Linking Garments to Knowledge: TextileBase as an Interdisciplinary Graph for Dress and Textile Research

Abstract

The article demonstrates how dress history and textile-related research can be enhanced through the interoperability of knowledge provided by a knowledge graph. The growing availability of digital cultural and historical data is not matched by a similar increase in their usability. Therefore, expanding the search radius to collections across disciplines and countries requires harmonisation and interoperability of knowledge. Reprex has created TextileBase – a knowledge base fully interoperable with libraries, archives, museums, the open knowledge system Wikidata, and open science repository systems. The article highlights key considerations when formulating searches and addressing terminology dissimilarities to ensure that data providers working across country, language, or disciplinary boundaries understand the intended meaning. To improve and streamline searchability in libraries for textual sources mentioning relevant historical garments, archives for their contemporary depictions, and museum collections for new artefacts, TextileBase transforms data and metadata into knowledge statements, links terms to an international controlled vocabulary, and carefully compares the works of various research and collection institutions.

Keywords: dress history, digital humanities, research data, data curation, knowledge base

Introduction

Clothing and textiles are interdisciplinary fields drawing on economics, technology, fashion, social anthropology, and history, to name just a few. Since 1980s the study of clothing history encompasses not only the evolution of patterns and designs over time but also the broader context of their creation and use [Welters and Lillethun 2018: 58–61], such as craftsmanship, trade, employment, gender roles, behavioural norms, information exchange, moral norms, prestige, and mythological beliefs, among many other aspects. Even when historical communities are poorly documented, the preserved clothing can provide dress historians with valuable insights regarding the society at a particular time and place.

As diverse as the history of clothing is, so too are the types of knowledge sources and data recorded to their study. The three main categories of sources in dress history from a knowledge carrier point of view are material, visual, and written [Taylor 2002]. The work of a researcher and a cultural and dress historian specializing in the late modern Baltics depends on

the availability of these sources and the researcher's ability to produce, evaluate, and connect data derived from them.

When studying these sources, one should aim to gather as much data as possible so that it can later be structured, organised, and compared with other data from a wide range of sources. Ideally, this comparison extends across different regions and time periods. Since these sources are housed in the collections of various memory institutions, it is the researcher's task to connect them to gain a comprehensive understanding of both the sources themselves and the data they provide. Different types of data emerge from different sources—material, written, and visual sources do not yield identical information. Only by integrating all available sources can a researcher aim to construct a complete picture. Therefore, it is essential to establish methods for connecting data as effectively and compatibly as possible.

Although much of the source examination and data collection still requires physical access, an increasing amount of data is being published online, allowing more research to be conducted digitally. The growing data availability is not matched by a similar growth in usability. Ideally, a researcher should ask the same question from all the relevant museums, archives, and libraries where a source may answer the question. But cross-institutional queries are yet more a promise than an available possibility because museums, libraries and archives even after decades of investment into interoperability store knowledge in different ways. In Latvia, asking all written and textual sources about a particular material is still mainly a future promise. This article demonstrates practical steps toward making this a reality: it requires moving beyond simple inventory searches to more semantically rich methods.

The same interoperability enhancements, if carefully designed, can broaden research beyond linguistic and national boundaries. However, as we increase our search radius to collections of different disciplines in different countries, harmonisation of knowledge becomes even more pressing. Location names follow different conventions, not only linguistic differences. Time appellation also has different formats in countries and disciplines, so a seemingly simple search for the first half of the 19th century related to the Latgale region requires a very careful elaboration of these quantifiers in German, Swedish or Russian-language collection of fashion, ethnography, or visual history. In this article, using examples of datasets based on original research in Latgale, Livonian Coast, and Setomaa, we show our conceptual system that can solve many of these problems.

Combining the expertise of the Reprex team and a dress historian, we began developing TextileBase [*TextileBase*]—an expandable graph database that brings together information on

garments from diverse humanities collections and related scholarly research. Our aim was to make our methods and findings reusable across disciplines, particularly for textile research and art history. How can the knowledge of these researchers be made interoperable with the research processes of adjacent disciplines that work with knowledge of historically used textile production, for example, in searching for more sustainable contemporary fashion uses? How can we ensure better and more efficient searchability in libraries for textual sources mentioning historically relevant garments, archives about their contemporary depictions, and various museum collections for new artefacts? Would it be possible to test simple or more complex hypotheses with an intelligent system that can sift through thousands or more catalogue items in various knowledge systems in the world? The aim of this article is to show our answers to these questions.

State of the art and our approach

We started working in 2024 with a single dataset made during numerous visits to the collections of the Latvian National Museum of History and the Ethnographic Open-air Museum of Latvia in 2018-2020, to see how modern advances in digital humanities can support the continuation and expansion of the research agenda behind the *Development of Folk Dress in Latgale in the 19th Century* (1.1.1.2/VIAA/1/16/092) project. The initial garment datasets from 19th-century Latgale, compiled in simple spreadsheets in a single file by Ieva Pigozne, a textile researcher and cultural historian, contained several tabular datasets on shirts, skirts, aprons, and other garments.

In the broader field of humanities research, scholars frequently work with data from GLAM institutions (galleries, libraries, archives, and museums), which are typically managed using relational databases. These systems connect, for example, an inventory number to a garment in a museum's collection and can return associated metadata such as acquisition dates or provenance. Relational databases are highly performant and well-suited for managing structured data at scale—such as the millions of catalogue queries processed daily by large libraries.

However, relational databases are less effective when data is distributed across diverse, heterogeneous systems. In such cases, graph databases offer a more flexible approach. They allow researchers to model and query complex relationships and interconnected concepts, even when the data originates from incompatible sources. While graph databases may be less optimized for high-volume transactions, they excel at uncovering patterns, associations, and hierarchies within and across datasets.

For our purposes, a graph-based knowledge base was the natural choice.

