

Decolonising digital Cultural Heritage Institutions

Harmful content and contentious language : the case of
Europeana

Aida FADIOUI

Supervisor: Prof. F. Truyen
KULeuven

Co-supervisor: Jolan Wuyts
Europeana Foundation

Thesis presented in
fulfillment of the requirements
for the degree of Master of Science
in Digital Humanities

Academic year 2021-2022

Preface

This internship report delves into the findings of research done on the topic of decolonising cultural heritage institutions. I would like to thank my supervisor Prof. Frederik Truyen and the KULeuven Digital Humanities faculty for their help during this internship. I would also like to express my deepest appreciation to everyone at the Europeana foundation for welcoming me and in particular to the Collections Engagement and the Diversity and Inclusion cross-team with whom I've had the great pleasure to collaborate for the invaluable knowledge and expertise they have imparted on me. Finally, I'm extremely grateful to my supervisor [Jolan Wuyts](#) for his constant support, feedback and assistance at every stage of this internship.

Summary

Most cultural heritage institutions are currently taking steps to towards decolonisation, whether it is on a structural level or for the collections they hold. The open data meta-aggregator and display space Europeana is currently laying the foundations of this work which among other things includes looking at what their partners and other institutions worldwide are doing. From editing and creating new metadata fields for their records to creating content and language warnings as well as context notices through the creation of clear user feedback mechanisms, what emerges from this research is the need to not only work closely with providers and local minoritised communities, but also the necessity of hiring and training people to deal with these issues.

Introduction	4
Quantitative data analysis	9
Metadata standards and editing	21
Contextualising objects	26
Content warnings and harmful language statements	28
Community engagement/co-creation	31
Challenges	34
Conclusion	35
Appendice 1 : Contentious/outdated terms list	36
Appendice 2 : Contentious terms API query	40
Appendice 3 : draft for Europeana's Statement on Potentially Harmful and Offensive Language and Content	41
References	46

Introduction

This report aims to delve into the findings of research done on the topic of decolonising cultural heritage institutions and harmful content and will try to outline some possible steps Europeana as an institution can take in that direction. While interning remotely at Europeana, I was able to work closely with the Collection Engagement team and the Diversity & Inclusion cross team. Established in January 2021¹, this cross team's goal is to explore and design changes to the Europeana Foundation structures in order to make everyone feel welcome, represented and safe - be it within the foundation or at the different events they organise, as well as to identify areas of concern and to understand the unconscious biases within the foundation. My tasks during this internship ranged from desk research to connecting with representatives of cultural heritage institutions, and the occasional writing of blog posts.

Europeana is an open data meta-aggregator and display space created by the Europeana Initiative with the support and funding of the European Commission, giving cultural heritage enthusiasts, professionals, teachers and researchers access to the digitised heritage collections of more than 4000 institutions across the continent². The platform holds records of over 50 million objects, and to achieve this Europeana works with a [network of aggregating partners](#) in order to collect, check and enrich the data. Europeana is funded by the European Union's Connecting Europe Facility and is operated by a consortium led by the non-profit Europeana Foundation³. Europeana provides several services, one of them of course being the [Collections portal](#), but there is also a series of [APIs](#) available.

¹ <https://pro.europeana.eu/post/meet-the-europeana-foundation-diversity-and-inclusion-cross-team>

² <https://www.europeana.eu/en/about-us>

³ <https://pro.europeana.eu/about-us/mission>

To be able to discuss harmful content and the possible guidelines that could be followed to deal with it, we first need to take a step back and understand the larger discussion happening right now in cultural heritage institutions (CHIs) around diversity, inclusion and equity, as well as the history of said institutions. CHIs today have an undeniable role in education and research, yet just like many other institutions, they are finding themselves in a crisis in societies more and more concerned with social and climate justice, intersectionality and other social causes (Okello Abungu, 2019, p. 64). Most of these institutions did however originate in the context of the 18th and 19th century, a period of rapid industrialisation and colonial enterprise and have a long and bloody history of plundering and spoliation embedded in their very structures⁴, and the very first step towards decolonisation is admitting this. Our institutions were built through the conquest of lands, the subjugation of people and the displacement of said people and what was theirs (Vawda, 2019, p. 74).

In many ways, a museum functions as the holder of accumulated privatised objects plundered from 'primitive' societies. It is a 'technology of the Western gaze, where the active and dominant Western subject defines 'others' as passive objects' (Hamm & Schönberger, 2021, p.34), which is problematic in many ways: first, because these artefacts once removed and abstracted from their original environments can no longer be understood in relation to their context, stories and people⁵. Second, and this transpires especially in the discussions surrounding repatriation and restitution, the belief that European and 'Western' institutions are more capable of preserving these artefacts than the cultures from which they were taken is still prevalent. All these objects are relegated to the past, viewed 'along a singular, linear concept of time in which Western modernity can invent tradition (in the singular) by naming its stages of progress and development' (Cushman, 2013, p. 119). They are not only cut off from their roots, but it is also assumed that there are

⁴ (Mineo, 2020)

⁵ (Cushman, 2013, p. 120)

no longer any living rightful inheritors or stewards of these objects, while these very people whose objects are being held by these institutions are also still exposed to violence.

We have the good fortune to be working here with a new type of institution and collections : digital ones. Digital archives and repositories afford us new possibilities not only to contextualise these objects in more meaningful ways, but also 'to disrupt, review, question, and revise the colonial knowledge regime that informs the archives from which we draw most of our materials' (*Decolonizing the Archive: Digital Remix and Reassembly*, ECDA). While artists and heritage providers are usually presented as the only legitimate experts on these objects, digital storytelling allows everyone, including the long silenced cultural descendants of those who were spoiled to tell the world about their history.

If many have been calling for the abolition of the museum as it is, because just as prisons and universities it is a settler institution and it produces the same dynamics⁶, it is still possible to engage with these structures in hopes of bettering them. So what exactly does decolonisation look like ? Well, the good thing is that there is no blueprint for these practices, no end goal : it is a process and each institution has to figure out what it looks like for itself . What we can do however, is learn from those who have started this process, which is why we will start this report by looking at some of the best practices found in the field. Next, we shall take a closer look at the Europeana API and the harmful content one can find by using it, and discuss points more specific to this institution, while trying to give a set of recommendations for the next steps we can take.

Now some might wonder, why should we care ? Why does it matter, especially when so many people see decolonisation as a fleeting trend? Simply put, because this is the only way forward. As Albert Memmi stated in his portrait of *The Coloniser and the Colonised*, colonisation materially and spiritually kills the colonised, but it also

⁶ (Azoulay, 2021)

'distorts relationships, destroys or petrifies institutions, and corrupts men, both colonisers and colonised' (Memmi, 2003, p.195). To think colonisers and colonised went through trauma of this magnitude and that it left us, our structures and institutions completely unscathed would be foolish, and to believe it is not affecting each and every one of us to this day would be dangerous.

An element that has long hindered CHI's efforts on the subject is the matter of responsibility. While more and more cultural organisations are slowly shifting from seeing themselves as stewards of the objects they hold rather than owners⁷, the definition of stewardship itself is still largely discussed⁸. Essentially, stewardship means being responsible for the preservation of the value of a resource or object; in the case of cultural stewardship it refers to a set of practices put in place to preserve works, objects, and events and provide *'the means by which people may take future opportunities to benefit from that value'* (Furner). It is about taking adequate care, which is defined by psychologists Anjali Dutt and Danielle Kohfeldt : *'a species of activity that includes everything we do to maintain, contain, and repair our 'world' so that we can live in it as well as possible'*⁹ and responsibility. It is also understanding that the way things have been done until now was often at the expense of community-based or Indigenous frameworks and ways of care.

Because institutions have long permitted discriminatory archival practices, every collection and archive is 'embedded in systems of power that shape what counts as knowledge and non-knowledge' (*Decolonizing the Archive: Digital Remix and Reassembly*, ECDA). While we cannot be sure that what we define as an objective judgement today will still be considered as such in the future, we can at least try to make the changes we believe necessary such as changing some terms in order to separate people's identities from their circumstances (eg. 'slave' to 'enslaved

⁷ (Mineo, 2020)

⁸ (McKechnie, 2014)

⁹ Anjali Dutt & Danielle Kohfeldt (2018) as cited by Francois et al., (2021)

person')¹⁰. While debates are still ongoing on how exactly this should be done, the consensus currently is on replacing terms or contextualising them, not erasing them. Not only because they may be critical resources for research but also because erasing these narratives will not solve anything : we need to know them in order to learn from them.

