on fabrics but also on objects surrounding them, such as books, blogs, and websites. However, as indicated by its name, its main objective is the standardization of fashion vocabulary. Other monolingual vocabularies related to textile heritage include the *Lemmario per la Catalogazione dell'Abito e Degli Elementi Vestimentari*. This vocabulary focuses on fashion and includes typologies from the 18th, 19th, and 20th centuries, structural and decorative components, and techniques.

There is a gap in research on silk textiles' standardization, and to the best of our knowledge, no specialized silk heritage thesauri cover not only weaving techniques but also materials, objects, colors, or even motifs. For this reason, SILKNOW has built a multilingual thesaurus dedicated to the specific terminology of historic silk textiles, which also includes local term variants [28]. In addition to giving a chance to preserve and transmit heritage, we seek to give a useful tool to the target institutions which, at the moment, are using different terms to describe their objects [29]. A thesaurus is a controlled vocabulary, but also a hierarchical tool, one that incorporates the relationships between terms: hierarchical, equivalence, association, etc. Terms related to silk textiles, productive crafts, motifs, etc. can vary greatly from one context to another, which makes information related to them less accessible. These local and traditional terms are usually forgotten or ignored since they are only present in archival records. However, such a specialized lexicon is in use among practitioners, especially in the domains of traditional knowledge and handicrafts, and it needs to be collected for its proper preservation. Documenting and standardizing this terminology is thus a great help for professionals, students, and researchers.

The SILKNOW thesaurus also covers the already noted need for a common framework, a standard tool that could gather as many terms as possible, with all their variants or synonyms, for independent institutions (<https://skosmos.silknow.org/> (accessed on 12 September 2023)). Nowadays, collaboration among museums and collections requires tools that foster data interoperability. A multilingual thesaurus not only facilitates information exchanges across collections and institutions but also makes heritage accessible to non-specialist audiences by lowering language barriers [30]. Making it freely accessible, reusable and linkable to other resources is the way to go if it is intended as a sustainable tool. In fact, the SILKNOW thesaurus has been built as an extension of the most widely used thesaurus within the cultural heritage community, the Getty Foundation's Art and Architecture Thesaurus (or AAT—<https://www.getty.edu/research/tools/vocabularies/aat/> (accessed on 12 September 2023)).

As of July 2023, the latest version of the thesaurus has 666 preferred terms and more than 600 alternative terms in the four languages in which it was developed. Its validation [31] was carried out following a coverage analysis which permitted the validation of textual data of online resources on the full thesaurus in all four languages it covers. We calculated the frequency of the individual thesaurus concepts that are present in the data coming from collections included in SILKNOW. Later, we compared two words and determined whether they had the same stem. Additionally, the thesaurus was evaluated among domain experts who compared data from SILKNOW ontology with the concepts included in the thesaurus. The result is an interdisciplinary and multilingual thesaurus that covers not only the most frequent concepts used in museums but also those that are used in academic papers and in the current and traditional silk industries. Hence, not only protecting tangible and intangible heritage but standardizing silk heritage language. The SILKNOW thesaurus has emerged as a crucial tool for preserving this heritage, not only by assigning appropriate names but also by facilitating connections among collections across time and space as can be seen in the following sections.

3.2. The SILKNOW Ontology

The data collected by SILKNOW is by nature heterogeneous. Indeed, the implementation of successive technologies over the decades means that museum metadata are scattered across multiple databases, spreadsheets and even structured texts. Old and new technologies often coexist. Traces of the old tools can often be found in the content that has been migrated to the new applications. Moreover, each institution has its own cataloging practices, and these practices may have evolved over time. The resulting metadata can, therefore, vary greatly. The inherent heterogeneity of these data results in the creation of data silos that are incompatible with each other, and therefore, mutually incomprehensible [32]. Moreover, data heterogeneity is further increased by the multiplicity of languages used. This makes the discovery of these data all the more difficult, as it requires users to master various languages and very different information management systems, as well as explicit or implicit data models. As a result, despite the prospects opened up by linked data and the semantic web for creating relevant links between disparate objects, the reality is still a long way from the promise.

To overcome the problem posed by data heterogeneity, SILKNOW uses a formal ontology, to formalize these cultural heritage data in a logical language made up of classes and relations [33]. This coherent and uniform representation of information will facilitate the discovery of information which, until then, was difficult to access due to its fragmentation in incompatible data silos.

To develop an ontology that will thus enable better data integration, we have chosen to use the CIDOC Conceptual Reference Model or CIDOC CRM, developed to express the underlying semantics of cultural heritage documentation [8]. Recognized by both the museum and ICT worlds, the CIDOC CRM is also an international standard, recognized as an ISO standard for version 5.0.4. It should be noted that the latest version of the CIDOC CRM is version 7.2.2 published in October 2022 (<http://www.cidoc-crm.org/versions-of-the-cidoc-crm> (accessed on 12 September 2023)); but we are currently using version 6.2, which was the latest version published at the time this work began.

The creation of this ontology required different steps to be taken. First, we analyzed and compared numerous records coming from different cultural heritage institutions—such as the Victoria and Albert Museum, the British Museum, the Musée des Tissus in Lyon, the Garin collection at the Museu de la Seda in Moncada, the Musée des Arts Décoratifs in Paris, and various French museums via the Joconde database, and consequently respecting different data models and cataloging standards. We also relied on the standards and documentation used by these institutions to produce metadata, notably the inventories in French museums [34], the HADOC Harmonized model for cultural data production [35], the ICOM guidelines for Museum Object information [36] and the Europeana data model [37]. We then drew up a list of the descriptive fields most commonly used by cultural heritage institutions to describe the textiles they preserve, eliminating those that are not of interest to the SILKNOW project—in particular we have not selected the information concerning the administrative management of these artifacts. These descriptive fields were then grouped into “information groups”. We have, for example, defined an “Object acquisition information group” to identify the acquisition method and date, the previous owner of the object, its current owner, and additional information about the acquisition (see Table 1):

Table 1. Contents of the object acquisition information group.

Object Acquisition and Legal Status Information Group	
Definition: information about the acquisition and ownership of a cultural heritage object. Several such information groups can be available for one object depending on the history of the object.	
<i>Acquisition method</i>	
The method by which an object was acquired.	
Ex: gift; purchase	
<i>Acquisition time-span</i>	
The timespan or the date of acquisition of the object.	
ex: Before 1998; 1950	
<i>Previous owner</i>	
The name of the person from whom, or organization from which, the object was acquired.	
<i>New owner</i>	
The name of the person who, or organization that, acquired the object.	
<i>Acquisition complement</i>	
Any additional information about the acquisition of the object.	
<i>Acquisition note</i>	
If necessary, additional comment on the acquisition of the object	

These categories of information allow us to make the best use of the functional overviews provided by the CIDOC CRM official documentation (<http://www.cidoc-crm.org/functional-units> (accessed on 12 September 2023)). These functional overviews divide the CRM entities and properties into different categories of information, with their graph representation, thus offering technically neutral templates of modeling applied to the meta-data that describe cultural heritage artifacts. For example, in the functional overviews, we find the category “Acquisition information” which corresponds to the “Object acquisition information group”. This preliminary step is, therefore, particularly useful because it allows us to rely on the functional overviews to express the semantics of these “information groups” with the entities and properties of the CIDOC CRM. This categorization of information has made it easier to express the underlying semantics of these descriptive fields; it enabled the selection of CRM classes and properties capable of expressing the meaning of these categories. For example, the Object Acquisition information group was thus expressed with the following classes and associated properties (see Table 2):

Table 2. CRM classes and properties used to express information on acquisition.

Domain	Property	Range
E8_Acquisition	P14_carried out by	E39_Actor
E8_Acquisition	P22_transferred title to	E39_Actor
E8_Acquisition	P23_transferred title from	E39_Actor
E8_Acquisition	P24_transferred title	E22_Man-Made Object
E8_Acquisition	P7_took place	E53_Place
E8_Acquisition	P4_has time-span	E52_Time-Span

This first step is followed by the mapping process which consists of producing semantic data from the data produced by the cultural heritage institutions and stored in relational databases, giving them an equivalent semantic expression by means of the chosen formal ontology. The mapping process was produced manually by domain experts in collaboration with computer scientists. The method adopted is the one suggested in [38], which proposes to interpret each of the descriptive fields as entity-relationship-entity (e-r-e). More precisely,

- tables and columns in the relational database are interpreted as entities;
- complete records are interpreted as entity instances;
- field names are interpreted as both relationships and entities;
- and field contents are interpreted as entity instances.