According to the ISO/IEC 24707:2023 standard [*ISO*, 2023], a *knowledge base* is a repository that contains structured information along with inference rules, representing human expertise and experience within a specific domain. In self-improving or intelligent systems, a knowledge base may also include information derived from solving past problems. The terms *knowledge base* and *K-base* are standardized under ISO, emphasizing their role in supporting reasoning, decision-making, and semantic interoperability.

Knowledge bases are frequently implemented using graph databases because their structure—often a semantic graph—can capture rich relationships between concepts, entities, and contexts. To store our knowledge base, we adopted Wikibase, a powerful, open-source graph database system originally developed to support the multilingual alignment of Wikipedia content. Wikibase has become widely used in digital humanities for harmonizing heterogeneous data sources, connecting library services [Bianchini et al. 2021; Sardo and Bianchini 2022; van Veen 2019], museum services [Pfeiffer and Gayo 2021], and archives, and creating interoperability among the institutional silos of these three types of heritage collections [Alexiev et al. 2020; Rossenova et al. 2022]. EU institutions themselves also use it [*SEMIC 2020*]. Wikibase was originally designed to harmonise the knowledge behind more than 300 Wikipedias written in different languages. Wikidata, that serves this purpose, is the world's largest open knowledge graph that contains more than 1.5 billion knowledge statements, many of which benefit TextileBase, too.

TextileBase is following the data curation and governance model of the Slovak Comprehensive Music Database, which has solved similar legal, organisational, semantic and technical problems in the *Open Music Europe* project, albeit not on material, tangible heritage objects, but intangible cultural assets, such as musical works and their recordings and scores[1]. This precedent illustrates how diverse datasets can be harmonised across legal and organisational boundaries, and how to get inspiration from the European Interoperability Framework (which is applicable for publicly funded digital services) and the evolving concept of the data (sharing) space.

A dataspace is an emerging approach to data governance. It recognises that in large-scale integration scenarios involving many partners, it would be prohibitively expensive and time-consuming to obtain an upfront unifying schema across all sources or to come to a legal

[1] See the planned *Slovak Comprehensive Music Database* [Antal 2023], which can be federated to further countries [Antal 2024] created in the [*Open Music Europe* 2023] project.

agreement on using or exchanging the data. It is an intelligent application that allows a near-instantaneous exchange, processing, sharing and provision of data on an "as-needed" or "as-permitted" basis while retaining complete control of each data holder over the conditions (e.g., who, when, and under what condition) of access to their data [Curr, 2020; EBU and Gaia-X 2022: 16; Nagel and Lycklama 2021]. TextileBase does not require a continuous online connection with different data providers: it consolidates garment-related knowledge periodically.

TextileBase is a semantic research infrastructure: it is a knowledge base supported by a graph database and a SPARQL API, which is designed to curate data and provide it for comparative textile and dress history research, with an evolving scope. Its current focus is on textiles from approximately 1820 to 1920 and from specific regions in present-day Latvia and southeastern Estonia. TextileBase integrates structured data and, where possible, digital surrogates (e.g., photographs, 3D visualisations) of garments and related artefacts to enable detailed cross-regional and diachronic analysis. It follows interoperability requirements with the European Open Science Cloud, Europeana and the evolving European Cultural Heritage Cloud.

It contains an increasing number of datasets, and some of the same datasets—especially those concerning Seto and Livonian garments—are also curated into a technically compatible infrastructure: the Finno-Ugric Data Sharing Space (FUDSS). While TextileBase is dedicated to textile-specific research questions, the FUDSS supports interdisciplinary research in Finno-Ugric studies, placing garments and visual culture in the context of language, music, ethnography, and regional history, allowing us to broaden our experience with methodology, and also to gain interdisciplinary feedback into textile research questions.

Both data sharing infrastructures follow a linked open data model, emphasising the creation of rich, extensible, and interoperable structured data. They also support the gradual accumulation of multi-modal data and enable cross-institutional collaboration through their flexible and federated data space governance model. Together, they aim to foster both domain-specific and interdisciplinary research across the Baltic and Finno-Ugric cultural areas.

Each artefact or secondary source—such as a garment, photograph, publication, video recording, or other knowledge object—is represented as a modular, entity-centered RDF dataset in a Wikibase document database, available in all five standard RDF serialisations (Turtle, JSON, JSON-LD, N-Triples, RDF/XML). For example, the *traditional Livonian women's festive woollen skirt*, collected by A. O. Heikel in 1902 in Miķeļtornis from the

Antell's Collection and held at the National Museum of Finland, is described at https://reprexbase.eu/textilebase/Item:Q347 and can be downloaded in Turtle format at https://reprexbase.eu/textilebase/Special:EntityData/Q347.ttl .

Whenever possible, such datasets include the link to a downloadable, viewable CC-BY licensed digital surrogate—a photograph, video, or 3D representation—embedded directly in the record. When full images cannot be shared, we provide either a thumbnail preview or a URL pointing to an external surrogate on a museum's website. This model enables flexible knowledge extension: new findings or new depictions of an artefact can be integrated into its dataset as elementary RDF statements. These modular datasets are the building blocks of growing semantic knowledge graph.

We organise these RDF records into regional databases, each associated with a virtual collection of digital surrogates when available. So far, we have created:

- a database of garments from Latgale (based on primary field research) [Pigozne et.al 2025],
- a database of garments attributed to Livonian users in Northern Kurzeme,
- a Seto garment database (blending linked and manually curated data),
- and an initial Riga metropolitan dataset.

At the time of writing our manuscript, with the help of a small grant from Finno-Ugric researchers and the support of the Estonian National Museum, we are also piloting a small Mari database to increase our geographical coverage, and gain experience with an increasing number of primary source languages. This dataset, like the Livonian and Seto data, are also shared via the Finno-Ugric Data Sharing Space, reflecting their relevance to Finno-Ugric cultural and linguistic research. These artefacts are mainly held in institutions in Estonia, Finland, and Hungary, with some Livonian artefacts stored in Latvia.