Quantitative data analysis

Based on resources such as Words Matter (Tropenmuseum, 2018), Carissa Chew's [Inclusive Terminology Glossary for cultural heritage professionals](#) (2022) and an internal document on the Rijksmuseum's research on replacing contentious terms, as well as personal research, I have first constructed a list of contentious and/or outdated terms that can be found in Europeana. This list only catalogues terms in English, French and Dutch (see appendice 1) and is of course non-exhaustive. This first draft was then used to conduct an exploration of the Europeana database, where the first step was to try and find out the frequency at which these terms appear (the detailed code for this can be found in appendice 2) :

```
{'handicaped': 3,  
'barbaar': 14,  
'inboorling': 15,  
'bedeltje': 17,  
'bohémiennes': 32,  
'nikker': 46,  
'Koelie': 62,  
'Kaffer': 75,  
'Événements d'Algérie': 78,  
'infirmes': 79,
```

¹⁰ (Andia, 2021)

'Mulatto': 82,
'Bohémien': 104,
'infirmes': 120,
'Mohammedan': 133,
'cripples': 134,
'Gitan': 148,
'negroes': 158,
'handicapé': 185,
'mûlatre': 217,
'berbers': 235,
'handicapés': 242,
'eskimos': 243,
'bohémienne': 260,
'négresse': 310,
'barbares': 344,
'Indiaan': 388,
'barbare': 420,
'gitans': 433,
'Politionele Actie': 445,
'Coolie': 541,
'crippled': 542,
'Kaffir': 548,
'bohémiens': 695,
'primitif': 775,
'Hottentot': 835,
'berbère': 871,
'Maure': 911,
'Zigeuner': 925,
'berbères': 943,
'aboriginal': 1076,

'talisman': 1095,
'barbarian': 1143,
'invalide': 1226,
'orientaux': 1230,
'gypsies': 1242,
'Lapon': 1295,
'neger': 1304,
'eskimo': 1386,
'Lappen': 1525,
'cripple': 1573,
'amulette': 1613,
'orientals': 1676,
'indigène': 1979,
'Mauresque': 2026,
'savage': 2068,
'berber': 2312,
'slaves': 2451,
'indiens': 2749,
'Planter': 3021,
'Gypsy': 3565,
'disabled': 3760,
'primitive': 3775,
'slave': 3980,
'Indians': 4271,
'invalides': 5064,
'Bombay': 6784,
'amulet': 8835,
'orientales': 9300,
'Lapp': 9640,
'sauvage': 11885,

'Colored': 18510,
'Moor': 26311,
'Laponie': 34155,
'Lapland': 37827,
'nègre': 54547,
'Indian': 55507,
'Indien': 104463,
'Orientale': 120716,
'Oriental': 137836,
'negro': 331462}

Of course these results have to be nuanced : for the terms 'negro' and 'nègre' many results refer either to the colour black in Spanish or Catalan, different regions in South America but also they can also be family names. For 'Oriental' and 'Orientale', it also often refers to a specific geographical location ; in several cases the terms 'Indian' and 'Indien' do refer to India and not to Indigenous North Americans.

We have, for example, 2451 results for the term 'slaves'. It would be interesting now to find out in which institutions' collections the word is more frequent. In order to make the results more readable here we will start with the institutions where we find 10 instances or more :

{'RBB AIT DISMARC': 10,
'DFF – German Film Institute & Film Museum': 10,
'Heidelberg University Library': 11,
'TV3 Catalan Television': 11,
'GESIS - Leibniz Institute for the Social Sciences. Library Cologne': 11,
'Institute for War-, Holocaust- and Genocidestudies': 12,
'American Jewish Historical Society': 12,
'Ghent University Library': 12,

'Aikaterini Laskaridis Foundation': 13,
 'PMR Maeyaert': 13,
 'Digital Repository Research Institutes': 13,
 'Horniman Museum and Gardens': 15,
 'The Portable Antiquities Scheme': 15,
 'Documentary Repository of the University of Salamanca': 15,
 'Film Library of Bologna': 16,
 'Institute of Historical Research': 17,
 'Uppsala University': 18,
 'National Library of Scotland': 19,
 'National Library of Poland': 20,
 'Luce Institute': 23,
 'Jagiellonian Digital Library': 24,
 'Library of the Czech Academy of Sciences': 27,
 'Center of the Judeo-Moroccan Culture': 29,
 'National Library of Israel': 30,
 'Wellcome Collection': 33,
 'Bavarian State Library': 38,
 'Royal Museums Greenwich': 47,
 'Rijksmuseum': 48,
 'Ignacio Larramendi Foundation': 72,
 'Bodleian Libraries, University of Oxford': 108,
 'National Library of France': 123,
 'Greater Digital Library': 124,
 'Promoter Digital Gallery': 173,
 'National Museum of World Cultures Foundation': 274,
 'The British Library': 735}

As for the singular 'slave', we have 3980 results which are distributed as follows :

{'RBB AIT DISMARC': 10,
 'TV3 Catalan Television': 10,
 'Allard Pierson': 10,
 'DFF – German Film Institute & Film Museum': 10,
 'Ghent University Library': 10,
 'Hellenic Broadcasting Corporation': 11,
 'National Library of the Netherlands': 12,
 'City of Music': 12,
 'Internet Culturale': 12,
 'National Gallery of Denmark': 12,
 'Royal Botanic Gardens Kew': 12,
 'The Royal Library: The National Library of Denmark and Copenhagen
 University Library': 13,
 'Deutsche Fotothek': 14,
 'Documentary Repository of the University of Salamanca': 14,
 'Bristol Museums, Galleries & Archives': 14,
 'Research deposit of the University of Seville': 15,
 'Egypt Centre': 16,
 'Royal Museums of Art and History, Brussels': 18,
 'Deutsche Welle': 19,
 'IMSLP/Petrucci Music Library': 20,
 'National Library of Spain': 20,
 'National Library of Poland': 21,
 'The Portable Antiquities Scheme': 22,
 'Romanian Academy Library': 22,
 'National and University Library of Slovenia': 24,
 'Austrian National Library': 24,
 'Braidense National Library': 24,
 'Heidelberg University Library': 25,
 'Rijksmuseum': 26,

'Luce Institute': 26,
 'Italian Geographical Society Photo-Archive': 26,
 'National Library of Scotland': 29,
 'The Digital Network of Museum Collections in Spain': 29,
 'Film Library of Bologna': 30,
 'Naturalis Biodiversity Center': 32,
 'Wellcome Collection': 38,
 'Uppsala University': 39,
 'National Library of Israel': 40,
 'National Audiovisual Institute France': 40,
 'Jagiellonian Digital Library': 41,
 'Library of the Czech Academy of Sciences': 41,
 'The Language Archive': 43,
 'Institute for War-, Holocaust- and Genocidestudies': 48,
 'Bavarian State Library': 61,
 'Ignacio Larramendi Foundation': 72,
 'Open Society Archives at Central European University': 72,
 'Greater Digital Library': 75,
 'Aikaterini Laskaridis Foundation': 82,
 'GESIS - Leibniz Institute for the Social Sciences. Library Cologne': 84,
 'Royal Museums Greenwich': 102,
 'Academy of Athens': 116,
 'Institute of Historical Research': 119,
 'National Museum of World Cultures Foundation': 140,
 'Internet Archive': 147,
 'Schola Graphidis Art Collection, Budapest': 158,
 'National Library of France': 259,
 'Bodleian Libraries, University of Oxford': 491,
 'The British Library': 557}

For both terms, the highest results are found in British institutions, the British Library and the Bodleian Libraries, which is not a surprise as they are after all English terms. For the plural 'slaves', the Greater Digital Library, a Polish institution, is another institution where we see a high frequency of the term. All the results are however from a bilingual (Polish and French) periodical dedicated to Slavic cultures ('slave' is French for 'Slavic'). The term is also present 173 times in the Promoter Digital Gallery's collection, all occurrences being sketches by Situ Zhaoguang for a state-commissioned sculpture named 'Anger of the Farming Slaves'.

The other institutions with the highest frequencies of the word 'slave' in their collection here on Europeana are the Bibliothèque Nationale de France and the Schola Graphidis Art Collection in Budapest, which once again are objects referencing Slavic cultures : for the Graphidis Art Collection, all the results are prints of ornamental art seemingly from the same manuscript whose description is in French.

While having these numbers gives us a good first insight into the presence of contentious and outdated terms in the Europeana collections, having a person go through each instance to know exactly how they are used seems necessary. As an illustration, we will go through the instances of the word 'bohémien' as well as the feminine and plural forms of the word (bohémiens, bohémienne(s)) which in French is a misnomer referring to Roma people, as they were thought to be originating from Bohemia¹¹. The word is still very commonly used, not only to refer to Roma people but especially to refer to an artistic and unconventional way of living. The use of the term started in 19th century France where the movement emerged in conjunction with Romanticism which was a more aristocratic or bourgeois movement. It was about living in poverty and hunger but also a certain carelessness and a rejection of industrial society¹². While 'bohémien' may generally have a positive connotation when describing a romanticised vagabond lifestyle, especially as it later evolved into 'boho'

¹¹ The term is defined by the CNRTL as a "*tzigane, member of a wandering tribe that were believed to be originating from Bohemia*".

¹² (Académie française, 2013)

and 'bobo' (bourgeois-bohème), the term is not the appropriate one to describe Roma people. On Europeana's website, we can find the different forms of the term 'bohémien' in the collections of 96 different institutions, with 51 having a frequency of more than 1 :

{'National Szechenyi Library': 1,
'Wellcome Collection': 1,
'Lucian Blaga Central University Library, Cluj-Napoca, Romania': 1,
'The University and National Library of Debrecen': 1,
'Uppsala University': 1,
'Örebro County Museum': 1,
'Digital Memory of Catalonia': 1,
'Association Films Plans-Fixes': 1,
'Historical Museum': 1,
'Central Institute for the Union Catalogue of Italian Libraries': 1,
'Under and Tuglas Literature Centre': 1,
'Bibliotheca Hertziana – Max Planck Institut for Art History': 1,
'Sancho el Sabio Foundation': 1,
'Turin Gallery for Modern and Contemporary Art': 1,
'Meise Botanic Garden': 1,
'Kunstbibliothek Berlin': 1,
'Modern Art Museum': 1,
'Luce Institute': 1,
'Austrian Broadcasting Corporation': 1,
'National Audiovisual Archive of Hungary': 1,
'GESIS - Leibniz Institute for the Social Sciences. Library Cologne': 1,
'National Heritage Institute, Bucharest': 1,
'Luigi Sturzo Institute': 1,
'Central Library of the Bulgarian Academy of Sciences': 1,
'Royal Museums of Fine Arts of Belgium': 1,

