

Metadata Analysis of Sakha Artefacts in Europeana^{*}

Saiyyna Vasileva^{1,*}, Daniel Zilio^{2,†}

¹University of Padua, Padua, Italy

²Department of Cultural Heritage, University of Padua, Padua, Italy

Abstract

This study positions itself within the growing field of research on the representation of displaced Indigenous cultural heritage in digital environments. This paper investigates how Sakha (Yakut) cultural artefacts are represented in Europeana, Europe's central cultural heritage aggregator. Using multilingual keyword queries and Europeana's Search and Record APIs, a dataset of 122 Sakha-related items was compiled and examined for metadata completeness, accuracy, and cultural attribution. The analysis reveals several inconsistencies: over half of the object-level metadata fields are missing, cultural identifiers are often absent or unstructured, and some records contain incorrect data. The study highlights the need for culturally informed metadata practices to improve representation of Indigenous heritage in large-scale digital library infrastructures.

Keywords

Digital Humanities, Indigenous Cultural Heritage, Europeana, Sakha, Metadata

1. Introduction

Digitisation has transformed access to cultural heritage. Through digitised collections, museums can engage audiences who were previously restricted by social or spatial limitations [1]. This expanded access enables cultural enthusiasts, independent researchers, students, and scholars to study cultural objects remotely. For displaced cultural artefacts, digitisation also offers the source communities a new opportunity to reconnect and engage with part of their heritage. In addition, digital aggregators increase the accessibility of museum collections by gathering materials from multiple institutions into a single platform. In Europe, the largest cultural heritage portal is Europeana. Because these digital resources are widely accessible, the accuracy and reliability of their metadata is essential.

Before a cultural artefact reaches a platform such as Europeana, it undergoes several layers of transformation. Therefore, what users access is a digital representation of the object that was removed from its original context, catalogued in a museum, and digitised through adjustment to specific standards and principles. Then Europeana collects data from more than 4,000 institutions through another tier of national, regional, or thematic aggregators [2, 3]. Those institutions and aggregators can use different metadata standards and knowledge organisation systems, creating inconsistencies that require additional steps of standardisation [4, 5, 6]. At each stage, information or metadata can change – it can be enriched, or some parts may be omitted. In cultural heritage institutions, enrichment occurs primarily through semantic enrichment, which uses Linked Open Data to add clearer meaning and connections to metadata from controlled vocabularies. It creates interconnected data, which in turn makes cultural heritage records more discoverable [7, 6]. But in these complex data transformation processes, some information can be lost, raising the question of "reliability" of the data [8].

Another issue of metadata standards and knowledge organisation systems is that they are based on Western perspectives [9, 10, 11]. All the benefits and drawbacks mentioned above of digitization particularly concern indigenous cultural heritage. Only in recent years has scholarship at the intersection of "digitisation" and Indigenous cultural heritage doubled between 2021 and 2024 (more than half of

IRCDL 2026: 22nd Conference on Information and Research Science Connecting to Digital and Library Science, February 19-20, 2026, Modena, Italy

*Corresponding author.

†These authors contributed equally.

✉ vasilyevasaina@gmail.com (S. Vasileva); daniel.zilioi@unipd.it (D. Zilio)

ORCID iD 0009-0002-7654-9203 (S. Vasileva); 0000-0001-7107-8858 (D. Zilio)



© 2026 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

the scholarship has been written between these years) [12]. The geography of knowledge production for Indigenous studies is uneven. The United States dominates with 24% of publications, followed by Australia (12%) and Canada (8%), reflecting their established leadership in Indigenous research. Within Europe, the United Kingdom (8%) and Germany (3.6%) stand out, likely due to their major ethnographic collections. In contrast, only 1.8% of the scholarship represents Arctic or Indigenous cultures in Russia [13], among them only two articles on Indigenous peoples of Siberia are available [14], and no articles on Sakha (Yakut) culture.

The paper therefore seeks to address that gap by focusing on the Sakha artefacts dispersed across European museums from a geographically remote region in northeastern Siberia. Although recent scholarship by Sakha scholars has advanced provenance studies and documentation of these collections, as of now, no study on the representation in digital aggregators has been done. This paper analyses the discoverability and metadata quality of Sakha cultural objects in Europeana. Using multilingual API queries and qualitative assessment, it identifies which objects can be retrieved, what metadata they contain, and how their digital representation reflects structural limitations of aggregate digital libraries. The study contributes empirical evidence to ongoing discussions on metadata reliability, Indigenous cultural representation, and the challenges of digitising ethnographic collections within European institutions.

2. Background: Sakha Heritage and Digital Representation

2.1. Displaced Sakha artefacts

Sakha, also known as the Yakut people, are an indigenous Turkic-speaking ethnic group that inhabits the Sakha Republic (Yakutia) in the North-Eastern part of the Russian Federation, the largest administrative region in Russia. There are two official names for the republic: the "Sakha Republic" and "Yakutia", as well as two terms referring to the people. Sakha is how Sakha people call themselves, while "Yakut" is believed to be a name given to the Sakha people by different ethnic groups, and it was the first term mentioned in official documents [15].

As the Sakha Republic attracted many curious travelers, ethnographers, and researchers as a "wild," vast and almost untouched place, and as it later was the home of millions of detainees from Imperial and Soviet Russia, the material culture of the Sakha people can be found in different parts of the world.

Outside of Russia, Sakha objects kept in museums are known to be in various countries in Europe and the US. The study of these collections by Sakha scholars began primarily in the 1990-s. In 1992, Vladimir Ivanov-Unarov and Zinaida Ivanova-Unarova undertook research on the Sakha holdings of the American Museum of Natural History in New York [16]. Shortly after, in 1993, historian Semen Gorokhov became the first Sakha scholar to discover a model of *Yhyakh* at the British Museum, publishing a small book dedicated to this artefact [17]. Later, the same model was closely studied by another Sakha scholar Tatiana Argounova-Low together with her colleague at the University of Aberdeen Alison Brown within the Narrative Objects project launched in 2015 [18], culminating most recently in a co-authored book published in 2025 [19].

One of the most important works done on the material culture from the Sakha Republic found in museums outside of Russia around the world is perhaps a multi-volume book *Material and Spiritual Culture of the Peoples of Yakutia in World Museums: (17th - early 20th centuries)* [20]. Representing the results of many years of research and collaboration between the National Art Museum of the Republic of Sakha (Yakutia) and different international institutions since 2006, the books provide detailed information and photo illustrations of the artefacts held in Western museums' collections by their ethnic attribution. According to the head of the project and editor of the books, Alexander Zhirkov, this project is an effort to record the artefacts that have left Yakutia as it is impossible to return them all back [21]. The project is bold and co-Indigenous-led, presenting objects with their original names and placing them within their cultural and historical contexts while at the same time helping Western museums fill the gaps. All of these studies focus on the provenance, history, and recreation of cultural contexts.