To illustrate our work and expectations, we provide two competency questions for a knowledge base that were guiding our methodological and technical research. In our work and in this article, we have chosen to focus on shirts as an essential garment of 19th-century traditional clothing.

Thus, the two questions are:

Q1. Where were linen shirts of traditional tunic cut, sewn by hand, still in use after 1850?

Q2. How can we search for information about shirts that match the criteria above in libraries, archives, and museums?

The two competency questions broadly set out our two methodological objectives in the design and creation of TextileBase as a research platform. We wanted to create knowledge base supported by a database management system that allows us to create datasets and then query them to investigate the use of traditional tunic cut, hand sewn shirts appearing in primary artefact sources or in secondary sources with time (and regional) limitations. TextileBase currently stores such information in RDF and allows such queries, therefore it can support a digital humanities research workflow. The second competency question relates to the ability to curate data from various museums, archives or libraries; considering also different modalities of still images or textual information; it allows us to ingest and store more and more structured knowledge so that we can gradually expand the time, geographical coverage, or the depth of answers that it can provide to research questions. This question relates to the problem that we do not have a central institution of textile research where we could go through all books, archival records, and historic garment artefact data which relate to the traditional tunic cut shirts. Museums, libraries, and archives still work in silos, with limited catalogue interoperability. Ieva Pigozne's prior work contained such information for the Latgale cultural region of Latvia. Our goal was to expand this dataset for comparative purposes across regions and contexts. Ideally, we would find textile researchers in, for example, Lithuania and Sweden who record the same data in the same tabular format. But in the absence of such researchers, we can also work with data created by different workflows of librarians, ethnographers, perhaps even researchers of sustainable fashion who physically inspect historical garments, recorded relevant information. They may call a shirt a chemise, and they may use many subcategories of linen, so content negotiation is necessary. But if they have such knowledge, we should be able to locate and curate it and then make it available for our primary target audience, the textile researcher.

Semantic interoperability and governance

As an answer to these competency requirements, we began building TextileBase, a linked open database around the first dataset. We defined TextileBase as a knowledge base that is fully interoperable with libraries, archives, museums, the open knowledge system Wikidata, and the open science repository systems (including EOSC and the fledgling European Cultural Heritage Cloud, EACH.) We chose the open-source Wikibase system, which has been used successfully many times in digital humanities and all essential GLAM collections. For

SPARQL queries, we import the database created by Wikibase into an Apache Fuseki Jena server; in the case of Finno-Ugric materials, we also place it into the Sampo semantic browser developed by Ikkala and the team [Ikkala 2021]; these functionalities can be tried out via the opening page of TextileBase on https://reprexbase.eu/textilebase/index.php?title=Main_Page. Our choice of the use of Sampo-UI was the excellent conceptual work of the Semantic Computing Research Group at the Aalto University working towards and extension of the 5-star FAIR model, often treated as a the "gold standard" in European open data towards an 8-star model that ensures a higher, cross-institutional interoperability, testing against semantic tools, and adding information on data quality [Hyvönen et al. 2024].

Semantic interoperability allows a shared understanding among different datasets to be joined and analysed together. It means that data arriving from different sources, such as Finna.fi or Muis.ee, or a researcher's CSV table, are mutually intelligible. This is not achievable without a critical revision of the workflows used by various museums, libraries, archives; and it is also not practically useful without legal interoperability. When different organisations create the datasets, adjustments, translations may be needed. And of course, we must consider the legal permission of data access and reuse. These issues go beyond the technical and semantic interoperability and require some form of data governance.

TextileBase follows the European Union's data sharing space model, inspired by the European Interoperability Framework, which integrates technical and semantic interoperability with broader considerations of legal and organisational compatibility. A data sharing space enables modular and federated data governance, where integration occurs on an as-needed, as-permitted basis. This model accommodates both large institutional partners and smaller, community-based collections.

For example, Finnish and Estonian institutions—particularly via Finna.fi and Muis.ee—publish well-structured, often RDF-compatible metadata and digital surrogates under CC-BY licenses, allowing TextileBase to import them. In contrast, Latvian national institutions generally do not provide reusable metadata or openly licensed images. Hungarian institutions may offer structured data and digital surrogates under various licenses.

While Finna.fi and Muis.ee provide reusable metadata and digital surrogates under open licenses, it is important to note that these platforms are designed as generalist, nationally representative infrastructures, not as domain-specific tools for textile research. Their structured metadata—while consistent and machine-readable—lacks textile-specific detail, such as structured garment typologies, information on weave structures, or material properties. For

instance, Muis.ee's RDF exports typically rely on free-text descriptions in Estonian and do not use textile-relevant ontologies or thesauri. We therefore do not simply ingest their RDF files or other structured data "as-is"; rather, we enrich and reinterpret the records, increasingly using linguistic models and domain-specific vocabularies to align them with comparative textile research needs. These generalist platforms offer a valuable foundation, but our work begins where theirs appropriately ends.

A key benefit of the data sharing space model is its support for heterogeneous curation workflows, including collaboration with smaller institutions. For example, the regional museums of Preiļi (Latvia) and Värska and Obinitsa (Estonia) provided open access and active and helpful collaboration, though they lack structured data services and are not yet integrated into Muis.ee. In these cases, the TextileBase team manually curated records from Excel files and created digital surrogates through fieldwork.

The federated, modular governance model accommodates diverse levels of digitisation, legal openness, and technical readiness—allowing TextileBase to function as a scalable, collaborative research infrastructure across institutional and national contexts.