'TopFoto': 1,
 'Centre for History and Analysis of Recorded Music, King's College London;
 United Kingdom': 1,
 'M Leuven': 1,
 'Historic Center Leeuwarden': 1,
 'Rotterdam City Archives': 1,
 'Europeana 1914-1918': 1,
 'Aikaterini Laskaridis Foundation': 1,
 'Naturalis Biodiversity Center': 1,
 'Madeira Arts Library': 1,
 'West Pomeranian Digital Library 'Pomerania': 1,
 'Silesian Digital Library': 1,
 'Pomeranian Digital Library': 1,
 'Mazowsze Digital Library': 1,
 'E-library of Warsaw University': 1,
 'Leo Baeck Institute, Center for Jewish History': 1,
 'Catalunya Film Archive': 1,
 'The Royal Belgian Film Archives': 1,
 'National Audiovisual Institute': 1,
 'Museum of City History Leipzig': 1,
 'Philharmonie de Paris': 1,
 'National Institute for Art History Library': 2,
 'Baden State Library': 2,
 'Museum Ludwig, Cologne': 2,
 'Jewish Historical Museum': 2,
 'Digital Art and Culture Archive Düsseldorf': 2,
 'National Audiovisual Institute France': 2,
 'Allard Pierson': 2,
 'Archives of North Holland': 2,
 'Institute for War-, Holocaust- and Genocidestudies': 2,

'National Library of Portugal': 2,
 'National Film Archive': 2,
 'Museum of Fine Arts Ghent': 2,
 'Deutsche Fotothek': 3,
 'MuseiD-Italia': 3,
 'City Archives' s-Hertogenbosch': 3,
 'Royal Museum of Fine Arts Antwerp': 3,
 'National Archives of France': 3,
 'Braidense National Library': 4,
 'The Royal Armoury': 4,
 'Rhône-Alpes Laboratory for Historical Research': 4,
 'Internet Culturale': 4,
 'MAK – Museum of Applied Arts': 4,
 'Hellenic Literary and Historical Archive - Cultural Foundation of the National
 Bank Of Greece': 4,
 'Interuniversity Health Library': 4,
 'Bodleian Libraries, University of Oxford': 5,
 'Institute of Cultural Heritage of Spain': 5,
 'Teylers Museum': 5,
 'Municipal Library of Prague': 5,
 'National Library of the Czech Republic': 6,
 'Deutsche Welle': 6,
 'German Documentation Center for Art History - Marburg Picture Index': 7,
 'Albertina': 7,
 'Ghent University Library': 8,
 'The Digital Network of Museum Collections in Spain': 8,
 'Rijksmuseum': 9,
 'Russian State Library': 11,
 'City of Music': 11,
 'National Library of Poland': 12,

'Herzog Anton Ulrich Museum': 12,
'Austrian National Library': 17,
'Heidelberg University Library': 17,
'National Library of the Netherlands': 18,
'Media Library of Architecture and Heritage': 18,
'Royal Institute for Cultural Heritage': 19,
'Czech Television': 22,
'Ministry of Culture': 22,
'IMSLP/Petrucci Music Library': 36,
'Bavarian State Library': 39,
'National Library of Spain': 43,
'ERIAC': 298,
'National Library of France': 312}

The highest frequency is found in the collection of the Bibliothèque Nationale de France. For some of the items such as in the [*Grandes Chroniques de France*](#) manuscript dated from the Middle Ages, the word 'bohémien' is found in the 'format' section of the metadata and it refers to a master illuminator from Bohemia, or more generally to an individual from Bohemia. For the other items which are either monographs, engravings or sound recordings, the term is found in the title of the work. Several items are royal ordinances on beggars and vagrants and thus by extension also concern 'bohemians', dating from the 17th and 18th century; in several prints the term is in the title while in the metadata we can see that the tag used in 'subject' is 'Tsiganes'. For the National Library of Spain, 5 items are prints that also have French titles which is the field where the term 'bohémien' is found, while the rest of the items, as is the case for the Bavarian State Library, are musical scores. For the IMSLP/Petrucci Music Library, all results are sound recordings while all the items from the French Ministry of Culture are prints representing Roma people with the term 'bohémien' in the title. Out of the 22 items, five specify that the title is either the one

[inscribed on the original item](#)¹³, that it was an [old title](#)¹⁴ or they [stipulate the origin of the title](#).

For the Czech Television collection which has the same frequency as the French Ministry of Culture, the term is found in the 'places' field of the metadata referring to the region of Bohemia. The 19 items from the Belgian Royal Institute for Cultural Heritage (KIK-IRPA) on the other hand all refer to Roma people : 8 items are pictures from the same photographer, Jacques Hersleven, the others are either paintings or ceramic ornaments. All but two objects, one painting representing two horses and ceramic figurines representing dancing people, have the term 'Bohémien' in the title and the term 'gitan' in the subject field, which is another misnomer for Roma people.

In the majority of the cases we have seen, the contentious term is found in the title of the object, which is given by the data provider. It also appears that tags in the 'subject' can be outdated or contentious terms. A large part of my internship was dedicated to doing desk research on harmful content guidelines in other CHIs. From this, I have compiled some examples of best practices and possible guidelines that can be followed to make the Europeana collections more appropriate, inclusive and representative, and the first ones we will explore involve metadata.

Metadata standards and editing

All the data that is being ingested, managed and published on Europeana follows the Europeana Data Model (EDM) which is a Resource Description Framework (RDF) based model. Because different cultural institutions use different data standards, the goal when developing EDM was to adopt an open, cross-domain Semantic Web-based framework rather than build a model based on a specific

¹³ Mention : (titre inscrit).

¹⁴ Mention : (ancien titre) or "[dit autrefois](#)".

community standard¹⁵. What is interesting is that EDM permits the distinction between the object that the structure is about, its digital representations and the intellectual and technical creation which is submitted by the provider¹⁶. Adding metadata fields would thus be a viable option for Europeana.

The Digital Library North (DLN) project, which is still ongoing, strives to create a digital library infrastructure that can address the information needs in Canada's northern regions (Farnel et al., 2017). It is a great example of a community driven initiative that led to the development of a metadata framework that can accurately and adequately describe cultural resources. The framework was informed by interviews and surveys done with the Indigenous communities of Canada's northern regions but also user evaluation of the platform. Choosing the Dublin Core Metadata Element Set as well as creating custom elements allowed them to have a certain flexibility while assuring the interoperability of their metadata.

For researcher Kate Holterhoff (2017), who worked on how digital image archivists should approach the curation of potentially harmful objects, the best available means to achieve this presently is *"improving database search functionality through heavy editing—metadata that is voluminous, polyvocal, and critical"*. She also emphasises the importance of contextualising the digitised *"hate-promoting primary documents"*, which we will explore further in another part of this report.

As for EDM, the fields 'timestamp_created' and 'timestamp_update' are the only fields in the metadata relating to time and they indicated when the object was created, but there is no specific field that would allow providers to give information about when and why a harmful description would have been changed. The 'Dates' field in the Getty AAT (Art and Architecture Thesaurus) for example comprises 'Display Date', which indicates when the term was of common usage, as well as 'Start Date' and 'End Date', which are years that delimit that specific span of time. It also has a field 'historical flag' where it is indicated whether the term used is 'current' or

¹⁵ (Europeana, 2013)

¹⁶ (Europeana, 2013)

‘historical’, meaning it was typically used in the past but we now describe the concept with a different term¹⁷.

The addition of [Traditional Knowledge](#) (TK) labels could also prove to be interesting when dealing with material from a specific community with whom Europeana or the data providers might collaborate. Developed by the Local Contexts organisation through a partnership with several Indigenous communities across the globe, these labels allow the expression of local and specific conditions for sharing and engaging with their knowledge. The labels, available in English, French, Spanish and Māori, are divided into three classes :

- Provenance labels which *“identify the group or sub-group which is the primary cultural authority for the material, and/or recognizes other interest in the materials”*.
- Protocol labels that are used to *“outline traditional protocols associated with access to this material and invite viewers to respect community protocols”*. E.g., the material is only meant to circulate between women of the community; it contains secret or sacred information...
- Permission labels indicating *“what activities the community has approved as generally acceptable. Other uses require direct engagement with primary cultural authorities¹⁸”*. E.g., community use only, non-commercial, open to collaboration ...

The EDM is a component of the Europeana Publishing Framework, a quality standard for digital content which was created in 2015 and in 2019 and later completed with a quality standard for metadata. While the content and metadata that partners can provide depend on their capabilities as well as their own agendas, all data providers have to follow the minimum requirements; the quality of their material will place their metadata and content in different tiers :

¹⁷ (Getty Foundation)

¹⁸ (Local Contexts)

Europeana Publishing Framework: Metadata

Here you see how the different metadata tiers of the Publishing Framework work.