The whole scheme is decomposed into e-r-e's, and each e-r-e is aligned with the CIDOC CRM [39]. In other words, the mapping consists of interpreting these entities and relationships and expressing them in CIDOC CRM semantics. In doing so, we aim to preserve as far as possible the original meaning of the data. Concretely this process produces triplets that link nodes together through properties, forming a network of human- and machine-readable data and enabling information exchange and integration.

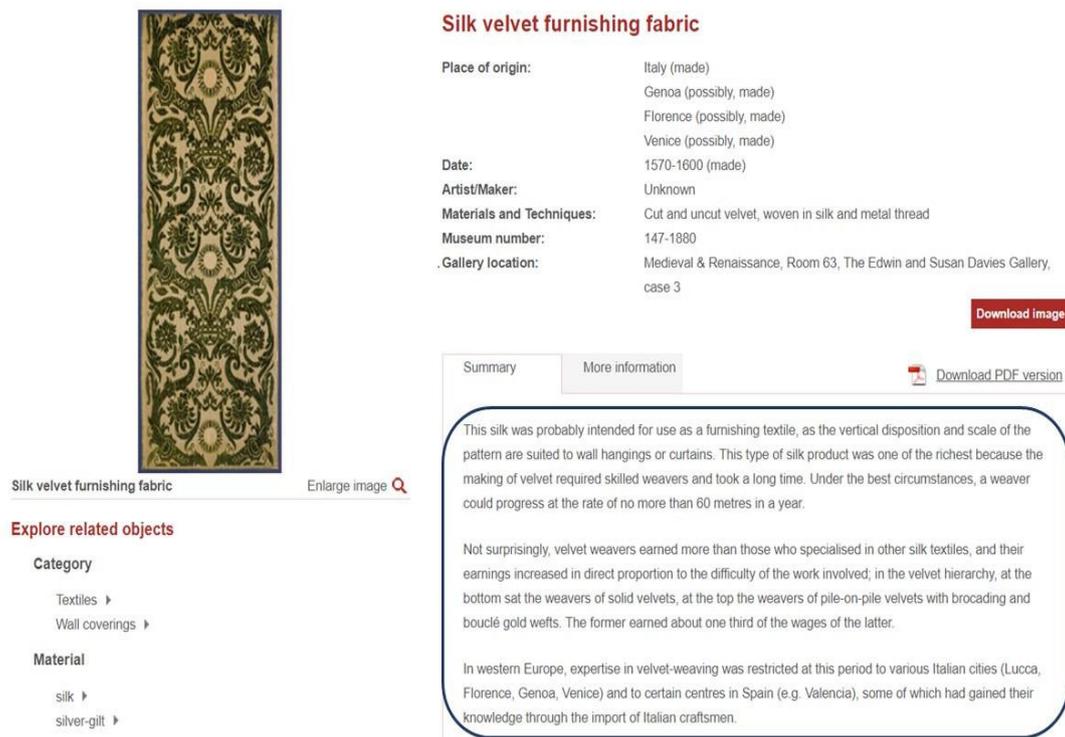
Given the data heterogeneity, carrying out this mapping process implied precisely understanding their meaning. After studying the structure of the different catalogs from which the data were extracted, we analyzed their contents to understand what information was expressed in them and also to assess their internal consistency. Cataloging practices within the same institution may have varied over time, and the meaning given to these descriptive fields may also change, depending on the practices adopted. As a result, the consistency of the content is generally weak. Based on the categorization of the data carried out previously, we not only selected the classes and properties most likely to express the semantics of this information, but we also refined this initial selection by adding new classes and properties and removing those that proved to be useless.

To understand the mapping process, it is necessary to mention that we have chosen to use the CRM class *E22_Man-Made Object* to model the artifact preserved, and therefore, described by cultural heritage institutions. Indeed, the CIDOC CRM uses this class to model “physical objects purposely created by human activity”. This class is, therefore, at the center of the SILKNOW ontology. In the example in Table 3, from the collections of Museu de la Seda in Moncada, the field “Denominación principal” contains the title given by the heritage institution to the object kept in its collections. We can express the underlying semantics as follows: the title of the artifact is “Abundancia” in the database. This means that we can interpret this field as a title, modeled with the class *E35_Title* in CIDOC CRM. The field name describes the relation that exists between the object (*E22_Man-Made Object*) and its title, which implies interpreting it with the property *P102_has title*.

Table 3. Information contained in the descriptive field “Denominación principal” and its mapping in CIDOC-CRM.

Fieldname	Content	Path
Denominación principal	Abundancia	E22_Man-Made Object P102 has title E35_Title

The SILKNOW ontology, consisting of the selected classes and properties, is publicly accessible (<https://ontome.net/profile/7> (accessed on 12 September 2023)) and documented via OntoMe [40], an ontology management system developed by the LARHRA research center into which the CIDOC-CRM documentation has been imported. To model the data collected by the SILKNOW project we have, therefore, used part of the classes and properties proposed by the CIDOC CRM model, but also those offered by an extension of this model, the Scientific Observation Model (CRMsci) [41] which is a formal ontology elaborated to integrate metadata about scientific observation. We have more particularly used the class *S4_Observation*, defined as “the activity of gaining scientific knowledge about particular states of physical reality gained by empirical evidence, experiments and by measurements”. This class seemed to us quite appropriate to model the historical and technical analyses resulting from the observation of ancient fabrics, usually contained in descriptive fields such as the one highlighted in Figure 1.



Silk velvet furnishing fabric

Place of origin:	Italy (made) Genoa (possibly, made) Florence (possibly, made) Venice (possibly, made)
Date:	1570-1600 (made)
Artist/Maker:	Unknown
Materials and Techniques:	Cut and uncut velvet, woven in silk and metal thread
Museum number:	147-1880
Gallery location:	Medieval & Renaissance, Room 63, The Edwin and Susan Davies Gallery, case 3

[Download image](#)

Summary | [More information](#) | [Download PDF version](#)

This silk was probably intended for use as a furnishing textile, as the vertical disposition and scale of the pattern are suited to wall hangings or curtains. This type of silk product was one of the richest because the making of velvet required skilled weavers and took a long time. Under the best circumstances, a weaver could progress at the rate of no more than 60 metres in a year.

Not surprisingly, velvet weavers earned more than those who specialised in other silk textiles, and their earnings increased in direct proportion to the difficulty of the work involved; in the velvet hierarchy, at the bottom sat the weavers of solid velvets, at the top the weavers of pile-on-pile velvets with brocading and bouclé gold wefts. The former earned about one third of the wages of the latter.

In western Europe, expertise in velvet-weaving was restricted at this period to various Italian cities (Lucca, Florence, Genoa, Venice) and to certain centres in Spain (e.g. Valencia), some of which had gained their knowledge through the import of Italian craftsmen.

Explore related objects

Category

- Textiles ▶
- Wall coverings ▶

Material

- silk ▶
- silver-gilt ▶

Figure 1. Example of historical and technical analyses from the Victoria and Albert Museum catalogue.

The quality of the data model was assessed by providing mapping rules between cultural heritage institutions' records and the SILKNOW ontology. We observed that all fields can be represented using the existing classes and properties of the SILKNOW ontology. In practice, we selected two representative records from each dataset and provided mapping tables and associated RDF graphs. The RDF graph in Figure 2 shows triplets that we have systematically created to model crucial information about the object described. On this graph, we visualize the triplets modeling the information about the production of the artifact (*E12_Production P108_has produced E22_Man Made Object*): when it was produced (*E12_Production P4_has time-span E52_Time-Span*), where it was produced (*E12_Production P8_took place on or within E53_Place*) and by whom it was produced (*E12_Production P14_carried out by E39_Actor*). As we are studying ancient fabrics using silk and specific manufacturing techniques, it is also essential to model information about the material(s) used (*E12_Production P126_employed E57_Material*) and the techniques employed (*E12_Production P32_use general technique E55_Type*)—information that is usually detailed in historical and technical analyses (*S4_Observation O8_observed E22_Man-Made Object*).

SILKNOW has also developed text and image analysis methods which, from the data describing silk-related objects, infer new properties on these objects, and ultimately enrich the existing metadata. We have thus modeled the integration of the new data produced by these analyses. The modeling should make it possible to make a clear distinction between these predictions and the original data and to provide the ADASilk users with information on the degree of reliability of this information. For this, we have chosen to use the Provenance Data Model (Prov DM) [42], recommended by the W3C (see Figure 3). Image or text analyses are represented in the form of 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 text—*E62_String* (text analysis)—as input (*prov:used*) and produces two statements as output (*prov:WasGeneratedBy* properties). Each of these declarations has an *E54_Dimension*. The date of the analysis can be specified (*prov:AtTime*). If necessary, we can specify the analysis module with a *prov:Agent class* (of type Software Agent) and document it (*E31_Document*).