In some cases, institutional or geopolitical constraints restrict data access. Both Latgale and Setomaa border the Russian Federation, and historically significant areas of Setomaa—including Pechory (Petseri) and its other parts from the Pskov region—are now within Russian territory, just like the Mari El Republic. Many relevant artefacts and archival materials are housed in Russian collections. Since the onset of the war in Ukraine in 2022, collaboration with Russian museums and research institutions has become impossible. Nonetheless, the federated architecture of TextileBase allows for the inclusion of basic metadata about known Russian holdings, preserving the possibility of future reintegration and maintaining a broader view of textile heritage in the region.

Usability for textile research

A typical first question from a dress historian might concern when and where a garment type was used. Using our example, which contains data on 19th-century linen shirts, a dress historian might want to explore where and when similar shirts were worn. However, to get meaningful results, the question needs to be more specific. A better query might be: "Where were linen shirts of traditional tunic cut, sewn by hand, worn in 1850–1900?" The responses to this query would be specific enough to help researchers address key issues in dress history today.

Linen tunic shirts were long-used undergarments in Europe, and by the late 19th century had become largely restricted to peasant dress, as they were no longer worn by the upper

classes. Peasants, the largest social group in most European countries at the time, were the last to abandon traditional clothing. Many elements of their attire were still handmade at home, and as the precisely dated garments, interviews, and photographic evidence shows, linen shirts were among the garments crafted domestically for the longest time. However, from the mid-19th century onward, peasants gradually transitioned to clothing that reflected urban fashion, often made from industrially manufactured fabrics and cut in styles influenced by towns.

Tracing the decline of traditional garments, such as hand-sewn linen shirts of tunic cut, is particularly informative. These shirts disappeared earlier than machine-sewn linen shirts and those with more modern cuts. Observing the spread of this innovation can help researchers identify where fashion changes first emerged and were adopted most rapidly. When combined with research results or historical records from other areas, such as history of agriculture, transportation, and education, these insights can lead to a deeper understanding of rural life in the second half of the 19th century, peasant interactions with towns, and possibly even patterns of trade and information exchange.

Moreover, such research allows for comparisons across communities, regions, and countries. For example, in Latvia, rural populations near Riga adopted fashion novelties earliest, while the people living on the Livonian Coast or Latgale followed a decade or two later – northern and eastern Latgale being slower than its southern and western parts. Adjacent territories in present-day Belarus, which belonged to the same Vitebsk Governorate of the Russian Empire as Latgale, or Setomaa which belonged to the neighbouring Pskov Governorate, underwent similar changes yet another couple of decades later. Extending these comparisons to other European countries would provide valuable insights into rural communities' transformation, including clothing changes, on a much broader scale. This, in turn, would lead to new discoveries and a deeper understanding of historical developments.

Without going into technical details of our data model, suffice to say that to answer question No.1, we needed to organise data in a way that contains information about the five properties of the garment.

- Garment type: we are looking for shirts, but mindful that a chemise is not the same as *krekls*. Therefore, we use a hierarchical taxonomy that is internationally and widely used for garments and other aspects of art history.
- Cut of the shirt: tunic, meaning that the front and back of the shirt are made of one piece of fabric which is not cut at the shoulders and with a hole left for the head.

- Fabrication method: we are looking for references of being sewn by hand, as opposed to machine-sewn and industrially produced.
- Used in the timespan of 1851-1900: careful distinction between use date, collection acquisition date, and fabrication date.
- An appellation for place of use: Latgale itself is a very fluid term, but we need to ensure that whatever contemporary or current settlement, parish, region or country is mentioned, we understand it correctly.

The dress historians input in designing the database are crucial to understand the nuances between a Latvian *krekls,* English *shirt,* and French *chemise* (a term also used in English), to find a useful category for the cut of the shirt, but for contemporary and historical place appellations we can rely on geographers, or for categorisations on the work of generations of art historians and their librarians embedded in the making of the Art & Architecture Thesaurus (AAT) [*Art & Architecture Thesaurus*] to provide a consistent terminology about the fields of art and architecture.

One might assume that translating a simple Latvian word like *krekls* is straightforward, but this is not the case. In English, 19th century historical shirts have two distinct names: shirt for men's garments and chemise for women's. However, this distinction cannot directly be applied to traditional Latvian, Estonian or other peasant clothing of Northern and Eastern Europe. The chemise, as known in English, differs significantly from peasant women's traditional shirts.

Many collections around the world use the English-based but widely translated AAT thesaurus to describe garments. This hierarchical taxonomy is a very useful start to cast our net to find potential Latvian *krekls* in a foreign collection. Latvian men's and women's shirts were similar, with subtle distinctions in length and decor. While they could feature decorative elements, they could also be undecorated, making the differences minimal and sometimes barely noticeable. TextileBase can extend AAT with custom categories to better bridge cultural garment types, for example, distinguishing chemise and the *krekls* by adding its wearer – Latgalian or Livonian women. From now on, we can go hunting for further female shirt data.

Figure No. 1. Fragment of the female festive shirt (CVVM 12753) from Šķilbēni municipality in Latgale, Latvia. Collection of the National History Museum of Latvia. Photo by Ieva Pigozne.

From data to knowledge

When we build a knowledge graph, we want to convert data to knowledge statements.

The following data points are not very informative in themselves:

data: “Rēzekne”, “Latvian National Museum of History”, “CVVM 12667”, “festive shirt”, “woman”.

Metadata makes a potentially informative data point much more usable:

metadata: “place of find”, “holding collection institution”, “inventory number”, “garment type”, “intended user”.

For the researcher who created the original fieldwork journal and its tabular spreadsheet version, it is clear that “Rēzekne” is a location that does not refer to the “Latvian National Museum of History” but the place where the “festive shirt” was found. To make the research diary usable for another dress researcher familiar with the Baltic region, we should provide the metadata at least; ensuring metadata attached enhances interoperability.