TIER	LANGUAGE	ENABLING ELEMENTS*	CONTEXTUAL CLASSES**	BENEFITS (CUMULATIVE)
A Europeana as a basic search platform. <i>'I want to find a specific item that I'm looking for.'</i>	At least 25% of the provided EDM metadata fields that are relevant have at least 1 language qualified value	At least one enabling element taken from one of the 'Discovery scenario' groups	None	Findability - material is available on Europeana Collections, indexed by search engines, and linked data technology Web traffic - click-throughs to your site
B Europeana as an exploration platform. <i>'I want to browse and explore Europeana even if I'm not sure what I'm looking for.'</i>	At least 50% of the provided EDM metadata fields that are relevant have at least 1 language qualified value	At least three distinct elements taken from two of the 'Discovery scenario' groups	At least one contextual class with all minimum required elements, OR link to LOD vocabulary	Greater potential for: Use in thematic collections - providing context and relation to other collections More marketing through Europeana
C Europeana as a knowledge platform. <i>'I want to search and browse in a more precise way, by named authors, specific subjects or topics.'</i>	At least 75% of the provided EDM metadata fields that are relevant have at least 1 language qualified value	At least four distinct elements taken from two of the 'Discovery scenario' groups	At least two distinct contextual classes with all minimum required elements, OR links to LOD vocabularies	Greater potential for: Use in Europeana's existing partnerships and projects , in e.g. education and research Use in apps and services Use by creative industry
<p>*Discovery scenario Browse by date or time-span Browse by subjects and types Browse by agents Browse by places</p> <p>**Contextual class edm:TimeSpan skos:Concept edm:Agent edm:Place</p> <p>Notes on contextual classes The contextual classes (listed above) can be included in the metadata at source, but the use of Linked Open Data vocabularies is preferred. References to contextual classes created as a result of Europeana's own Semantic Enrichment do not count for the tier classification.</p>				
<p>Enabling elements dcterms:created, dcterms:issued, dcterms:temporal, edm:hasMet (with a time-span - edm:TimeSpan class) dc:subject (with a concept - skos:Concept class), dc:format, dc:type, dcterms:medium dcterms:contributor, dc:publisher, dc:subject (with an agent - edm:Agent class), edm:hasMet (with an agent - edm:Agent class) dc:subject (with a place - edm:Place class), dcterms:spatial, edm:currentLocation</p> <p>Minimum required metadata elements edm:begin, edm:end skos:prefLabel and skos:note (or skos:broader, skos:narrower, skos:exactMatch, skos:closeMatch, skos:related) skos:prefLabel and rdaGr2:professionOrOccupation (or edm:begin, rdaGr2:dateOfBirth, edm:end, rdaGr2:dateOfDeath, rdaGr2:placeOfDeath, rdaGr2:placeOfBirth) skos:prefLabel, wgs84_pos:lat, wgs84_pos:long</p>				



EUROPEANA PUBLISHING FRAMEWORK: METADATA v0.8
Quick Summary

See [Europeana Publishing Framework](#) and [Europeana Publishing Guide](#) for more detailed information.

The tiers are based on three elements : language, enabling elements and contextual classes. *Language* refers to the use of language tags, which allows automatic linking and translation processes to be implemented. The usage of *enabling elements* permits a greater information retrieval and *contextual classes* improve objects' findability and accuracy. The more of these elements are provided by the partners, the higher up the tiers their collections are and thus the more benefits they have.

Europeana Publishing Framework: Content

Here you see how the different content tiers of the Publishing Framework work. In this example, we use images. For video, audio, text and 3D, please see the Europeana Publishing Guide.

TIER	DIRECT LINK TO OBJECT PROVIDED	RIGHTS STATEMENTS	EUROPEANA API DISPLAYS	BENEFITS (CUMULATIVE)
1 Europeana as a search engine . 'I want to search and browse collections online.'	Yes. Minimum of 0.1 megapixel in size	Any from the Europeana licensing Framework	Metadata plus direct link to object	Findability - indexed by search engines, linked data technology Web traffic - click-throughs to your site
2 Europeana as a showcase . 'I want to be guided through collections online.'	Yes. Minimum of 0.42 megapixel in size	Any from the Europeana licensing Framework	Metadata plus direct link to object	Use in thematic collections - providing context and relation to other collections on Europeana More marketing through Europeana
3 Europeana as a distribution platform for non-commercial reuse. 'I want to find, view and use collections in my own non-commercial projects.'	Yes. Minimum of 0.95 megapixel in size	Any from the Europeana licensing framework that allow for some re-use	Metadata plus direct link to object and can filter to show only re-usable objects	Impressions - collections viewed on platforms outside of Europeana Use in Europeana's existing partnerships and projects in e.g. education and research Use in apps and services by third parties
4 Europeana as a free reuse platform . 'I want to find, view and use collections in whatever way I choose.'	Yes. Minimum of 0.95 megapixel in size	Any from the Europeana licensing framework that allow free re-use	Metadata plus direct link to object and can filter to show only re-usable objects	Use on open platforms like Wikimedia Use by creative industry Use in commercial apps and services



EUROPEANA PUBLISHING FRAMEWORK: CONTENT
Quick Summary

The content tiers are also based on three elements : *direct link to object provided* which varies depending on the quality of said link; *rights statements* which have different levels of re-usability; and *Europeana API displays*. The minimum requirements are also applied to legacy material i.e. material that is already published on Europeana : removal of metadata and content that does not meet the standard or its improvement will then happen over time in order to ensure a consistent improvement in data quality¹⁹.

On the level of the publishing framework, two options could be explored :

- For the metadata, adding a condition on harmful/contentious terminology, either having it changed to more accurate terms or when it is in a title or description that cannot be changed, using quotation marks or specifying that the term or expression is no longer adequate.

¹⁹ (Scholz & Ciocoiu, 2022)

- For the content, ensuring that diversity and inclusion are taken into account when the material is picked. CHIs could be given the option to either work with or employ local minoritised communities, a sensitivity reader or a cultural consultant to curate and filter the content.

Contextualising objects

It can occasionally be difficult to see how some particular content is contentious : what is offensive to some isn't to others, especially if we don't have the keys to understand the context in which content was produced. We will take the example of colonial photography.

Photography was a very important instrument for colonial powers, as Teju Cole puts it : *“The dominant power decided that everything had to be seen and cataloged, a task for which photography was perfectly suited. Under the giant umbrella of colonialism, nothing would be allowed to remain hidden from the imperial authorities²⁰”*. Portraits taken by European photographers are legion in historical archives, especially in the collections of anthropological, ethnographical and geographical institutions. While we rarely know the exact context behind specific pictures, they are hardly ever authentic documentations of cultures but rather representations of the West's idea of the other. In *Staging the Egyptian Harem for Western Eyes* (2017), Felix Thürlemann explains how a specific *mashrabiya*, a traditional decorative enclosed balcony allowing the people inside the house to see the streets without being seen, led him to find out that 19th century photographs claiming to depict everyday life were in fact completely staged. By comparing them with other documentaries, he realised that the very same wooden screen was present in hundreds of different pictures.

²⁰ (Cole, 2019)



Femmes marocaines by Flandrin - Center of the Judeo-Moroccan Culture, Belgium - In Copyright - Educational Use Permitted.
<https://www.europeana.eu/en/item/271/https://www.judaisme.marocain.org/30618>



kvinnor, fotografi, photograph by Lekegian, Gabriel - 1890 - Museum of Ethnography, Sweden - CC BY-NC-ND.
https://www.europeana.eu/en/item/91627/SMVK_EM_fotografi_2079904

This is of course not to say that all these photographers were ill-intentioned, but what we consider as authentic ethnographic material today might actually have been advertisements aimed at European and North American tourists. And this is exactly why context is important : although editorial posts on the Europeana do a great job bringing attention to marginalised groups, and presenting objects in the collections within the wider context in which they were created, such as the [China in Perspective exhibition](#), or the blog on [the queer art of GAN](#), all those objects are still accessible without this context through the portal.

One solution that could be implemented would be to create context notices that pop-up on the item pages, giving a short explanation of the socio-historical context of the object and linking to further resources on the subject. In the next part of this report we will see an example of these notices implemented on Trove, an

Australian database aggregator. These notices could be linked to specific metadata tags, especially ones in the subject, object and geographical location fields.

Content warnings and harmful language statements

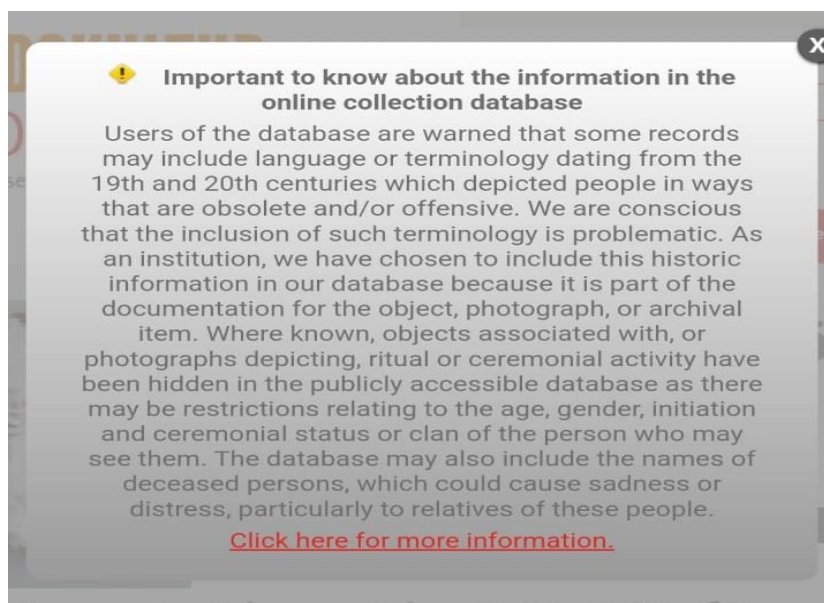
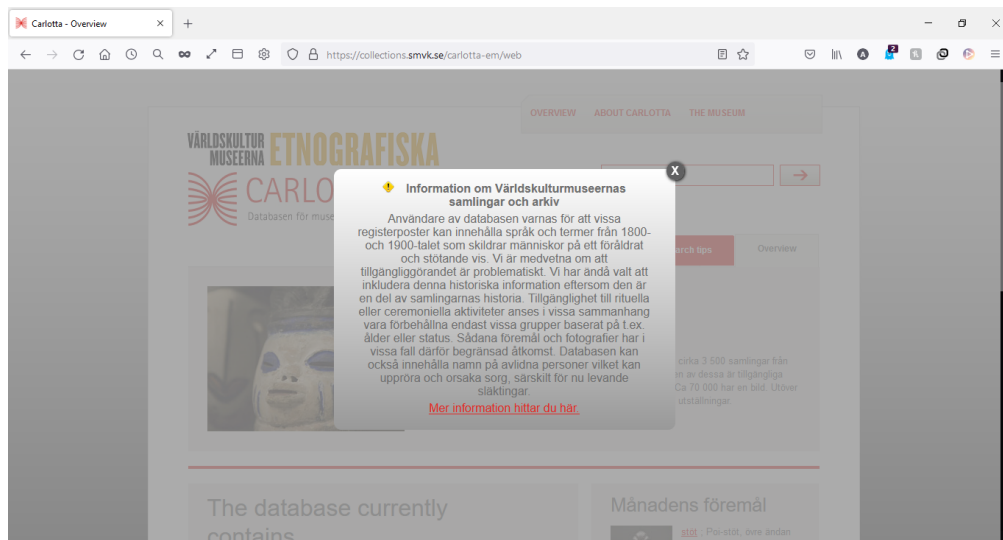
Another common practice for cultural heritage institutions online is using content warnings. Editorial content published on the Europeana can sometimes link to items of a sensitive nature, and since the beginning of this year a content warning has been introduced that when included by editors gives readers the option to proceed with viewing the editorial²¹. For example, the [Black people in European art](#) gallery comes with the following warning :