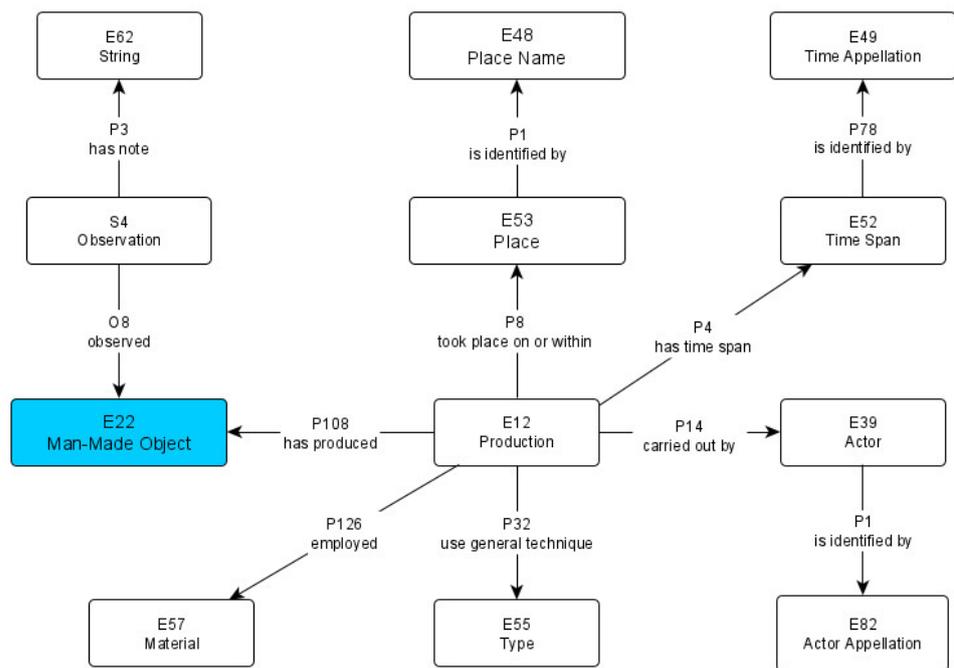


Figure 2. RDF graph: weaving process modeled with CIDOC CRM and CRMsci.

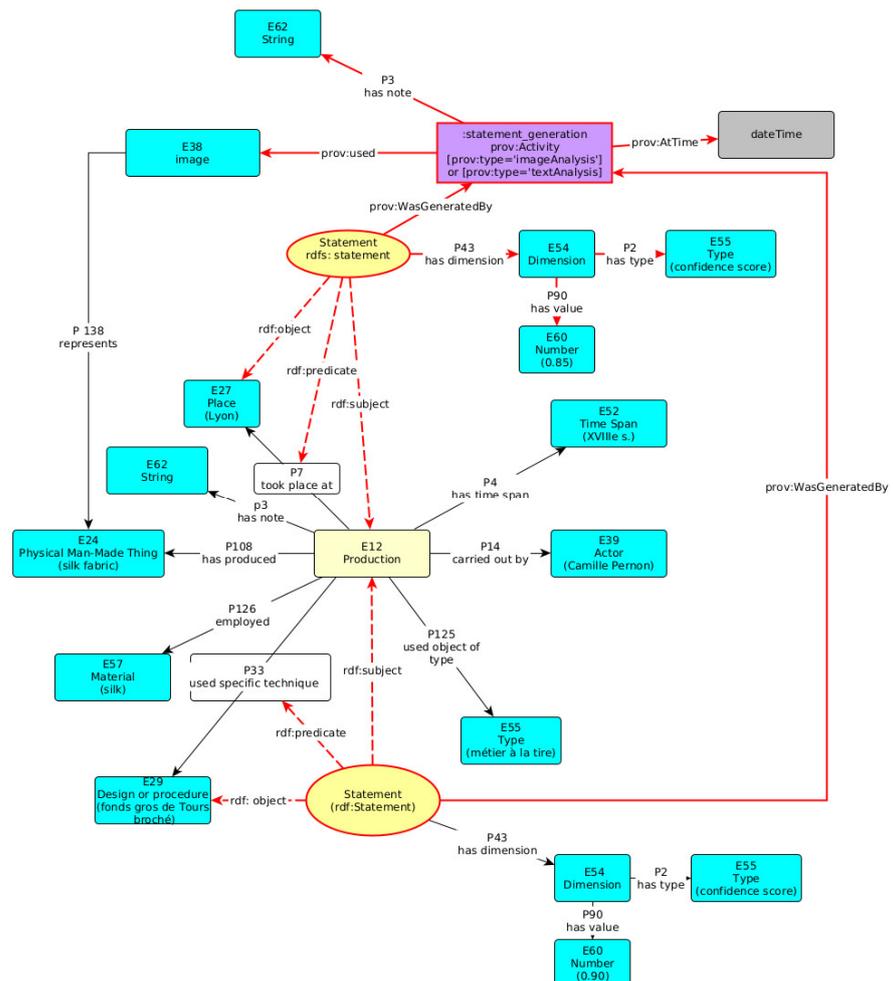


Figure 3. Integrating data from the image analysis module with the Provenance Data Model.

The metadata describing the textile artifacts are also very rich and this first mapping, aimed to store these metadata “as they were”, cannot fully reflect this richness. In particular, we note the use of free text to analyze the structure and decoration of the fabrics or to present the historical context of their production or even their use. During the first step of this work, this information has been stored as a “note” or using the CIDOC CRM and Scientific Observation Model: *S4_Observation P3_has note E62_String*. Table 4 provides examples from the collections of the Chiesa Madre di Caccamo (Sicily). Free text is used here to describe the complex construction of the patterned fabric:

Table 4. Information contained in the field “Costruzione” modeled with S4_Observation.

Fieldname	Content	Path
Costruzione	fondo in raso da 5, diffalcamento 2, faccia ordito, prodotto da tutti i fili e da tutte le trame di fondo. Opera creata dal raso da 5, diffalcamento 3 faccia trama prodotto da tutti i fili e da tutte le trame di fondo, unitamente a 2 trame braccate [...].	S4_Observation O8_observed E22_Man-Made Object S4_Observation P3_has note E62_String S4_Observation P2_has type E55_Type (Costruzione)

The extraction of information from these textual data (see Section 3.3) shows the extent to which these observations produce detailed technical analyses and new historical perspectives on these artifacts. It is thus possible to have access to information on the description of patterns and weaves, weaving techniques, or styles. By choosing to model this information with a simple note, however, it is not possible to fully reflect the semantics of this information, nor to provide easy access to it. Indeed, users will not be able to formulate fine queries on this data, which nevertheless offers particularly interesting information.

Fortunately, CIDOC CRM is a very flexible and extensible model. This means that, if necessary, it is possible to create new classes and properties to express new types of information, without modifying the basic structure of the model. This allows the development of more specialized extensions—such as the Scientific Observation Model, for example, or FRBRoo (Functional Requirements for Bibliographic Records) for the process of creation, production and expression in literature and the performing arts. In line with these compatible models we have, therefore, created a CRM extension designed to formally describe the process of creating and producing textile artifacts.

We have created 23 classes and 12 properties, accessible via Ontome (<http://ontome.dataforhistory.org/namespace/36> (accessed on 12 September 2023)). We adopted a “bottom-up” approach, first of all based on the collected data. We also worked closely with domain experts and ICT experts to verify that these classes and the properties we proposed to create were useful and meaningful. For example, we created the class *T1_Weaving*, that is a subclass of *E12_Production*, to easily express how a *T7_Fabric*, that is a subclass of *E22_Man-Made Object*, was woven. We also created the class *T8_Part Weaving* to express the fact that the weaving process can include different but simultaneous actions—especially in the case of complex fabrics such as patterned fabric—using various techniques as well as several warps and wefts. We have then created classes and properties to accurately model this complex process, which often involves the use of several *T25_Weaving Technique*, therefore, various *T21_Weave*, and different *T17_Weft* and *T16_Warp* (see Figure 4).