2.2. History and Digitisation of indigenous artefacts

From ancient and medieval worlds up to the nineteenth century, objects circulated hand-in-hand through trade, wars, migrations, and many other historical processes and events [22]. In the nineteenth century during New Imperialism, this circulation was further shaped by the intensified collecting of the period [23], when museums and private collectors gathered objects from around the world. This is the time when ethnographic museums were established and indigenous artefacts became part of their collections [24], where they were often referred to and displayed as “primitive”, as objects of “static” cultures that need to be saved [25, 26, 24], or as colonial trophies [27]. In other cases the issues are lack of records or their inaccuracy [28]. It shows that the process of museum collecting in general has often involved removing objects from their original contexts and putting them into the new contexts, i.e. recontextualisation [29, 30]. Thus, studies highlight that reevaluating collections shaped by colonial influences is necessary to prevent the uncritical continuation of outdated scientific approaches in the description, classification, and interpretation of collection objects [31].

Digitisation initiatives in museums began as early as the end of the 1960-s [32]. Although, the same challenges and concerns that were addressed at the beginning, were still relevant in the 2000-s. This is because the main metadata standards such as Dublin Core [33] and CIDOC CRM [34] emerged in the 1990-s, followed later by other systems. FAIR Guiding Principles were introduced in 2016 to improve the Findability, Accessibility, Interoperability, and Reuse of digital records [35].

The digitisation processes started and developed hand in hand with the decolonial movements and emergence of its theory. The 1960-s saw a rise in repatriation, when previously colonised states gained independence and began advocating for the return of artefacts taken during colonial rule [36], with international conventions such as the 1970 UNESCO Convention on the Means of Prohibiting and Preventing the Illicit Import, Export and Transfer of Ownership of Cultural Property [37]. In the 1980-s traditional museology was widely influenced by “New museology” [38], which challenges old practices by advocating critical reflection on colonial legacy, ethical collection practices and more inclusive narratives that involve source communities. The New Museology and the decolonisation discourse influenced ICOM to develop a Code of Ethics for Museums in 1986 [39]. The CARE (Collective Benefit, Authority to Control, Responsibility, and Ethics) principles were developed in 2019 with the aim of shifting towards approaches that align data governance with indigenous worldviews and knowledge systems, ultimately serving the interests of Indigenous Peoples [40].

Because of this parallel development, the digitisation sometimes came before the right treatment of the indigenous artefacts and their metadata. Unfortunately, the initial goal of digitisation in general may have been on maximising the number of digital copies available online rather than on their quality [41]. As a result, many of such digital records were considered inconsistent: among the issues of digitisation of indigenous cultural heritage were mentioned difficulty in searching using indigenous terms [42], “neutral” broad categories [43], lack of cultural context [30], use of outdated or offensive terms, exoticisation, lack of attribution to Indigenous contributors in metadata [44]. An example to overcome these issues is the development of a custom indigenous vocabulary for metadata through collaboration with the source communities [44, 43, 10], and a general shift in approach [45]. In a modern world, digitisation should also be part of decolonisation [10].

Community or independent archives have also emerged as a response to existing issues. These community archives are closely linked to Indigenous self-determination, empower communities and contribute to larger social change [46]. Ideally, such projects are led by the communities themselves as Australian Institute of Aboriginal and Torres Strait Islander Studies (AIATSIS)¹, but more often they are developed in collaboration with community members, such as the projects Mootookakio’ssin² and the Digital Museum of the North³. The latter is a recent initiative launched by the end of 2024 that gathers and showcases cultural heritage of Siberian Indigenous cultures (Nenets, Khanty, Sakha, Evenk, Chukchi, Koryak, Itelmen, Even, Uilta, Ainu, and Nanai) [47]. Currently, there are 227 objects

¹<https://aiatsis.gov.au/about/what-we-do/about-collection>

²<https://mootookakiossin.ca>

³<https://mu-north.org>

gathered through consultation with the Indigenous communities by allowing them to choose the “most meaningful” and including explanation videos by local experts [48].

3. Europeana as a Cultural Heritage Aggregator

Europeana is “the reference point for Europe’s online cultural heritage” as described by an advisory group appointed by the European Commission [49], and it is true. In simple words, Europeana is a gathering place for the digitised materials of cultural heritage institutions across Europe.

When an institution shares its data with Europeana, it should follow the principles of the European Data Model (EDM)⁴, a data model that provides a structured way to organize and connect cultural heritage data from different institutions [50]. The EDM was specifically designed to integrate and align with established standards like Dublin Core, Encoded Archival Description, and LIDO, which are widely-known and used in the cultural field. Europeana encourages activities to reuse their digital collections by supporting different projects and providing an API [51].

There are different types of API for accessing the Europeana collection, such as Search API (search for items), Record API (get all metadata available for a given item), Entity API (search and get metadata for entities), Recommend API (returns related items, user sets or entities) among many [51]. The search API acts as a search engine, where a keyword or phrase is used to search the collection as it is done on the Europeana.eu website without API.

A library specifically developed for using Europeana APIs with Python is called PyEuropeana, as defined, it “has comprehensive and clear documentation” [51]. The package interacts with almost all the main APIs of Europeana: the current working version of the Search API, Record API, Entity API; it was last updated in 2023 [52].

4. Accessing Europeana API

Defining a search request for the Search API is one of the most crucial steps that requires an understanding of the EDM as it has many different parameters. First of all, Sakha Republic has a synonym - Yakutia. In this case, many words can be used to describe objects belonging to Sakha culture, such as “Sakha”, “Yakut”, “Yakutia”, “Yakutian” etc. Those keywords can be found not only in the title of digital records but also in other fields of metadata, such as description, place of belonging, other geographical information, cultural provenance etc. In order to find suitable search criteria, different scenarios for the search request were created and tested to find artefacts related to Sakha culture.

Since the study focuses on material objects related to the Sakha Republic and not texts, audio, or video contents, the filter to keep only the “image” type of the digital media on Europeana was used.

A simple keyword search with the terms “Sakha” and “Yakutia” returned 105 and 123 results accordingly. The use of “wildcard search” with tilde and asterisk signs – “Yakutia~” to find words with similar spelling and “Yakut*” to find words with different endings – among many, returned results that were not related to the object of the study (such as inclusion of Lithuanian given names and botanical records). For this reason, these types of search are not used as a search request.