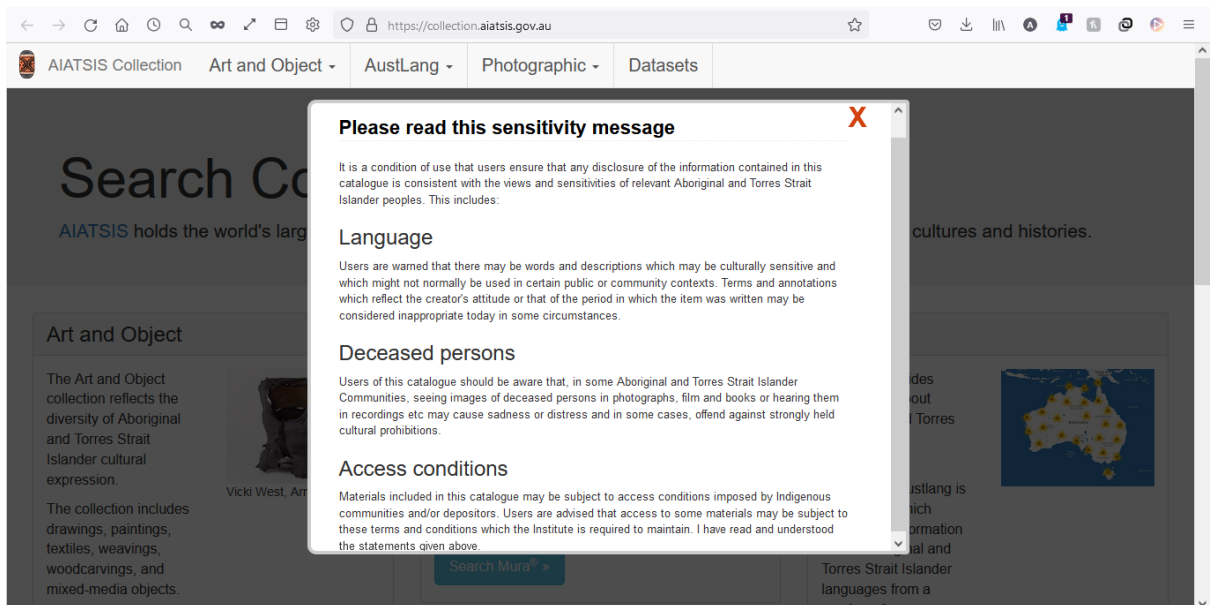
The image shows a grey rectangular banner with the word 'GALLERY' in small blue capital letters at the top left. Below it, the title 'Black people in European art' is written in a large, bold, black sans-serif font. Under the title, there is a paragraph of text in a smaller black font: 'Black people have been portrayed in European art for centuries - often exoticised, subjugated and minoritised - reflecting that these depictions were made by white artists.' At the bottom of the banner, a line of italicized text in a smaller black font reads: 'Content warning: some of these artworks' descriptions of people are outdated, derogatory and racist.'

Extending this directly to contentious individual objects might not be easy, but having a general pop-up content warning in the form of a banner or a layer is not unheard of. On [Carlotta](#), the information system owned by the Swedish National Museums of World Culture, this is what appears at the beginning of each browser session when one accesses the collections of said museum :

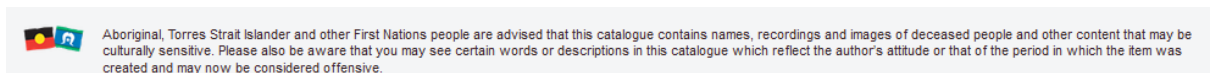
²¹ (Verloop & Moskalenko, 2022)



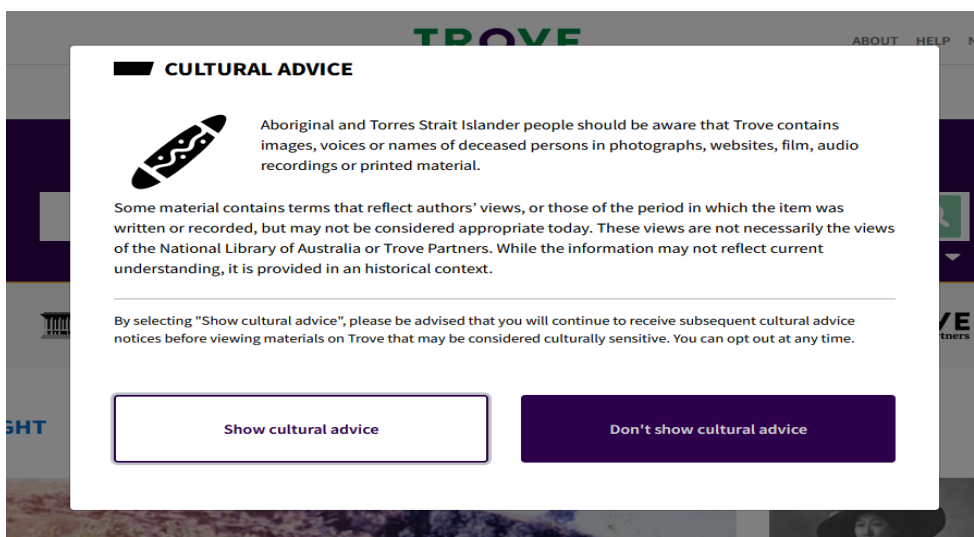
When we follow the red button for more information, we are redirected to the [Search the collections page](#), on which more clarifications on the available information is given. Similarly, the Australian Institute of Aboriginal and Torres Strait Islander Studies (AIATSIS) has a sensitivity message appearing as a pop-up informing users on the language used in resources :



On the National Library of Australia's [website](#), a statement is displayed at the bottom of every catalogue page :



On the Australian database aggregator [Trove](#), a cultural advice popup at the beginning of each browser session gives users the option to either receive more cultural advice notices while they browse the platform or to hide them :



Trove also has a [page](#) dedicated to cultural safety for First Australians on which there is first a section acknowledging First Australian peoples as the Traditional Custodians of their country as well as a cultural safety statement that expands on the cultural advice notices and an explanation of the terminology used on the website.