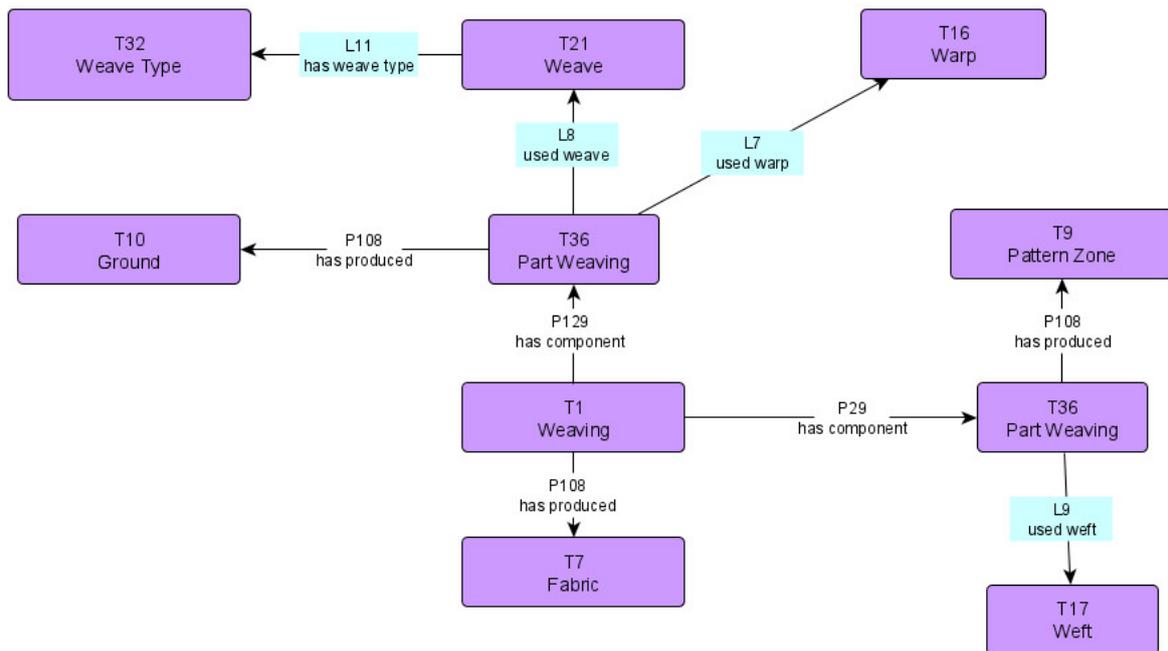


Figure 4. New classes and properties to model the weaving process.

Some of these classes also make it easier to create links between these data and the definitions provided by the SILKNOW thesaurus. This thesaurus provides additional information that users can access directly from the data they are currently studying. Thanks to these classes it is, therefore, possible to create links between the data, regardless of the language in which it is expressed, and the thesaurus, which not only enriches the user’s experience but also provides useful contextual information for a better understanding of the data itself. For example, the class *T32_Weave Type* makes it possible to create links between the types of weaves described in the technical analyses and the thesaurus (see Figure 5).

PREFERRED TERM	Derived Weave
DEFINITION	It refers to those weaves that are obtained by extension or modification of the fundamental ones, or by the application therein based on certain or appropriate evolutions. They are classified as derivatives of plain weave, twill and satin.
BROADER CONCEPT	Plain weaving
NARROWER CONCEPTS	Dice-cloth Extended tabby
RELATED CONCEPTS	Batavia Twill Burat Composite twill Cross twill Roman Twill Satina twill Stitched sateen

Figure 5. Examples of type of weaves (modeled as *T32_Weave Type*) defined in the thesaurus.

3.3. Towards Automated Annotation through AI: Text Analysis

A large number of culturally significant historical artifacts have been digitized and made available online. This means that experts in cultural heritage, and often the general public, now have the ability to search for and access information about artifacts instantly

even when these are stored in distant parts of the world. However, in order to make the digitized information truly sustainable in the long run, certain challenges to exploring and accessing it remain, two of which we will address now. The first challenge is, obviously, language. For example, the European Union, which includes most of Europe, a continent with a very closely linked cultural history, has 24 official languages. The second challenge is the lack of standardized representation of knowledge across the different archives or catalogs, i.e., the lack of a common ontology. Important data, such as the production technique or material used to create an object, is either not specified categorically, or when specified, it does not necessarily use the same term as in a different archive. This difference in terminology has a negative impact on the ability of an expert to find and understand related artifacts across archives. It also makes it harder for automated methods to provide useful features for the exploration of large groups of artifacts such as suggesting similar artifacts and providing visualizations of groups of artifacts along their properties. Most catalogs, for each digitized artifact, often have only a title, a short text description of the artifact, an image, and less often, an incomplete set of arbitrarily defined categorical descriptions in non-standard terminology, although not all of these are necessarily present for different artifacts. Here, we use the available text, in whatever language it is written in, both title and short description, to infer categorical properties of the underlying object, through the use of supervised text classification. The properties we infer are intended to be useful for cultural heritage experts. These properties can be aligned with a specific ontology (such as our extension of CIDOC-CRM) or thesaurus which can be cross-lingual (such as our silk heritage thesaurus).

3.3.1. Data Description

A supervised text classification approach requires a labeled dataset. Our dataset was obtained by crawling online catalogs relevant to silk cultural heritage. These included the following: Victoria & Albert Museum, London (VAM); Boston Museum of Fine Arts (MFA) and the Red Digital de Colecciones de Museos de España (CERES)—a catalog of multiple museums. Only pages containing information regarding silk fabric artifacts were retrieved. From these, the title, text description of the artifact, and any categorical fields present were extracted. These categorical fields were then normalized, that is, converted to a standard representation in English, defined by domain experts. The categories, their possible values, and the total number of samples for each can be seen in Table 5.

Table 5. Summary of the data: the categories we attempt to infer, the list of their possible values given our data, and the total number of samples for each variable.

	Technique	Material Used	Production Place (Country)	Production Date (Century)
Values	brocading, embroidering, knitting, lace, printing, sewing, velvet, weaving	cotton, leather, linen, metal_thread, wool, printed, other	Africa, AT, AZ, BE, UK, CN, FR, DE, GR, IR, IT, JP, NL, PT, RU, ES, SY, TR, US, South Asia	10, 14, 15, 16, 17, 18, 19, 20
Number of Samples	3783	4058	8116	7765

3.3.2. Methodology

Our methodology to infer the properties of a silk fabric from its short text description is based on a supervised text classification approach using a machine learning algorithm. This entails several steps which we will describe presently. We start by converting text into a normalized form and segmenting it into tokens which mostly correspond to words. Individual words are then converted, via a lookup table called an embedding layer, into (word) vectors also known as word embeddings. These vectors are learned, typically through co-occurrence, in a way that captures both the semantic and syntactic properties of words. We use multilingual aligned word vectors, where vectors that represent words in

one language are aligned with vectors that represent the same words in other languages. This means that to our learning algorithm, the same word in different languages will look similar (e.g., the English word “silk” will look similar to its Spanish translation, “seda”). In particular, we use the pre-trained multilingual aligned embeddings described in [43]. Finally, these vectors are fed into a classifier, a Convolutional Neural Network which outputs a predicted class value (1 class out of N possible predefined choices) via a softmax layer.

The architecture of our Convolutional Neural Network, shown in Figure 6, follows from previous work in applying CNNs to text [44,45]. The word embeddings are concatenated and a predefined number of convolutional filters (feature maps) with different fixed window sizes (kernel sizes) are applied to each possible window of words to extract “features”. These are then passed through a non-linearity and a max-pooling operation. The idea is to capture the most important feature for each feature map. Pooling over time (1d max-pool) deals with variable text lengths, we used a fixed maximum of 300 word-tokens, determined from analysis of the data. After the pooling, the different features for each window are concatenated together, regularized by a dropout layer and put through a final fully connected output layer with a softmax activation to give a distribution of probabilities over the classes. The general intuition behind the algorithm is that each window of size $h = 2, 3, 4$ learns to extract something similar to word n -gram features where $n = h$. In this work, the sequence of operations consisting of Convolution, Activation, and Max Pool form a convolutional block. A single convolutional block handles a single window size. We use three blocks in parallel, corresponding to the window sizes $h = 2, 3, 4$. We use the Gaussian Error Linear Unit (GELU) [46] as our activation function and the Alpha Dropout [47] variant of dropout. The activation function, dropout variant, dropout probability, convolutional kernel sizes (window sizes), and the number of filters were all treated as hyper-parameters and selected through hyper-parameter tuning on a subset of the data. In all experiments, the network was trained for 300 epochs, using mini-batch stochastic gradient descent, with a batch size of 64, and an initial learning rate of 0.005.