In addition to simple parameters, Europeana also aggregated several fields together to make it easier to search records: what, when, where, who, and subject (they also represent contextual classes of the EDM) [51]. “What”, a field that gathers concepts, which are categories from organised knowledge bases, returned 6 results; “where”, a spatial parameter, returned 112 results; and finally, “subject”, a general topic⁵ plus spatial and temporal coverage⁶, returned only 17 results mainly related to natural history subject.⁷

⁴<https://pro.europeana.eu/page/edm-documentation>

⁵<https://www.dublincore.org/specifications/dublin-core/dcmi-terms/terms/subject/>

⁶<https://www.dublincore.org/specifications/dublin-core/dcmi-terms/terms/coverage/>

⁷Both terms “Sakha” and “Yakutia” were used for the search.

This study began as a thesis project in the late 2024, and as of February 2025, “Sakha” as a defined place category in the Europeana database referring to a “Federal subject of Russia, republic of Russia” returned only 12 results [53], while as of November 2025 there are 70 items related to the category. It shows that the metadata of items are being constantly updated on Europeana.eu.

4.1. Terms selection for search query

Multilingual search is supported on Europeana, which means that if a user searches in one language, the search query translates into another, when possible [54]. It also depends on the metadata quality of a specific digital item, since the translation is possible when the fields of metadata were constructed well by the data providers - cultural institutions. To test the multilingual search, both “Yakutia” in English and “Jakutien” in German were placed separately in the search field. The test showed that almost 20% of the results in German were not available in the results in English language⁸. It is a fairly high percentage of missing results; therefore, the terms “Yakut” (as an adjective or a person), “Yakuts” (as people) and “Yakutia” (as a place) were translated into the 24 official languages of the EU plus Russian [55]. In addition, a name of capital of the republic, Yakutsk, was included in the search. Since the term “Sakha” is widely recognised in its original form in many languages, including most of the EU languages, it is included in the search query without translation. Of 134 terms in 24 official EU languages and Russian, 43 duplicates were found and 91 unique terms related to Sakha were filtered.

A search with all the unique terms combined with the logical ‘OR’ gives more than a thousand results. More than half of the terms (47 of the 91) that did not give any result were excluded from the list. Remaining terms were manually checked, after which 6 were removed to avoid confusion (Table 1), leaving 38 terms to be included in the search request.⁹

Table 1
Deleted terms with results and reasons for deletion

Unique terms	Search results	Reasons for deletion
Jakut	389	Translates as “jacket” in Finnish
Iacúit	10	Latin inscription “sic iacuit” - laying
Jakuts	3	A given name (Jakuts László)
Jakutski	11	A name of a bird in Slovenian (Jakutski škurh)
Jakutok	1	Translates as “jacket” in Hungarian
Jakuci	2	A surname name in Croatia

4.2. Dataset creation, filtering, and cleaning

A total number of 415 items were found through the Search API. Using PyEuropeana in Python the Search API query looks as following:

```
result = apis.search(query = terms, sort = 'europeana_id', qf = 'TYPE:IMAGE',
profile='rich', rows = 1000),
```

where Result stores the JSON response; terms is a list of 38 terms separated with a logical term OR; TYPE: IMAGE filters results with image type removing the video, audio, text, and 3D files; sort = 'europeana_id' is for sorting the results by their ID in Europeana; rows is for controlling a maximum number of results.

⁸“Yakutia” and “Jakutien” returned in 125 and 106 items accordingly, the results were then checked for a match – 21 of “Jakutien” was not in the “Yakutia” result.

⁹Sakha, Yakut, jakuut, Jakute, Iacuto, Yakuto, Якут, Yakuts, Jakuutit, Jakuten, Jakuterna, Якуты, Якутия, Jakutija, Jakutsko, Jakutien, Jakoetië, Yakutia, Jakutia, Iakoutie, Γιᾱκουτία, Jakutföld, Jacuzia, Jakucja, Iacútia, Iacutia, jakutisk, Jakutisch, Якутский, Yakoute, Jakutin, Jakuté, Yakuta, Yakutiya, Yakutsk, Jakutsk, Якутск, Саха.

PyEuropeana has a function that transforms a search result into a readable and clean dataset, but it eliminates some of the metadata fields. Thus, the Record API was used along with the creation of new functions to get more metadata fields. The full dataset contained 34 columns.

Table 2
Reasons and number of deleted items

Reason of exclusion	N. of items	Share, %
Thematically unrelated historical content	131	31.6
Outside the scope of cultural heritage	107	25.8
Homonymy error	37	8.9
Related to other culture	10	2.4
Duplicate	8	1.9
Total	293	70.6

Not all 415 items were related to the Sakha culture or people. After a manual check, 293 items were deleted from the final dataset, leaving only 122 relevant ones (29.3%). Table 2 summarises the reasons for exclusion. Many digital records were thematically unrelated (31.6%), including photographs of Soviet exiles, materials related to Soviet culture and politics, and other non-relevant photographs. 25.8% of the results were outside the scope of cultural heritage, consisting of natural history collections and geographical maps. Additional records were removed due to homonymy errors, relation to other cultures, and duplications.

5. Dataset description: Sakha cultural artefacts in Europeana

Most of the items are located in museums in Germany (86%), few objects are in Sweden (6.6%), Finland (4%), Estonia (1.6%), whereas Lithuania, Spain, and Italy have only 1 item each. More than half of the objects (60.7%), belong to the collection of Grassi Museum of Ethnology in Leipzig (Figure 1).

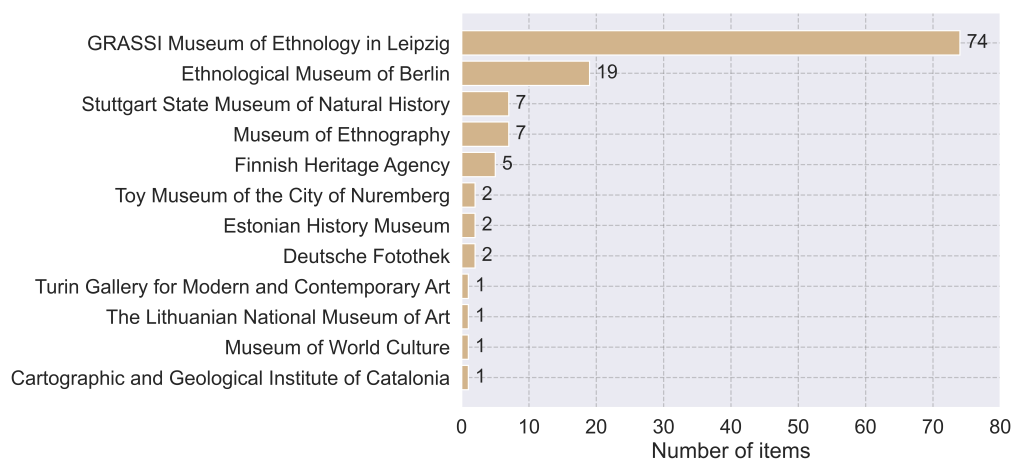


Figure 1: Number of Sakha items in the dataset by providing institutions

It is possible to see when the digital records were added on Europeana and when they were updated. Figure 2 shows the number of records created by year - more than half of the objects (61%) were added in 2024. Almost all of them belong to the collection of Grassi Museum of Ethnology in Leipzig. 87% of all objects were added starting from 2020, which is quite recent considering that Europeana was launched in 2008. The year of update differs from the year of creation: 72% of the digital records were last updated in the last 2 years (Figure 2) – 82 objects (67%) were updated in September 2025. It mainly shows a very recent development of the digitising activities of Sakha cultural heritage by museums or their sharing activities with Europeana.