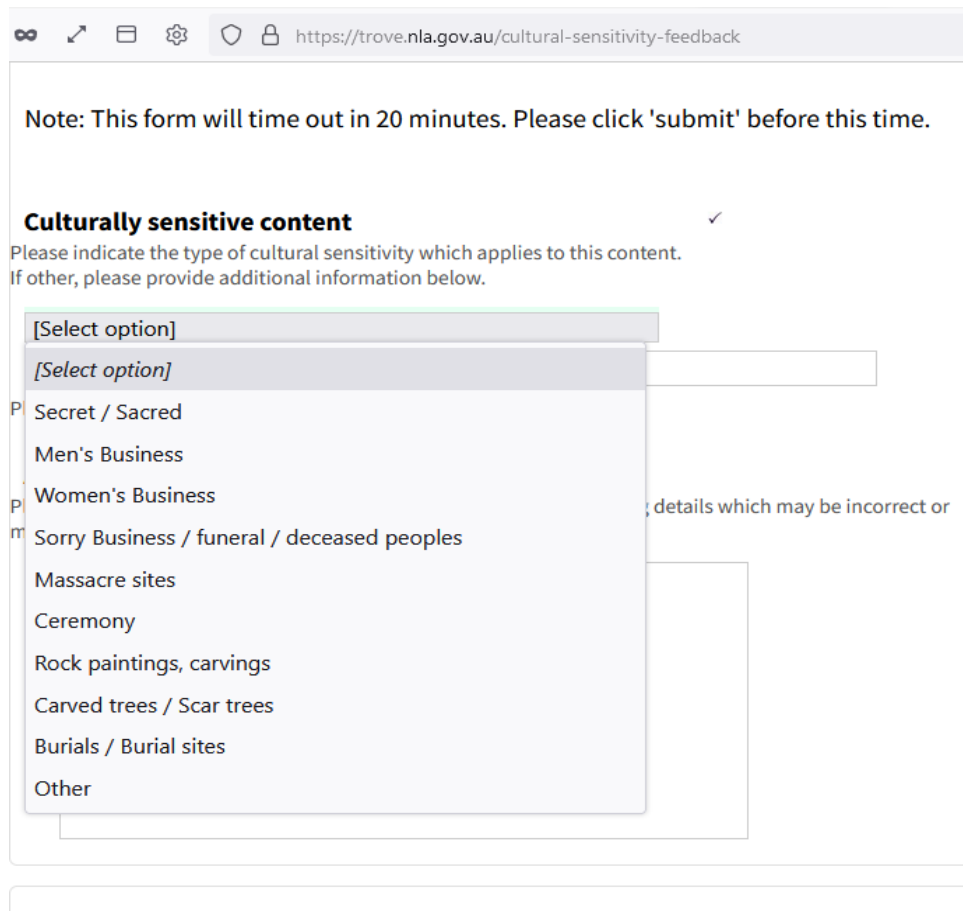
Other institutions have elected to publish statements on their websites such as the DPLA's [Statement on Potentially Harmful Content](#), the Morgan Library & Museum [Statement on Critical Cataloging](#) or the Digital Collections and Archives [Statement on Potentially Harmful Language in Archival Description](#). A regularly updated list of statements on bias in library and archive can be found on the Cataloging Lab's [website](#). Building on several of these statements I have drafted a 'Statement on Potentially Harmful and Offensive Language and Content' which can be found in appendice 3. The main issue that can be raised at this point is the one of flagging said harmful and/or offensive content and language : going over more than 50 million records of objects is a herculean task that simply cannot be a single person or even a team's work. Taking the route of automating this task by using AI and relying on NLP techniques is also often being considered, and while it is out of the scope of this thesis we are aware that some institutions have experimented with these tools. Machine learning experiments are however constantly coming under fire for the biases they pick up : be it the [racist AI beauty pageant judge](#) and programs used by [US courts being biased against black prisoners](#) (autre ex). While AI biases are anomaly an output of machine learning algorithms²², they remain very frequent because ingesting prejudiced data leads machines to making prejudiced assumptions.

Community engagement/co-creation

This is where users can prove to be one of the greatest resources we have. The majority of the statements we talked about previously contain a call to action, asking the platform users for feedback. On the DPLA's website, readers can report harmful content by sending an email, as is the case for the Morgan Library & Museum. On the Trove website, a form is available for users to inform Trove that they

²² (Dilmegani, 2020)

think some content should be marked as culturally sensitive. Different options are given pertaining to the nature of the content :



The screenshot shows a web browser window with the URL <https://trove.nla.gov.au/cultural-sensitivity-feedback>. At the top, a note states: "Note: This form will time out in 20 minutes. Please click 'submit' before this time." Below this, the section is titled "Culturally sensitive content" with a checkmark icon. The instructions read: "Please indicate the type of cultural sensitivity which applies to this content. If other, please provide additional information below." A dropdown menu is open, showing the following options: "[Select option]", "[Select option]", "Secret / Sacred", "Men's Business", "Women's Business", "Sorry Business / funeral / deceased peoples", "Massacre sites", "Ceremony", "Rock paintings, carvings", "Carved trees / Scar trees", "Burials / Burial sites", and "Other".

Users can then add any additional information they want on the content, add attachments and leave their contact details. The Europeana platform also has a feedback button at the bottom-right of every item page :



This however is not specifically designed to flag harmful/contentious items or language : creating a clear feedback mechanism for users would be crucial. This mechanism should also communicate said feedback to the content providers as they are ultimately the owners and experts on said content.

There are many other ways to involve users in dealing with this content and try to make the Europeana collections more diverse and inclusive, such as having the community from where an object originates tell its story. When cultural objects are removed from their original context and exhibited in cultural institutions, they are made into artefacts²³ : having members of the community tell these stories allows them to revoice their own narratives that have often been erased or pushed aside. Users of Europeana do have the opportunity to [share their very own stories](#) about several themes, [submit blog ideas](#) and write for the platform but also take part in the [Digital Storytelling Festival](#) initiative. And we have seen CHIs working closely with different communities such as the Pitt Rivers Museum at Oxford University, which has been [working with representatives from the Maasai cultures](#) whose objects they hold in order to discuss the care of these objects. The representatives have also expressed how the exhibition of their cultural objects was framing them as a culture of the past, while different “*Maasai groups in Kenya and Tanzania are actively resisting the erosion of their land and culture*”²⁴. Initiatives such as this are part of a slow shift happening in cultural institutions from “linear ‘*material approach*’, to a multi-variant ‘*people centered*’ approach that deeply appreciates the past in informing the present and future, while embedding in the museum the workings of human diversities and human faces” (Okello Abungu, 2019, p.64).

²³ (Hollowell & Nicholas, 2009, p.150)

²⁴ (Pitt Rivers Museum, 2020)

Challenges

The sheer quantity of objects published on the Europeana platform is probably the largest obstacle we are facing. Creating metadata is a time and labour intensive enterprise, and doing so for over 50 million records would necessitate substantial financial and human investment. Staff needs to be hired and/or trained, and even if we consider the automatisisation route which some institutions such as the [DPLA](#) and [Cultural-AI](#) have been experimenting with, human validation and checking is still a necessity.

Conclusion

Through this report, we have explored what it means to decolonise cultural heritage institutions by looking at harmful and contentious content and language. While there is no one-size-fits-all approach to dealing with these issues, we have seen that throughout the cultural heritage sector, institutions are taking steps and experimenting with different tools and methodologies. We have outlined different possibilities that can be followed in the context of Europeana such as metadata editing and the creation of new descriptive fields; introducing a general content and language warning on the website; creating context notices to give users a better understanding of materials they are accessing; and having on the Europeana platform, a statement on the harmful and/or contentious content and language. And as we have established in this report, one person might be able to do enough research to give us good first insights into the presence of contentious and outdated terms in Europeana's collections, but a lot more people are needed to go through each and every instance to find out how they are used and what the viable alternatives are. None of the suggestions cited are quick nor easy, and Europeana is only laying the foundation of this work. However, as an institution that has an integral part to play in education and research across Europe, decolonising these digital cultural heritage collections is imperative moving forward. Involving the various communities that could potentially be harmed by this content is crucial for every step we would wish to take.

Appendice 1 : Contentious/outdated terms list

English	Possible replacement	Nederlands	Vervangings mogelijkheden	Français	Options de remplacement
Aborigine (n) Aboriginal (adj)	Terminology used by the people themselves, Indigenous	Inboorling (n), inlander (n) Inlands (adj)	Inheemse bevolking inheems	Indigène (n and adj)	Autochtone
Amulet, talisman	Use specific terms	Amulet, talisman,	Specifieke term gebruiken	Amulette, talisman	Utiliser des termes spécifiques
Barbarian	No suitable alternative, if unavoidable use quotation marks	Barbaar	Geen alternatief, als onvermijdelijk aanhalingstekens gebruiken	Barbare	Pas d'alternative, si inévitable utiliser des guillemets
Berber	Amazigh (sg), Imazighen (pl)	Berber (n) Berbers (adj)	Amazigh (adj and sg n) Imazighen (pl n)	Berbère	Amazigh (sg), Imazighen (pl)
Bombay	Mumbai	Bombay	Mumbai	Bombay	Mumbai
Colored	Person/people of colour	Kleurling, gekleurd	Iemand van ..		
Coolie	Use specific terms : indentured labourer, low-wage labourer...	Koelie	Specifieke term gebruiken : contract arbeider, mijnwerker, drager	Coolie	Utiliser des termes spécifiques : travailleur agricole, travailleur sous contrat...
Disabled/Handicapped/Crippled	Disabled people, person with disabilities,	Gehandicapt en	Mensen met een beperking	Handicapé, infirme, invalide	Personne ayant un handicap, personne en

	differently abled				situation de handicap
Eskimo	Terminology used by the people themselves	Eskimo	Terminologie die door de mensen zelf wordt gebruikt	Eskimo	Terminologie employée par les peuples pour se décrire eux-même
Gypsy	Roma person/people	Zigeuner(s)	Roma, Rondtrekken de groepen/ge meenschappen	Gitan/Bohémien	Rom, Roms
Hottentot	Khoisan (people)	Hottentot	Khoikhoi	Hottentot	Khoïkhoï
Indian	Terminology used by the people themselves, Indigenous, Native American	Indiaan	Namen die de oorspronkelijke bewoners zelf gebruiken	Indien	Terminologie employée par les peuples pour se décrire eux-même
Kaffir	No suitable alternative, if unavoidable, use quotation marks. Appropriate when referring to the self- identified Sri Lankan Kaffirs.	Kaffer	Geen alternatief, als onvermijdelijk aanhalingsteken gebruiken		
Lapp, Lapland	Sámi, Sápmi	Lappen, Lapland	Sámi, Sápmi	Lapon, Laponie	Sami, Sápmi
Mohammedan	Muslim	Mohammedan	Moslim		
Moor	No suitable alternative, if	Moor	Geen alternatief,	Maure/Mauresque	Pas d'alternative,

	unavoidable use quotation marks		als onvermijdelij k aanhalingste ken gebruiken		si c'est inévitabe utiliser des guillemets
Mulatto	No suitable alternative, if unavoidable use quotation marks	Mulat	Geen alternatief, als onvermijdelij k aanhalingste ken gebruiken	Mulâtre	Métisse
Negro,negro es	Black person/peop le	Neger, negers nikker	Zwarte, zwarte mensen	Nègre, négresse	Personne noire/noir
Oriental	Use specific terms : Chinese, Japanese, Egyptian, Palestinian...	Oosters, Oosterling	Specifieke term gebruiken : Chinees, Japans, Egyptenaar, Palestijns...	Oriental	Utiliser des termes plus spécifiques : Chinois, Japonais, Egyptien, Palestinien...
		Planter	Plantage- eigenaar, plantage directeur		
Primitive	No suitable alternative, if unavoidable use quotation marks	Primitief	Geen alternatief, als onvermijdelij k aanhalingste ken gebruiken	Primitif	Pas d'alternative, si c'est inévitabe utiliser des guillemets
Slave	Enslaved person	Slaaf			
Savage	No suitable alternative, if unavoidable use	Wilde (n)	Geen alternatief, als onvermijdelij k	Sauvage	Pas d'alternative, si c'est inévitabe

	quotation marks		aanhalingstekens gebruiken		utiliser des guillemets
				Événements d'Algérie	Guerre d'Algérie, révolution algérienne, guerre d'indépendance algérienne
		Polititionele Actie	Indonesische Onafhankelijkheidsoorlog		

- Acquired, obtained : euphemistic terms usually used to distort the fact that an item was looted or stolen.
- Discovered, found : perpetuates Eurocentrism for scientific discoveries, geographically perpetuates the narrative that lands could be claimed when inhabited by non-Christians/Europeans (terra nullius principle). Appropriate possible replacement : “the first European person who traveled to”.