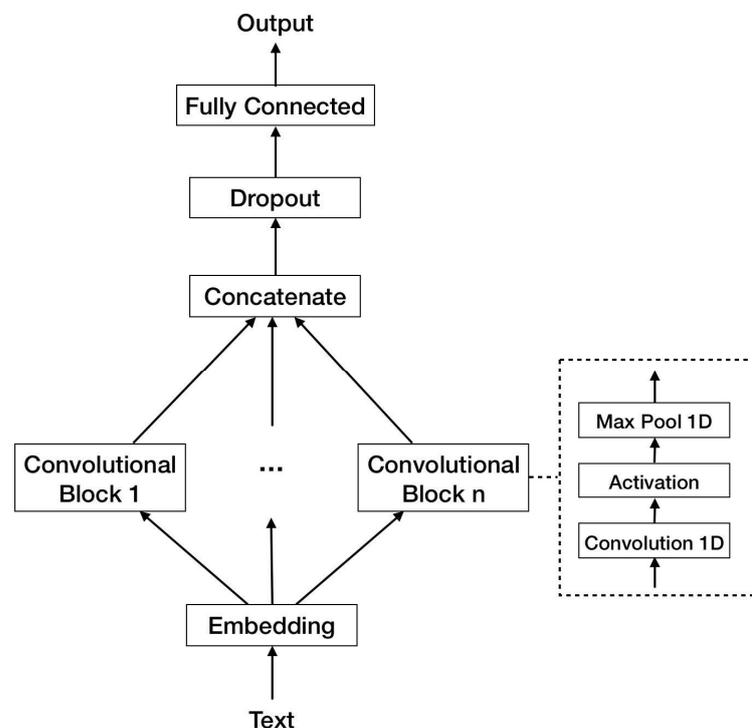


Figure 6. Convolutional Neural Network Architecture for text classification.

3.3.3. Experiments and Results

Our experiments focus on answering specific research questions. Our research questions are posed in terms of specific application scenarios:

1. Given labeled data (digitized artifacts) from one catalog (e.g., a museum), can we infer those labels (properties) in non-labeled data in the same catalog? The practical applications of this include the ability to infer properties in a catalog from a subset of that catalog's data which was semi-automatically or manually labeled, filling missing data, and semi-automatic conversion to a different ontology.
2. Given labeled data from one catalog, can we infer the labels of non-labeled data in a different catalog? Practical applications include aligning the ontologies of two or more different catalogs, and, if one can be labeled with a standard ontology then that effort can be leveraged to provide those categorical labels to other catalogs.
3. Given labeled data from one catalog, can we infer the labels of non-labeled data in a different language catalog? Applications are the same as in the previous case, but cross-lingual.

For each of these questions, we created a corresponding experimental evaluation scenario. For the first scenario, we use the data we collected from VAM and split it into separate train and test sets (Scenario 1). This artificial split of the data was performed using random stratified splitting, a technique that randomly selects the examples in each set while preserving the distribution of the labels from the original set; 80% of the examples are used as training and 20% as test examples. In the second scenario, we used the VAM catalog as the training set and the MFA collection as the test set (Scenario 2). In the third scenario, we again use VAM, which is in English, for training but we use CERES, in Spanish, as the test set (Scenario 3). Note, that we use VAM as a training set in all scenarios purposefully to enable better comparison between the results.

Our results given in Table 6 clearly show that it is possible to infer properties of silk fabrics from a short description of them, even across catalogs and across languages. The best results are obtained when labeled data from the same catalog is used (Scenario 1). The biggest challenge faced by a text classification algorithm when dealing with short descriptions, in the context of digitized artifacts, is how different these texts are across catalogs, in both form (syntax and length), content (semantics), and in the objects they describe (e.g., museums are not random collections of objects but rather curated, often thematically and locally). We can clearly see the difference in results with regards to "production place", and "production date" between Scenario 1 and 2. MFA descriptions rarely contain any words relevant to these two properties, while VAM often explicitly mentions regions, cities, and even countries with regard to one and dates with regard to the other. Thus, a text classifier trained on VAM descriptions is ill-prepared to handle MFA descriptions. MFA descriptions, though much shorter than VAM descriptions, usually do include techniques and materials, making the results in Scenario 2 for these properties much closer to Scenario 1. In the cross-lingual Scenario 3, we can see a further performance drop attributable, in part, to the difference in language. Descriptions in CERES, while focused primarily on depictions, often explicitly mention locations (e.g., cities) which help explain the better-than-expected results for "production place". City names are easier to align in pretrained embeddings than very domain-specific techniques and materials since these embeddings are primarily trained on Wikipedia and aligned through the use of dictionaries that are not domain-specific. With regards to dates, when explicit, VAM tends to express them using Arabic numerals and ranges (e.g., "1740–1800") while dates in CERES tend to use roman numerals (e.g., "siglo XIX") which is responsible for part of the difference in accuracy.

Table 6. Evaluation results (accuracy) for the different scenarios.

	Technique	Material Used	Production Place	Production Date (Century)
Scenario 1 (within museum)	97.6%	91.4%	97.4%	88.6%
Scenario 2 (across museums)	88.3%	77.7%	24.22%	48.2%
Scenario 3 (across museums and languages)	54.9%	59.8%	86.4%	20.7%

3.3.4. Text Analysis: Conclusions

We have shown that it is possible to infer, from a short text description of a silk fabric, properties relevant to the cultural heritage domain. We have also shown that this is possible in a cross-catalogue and cross-lingual setting. Applications of this development include but are not limited to, machine-aided improvement of categorical digitized data within an archive, changing ontologies of categorical properties in a catalog to align them with a different ontology or thesaurus, and helping a centralized resource (e.g., an open knowledge catalog that includes data from multiple museums) homogenize digitized artifacts across its sources.

The promise of this approach in the context of text descriptions of cultural heritage was already made evident in [48], as presented in their work on text descriptions of paintings from the Rijksmuseum. They used an Information Extraction approach rather than classification and thus are limited to extracting properties explicitly present in the text descriptions. These included all the properties we use: Technique, Material, Date, and Place, plus others such as Creator, Style, and Depiction. They used a total of 250 manually annotated texts. They reported an average F1 of 61.2% compared to a non-expert human average of 65.1%. Our work differs from this in two key areas. First, our classification approach is more generalizable, as it does not necessarily require information to be directly present in the text and more resilient to misspellings and non-standard grammar. Second, our work considers multiple archives and multiple languages and specifically evaluates the ability to learn across archives. In the broader context of machine learning, there is more similar work in image classification where [49] uses an image classification on photographs of silk fabrics to infer the exact same properties in this work, although with a much more limited set of labels and only within the context of a single archive. In the context of text classification, our approach is based on the text Convolutional Neural Network of [45] but uses multilingual aligned embeddings which is necessary for cross-linguality and minor architectural replacements.

We have provided a methodology for creating and evaluating a text classifier that can handle the challenge of working with highly heterogeneous data. Several interesting avenues of future work remain open, especially along domain adaptation. For example, it would be interesting to have pre-trained embeddings that are more tuned to the cultural heritage domain. The challenge to this lies in obtaining enough text to perform such an adaptation. A second avenue would be the adaptation of the multilingual alignment to include the use of such data as well as domain-specific dictionaries and thesaurus. The source code of the classifier is available online under an open-source license [50].

3.4. Spatiotemporal Visualization of Maps

The amount of openly accessible data about cultural heritage continues to grow sharply. Information available through institutional websites, journals and social media generates huge amounts of data. The visualization and analysis of this information has become an emerging field with extensive scientific activity [51]. For this reason, many old data visualization techniques have been redesigned, while new ones have been developed, too [52].

An important case in point is the visualization of spatiotemporal data [53]. Within it, cultural heritage information brings in additional complexity, due to the frequent uncertainty regarding both the time and space of historical events [54]. The STMaps tool [55], designed and developed in the SILKNOW project, aims at visualizing and analyzing spatiotemporal data stored and represented in a knowledge graph. It allows the interactive visualization of the data and the relationships between them, as well as their evolution over time. Using advanced techniques allows us to find unusual patterns and behaviors.

STMaps has been released into a GitHub repository [56], where it can be downloaded. This tool, however, is not designed for use in one single project or data domain. Rather, it pursues a more ambitious goal, as shown by its recent evolution into the STEVO framework [57]. It was designed and developed to represent specified data, in a model that defines how to visualize and interact with the data. Thereafter, a model that formalizes how to visualize this information is also proposed. The design of an ontology that implements this model -based on previous work- is also outlined in this section. Finally, the design and development of a software framework that allows the visualization of this information through a web application is presented. All these tasks were designed with the ultimate goal of providing any CHI with an innovative and freely available software resource, serving the need for spatiotemporal information, a very common requirement in the field.

3.4.1. Implementation

In order to configure the visualization aspect of the STMaps tool, we have used the Visualization Ontology (VISO) [58]. VISO is a generic approach, mostly related to two-dimensional space visualization. It was extended in order to manage virtual reality concepts and the data visualization techniques used in STMaps.