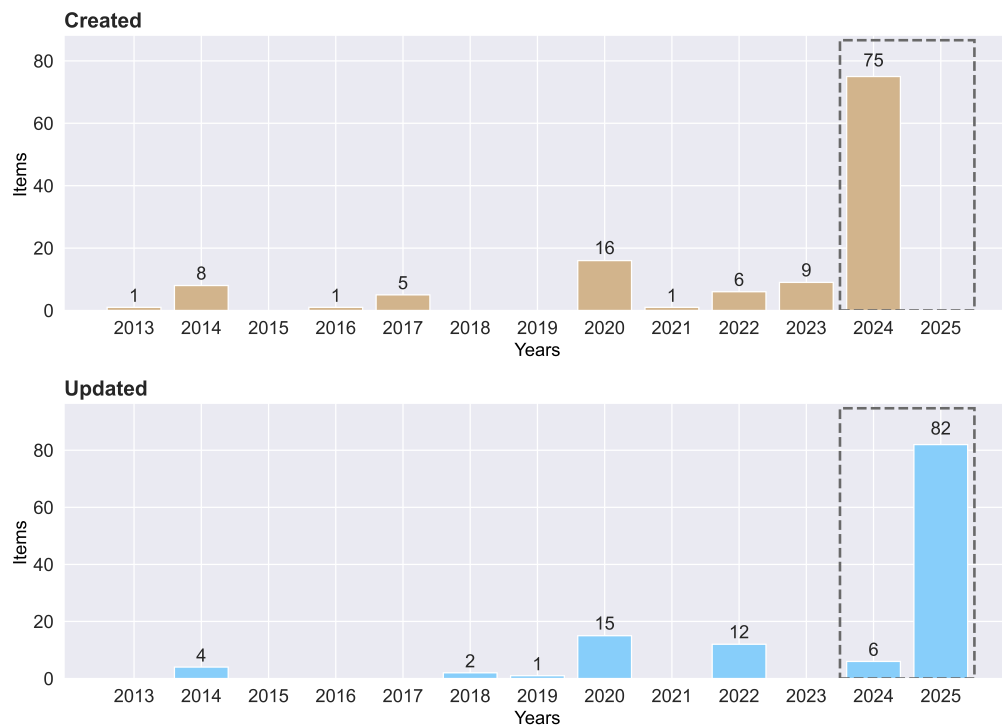


Figure 2: Number of Sakha items by the year of creation and update

Europeana has developed an internal measure of metadata completeness that is assigned to each item and ranges from 1 to 10 [51]. It is based on the availability of required and optional metadata fields for the items and is computed at the metadata enrichment stage [56]. There is no publicly available formula for completeness in the Europeana documentation. 14 objects in the dataset have 0 completeness, which should not be the case since the minimum value is 1. Without considering those 14 objects, an average completeness of all 122 items is 6 - just above average. To better understand the completeness per institution, the mean values were calculated for each of them (Figure 3). The Toy Museums of the City of Nuremberg, the Lithuanian National Museum of Art, and Museum of World Culture have the highest possible average completeness, but it is important to note that they have only one or two items.

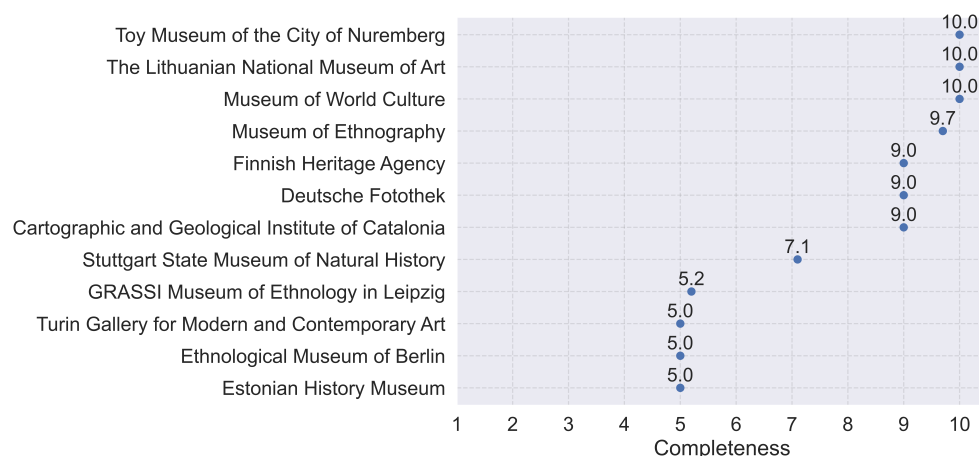


Figure 3: Average metadata completeness scores per providing institution (Completeness scores are calculated individually by Europeana for each item)

6. Metadata analysis and Discussion

Figure 4 shows the number of present and missing values for each column, i.e., for each field that describes an item. For example, every item has a title, but only a few have a description. As seen in Figure 4, the share of the "black holes", i.e. missing data, is 33% from all the data (calculated as a share of empty cells in a number of all cells in the data).