Different cases : colonial names, racist slurs, use of people-first language rather than identity-first language.

Appendice 2 : Contentious terms API query

To query Europeana's [search API](#), I have used Python 3 and the interactive computing platform [Jupyter](#) Notebook. A link to the notebook is available [here](#).

In order to research Europeana's collections and find the frequencies of the terms we are interested in we first need to download the necessary packages :

```
import os # this module provides functions for interacting with the operating system
import pandas as pd # for data manipulation and analysis
import json # standard format inspired by JavaScript for data exchange and data transfer as text format over a network
from pprint import pprint # utility module that you can use to print data structures in a readable, pretty way
import re # regular expression syntax
import pyeuropeana.apis as apis # Europeana's search API
```

After entering our API key, the first step will be to create an empty dictionary in which the terms and their counts will be stored, as well as a list of all the terms :

```
termscounted = {}
terms = ["aborigine", "aboriginal", "inboorling", "inlander", "indigène", "barbarian", "barbaar", "barbare", "barbares",
"berber", "berbers", "berbère", "berbères", "Bombay", "Colored", "kleurling", "gekleurd", "Coolie", "Koelie",
"disabled", "handicaped", "crippled", "cripple", "cripples", "handicapé", "handicapés", "gehandicapten", "infirmes",
"infirmer", "invalide", "invalides", "eskimo", "eskimos", "Gypsy", "gypsies", "Zigeuner", "Gitan", "gitans",
"Bohémien", "bohémiens", "bohémienne", "bohémiennes", "Hottentot", "Indian", "Indians", "Indiaan", "Indien",
"indiens", "Kaffir", "Kaffer", "Lappen", "Lapp", "Lapland", "Lapon", "Laponie", "Lapp", "Mohammedan", "Moon",
"Maure", "Mauresque", "Mulatto", "mulat", "mûlatre", "negro", "negroes", "neger", "negers", "nikker", "nègre",
"négresse", "Oriental", "oosters", "oosterling", "Orientale", "orientals", "orientales", "orientaux", "Planter",
"primitive", "primitief", "primitif", "slave", "slaaf", "slaves", "savage", "sauvage", "Événements d'Algérie",
"Politieele Actie", "amulet", "amulette", "talisman", "bedeltje"]
```

Now we can count the occurrences for each term and print the term and total:

```
for term in terms:
    response = apis.search(query=term)
    response.keys() #the result of the search is a dictionary
    termscounted[term] = response["totalResults"]
    print(term)
    print(response["totalResults"]) # the total number of items that the query returned
```

In order to make them more readable (see Quantitative data analysis) we will sort the dictionary :

```
dict(sorted(termscounted.items(), key=lambda item: item[1]))
```

The next step was to see in which institution's collections the term 'slaves' was more frequent, and then again sort the results. We set the limit of rows at 4000 because from the results of the previous query we know that out of the terms we will be researching the highest frequency is under that limit :

```

response2 = apis.search(query="slaves", rows=4000)
response2.keys()
count_data_provider = {} # create dictionary that will store the all the data providers and their count

print(len(response2['items'])) #we first perform a check and print the total result
for item in response2['items']:
    print(item['country']) # for each item we print the country
    print(item['dataProvider']) # and we also print the data priver
    dataprovider = item['dataProvider'][0]
    if dataprovider not in count_data_provider: #we add all the counts to the data provider dictionary
        count_data_provider[dataprovider] = 1
    else:
        count_data_provider[dataprovider] += 1

```

```
dict(sorted(termscounted.items(), key=lambda item: item[1]))
```

The same operation was repeated for the singular 'slave' :

```

response3 = apis.search(query="slave", rows=4000)
response3.keys()
count_data_provider = {}
for item in response3['items']:
    print(item['country'])
    print(item['dataProvider'])
    dataprovider = item['dataProvider'][0]
    if dataprovider not in count_data_provider:
        count_data_provider[dataprovider] = 1
    else:
        count_data_provider[dataprovider] += 1
print(len(response3['items']))

```

```
dict(sorted(termscounted.items(), key=lambda item: item[1]))
```

And one final time for the term 'bohémien' to which we added the character * signifying we are looking for all items with zero or more occurrences of the preceding element :

```

response4 = apis.search(query="bohémien*", rows=5000)
response4.keys()
count_data_provider = {}

print(len(response4['items']))
for item in response4['items']:
    print(item['country'])
    print(item['dataProvider'])
    dataprovider = item['dataProvider'][0]
    if dataprovider not in count_data_provider:
        count_data_provider[dataprovider] = 1
    else:
        count_data_provider[dataprovider] += 1

```

```
dict(sorted(termscounted.items(), key=lambda item: item[1]))
```

Appendice 3 : draft for Europeana's Statement on Potentially Harmful and Offensive Language and Content

Europeana is an open web portal offering access to the digitised heritage collections of more than 4,000 data providers across Europe. The platform holds records of over 50 million objects, and in order to achieve this, Europeana works with a network of aggregating partners which collect, check and enrich the data they receive from the different institutions. Our goal at Europeana is to contribute to an open, knowledgeable and creative society while advocating for better digital practices in the cultural heritage sector that supports openness, transparency, and reuse of digital cultural heritage.

Whether it is the content itself or the language used to describe it, we are deeply aware that what we share can have the potential to perpetuate systemic discrimination and cultural bias. If you encounter content you would like us to review or want to share any possible solutions, we welcome any feedback you might have.

This is why we are working to bring cultural heritage institutions together for discussions and foster constructive dialogue and action, while learning from each other on how to deal with such situations. Internally, our Diversity and Inclusion Cross-team, Europeana aggregators and network members are exploring and designing changes to our structures so that everyone feels welcome, represented and safe as part of the Europeana Initiative. We are also working on identifying areas of concern and understanding the unconscious biases within the Europeana Foundation.

What type of harmful or offensive language and content could you encounter on Europeana ?

For some items, the objects themselves may contain harmful stereotypes or depictions. For others, some metadata fields such as the subject, object or geographical location for example may contain harmful language. Sometimes the editorials we publish may link to objects that are harmful, and some user-generated content like user galleries could have harmful descriptions or titles. More specifically, you might encounter :

- Objects or descriptions reflecting discriminatory views on sexualities, genders, religion or sexist, racist, ableist, misogynistic/misogynoir, and xenophobic beliefs and opinions;
- Objects or descriptions reflecting colonial or violent views;
- Graphic historical content such as medical procedures, nudity, violent deaths, wars and more.

How are the descriptions of these 50 million objects created ?

Archivists, curators and other cultural heritage professionals often re-use language provided by creators or former owners of the material when describing it. A lot of this material is also historical in nature, which means that some terms that may be perceived as outdated or contentious from our current perspective can be found in collections catalogues. Some objects can also express views now considered problematic, and are sometimes collected and preserved precisely because of their historical significance. Many cultural heritage institutions are actively reviewing and reconsidering the wording of their data to reduce harm and to reflect current societal views, but this is a long and often complex process. Actions taken in this work range from establishing new thesauri and glossaries to contextualise harmful language and provide alternatives, put outdated terms in quotation marks or provide an explanatory text contextualising objects expressing harmful views or terms. And while replacing outdated terms with current and accepted ones is also a process that was undertaken by many institutions, they also try to preserve and contextualise

transcribed titles, captions and inscriptions as a way to ensure that these now outdated attitudes and viewpoints aren't simply erased but rather considered critically.

We also acknowledge that some of the controlled vocabularies and thesauri that we or our providers use, i.e. standardised classification systems such as the Library of Congress Subject Headings (LCSH) or the Getty AAT, also still contain potentially outdated or offensive terms. For an overview of which vocabularies Europeana uses, you can consult [this spreadsheet](#).

What about Europeana's Editorial?