The tool is implemented in Unity, a technology that allows the development of a cross-platform application with state-of-the-art graphics. It can be used by embedding a WebGL plugin into an HTML web page. The WebGL plugin technology is executable in most of the operative systems. Before the rendering process starts, access to the domain knowledge graph must be defined, in addition to how to visualize the data and which data to visualize. Figure 7 shows a schema of a system embedding STMaps.

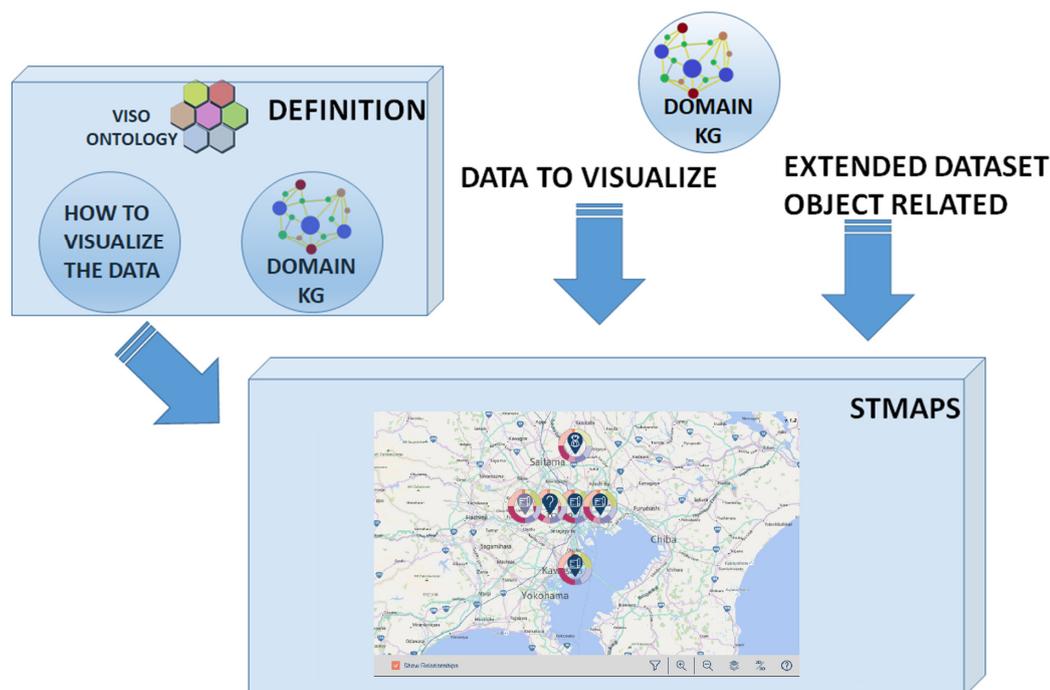


Figure 7. STMaps system integration schema.

3.4.2. Functionality

STMaps obtains the map images needed to represent the objects of the domain data from their built-in spatial coordinates. It also creates a quad-tree-based representation, which splits the map into clusters and groups the data according to the zoom level. Thus, depending on the active user zoom level, either cluster points are depicted by a representative icon or single points are directly displayed. This means that, by zooming in, clusters are divided into subclusters, until they are replaced by independent points. The cluster zoom levels and the icon aspect are determined in the configuration file. This clusterization is essential to keep the map readable without losing information.

The tools incorporate uncertainty related to space and time, a frequent problem in cultural heritage data. To deal with it, STMaps represents these objects with uncertain data by using special, different icons, and also by displaying data about alternative instances.

STMaps offers two possibilities to visualize the existing relationships (in properties such as subject, technique, and chronology. . .) between the displayed objects on the map. The first one is a classic, basic style, just connecting the related objects by colored lines. By displaying a window with extended information related to a data point, the user may select a set of object properties. With this selection, the tool depicts the relationships from this object to other objects with the same value in the selected properties.

The second way to display on the map the relationships between the objects is an outer ring, that shows segments filled with different colors. The size of these segments is proportional to the percentage of points with the same value for this given property of the object. For example, if the red segment covers 25% of the ring it means that 25% of the points have the same value in the property represented by the red color. This option is an easy, graphical way to detect objects with no relationships or a high number of relationships.

Both ways to show the relationships within STMaps are displayed in Figure 8.

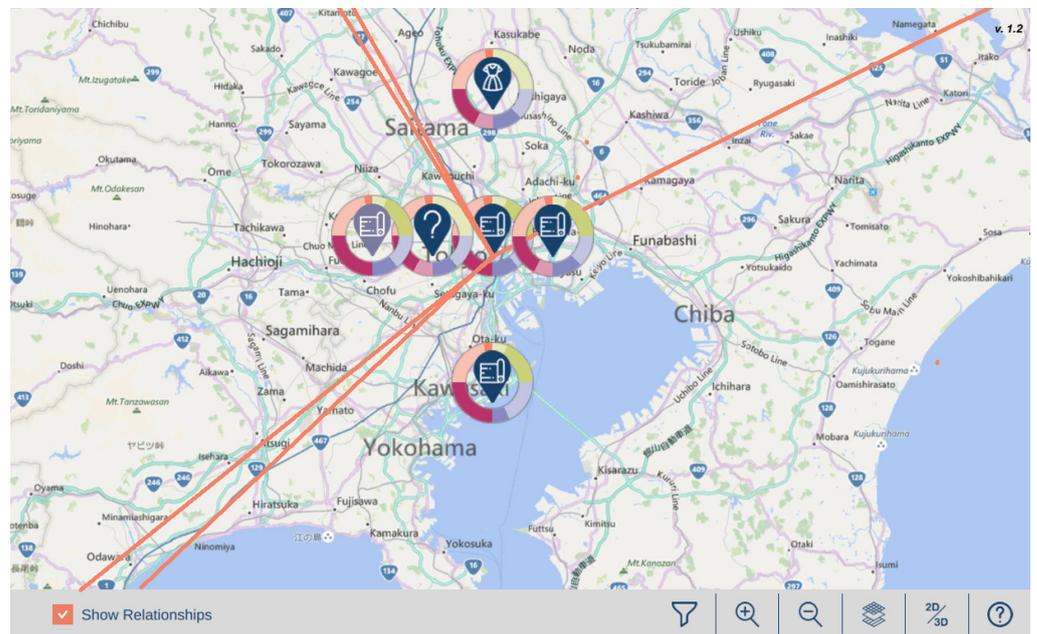


Figure 8. STMaps screenshot, showing the two possibilities to represent relationships between dataset objects.

STMaps also offers two ways to display the changes in data across time. The first is a classic timeline, showing a time interval with a time slider. The user can drag this slider in order to see the data status at a specific moment in time, according to the time resolution defined in the configuration file.

The second option is a time layer. With this functionality, the user may define a number of layers to visualize (from 2 to 4). Then, a time interval is associated with each layer and

the application represents in each layer all the data related to the corresponding chronology. The layers and the data are displayed in a 3D environment. A user interface allows the adjustment of any desired layer, for better visualization. This second option incorporates simultaneity, as the user can visualize data from different time lapses, at the same time. Figure 9 shows a screenshot of STMaps with the time layer functionality activated.

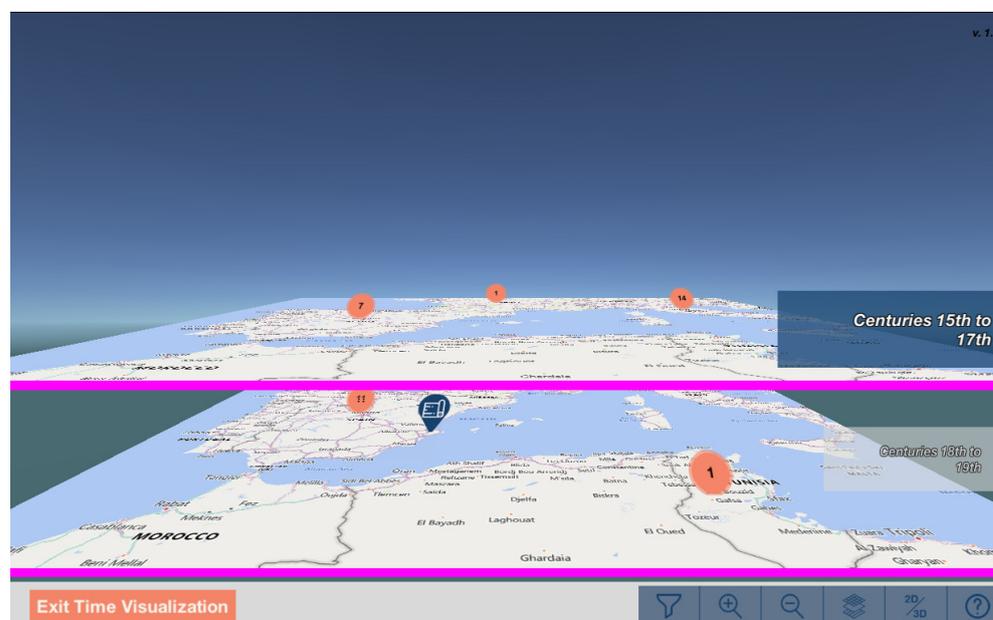


Figure 9. STMaps screenshot where the time layer visualization mode is active.