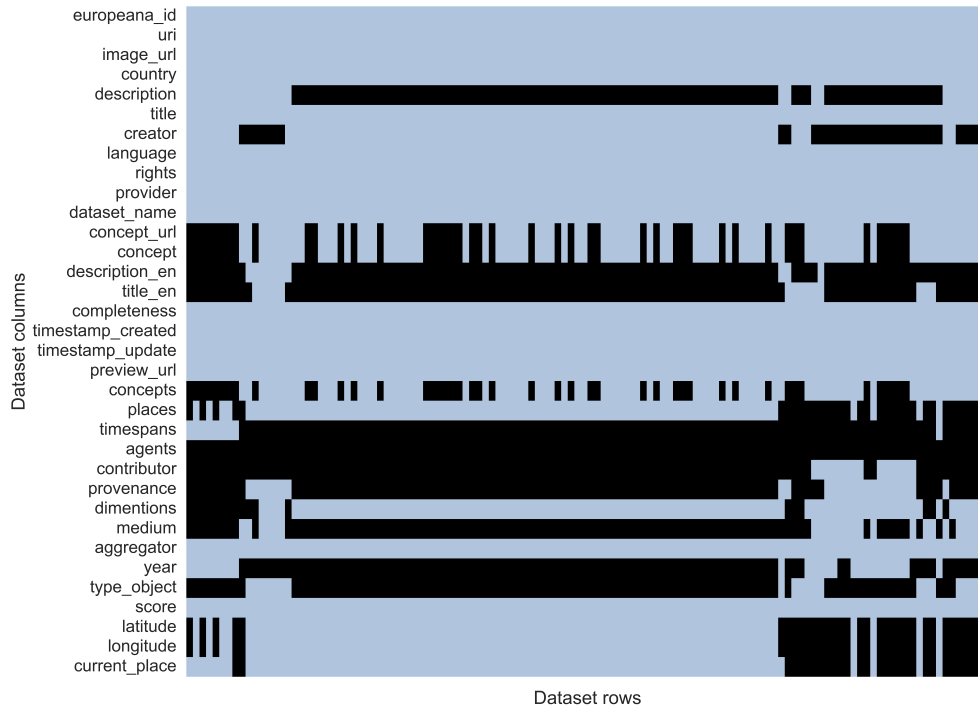


Figure 4: Visualisation of missing values (missing values are in black)

To see separately how the share of missing values differs for metadata fields only related to a cultural object, the fields were classified into three types: fields related to a cultural object, cultural institution (museum), and digital record. These fields, to some extent, are based on the EDM three core classes [50]. The digital record information, such as ID, links, language, dataset name, completeness value, days of creation and update, is fully present, i.e. for every cultural object. The fields that describe a museum, such as the name, city and country of the institution, rights for reuse, and aggregator name, are almost full, having only 4.6% of missing values (Table 3). And finally, more than half of the data (57.9%) are missing for the fields describing cultural artefacts (title, description, creator, related concepts and subjects, timespans, year of creation, provenance, etc.) This can indicate general limited information available about the Sakha artefacts in European museums.

Table 3
Share of missing data by field types

Field types	Number of fields	Missing data, %
Object	19	57.9
Institution	5	4.6
Digital record	10	0
Total	34	33

6.1. General metadata limitations

Table 4 presents the general limitations found during metadata assessment. It was observed that the main descriptions of most of the items (63.9%) were missing in Europeana, but were present in the original source. This issue occurs because the description field is not mandatory if the title is present[57]. Similarly, some fields' information was missing for 9 records. Wrong information was present in fields in 9.8% of the cases: five records contain "Sibirien" the field "places", which refers to a place in Northern Trøndelag in Norway; other cases include the attribution of wrong subjects (concepts); in one case, a providing country was labelled as Europe. The values of a field were put by mistake in another field for 12 items: for 10 items the Russian Museum of Ethnography in St Petersburg was put in a "creator" field instead of a "provenance" one, since the artefacts were previously part of the museum's collection. Other inconsistencies include the absence of links to the original source and duplicates made by different aggregators in different years. In general, missing data and value mismatch are related to structural differences between the EDM and metadata models of the providing institutions. Instead, the wrong values appear as a result of a data enrichment stage, as they do not come from the original source.

Table 4
Metadata assessment – general limitations

General limitations	N. of items	Share, %
Description is fully or partly missing in Europeana metadata	78	63.9
Wrong information/value in a field	12	9.8
A value of one field put in another	12	9.8
A field's value is missing in Europeana metadata	9	7.4
No link to original source	7	5.7
Has a duplicate (duplicates)	3	2.5

Moreover, in 57 cases, "Jakutien (Republik Sacha), Russland" appears in the same field as the creation century under "creation date," which is misleading and suggests a metadata issue. However, a closer look at the API responses, showed that these elements are stored separately. "Jakutien (Republik Sacha), Russland" is placed in the "isPartOf" field, which represents a larger unit to which a cultural object belongs – this can be a place, a time period, or an associated agent. Therefore, its appearance in the front-end "creation date" field is the result of several fields' aggregation. This creates a confusion, also because the "isPartOf" field is similar to existing fields such as "concept," "subject" (type), and "places."

These limitations were also previously observed by the Europeana Data Quality Committee¹⁰, as well as by scholars [58, 56]. Katharina de Burger in her master's thesis explored metadata management at Europeana, she highlighted the following issues: broken links, incompatible vocabulary, enrichment issues, missing data, limitations of access and discoverability [56], which are also present in the case of the Sakha objects. Another study that assessed the quality of metadata of records in Europeana highlighted that it was generally lower than in the collections of other cultural institutions, with issues such as missing data and unnecessary repetition of information [58].

6.2. Metadata limitations specific for Sakha artefacts

As mentioned earlier, the differences between the data models lead to data losses during aggregation. The Sakha case provides a concrete example of how such structural differences lead to the loss of culturally specific categories. For example, originally Sakha (Yakut) people were created as a separate category for ethnic groups in the online collections of the Ethnological Museum of Berlin and Museum of Ethnography in Stockholm, making it easier to search and filter objects belonging to the Sakha culture. However, the EDM has no designated field for ethnic or cultural belongings, except that, in theory, this information can be present in concepts or subjects. But in the case of the Sakha records in this study neither Sakha nor Yakut were present in those fields.

¹⁰<https://pro.europeana.eu/project/data-quality-committee>

As the dataset shows, in 15.6% of cases the ethnic relation is completely lost or present as plain text rather than as a filterable category (Table 5). This can lead to the loss of a considerable amount of items while performing a search in Europeana dataset, particularly in cases where an item does not contain the words “Sakha” or “Yakutia” in the description, but actually belongs to the culture. The only filterable category related to Sakha objects is the “Places” field as mentioned earlier, but 42.6% of the objects in the dataset still do not have it (Table 5). Moreover, the field “Places” can represent a place of origin, institution, or any associated place, confusing users.

Table 5

Metadata assessment – limitations related to specific for Sakha artefacts

Related to Sakha	Number of items	Share, %
Sakha as a place is not a filterable category	57	46.7
The category of Sakha as an ethnic group is lost from the original source	19	15.6
Sakha is only mentioned in the title or description	13	10.7
Only Russia or Siberia is mentioned in the field “place”	10	8.2
The category of Sakha as a place is lost from the original source	8	6.6
Possibly related to other culture	5	4.1
Relates to other culture (Even, Khanty)	3	2.5
Sakha Republic is mentioned in provenance field	3	2.5
Use of “primitive”	1	0.8

In 10.7% of the cases, the relation to Sakha culture was mentioned only in the title or description, indicating that the cultural relation is not treated as a structured metadata element. 8.2% of digital records mention only Russia or Siberia in the “places” field. This puts Sakha culture under the broad umbrella of geographical references, which are inhabited by many other cultures. It also limits the discoverability of Sakha artefacts in digital searches.