An even higher level of reusability and interoperability can be reached if we transform the data and the metadata into knowledge statements. When we start to collaborate with researchers who do not know the Baltic region, or do not speak the local language, or work in an adjacent field of sustainable fashion research, we must be more specific. Is “CVVM 12667” identifying the shirt, the museum, the village, or the woman?

- “CVVM 12667” is issued by “Latvian National Museum of History”.
- “Ieva Pigozne” assigned the following characteristics (based on her inspection of the original shirt: it has “tunic cut” design, it is made of “linen” and “cotton”, and was intended for “women”, and it used “hand sewing” on “plain weave”.

Such a transformation makes data more reusable by reducing ambiguity. And we can enhance this reusability even further.

Linking the “identified by” property to controlled vocabularies and their carefully reviewed translations, we can convey the message that “CVVM 12667” is a unique identifier of the garment under observation. Linking the word “women” to an international controlled vocabulary, we can ensure that users who do not speak English but look for similar garments can replace “women” with “woman”, and then “woman” with “*kvinne*” or “*nő*”.

Knowledge statements carry truth values and are verifiable. The claim that “CVVM 12667” is issued by “Latvian National Museum of History” can be verified easily to be true. But much scientific advancement happens through proofs and refutations. Adding provenance

statements allows the next users to determine how much they trust the information about "CVVM 12667". Was it inspected by a general museologist, a dress historian, or a researcher of sustainable fashion? Did the inspection happen recently or in the last century? Are their inspection reports available?

Does the inspection report of "Ieva Pigozne" corroborate the "Daniel Antal" inspections or cast doubt on them? Daniel Antal can measure length (corroborating such a claim by Ieva Pigozne) and can create colour-accurate photographs and assert the accurate description of the colours of the linen or decorations, but he may not see the difference between a chemise and a *krekls* (his observations on garment classification are useless.) Ieva Pigozne is a competent dress historian with no graded colour scale (therefore the other author's evidence may add further detail, but only in terms of defining the colours of parts of the garment.) Adding provenance information and competency constraints make data reuse more precise and responsible.

If the authors, inspecting a shirt embroidered with wool yarn, made a particular claim about the type of wool from a particular variety of sheep, is this really a credible claim? Not all observations require disciplinary expertise — but provenance clarifies when they do. Perhaps neither of the two authors would be qualified to make this distinction; however, unless somebody wants to bring our observations to a faraway disciplinary field (farming history and sheep varieties reconstructed from historical wool species [Ryder 1984]), our observation about the wool is sufficient. If a farming historian wants to know more about the wool, that researcher at least knows where to find another wool specimen from the 19th century Latgale but cannot rely on our observations regarding breed history.

The transformation of data to knowledge statements supports structured, qualified reuse across disciplines. The best way to show the advantages of this increased reusability is by improving how we handle geographic labels — leveraging tools from historical geography. Place appellation is particularly difficult in the Baltic region because place names have changed significantly over time and between languages. Places were differently named by Russian, German, and Polish authorities in the Russian Empire, renamed and reorganized by Latvians during the interwar period, by the Soviet authorities, and in the modern independence era.

We can illustrate our solution to this issue using examples from the Livonian Coast. To begin with, place names in heritage data are often ambiguous. A single settlement may be recorded in multiple languages, scripts, or historical forms. For example, *Mazirbe*—a coastal village in western Latvia—may appear in records as *Irē* (Livonian), *Мазирбе* (Russian), *Klein-*

Irben (German), or *Vähä-Irben* (Finnish-German hybrid). A museum object tagged as originating from "Irben" may therefore go unrecognized unless these variants are reconciled.

To address this, we developed a multilingual gazetteer for the Livonian Coast [Antal et al. 2025a; Antal et al. 2025b]. It aligns historical and contemporary names using canonical identifiers, geographic coordinates, and language-tagged aliases. The region was chosen for its compact scale—roughly a dozen villages—yet complex multilingual and administrative history. We cross-referenced open datasets (e.g., GeoNames, Wikidata) with scholarly sources on Finno-Ugric linguistics, ensuring accuracy across Livonian, Latvian, Russian, German, Finnish, and other naming traditions.

Instead of storing names as isolated strings, we represent each location as a knowledge node, enriched with alternate names and linked to stable identifiers (e.g., https://www.geonames.org/458007/lielirbe.html for Lielirbe). These nodes also include geographic coordinates and relationships to current administrative units. This allows us to infer, for example, that an artefact labeled from "Gross-Irben" in a Finnish archive likely corresponds to modern-day *Lielirbe*, even if the term has fallen out of use.

This model enhances recall and precision in data discovery—especially in heritage collections where place names reflect changing borders and multilingual cataloging practices. It builds on prior work in Wikibase-based reconciliation [Bianchini et al. 2021; Fagerving 2023] and sets the foundation for culturally sensitive geographic reference in Finno-Ugric contexts.

Breaking institutional silos with organisational interoperability

We can envision that a dress historian would seek information about shirts matching the criteria described above in digitised databases and catalogues of libraries, archives, and museums. The most valuable information on historical topics comes from examining the main dress history sources—material, visual, and written. Therefore, we can assume that the researcher would focus on these sources or their published editions.

The specialization of memory institutions determines which sources of dress history they primarily store. Museums typically preserve material and visual sources, including garments, accessories, textile fragments, drawings, paintings, lithographs, and photographs. Archives focus on written records and visual materials, mainly in paper formats, such as documents, manuscripts, drawings, and photographs. Libraries primarily house books but may also hold visual and written sources, like archives. While museums may contain some written

sources, neither archives nor libraries are likely to store material artefacts. This specialization influences how a dress historian approaches searches within the catalogues of these institutions.

Digital humanities researchers and curators of digital collections have been working in the past two decades to break these silos and allow us to search parallel for information about Latvian traditional cut, hand-sewn, linen-made shirts in libraries, archives, and museums. Such parallel knowledge extension overlaps with the AI use of knowledge graphs.