4. Conclusions: Heritage Institutions Need to Focus on Information Sustainability and Open Access

In conclusion, this case study shows the possibilities and demands that information sustainability places on heritage institutions, based on a collaborative project such as SIL-KNOW. Among the possibilities, interoperability does not only offer users the opportunity to discover data on repositories shared across institutions, instead of having to knock on every museum's door (or website) to find the data. It also allows the application of algorithms to the massive amounts of data that these repositories hold (or can arrive to hold), facilitating cross-lingual access to the information, searching for unexpected patterns and matches among previously unrelated pieces, or suggesting automated annotations for poorly cataloged objects, based on the information within similar objects' records. Data visualization is a new field that enhances our understanding of massive amounts of information, for instance, thanks to spatiotemporal maps. Crucially, it ensures the long-term sustainability of heritage information, since the efforts to standardize it also entail improving its quality and ensuring its permanent availability.

The demands imposed by this process are also evident from the previous pages. It is important to have a sound knowledge of the existing information environment, in order to align with standards or platforms, and not just reinvent the wheel in every new project. Terminology standardization (and its correct application) is a key issue, one that is, however, usually forgotten in favor of other, more glamorous tasks. Multilingual thesauri are cornerstones for any cataloging effort that aims at producing information that can still be found and properly understood in the long term. Museum records are an almost untapped resource in our era, increasingly hungry for good-quality data. Nonetheless, bringing them into the semantic web is a complex task, one that involves their mapping to standards such as CIDOC-CRM, while also extending those standards, and paying attention to detail. Turning the promises of artificial intelligence into realities useful for

the sector of cultural heritage requires collaboration with computer scientists and a fair understanding of the opportunities and limitations of algorithmic tools.

An institutional commitment towards open access is both a previous requirement and a result, in enabling all these possibilities. Without that commitment, these endeavors lack data, the basic fuel that they need to develop. It is true that a more positive attitude towards open access is now frequent among museums, compared to previous years. However, many challenges remain unanswered in this area, since not everything depends on the mere will of decision-makers. Technical and organizational challenges are still important, especially for small and medium-sized institutions, and cannot be overlooked. However, as long as researchers and developers have open access to data, more and more studies will confirm that advantages surpass complications by far, in this kind of collaboration.

Author Contributions: Conceptualization, D.M.; Investigation, M.G.S. and M.P.; Resources, A.L.M.; Writing—original draft, J.S.L., J.S.P., P.V. and L.R.; Writing—review & editing, E.M.R.; Supervision, E.A.P. All authors have read and agreed to the published version of the manuscript.

Funding: Research leading to these results has taken place within the research project “SILKNOW. Silk heritage in the Knowledge Society: from punched cards to big data, deep learning and visual/tangible simulations”, funded by the European Union’s Horizon 2020 research and innovation program under grant agreement No. 769504; as well as within “ClioViz. Visualización avanzada de datos históricos a través de mapas multidimensionales y gráficos interactivos” (ref. no. PID2021-126777NB-I00), a research project funded by MCIN/AEI/10.13039/501100011033/FEDER-UE.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data supporting reported results can be found at <https://github.com/silkknow> (accessed on 12 September 2023).

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Bradley, K. Defining Digital Sustainability. *Libr. Trends* **2007**, *56*, 148–163. [CrossRef]
2. Stuermer, M.; Abu-Tayeh, G.; Myrach, T. Digital sustainability: Basic conditions for sustainable digital artifacts and their ecosystems. *Sustain. Sci.* **2017**, *12*, 247–262. [CrossRef] [PubMed]
3. Chowdhury, G. Sustainability of digital information services. *J. Doc.* **2013**, *69*, 602–622. [CrossRef]
4. Borkopp-Restle, B.; McNeil, P.; Martinetti, S.; Miller, L.; Riello, G. Museums and the Making of Textile Histories: Past, Present, and Future. *Perspective* **2016**, 43–60. [CrossRef]
5. Brovarets, T. Apocalyptic Motifs on Century-Old Ukrainian Rushnyks through Today’s Digital Folklore Communication. *Colloq. Humanist.* **2022**, 1–21. [CrossRef]
6. Haffner, D. A Textile Thesaurus—Merging and Enlarging the Existing Vocabularies. In Proceedings of the ICOM General Conference, Milan, Italy, 4 July 2016; Available online: <http://cidoc.mini.icom.museum/wp-content/uploads/sites/6/2018/12/haffner-textile-thesaurus.pdf> (accessed on 12 September 2023).
7. Hooland, S.; Verborgh, R. *Linked Data for Libraries, Archives and Museums: How to Clean, Link and Publish Your Metadata*; Facet Publishing: London, UK, 2014.
8. Le Boeuf, P. Le Modèle conceptuel de référence du, CIDOC. De la sémantique des inventaires aux musées en dialogue. *Cult. Et Musées* **2014**, *22*, 89–111. [CrossRef]
9. Dörr, M. CRM Family of Models. 2018. Available online: http://dataforhistory.org/sites/default/files/dfh20180525_doerr.pdf (accessed on 12 September 2023).
10. Rosenberg, D.; Grafton, A. *Cartographies of Time*; Princeton Architectural Press: New York, NY, USA, 2012.
11. Lippincott, K.; Eco, U.; Gombrich, E.H. *El Tiempo a Través del Tiempo*; Grijalbo: Barcelona, Spain, 2000.
12. Sebastián Lozano, J. Mapping Art History in the Digital Era. *Art Bull.* **2021**, *103*, 6–16. [CrossRef]
13. Nouvelles Notices Versées sur Joconde. 2023. Available online: <https://www.culture.gouv.fr/Thematiques/Musees/Les-musees-en-France/Les-collections-des-musees-de-France/Joconde-catalogue-collectif-des-collections-des-musees-de-France/Nouvelles-notices-versees-sur-Joconde> (accessed on 12 September 2023).
14. Lettre d’Information Publiée par le Bureau de la Diffusion Numérique des Collections. Available online: https://www.culture.gouv.fr/Media/Medias-creation-rapide/lettre_info_41.pdf4 (accessed on 12 September 2023).