The metadata structure in Europeana supports a field about provenance of cultural objects, an important information that aims to tell a story of an object’s journey - about how and when an artefact left its own home and found a new one. Only in 19.7% of cases the provenance information is present in the dataset, and only three records (2.5%) mention Sakha Republic or culture in field “provenance” (Table 5). In cases from the Ethnological Museum of Berlin and the German Photo Library the field contains only the institutions’ name, which can be misleading. Most of the objects from the Ethnological Museum of Berlin have their contributors listed as Vasily Priklonsky, Petr Shimkevich, Alexander Bunge, and Piotr Ostrovskikh, well-known Russian ethnographers and researchers that travelled to Siberia for their expeditions or work. For this reason, it is expected that the provenance information mentions them along with the time period and the type of ownership transfer. For example, it is known that Vasily Priklonsky donated some Sakha objects from his ethnographic collection, and Petr Shimkevich sold his photographs at the end of the 19th century to Adolf Bastian, then director of the Royal Museum in Berlin [20, 59]. The lack of the provenance information, however, if not structural, is on part of the museums, Europeana only reproduces it.

After a close examination, it was discovered that the cultural belonging of five items in the dataset is uncertain as the artefacts can be cultural heritage of other Indigenous groups in Yakutia or Siberia in general. It was also discovered that three objects – two hats and one apron in the collection of the Ethnological Museum of Berlin – belong to the Even culture. It shows the historical cultural exchanges in Yakutia, but at the same time it raises a question about the reliability of museum categorisations, and highlights how digitisation processes can perpetuate earlier misattributions rather than correcting them.

In one case, description of a photograph of a man with reindeer provided by Deutsche Fotothek contained a word “primitive” (Figure 5). Even if the adjective is related to skis, the use of the term in relation to an Indigenous culture can be offensive. A DE-BIAS project co-funded by the Digital Europe Programme of the European Union and supported by Europeana that aimed to contribute to the representation of Indigenous and minoritised communities in the digital space developed a vocabulary,

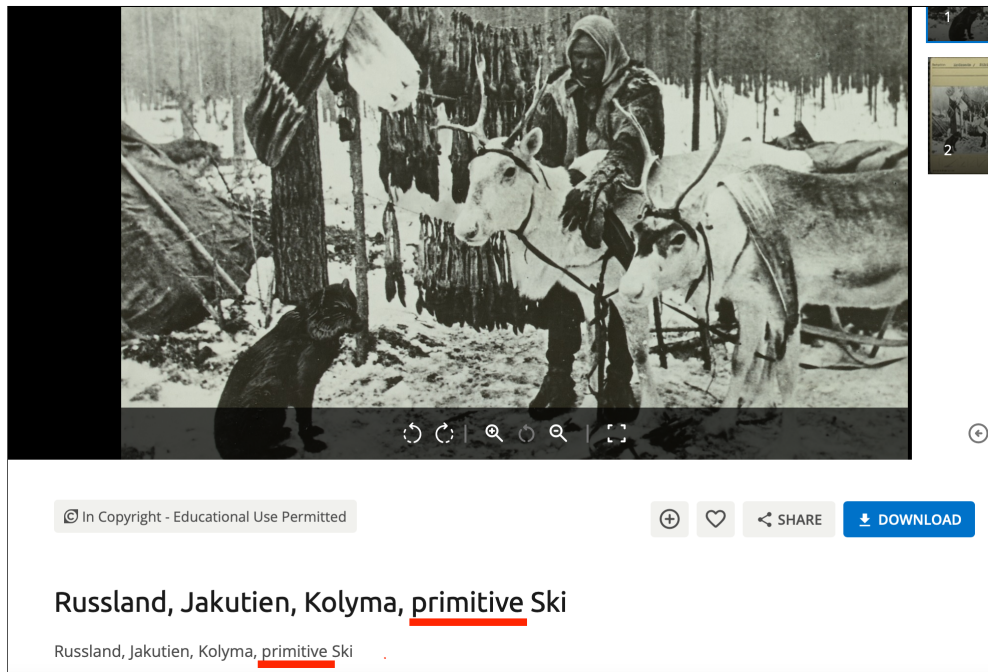


Figure 5: Screenshot of a digital record provided by Deutsche Fotothek on Europeana [61]

which flagged the term “primitive” as “synonymous with the racialised and temporal ‘Other’” not recommending the use of the term [60].

7. Conclusion

The study of Sakha cultural heritage held in museums outside of Russia is still relatively recent, and collaborations between Sakha scholars, communities, and museums remain mainly project-based. As a result, digital records of Sakha artefacts may appear before curators have updated or corrected cultural contexts, classifications, or provenance information at the institutional level. The analysis reflect this situation as many records were updated recently, but still the descriptions remain absent or incomplete, and metadata remain partly inconsistent.

Although the Europeana Data Model has many attributes and makes it possible for multiple categories to be included in one parameter, it has some limitations. There is no dedicated metadata field for ethnic group or cultural affiliation. Instead, such information can be placed under the category “concept,” alongside geography, materials, techniques that distinguish a culture, provenance, or object types. This is particularly concerning given that Europeana aggregates collections from thousands of institutions, including major ethnographic museums in Europe. In this context, the absence of structured cultural identifiers contributes to the continued invisibility of Indigenous cultures within large European data spaces.

Overall, the metadata analysis shows how digitisation both enables and constrains the representation of Sakha culture. This case study suggests that Europeana primarily functions as a discoverability layer for Sakha artefacts, displaying them in a large digital space. However, achieving a research-ready dataset requires significant additional processing, such as the creation of custom functions for data retrieval and manual data corrections to address the inconsistencies and missing contextual information. The extent to which discovered limitations apply to other Indigenous cultures, remains to be investigated. It is also important to mention that the 122 items found through the multilingual keyword search likely represent only a portion of all Sakha-related artefacts available on Europeana, pointing to the need for improved metadata curation.

Declaration on Generative AI

During the preparation of this work, the author(s) used ChatGPT in order to: Grammar and spelling check, Paraphrase and reword. After using this tool/service, the author(s) reviewed and edited the content as needed and take(s) full responsibility for the publication's content.