Archive catalogue searches using keywords like “linen” or “shirt” often yield little, especially with niche descriptors, like “linen”, “shirt” “traditional tunic cut”, “sewn by hand”, “1851-1900”. However, archival records (documents) often have much more precise dating than museum collections. Unless we are lucky to investigate a special dress history collection within an archive, we will likely search for images of people—for example, women, or more specifically, peasant women—from 1851 to 1900. Most images made of women in this period show the women fully clad. This is important, because we see their garments.

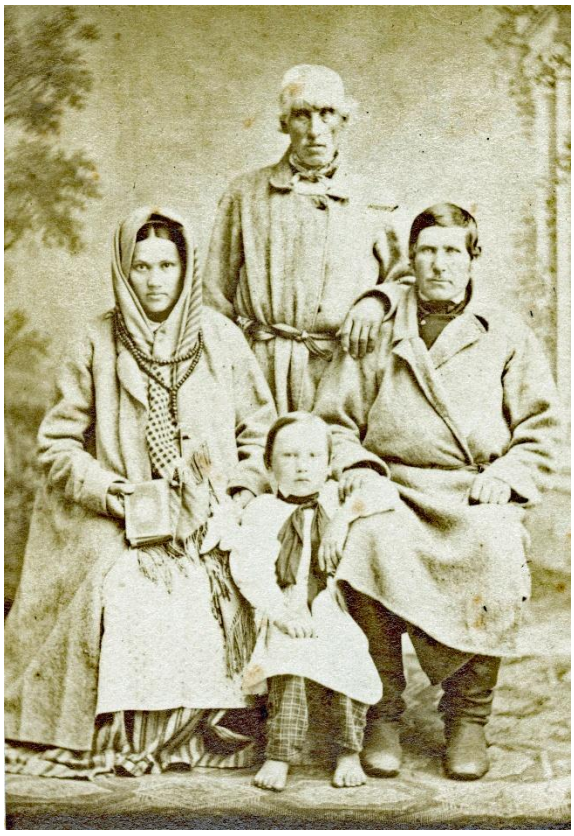

Figure No. 2. Latvians from Daugavpils area in 1869. A photograph published by Gustav von Manteuffel. [Manteuffel 1869: 43.]

Metadata about intended wearer, location, and usage period should link the shirt to other records across systems. From photographic evidence that is dated 1869, we can infer that at least one woman was wearing such a garment in 1869. From an artefact in a museum, we may

only be able to appellate for the second half of the 19th century, as the museum does not have record of a particular event when a specific woman wore that particular garment.

In technical terms, this requires a careful design of classification of nodes and properties (edges) that work in museums and archives, for example, that can connect the concept of a Woman (as a human being with a female sex at birth or such a gender identity) and the intended user to a garment. Precisely this means that we design TextileBase to contain this vocabulary from Records in Context, the new archival international standard that superseded the previous international standards in 2023. Although not yet widely adopted, RiC's [Archives Expert Group on Archival Description 2023], backward compatibility with ISAD(G) allows us to ensure strong archival interoperability.

Libraries, with their extensive collections, also contain books that systematically publish sources, including catalogues issued by museums, archives, or private collectors. However, not all catalogues feature primary sources relevant to dress history. Therefore, the researcher's search would need to specify museums and collections containing historical dress artefacts from the 19th century. Institutions or collectors focusing on aircraft, musical instruments, or firefighting, as well as museums specializing exclusively in ancient history or the Soviet occupation, would be excluded from the search. Even within history museums, some catalogues may not feature textiles in their topic-specific collections. The researcher would need to determine, for example, which catalogues from each country's National History Museum include artefacts relevant to dress history. To improve library search interoperability, TextileBase incorporates DCTERMS, the modern ontology-based extension of Dublin Core. [*Dublin Core* 2020].

Museums not yet visited on fieldwork may provide access to further surviving linen shirts and their fragments from the second half of the 19th century. However, such artefacts are often imprecisely dated. Collectors typically recorded where an artefact was acquired, which usually made it possible to determine its origin. Original owners rarely knew the exact date of an item's creation, particularly if it was a family heirloom, and the most artefacts have since been dated by museum curators using relative dating based on their expertise and the characteristics of the object. The qualifications of those involved in collecting, preserving, and studying these artefacts vary, as many museums do not employ dress historians. As a result, the descriptions of garments may not always use professional terminology. A researcher's search strategy may focus on garment type (e.g., *shirt*) and material (if indicated).

Chronological criteria could be refined using search terms such as *19th century*, *second half of the 19th century*, *1875*, or *1860s*.

In technical terms, we designed TextileBase to understand the vocabulary of CIDOC-CRM, the conceptual model of museum systems [Bekiari et al., 2024]. CIDOC-CRM and RiC are compatible, and together with DCTERMS, they enable efficient, automated cross-institutional discovery of potential garment artefacts, for example, potential peasant shirts, in much higher number, search accuracy and with far less effort than traditional methods. Of course, to retain human control, expert review by dress historians and curators remains essential for validation and classification.

The previous examples also highlight the importance of differing data quality and the importance of checking the provenance of the data that we import. Is it certain that we see a peasant shirt on the picture? Is the time or place appellation of garment certain? To improve data trust, TextileBase includes provenance metadata: who observed or measured the item, when, and with what qualifications. Who measured the length of an artefact and when? For adding such information to TextileBase, we use the foundational PROV model and the PROV-O ontology [Gil et al. 2013; Lebo et al. 2013]. PROV-O is widely used by data stewards in museums, libraries, archives and open science repositories.