15. D’Agnelli, F.M.; Rizzo, M.T. Raccontare il Patrimonio Religioso: Identità ed Etica Nella Restituzione sul Portale Beweb. In *Nessuno Poteva Aprire il Libro. . . Miscellanea di Studi e Testimonianze per i Settant’anni di fr. Silvano Danieli, OSM*; Guerrini, M., Ed.; Firenze University Press: Firenze, Italy, 2019; pp. 113–130.
16. Brovarets, T. A Grave Cross on Eastern-Slavonic Ritual Towels. *Eikon/Imago* **2021**, *10*, 43–49. [CrossRef]
17. Renovamos IMATEX. 2021. Available online: https://cdmt.cat/es/renovem-imatex-10_05_2021/ (accessed on 12 September 2023).
18. Castells, M. *The Rise of the Network Society*, 2nd ed.; Oxford Blackwell Publishers: Oxford, UK, 2000.
19. Wallace, A. Words Mean Things (A Glossary). Open GLAM 2020. Available online: <https://openglam.pubpub.org/pub/the-glossary/release/1> (accessed on 12 September 2023).
20. Wallace, A. Clarifying “Open”. Open GLAM 2020. Available online: <https://openglam.pubpub.org/pub/clarifying-open/release/1> (accessed on 12 September 2023).
21. Pekel, J.; Nilsson, K. Making Impact on a Small Budget. How the LSH Museet Shared Their Collection with the World. *Europeana Pro* 2015. Available online: <https://pro.europeana.eu/post/making-impact-on-a-small-budget> (accessed on 12 September 2023).
22. Patti, E.; Quiviger, F. “Linking Venus”. *New Technologies of Memory and the Reconfiguration of Space at the Warburg Library. Between* **2014**, *4*, 1–29.
23. Smeets, R. Language as a Vehicle of the Intangible Cultural Heritage. *Mus. Int.* **2004**, *56*, 156–165. [CrossRef]
24. Schreiber, G.; Amin, A.; Aroyo, L.; van Assem, M.; de Boer, V.; Hardman, L.; Hildebrand, M.; Omelayenko, B.; van Osenbruggen, J.; Tordai, A.; et al. Semantic annotation and search of cultural-heritage collections: The MultimediaN E-Culture demonstrator. *J. Web Semant.* **2008**, *6*, 243–249. [CrossRef]
25. Gunzburger, C.A. Talking about Textiles: The Making of The Textile Museum Thesaurus. In *Textile Society of America Symposium Proceedings*; Digital Commons University of Nebraska: Lincoln, NE, USA, 2006; Volume 302, pp. 72–78. Available online: <https://digitalcommons.unl.edu/tsaconf/302/> (accessed on 12 September 2023).
26. Gunzburger, C.A. *The Textile Museum Thesaurus*; Textile Museum: Washington, DC, USA, 2005.
27. Van Steen, N. *Europeana Fashion Thesaurus v1. Deliverable 2.3. 2012*. Available online: <https://cordis.europa.eu/docs/projects/cnect/7/297167/080/deliverables/002-EuropeanaFashionDeliverable23EuropeanaFashionThesaurusv1.pdf> (accessed on 12 September 2023).
28. Calderón, P.O.; Puerto, F.P.; Verhagen, P.; Prieto, A.J. Designing a Thesaurus about Historical Silk for Small and Medium-Sized Textile Museums. In *Science and Digital Technology for Cultural Heritage—Interdisciplinary Approach to Diagnosis, Vulnerability, Risk Assessment and Graphical Information Models*; CRC Press: Sevilla, Spain, 2009; pp. 187–190.
29. Owens, L.A.; Cochrane, P.A. Thesaurus Evaluation. *Cat. Classif. Q.* **2004**, *37*, 87–102. [CrossRef]
30. Isaac, A.; Zinn, C.; Matthezing, H.; Van de Meij, H.; Schlobach, S.; Wang, S. The Value of Usage Scenarios for Thesaurus Alignment in Cultural Heritage Context. 2007. Available online: <https://api.semanticscholar.org/CorpusID:14684772> (accessed on 12 September 2023).
31. Alba, E.; Gaitán, M.; León, A.; Mladenčić, D.; Brank, J. Weaving words for textile museums: The development of the linked SILKNOW thesaurus. *Herit. Sci.* **2022**, *10*, 59. [CrossRef] [PubMed]
32. Halevy, A. Why Your Data Won’t Mix: New tools and techniques can help ease the pain of reconciling schemas. *Queue* **2005**, *3*, 50–58. [CrossRef]
33. Guarino, N. Understanding, building and using ontologies. *Int. J. Hum. Comput. Stud.* **1997**, *46*, 293–310. [CrossRef]
34. Arrêté du 25 mai 2004 fixant les normes techniques relatives à la tenue de l’inventaire, du registre des biens déposés dans un musée de France et au récolement. *J. Off. Lois Décrets* **2004**. Available online: <https://www.legifrance.gouv.fr/loda/id/JORFTEXT00000604037> (accessed on 12 September 2023).
35. Briatte, K. *HADOC Modèle Harmonisé Pour la Production des Données Culturelles*; Ministère de la Culture et de la Communication: Paris, France, 2012.
36. *International Guidelines for Museum Object Information: The CIDOC Information Categories*; International Committee for Documentation of the International Council of Museums: Paris, France, 1995. Available online: <https://cidoc.mini.icom.museum/wp-content/uploads/sites/6/2020/03/guidelines1995.pdf> (accessed on 12 September 2023).
37. Europeana. *Europeana Data Model*. Available online: <https://pro.europeana.eu/page/edm-documentation> (accessed on 12 September 2023).
38. Kondylakis, H.; Doerr, M.; Plexousakis, D. Mapping Language for Information Integration. Technical Report 385 FORTH-ICS. 2006. Available online: https://www.cidoc-crm.org/sites/default/files/Mapping_TR385_December06.pdf (accessed on 12 September 2023).
39. Doerr, M. *Mapping a Data Structure to the CIDOC Conceptual Reference Model*; ICS-FORTH: Heraklion, Greece, 2002.
40. Beretta, F. *OntoME, Ontology management environment*. In Proceedings of the 2nd Data for History Workshop, Lyon, France, 24–25 May 2018.
41. Definition of the CRMsci. An Extension of CIDOC-CRM to Support Scientific Observation, Version 1.2.8. 2020. Available online: <https://cidoc-crm.org/crmsci/ModelVersion/version-1.2.8> (accessed on 12 September 2023).
42. The PROV Data Model, W3C Recommendation. 30 April 2013. Available online: <https://www.w3.org/TR/prov-dm/> (accessed on 12 September 2023).

43. Joulin, A.; Bojanowski, P.; Mikolov, T.; Jégou, H.; Grave, E. Loss in Translation: Learning Bilingual Word Mapping with a Retrieval Criterion. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, Brussels, Belgium, 31 October–4 November 2018*; Association for Computational Linguistics: Stroudsburg, PA, USA, 2008; pp. 2979–2984.
44. Collobert, R.; Weston, J.; Bottou, L.; Karlen, M.; Kavukcuoglu, K.; Kuksa, P. Natural Language Processing (almost) from Scratch. *J. Mach. Learn. Res.* **2011**, *12*, 2493–2537.
45. Kim, Y. Convolutional Neural Networks for Sentence Classification. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), Doha, Qatar, 25–29 October 2014*; Association for Computational Linguistics: Stroudsburg, PA, USA, 2014; pp. 1746–1751.
46. Hendrycks, D.; Gimpel, K. Gaussian Error Linear Units (GELUs). *arXiv* **2020**, arXiv:1606.08415.
47. Klambauer, G.; Unterthiner, T.; Mayr, A.; Hochreiter, S. Self-Normalizing Neural Networks. *arXiv* **2017**, arXiv:1706.02515. [[CrossRef](#)]
48. Ruotsalo, T.; Aroyo, L.; Schreiber, G. Knowledge-Based Linguistic Annotation of Digital Cultural Heritage Collections. *IEEE Intell. Syst.* **2009**, *24*, 64–75. [[CrossRef](#)]
49. Dorozynski, M.; Clermont, D.; Rottensteiner, F. Multi-task deep learning with incomplete training samples for the image-based prediction of variables describing silk fabrics. *ISPRS Ann. Photogramm. Remote Sens. Spat. Inf. Sci.* **2019**, *IV-2/W6*, 47–54. [[CrossRef](#)]
50. SILKNOW Text Classification Code Link. 2021. Available online: <https://github.com/silknow/text-classification> (accessed on 12 September 2023).
51. Qin, X.; Luo, Y.; Tang, N.; Li, G. Making data visualization more efficient and effective: A survey. *VLDB J.* **2019**, *29*, 93–117. [[CrossRef](#)]
52. Wang, J.; Hazarika, S.; Li, C.; Shen, H.W. Visualization and Visual Analysis of Ensemble Data: A Survey. *IEEE Trans. Vis. Comput. Graph.* **2018**, *25*, 2853–2872. [[CrossRef](#)]
53. Alam, M.M.; Torgo, L.; Bifet, A. A Survey on Spatio-temporal Data Analytics Systems. *ACM Comput. Surv.* **2022**, *54*, 219. [[CrossRef](#)]
54. Windhager, F.; Filipov, V.A.; Salisu, S.; Mayr, E. Visualizing Uncertainty in Cultural Heritage Collections. In *EuroVis Workshop on Reproducibility, Verification, and Validation in Visualization (EuroRV3)*; The Eurographics Association: Saarbrücken, Germany, 2018. [[CrossRef](#)]
55. Sevilla, J.; Casanova-Salas, P.; Casas-Yrurzum, S.; Portalés, C. Multi-Purpose Ontology-Based Visualization of Spatio-Temporal Data: A Case Study on Silk Heritage. *Appl. Sci.* **2021**, *11*, 1636. [[CrossRef](#)]
56. STMAPS Github Repository. 2021. Available online: <https://github.com/silknow/spatio-temporal-map> (accessed on 12 September 2023).
57. Sevilla, J.; Samper, J.J.; Fernández, M.; León, A. Ontology and Software Tools for the Formalization of the Visualisation of Cultural Heritage Knowledge Graphs. *Heritage* **2023**, *6*, 4722–4736. [[CrossRef](#)]
58. Polowinski, J.; Voigt, M. VISO: A Shared, Formal Knowledge Base as a Foundation for Semi-automatic InfoVis Systems. In *Proceedings of the CHI '13 Extended Abstracts on Human Factors in Computing Systems, Paris, France, 27 April–2 May 2013*; pp. 1791–1796. [[CrossRef](#)]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.