References

- [1] O. C. Frost, *When the Object is Digital: properties of digital surrogate objects and implications for learning*, Routledge, 2010.
- [2] E. Foundation, Statistics dashboard, 2025. URL: <https://metis-statistics.europeana.eu/en/data/dataProvider>, accessed: 2025-11-27.
- [3] Europeana Foundation, Europeana aggregators forum, 2024. URL: <https://pro.europeana.eu/page/aggregators>, accessed: 2025-12-12.
- [4] A. Dressen, Cultural heritage data for research – an introduction, *Open Library of Humanities* 11 (2025). URL: <https://olh.openlibhums.org/article/id/18710/>. doi:10.16995/olh.18710.
- [5] D. Belteki, A. J. Rees, A.-M. Sichani, Datafication and cultural heritage collections data infrastructures: Critical perspectives on documentation, cataloguing and data-sharing in cultural heritage institutions, *Journal of Open Humanities Data* 11 (2025). URL: <https://openhumanitiesdata.metajnl.com/articles/10.5334/johd.277>. doi:10.5334/johd.277.
- [6] S.-J. Chen, Semantic enrichment of linked archival materials, *Knowledge Organization* 46 (2019) 530–547. doi:10.5771/0943-7444-2019-7-530.
- [7] M. L. Nappi, M. Buono, C. Chivăran, R. M. Giusto, Models and tools for the digital organisation of knowledge: accessible and adaptive narratives for cultural heritage, *Heritage Science* 12 (2024) 112. doi:10.1186/s40494-024-01219-z.
- [8] N. Barbuti, Thinking digital libraries for preservation as digital cultural heritage: by r to r4 facet of fair principles, *International Journal on Digital Libraries* 22 (2021) 309–318. doi:10.1007/s00799-020-00291-7.
- [9] Z. Manžuch, Ethical issues in digitization of cultural heritage, *Journal of Contemporary Archival Studies* 4 (2017). URL: <https://elischolar.library.yale.edu/jcas/vol4/iss2/4>.
- [10] H. Whaanga, D. Bainbridge, M. Anderson, K. Scrivener, P. Cader, T. Roa, T. T. Keegan, He matapihi mā mua, mō muri: The ethics, processes, and procedures associated with the digitization of indigenous knowledge—the pei jones collection, *Cataloging & Classification Quarterly* 53 (2015) 520–547. doi:10.1080/01639374.2015.1009670.
- [11] J.-E. Mai, Ethics, values and morality in contemporary library classifications, *KNOWLEDGE ORGANIZATION* 40 (2013) 242–253. doi:10.5771/0943-7444-2013-4-242.
- [12] Elsevier Scopus, Bibliographic dataset: Digitised indigenous heritage, 2025. URL: <https://www.scopus.com>, retrieved on 2025-11-28.
- [13] Elsevier Scopus, Bibliographic dataset: Digitised indigenous heritage - arctic or indigenous peoples of russia, 2025. URL: <https://www.scopus.com>, retrieved on 2025-11-28.
- [14] Elsevier Scopus, Bibliographic dataset: Digitised indigenous heritage - indigenous peoples of siberia, 2025. URL: <https://www.scopus.com>, retrieved on 2025-11-28.
- [15] V. L. Priklonsky, *Letopis' Yakutskogo kraya, sostavlennaya po official'nyim i istoricheskim dannym* [Chronicle of the Yakut region, compiled from official and historical data], Izdanie G. V. Yudina, Krasnoyarsk, 1896. URL: <https://new.nlrs.ru/open/2643>.
- [16] V. Ivanov-Unarov, Z. Ivanova-Unarova, The Revitalization of the Traditional Culture of Northeast Siberian Peoples, volume 4 of *Contributions to Circumpolar Anthropology*, Arctic Studies Center, National Museum of Natural History, Smithsonian Institution, Washington, 2003, p. 336–347.
- [17] S. N. Gorokhov, Ysyakh. Iakutskii Ekspонат v Kollektzii Britanskogo Muzeia [Ysyakh. A Yakut Artefact in the Collection of the British Museum], Sitim, Yakutsk, 1993.

- [18] A. K. Brown, T. Argounova-Low, Narrative objects the sakha summer festival and cultural revitalization, *Museum Worlds* 3 (2015) 128–141. doi:10.3167/armw.2015.030109.
- [19] T. Argounova-Low, A. K. Brown, *Narrative Objects: Museums, the Sakha Summer Festival, and Cultural Revival in Siberia*, Routledge, London, 2025. doi:10.4324/9780429456398.
- [20] A. N. Zhirkov, Z. I. Ivanova-Unarova, A. L. Gabysheva, S. R. M. kultury i dukhovnogo razvitiia, P. F. on Indigenous Issues, I. C. of Museums Rossiiskii komitet, A. nauk Respubliki Sakha, N. khudozhestvennyi muzei Respubliki Sakha, Material and Spiritual Culture of the Peoples of Yakutia in World Museums: (17th - early 20th Centuries), Material and Spiritual Culture of the Peoples of Yakutia in World Museums, Bichik, Yakutsk, 2017.
- [21] National Art Museum of the Republic of Sakha (Yakutia), Sostojalas torzhestvennaja prezentacija alboma-kataloga "materialnaja i duhovnaja kultura narodov jakutii XVII–nach. XX vv. v muzejah mira" [the grand presentation of the album-catalogue "material and spiritual culture of the peoples of yakutia in the 17th–early 20th centuries in the museums of the world" took place], 2023. URL: <https://sakhamuseum.ru/vystavki-i-sobytiya/novosti/SostoyalastorzhestvennayaprezentatsiyaalbamakatalogaMaterialnayaidukhovnayakulturanarodovYakutiiXVII/>.
- [22] A. Gerritsen, G. Riello, *The Global Lives of Things: Material culture in the first global age* 1, Routledge, 2015.
- [23] D. Maleuvre, *Museum Memories: History, Technology, Art*, Stanford University Press, Stanford, 1999.
- [24] A. A. Shelton, *Museums and Anthropologies*, John Wiley & Sons, Ltd, 2006, p. 64–80. URL: <https://onlinelibrary.wiley.com/doi/abs/10.1002/9780470996836.ch5>. doi:10.1002/9780470996836.ch5.
- [25] S. A. Crane, *The Conundrum of Ephemerality*, John Wiley & Sons, Ltd, 2006, p. 98–110. URL: <https://onlinelibrary.wiley.com/doi/abs/10.1002/9780470996836.ch7>. doi:10.1002/9780470996836.ch7.
- [26] S. Conn, *Science Museums and the Culture Wars*, John Wiley & Sons, Ltd, 2006, p. 494–508. URL: <https://onlinelibrary.wiley.com/doi/abs/10.1002/9780470996836.ch30>. doi:10.1002/9780470996836.ch30.
- [27] D. Garneau, *From Colonial Trophy Case to Non-Colonial Keeping House*, Routledge, 2022.
- [28] G. Sculthorpe, *Provenance and Place in Indigenous Australia*, Rowman & Littlefield, Lanham, 2019.
- [29] S. Macdonald, *Collecting Practices*, John Wiley & Sons, Ltd, 2006, p. 81–97. URL: <https://onlinelibrary.wiley.com/doi/abs/10.1002/9780470996836.ch6>. doi:10.1002/9780470996836.ch6.
- [30] A. Muñoz, *From Curiosa to World Culture. The History of the Latin American Collections at the Museum of World Culture in Sweden*, Ph.D. thesis, University of Gothenburg, 2011. URL: <https://gupea.ub.gu.se/handle/2077/25554>.
- [31] C. Bründlmayer, "Out of Context" – Translocation of West African Artefacts to European Museums: The Case of the Leo Frobenius Collection from Mali, Routledge, 2023.
- [32] R. Parry, *Museum/Computer: A History of Disconnect*, Routledge, 2007.
- [33] The Dublin Core Metadata Initiative, Dcmi history, 2024. URL: <https://www.dublincore.org/about/history/>, accessed: 2025-01-15.
- [34] Documentation Standards Working Group, The cidoc crm, 2024. URL: http://old.cidoc-crm.org/previous_releases_cidoc.html, accessed: 2025-01-15.
- [35] GO FAIR, Fair principles, 2016. URL: <https://www.go-fair.org/fair-principles/>, published in 2016. Accessed:2025-12-12.
- [36] V. Zaiden, A. Avelar, Cultural heritage as a reparation mechanism: Germany and brazil, *ICOFOM Study Series* (2021) 228–241. doi:10.4000/iss.3959.
- [37] S. The United Nations Educational, C. Organization, *Fight illicit trafficking (1970 convention)*, 1970. URL: <https://www.unesco.org/en/fight-illicit-trafficking/about>, accessed: 2025-03-03.
- [38] A. Desvallées, F. Mairesse (Eds.), *Key concepts of museology*, Armand Colin, Paris, 2010.
- [39] I. E. Committee, *ICOM Code of Ethics for Museums*, ICOM, Paris, 2017.
- [40] O. L. Figueroa-Rodríguez, J. Holbrook, R. Lovett, S. Materechera, M. Parsons, K. Raseroka, The care principles for indigenous data governance, *Data Science Journal* 19 (2020). URL: <https://datascience.codata.org/articles/10.5334/dsj-2020-043>. doi:10.5334/dsj-2020-043.
- [41] N. B. Thylstrup, *The Politics of Mass Digitization*, MIT Press, 2024.