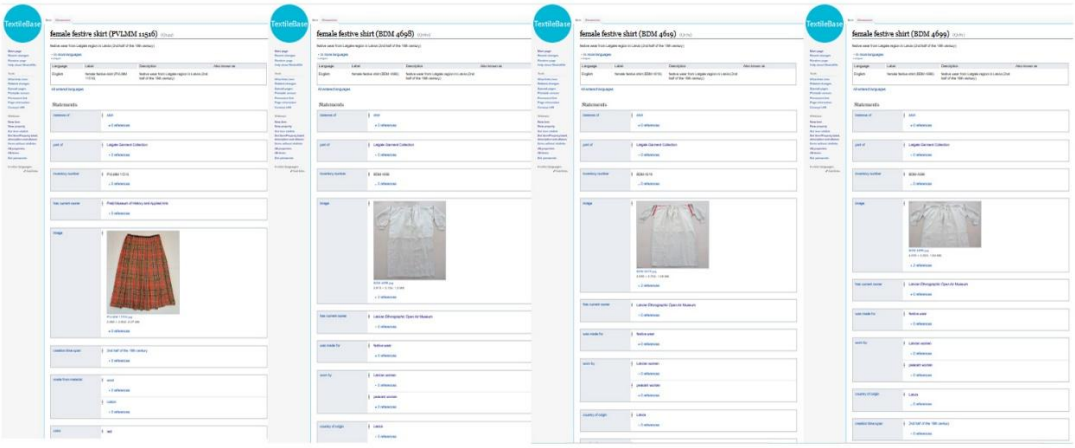

Figure No. 3. Entry examples on TextileBase.

Conclusion and further work

In this article, we presented the development of TextileBase, a new open knowledge graph. It operates as a data sharing space between various organisations that provide open or

on-request access to their databases enabling the gradual integration of textile and dress history research data.

Researchers constantly collect data and information in the language they know from the point of view of their research agenda, with their methodology. Libraries, archives, and museums organise potentially informative data in very different ways about garments as artefacts or actual outfits worn, for example, in early photographs. Organising interoperability among their datasets or databases is not only a legal and technical interoperability question.

Achieving true organisational interoperability requires analysing institutional workflows and metadata practices. Creating such a high level of interoperability is a challenge not only on the technical level but also requires a careful comparison of the works of various research and collection institutions, as we showed with the two competence questions. It also demands semantic interoperability well beyond FAIR-compliance. We must ensure that art historians, dress historians, contemporary and historical textile researchers, librarians, archivists, and museologists are able to interpret and align each other's data with shared understanding. Time and place references must be designed to integrate with geospatial and open data infrastructures. This process and semantic organisation take a long time, and unfortunately, it must be addressed early, before populating the knowledge base at scale.

TextileBase as a graph is modest at the point of writing, but we hope it will grow quickly. Next, we aim to link it with Europeana and the European Cultural Heritage Cloud (ECCCH). We also plan to expand the graph at a granular level adding new datasets and invite individual researchers to provide us with smaller but high-value datasets.

As the knowledge graph evolves, TextileBase increasingly supports the spatial, temporal, and typological queries needed to address these and more complex questions. This reflects the authors' shared interest in digital humanities methodology, but with a strong emphasis on practical usability beyond the "computer lab"—especially in collaborative and field-based settings.

TextileBase is a developing research tool, created not only to organise data but to support hypothesis-driven historical textile analysis. The technologies described in this paper offer significant opportunities for researchers in textile, dress, and broader cultural history. We hope that our work will serve as a positive example and inspire others to join with their collections and research data, and to pursue further cooperation in the form of joint projects.

Acknowledgements

The authors are very grateful to Anna Márta Mester for the data management of TextileBase and Asmah Federico for the curation of digital surrogates.

In its initial stage this work was supported by the Short-term Scientific Mission *Collecting, Evaluating, and Connecting Data for Dress History* carried out by Ieva Pigozne within the COST action *Europe through Textiles: Network for an integrated and interdisciplinary Humanities (Euroweb)* CA19131. Further Ieva Pigozne's work and contribution to this paper is supported by the *Recovery and Resilience Facility project Internal and External Consolidation of the University of Latvia* (No. 5.2.1.1.i.0/2/24/I/CFLA/007), grant *Contextualization of Traditional Clothing: Reassessing the Connection Between Riga and Latgale* (Grant No. LU-BA-PG-2024/1-0023).

References

Alexiev, V., Tarkalanov, P., Georgiev, N., Pavlova, L. (2020). Bulgarian icons in wikidata and EDM. *Digital Presentation and Preservation of Cultural and Scientific Heritage*. No. 10, p. 45–63.

Antal, D. (2023). *The planned Slovak comprehensive music database*. Available: https://doi.org/10.5281/zenodo.10392759

Antal, D. (2024). *A szlovák adatkicserélési tér magyarországi föderációjának lehetőségei. HUNGARNET Egyesület*. Available: https://doi.org/10.31915/NWS.2024.25

Antal, D., Pigozne, I., Mester, A. (2025a). *Remapping the Livonian Coast: A Multilingual Gazetteer of the Settlements of Northern Kurzeme.* Available: https://doi.org/10.5281/zenodo.15668712

Antal, D., Pigozne, I., Mester, A., Nagy, M. (2025b). *Dataset of the Multilingual Gazetteer of the Settlements on the Livonian Coast of Northern Kurzeme.* Available: https://doi.org/10.5281/zenodo.15723051

Archives Expert Group on Archival Description, I.C. on. (2023). *Records in contexts–conceptual model*. Available: https://www.ica.org/app/uploads/2023/12/RiC-CM-1.0.pdf (viewed: 23.01.2025.)

Art & Architecture Thesaurus. The Getty Research Institute. Available: https://www.getty.edu/research/tools/vocabularies/aat (Viewed: 09.02.2025.)

Bekiari, C., Bruseke, G., Canning, E., Doerr, M., Michon, P., Ore, C.-E., Stead, S., Athanasios, V. (Eds.) (2024). *Definition of the CIDOC Conceptual Reference Model*. Available: https://cidoc-crm.org//sites/default/files/cidoc_crm_version_7.2.4.pdf (viewed: 25.01.2025.)

Bianchini, C., Bargioni, S., Pellizzari di San Girolamo, C.C. (2021). Beyond VIAF Wikidata as a complementary tool for authority control in libraries. *Information technology and libraries*. No. 40 (2). Available: doi: 10.6017/ital.v40i2.12959

Curry, E. (2020). Dataspaces: Fundamentals, principles, and techniques. *Real-Time Linked Dataspaces: Enabling Data Ecosystems for Intelligent Systems*. Springer International Publishing, Cham, p. 45–62.

Dublin Core. (2020). DCMI metadata terms. Available: https://www.dublincore.org/specifications/dublin-core/dcmi-terms/2020-01-20/ (viewed: 11.01.2025.)

EBU, Gaia-X. (2022). Dataspace for cultural and creative industries. Position paper. v.2.0. Available: https://gaia-x.eu/wp-content/uploads/2022/10/EBU_position-paper_Media-Data-Space.pdf (viewed: 10.10.2024.)

Fagerving, A. (2023). Wikidata for Authority Control: Sharing Museum Knowledge with the World, *Digital Humanities in the Nordic and Baltic Countries Publications*. Available: https://doi.org/10.5617/dhnbpub.10665

GeoNames. Available: https://www.geonames.org/ (viewed: 09.02.2025.)

Gil, Y., Miles, S., Belhajjame, K., Deus, H., Garijo, D., Klyne, G., Missier, P., Soiland-Reyes, S., Zednik, S. (2013). *Prov model primer. W3C working group note*. Available: https://www.w3.org/TR/prov-primer/ (viewed: 04.12.2024.)

Hyvönen, E., Tuominen, J. (2024): 8-star Linked Open Data Model: Extending the 5-star Model for Better Reuse, Quality, and Trust of Data. In: *Joint Proceedings of Posters, Demos, Workshops, and Tutorials of the 20th International Conference on Semantic Systems co-located with 20th International Conference on Semantic Systems (SEMANTiCS 2024)*, https://ceur-ws.org/Vol-3759/paper4.pdf (viewed: 11.07.2025.)

Ikkala, E., Hyvönen, E., Rantala, H., Koho, M. (2021). Sampo-UI: A full stack JavaScript framework for developing semantic portal user interfaces. *Semantic Web*. 2021;13(1):69-84. doi:10.3233/SW-210428

ISO. (2023). ISO/IEC 2382:2015(en), information technology — vocabulary. Available: https://www.iso.org/obp/ui/en/#iso:std:iso-iec:2382:ed-1:v2:en (viewed: 19.09.2024.)

Lebo, T., Sahoo, S., McGuinness, D., Belhajjame, K., Cheney, J., Corsar, D., Garijo, D., Soiland-Reyes, S., Zednik, S., Zhao, J. (2013). *PROV-o: The PROV ontology. W3C*

recommendation. Available: https://www.w3.org/TR/2013/REC-prov-o-20130430/ (viewed: 17.12.2024.)

Manteuffel, G. (1869). *Polnisch-livland: Mit lithographirten karten, originalphotographien, holzschnitten und anderen beilagen.* N. Kymmel, Riga.

Nagel, L., Lycklama, D. (Eds.) (2021). *Design principles for data spaces. Position paper. Version 1.0.* Available: https://doi.org/10.5281/zenodo.5244997

Open Music Europe. (2023). Open Music Europe (OpenMusE) – An Open, Scalable, Data-to-Policy Pipeline for European Music Ecosystems. Available: https://doi.org/10.3030/101095295

Pigozne, I., Antal, D., Mester, A.M. (2025). *Linked Open Datasets on Garments from the Latgale Region.* Available: https://zenodo.org/records/15167825 (Viewed: 07.04.2025.)

Pfeiffer, M., Gayo, J.E.L. (2021). *Representing CIDOC-CRM in wikibase for the luxembourg shared authority file.* Available: https://swib.org/swib21/slides/05-03-gayo.pdf (viewed: 18.12.2024.)

Rossenova, L., Duchesne, P., Blümel, I. (2022). Wikidata and wikibase as complementary research data management services for cultural heritage data. *CEUR Workshop Proceedings. Presented at the The 3rd Wikidata workshop, workshop for the scientific wikidata community, @ISWC 2022.* Available: https://serwiss.bib.hs-hannover.de/frontdoor/deliver/index/docId/2573/file/rossenova_etal2022-wikidata_research_data_mgmt.pdf (viewed: 27.12.2024.)

Ryder, M.L. (1984). Medieval sheep and wool types. *The Agricultural History Review*, No. 32, p. 14–28.

Sardo, L., Bianchini, C. (2022). Wikidata: A new perspective towards universal bibliographic control. JLIS.it: *Italian journal of library and information science* 13, 291–311. Available: https://doi.org/10.4403/jlis.it-12725

SEMIC 2020. (2020). Providing sustainable data services through wikibase and wikidata. Available: https://interoperable-europe.ec.europa.eu/sites/default/files/custom-page/attachment/2020-11/Parallel-track-4_B-Fischer_J-Thill_A-Angjeli%20final%20ppt.pdf (viewed: 09.01.2025.)

Taylor, L. (2002). *The study of dress history, Studies in Design and Material Culture*. Manchester University Press.

TextileBase. Available: https://reprexbase.eu/textilebase (Viewed: 10.02.2025.)

van Veen, T. (2019). *Wikidata: From 'an' identifier to 'the' identifier. Information technology and libraries*. 38, 72–81. Available: https://doi.org/10.6017/ital.v38i2.10886

Welters, L., Lillethun, A. (2018). *Fashion history: A global view, Dress, body, culture*. Bloomsbury Publishing.