- [42] D. E. Marsh, Digital knowledge sharing: perspectives on use, impacts, risks, and best practices according to native american and indigenous community-based researchers, *Archival Science* 23 (2023) 81–115. doi:10.1007/s10502-021-09378-9.
- [43] S. Sánchez Membrilla, Indigenous Collections at the Museum of World Culture: Digitisation, Decolonisation and Other Stories, Ph.D. thesis, Linnaeus University, 2024. URL: <https://urn.kb.se/resolve?urn=urn:nbn:se:lnu:diva-127196>.
- [44] K. Kuoljok, Same Film Archive in the Digital Age: Challenges and Opportunities in Creating Metadata, Springer Nature Switzerland, Cham, 2025, p. 67–86. URL: https://doi.org/10.1007/978-3-031-76941-2_4. doi:10.1007/978-3-031-76941-2_4.
- [45] L. K. Gibson, Pots, belts, and medicine containers: Challenging colonial-era categories and classifications in the digital age, *Journal of Cultural Management and Cultural Policy / Zeitschrift für Kulturmanagement und Kulturpolitik* 6 (2020) 77–106. doi:10.14361/zkmm-2020-0204.
- [46] A. Flinn, M. Stevens, E. Shepherd, Whose memories, whose archives? independent community archives, autonomy and the mainstream, *Archival Science* 9 (2009) 71. doi:10.1007/s10502-009-9105-2.
- [47] F. for Siberian Cultures, Digital museum of the north, 2024. URL: <https://mu-north.org/>, accessed: 2025-07-23.
- [48] E. Kasten, N. Mamontova, D. Oparin, V. Solovyeva, L. Zdor, M. Zdor, Co-producing Knowledge About Western Museum Collections: An Avenue for Siberian Communities' Engagement, Fürstenberg/Havel: Kulturstiftung Sibirien, 2025.
- [49] E. Niggemann, J. D. Decker, M. Lévy, E. Commission (Eds.), The new renaissance: report of the Comité des Sages on bringing Europe's cultural heritage online, Publications Office, Luxembourg, 2011. doi:10.2759/45571.
- [50] Europeana Foundation, Edm – mapping guidelines – europeana knowledge base – confluence, 2024. URL: <https://europeana.atlassian.net/wiki/spaces/EF/pages/987791389/EDM+-+Mapping+guidelines>, accessed: 2025-12-12.
- [51] Europeana Foundation, Apis, 2024. URL: <https://pro.europeana.eu/discover-the-data/apis>, accessed: 2025-12-12.
- [52] J. E. Cejudo Grano de Oro, pyeuropeana: A python wrapper around europeana apis, 2022. URL: <https://github.com/europeana/rd-europeana-python-api>.
- [53] Europeana.eu, Sakha (place in collections), 2024. URL: <https://www.europeana.eu/en/collections/place/199230-sakha>, accessed: 2025-01-15.
- [54] A. Isaac, M. Marrero, N. Jarrett, Search europeana.eu in spanish, polish, romanian and hungarian, 2025. URL: <https://pro.europeana.eu/post/search-europeana-eu-in-spanish-polish-romanian-and-hungarian>, updated on Wednesday August 20, 2025.
- [55] E. Union, Languages, multilingualism, language rules | european union, 2019. URL: https://european-union.europa.eu/principles-countries-history/languages_en, (accessed Feb. 2, 2026).
- [56] K. de Burger, Unlocking cultural heritage: A case study of metadata management at europeana, Lunds universitet, 2024. URL: <http://lup.lub.lu.se/student-papers/record/9154811>, master's thesis.
- [57] D. Q. C. Europeana Foundation, Problem patterns, 2017. URL: <https://docs.google.com/spreadsheets/d/1atZr1w-h9AdWwWSBYLCCk6fAdJSxCNP56QRLNY1jLg>.
- [58] S. Kapidakis, Comparing metadata quality in the europeana context, in: Proceedings of the 5th International Conference on Pervasive Technologies Related to Assistive Environments, PETRA '12, Association for Computing Machinery, New York, NY, USA, 2012, p. 1–8. URL: <https://doi.org/10.1145/2413097.2413129>. doi:10.1145/2413097.2413129.
- [59] A. I. Savvinov, Priamurskie yakuty v materialakh petra shimkevicha: k istorii formirovaniya etnoterritorialnykh grupp v priamurie [the amur yakuts in the materials of peter shimkevich: the formation history of ethnic and territorial groups in the amur region], *Severo-Vostochnyy gumanitarnyy vestnik* 4 (2020) 45–57. doi:10.25693/SVGv.2020.33.4.005.
- [60] DE-BIAS, De-bias: Vocabulary – english, 2025. URL: <https://pro.europeana.eu/page/the-de-bias-vocabulary>.

- [61] Europeana, Russland, jakutien, kolyma, primitive ski, 2022. URL: https://www.europeana.eu/en/item/440/item_HHSWJTU5ULA7JFXA3C2I2JBT7V22GZYQ, (accessed Nov. 25, 2025).