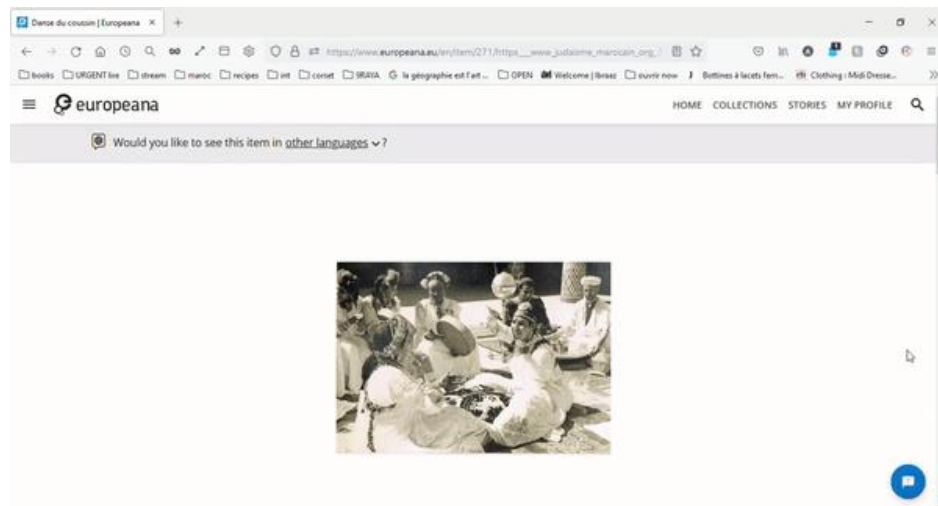
On the editorial side we are working on bringing attention to marginalised groups, who have long been disenfranchised, through the blogs and exhibitions we create and publish. These mediums aim to present different objects in our collections within the wider context in which they were created. All objects are however still accessible without this context through our portal.

How can I report harmful content or language ?

Europeana's mission is to make cultural heritage accessible for educational, research, creation and recreation purposes, which is why this content is made available on our platform. This is in no way an endorsement of the views expressed by this potentially harmful or offensive content. We deeply apologise for any harm the viewing of any potentially offensive material may cause, and would be grateful to receive your feedback on this matter. Our goal is to improve the accessibility of cultural objects by enhancing the metadata in a respectful and accurate manner, and we acknowledge that this will be a long and iterative process.

While we cannot alter the content, items can be taken down after examination. If you wish to report content you find harmful or offensive, or insensitive, offensive,

outdated, or inaccurate language, you can use the feedback button at the bottom-right of the item's page.



This statement draws on [similar statements by other cultural institutions](#) that have been striving to create more diverse, inclusive and respectful environments for all their users.

References

- Académie française. (2013, August 29). *Bohème et bohème* | Académie française.
Académie française |. Retrieved August 9, 2022, from <https://www.academie-francaise.fr/boheme-et-boheme>
- Altman, A. (2022, March 30). *Supporting curatorial work with machine learning* | DPLA. Digital Public Library of America. <https://dp.la/news/supporting-curatorial-work-with-machine-learning>
- Andia, L. (2021, November 29). *How do you remove offensive language without erasing history?* Medium. <https://medium.com/digipaul/how-do-you-remove-offensive-language-without-erasing-history-ab3efe258887>
- Azoulay, A. A. (2021). *DecolonizeThisPlace : Modernity Is an Imperial Crime A Conversation with Ariella Aïsha Azoulay, Dalaëja Foreman and Shellyne Rodriguez, facilitated by Nitasha Dhillon* [Transcript].
- Berry, D. (2018, August 2). *Digitizing and Enhancing Description Across Collections to Make African American Materials More Discoverable on Umbra Search African American History / Dorothy Berry – The Design for Diversity Learning Toolkit*. The Design for Diversity Learning Toolkit.
<https://des4div.library.northeastern.edu/digitizing-and-enhancing-description-across-collections-to-make-african-american-materials-more-discoverable-on-umbra-search-african-american-history/>
- Bragg, M. (2021, March 25). *An Intern's Investigation on Decolonizing Archival Descriptions and Legacy Metadata - Bitstreams: The Digital Collections Blog*. Duke University Libraries Blogs.

<https://blogs.library.duke.edu/bitstreams/2021/03/25/an-interns-investigation-on-decolonizing-archival-descriptions-and-legacy-metadata/>

Christen, K. (n.d.). *Mukurtu Software Preserves Indigenous Digital Heritage through Technologies of Today*. Institute of Museum and Library Services.

<https://www.ims.gov/grant-spotlights/mukurtu-software-preserves-indigenous-digital-heritage-through-technologies-today>

Cole, T. (2019, February 6). When the Camera Was a Weapon of Imperialism. (And When It Still Is.) (Published 2019). *The New York Times*.

<https://www.nytimes.com/2019/02/06/magazine/when-the-camera-was-a-weapon-of-imperialism-and-when-it-still-is.html>

Cushman, E. (2013, November). Wampum, Sequoyan, and Story: Decolonizing the Digital Archive. *College English*, 76(2), 115-135.

https://www.researchgate.net/publication/290331139_Wampum_Sequoyan_and_Story_Decolonizing_the_Digital_Archive

Dilmegani, C. (2020, September 12). *Bias in AI: What it is, Types, Examples & 6 Ways to Fix it in 2022*. AIMultiple. Retrieved August 16, 2022, from <https://research.aimultiple.com/ai-bias/>

DPLA Membership IDEAS Statement | DPLA. (n.d.). DPLA Pro.

<https://pro.dp.la/hubs/dpla-membership-ideas-statement>

DPLA's Statement on Potentially Harmful Content | DPLA. (n.d.). Digital Public Library of America. <https://dp.la/about/harmful-language-statement>

ECDA. (n.d.). *Decolonizing the Archive: Digital Remix and Reassembly*. Early Caribbean Digital Archive.

<https://ecda.northeastern.edu/home/about/decolonizing-the-archive/what-we-are-doing/>

Europeana. (2013). *Europeana Data Model Primer*.

https://pro.europeana.eu/files/Europeana_Professional/Share_your_data/Technical_requirements/EDM_Documentation/EDM_Primer_130714.pdf

Farnel, S., Shiri, A., Campbell, S., Cockney, C., Rathi, D., & Stobbs, R. (2017). A Community-Driven Metadata Framework for Describing Cultural Resources: The Digital Library North Project. *The Digital Library North Project, Cataloging & Classification Quarterly*, 55(5), 289-306.

<https://doi.org/10.1080/01639374.2017.1312723>

Francois, J., Zarzycki, L., Beaumont, E., Hernández, F., & Sheikh, S. (2021, March 17). *Decolonise this museum: can cultural institutions care? - Architectural Review*. The Architectural Review. <https://www.architectural-review.com/essays/decolonise-this-museum-can-cultural-institutions-care>

Furner, J. (n.d.). *FAQ – Jonathan Furner*. Jonathan Furner.

http://www.jonathanfurner.info/?page_id=49

Getty Foundation. (n.d.). *About the AAT (Getty Research Institute)*. Getty : Resources for Visual Art and Cultural Heritage. Retrieved August 13, 2022, from

<https://www.getty.edu/research/tools/vocabularies/aat/about.html#scope>

Hamm, M., & Schönberger, K. (Eds.). (2021). *Contentious Cultural Heritages and Arts: A Critical Companion*. Wieser Verlag.

- Hollowell, J., & Nicholas, G. (2009). Using Ethnographic Methods to Articulate Community-Based Conceptions of Cultural Heritage Management. *Public Archeology*, 8(2-3), 141-160. <https://doi.org/10.1179/175355309X457196>
- Holterhoff, K. (2017, November 3). *From Disclaimer to Critique: Race and the Digital Image Archivist*. digital humanities quarterly. <http://digitalhumanities.org/dhq/vol/11/3/000324/000324.html>
- Hughes-Watkins, L. (2018). Moving Toward a Reparative Archive: A Roadmap for a Holistic Approach to Disrupting Homogenous Histories in Academic Repositories and Creating Inclusive Spaces for Marginalized Voices. *Journal of Contemporary Archival Studies*, 5. <https://elischolar.library.yale.edu/cgi/viewcontent.cgi?article=1045&context=jcas>
- Jaffe, R. (2020). Rethinking Metadata's Value and How It Is Evaluated. *Technical Services Quarterly*, 37(4), 432-443. <https://doi.org/10.1080/07317131.2020.1810443>
- Leturcq, J.-G. (2008). La question des restitutions d'œuvres d'art : différentiels maghrébins. *L'Année du Maghreb*, IV, 79-97. <https://doi.org/10.4000/anneemaghreb.431>
- Liability of online platforms*. (2021). European Parliament. [https://www.europarl.europa.eu/RegData/etudes/STUD/2021/656318/EPRS_STU\(2021\)656318\(ANN1\)_EN.pdf](https://www.europarl.europa.eu/RegData/etudes/STUD/2021/656318/EPRS_STU(2021)656318(ANN1)_EN.pdf)
- Local Contexts. (n.d.). *TK Labels – Local Contexts*. Local Contexts. <https://localcontexts.org/labels/traditional-knowledge-labels/>

- Lunch, N., & Jenkins, T. (2020, August 11). *Living cultures: decolonising cultural spaces*. The Ecologist. <https://theecologist.org/2020/aug/11/living-cultures-decolonising-cultural-spaces>
- McKechnie, A. (2014, December 4). *Experts Discuss the Meaning of Ownership and Stewardship of Cultural Property at Drexel's Annual Museum Leadership Conference*. Drexel University. <https://drexel.edu/news/archive/2014/december/museum-leadership>
- Memmi, A. (2003). *The Coloniser and the Colonised*. Earthscan.
- Mineo, L. (2020, November 19). *Museums of Native culture wrestle with decolonizing*. Harvard Gazette. <https://news.harvard.edu/gazette/story/2020/11/museums-of-native-culture-wrestle-with-decolonizing/>
- Möntmann, N. (2021). Small-scale art organizations as participatory platforms for decolonizing practices and sensibilities. *Journal of Aesthetics & Culture*, 13(1). <https://doi.org/10.1080/20004214.2021.1972526>
- Museum für Völkerkunde zu Leipzig: Decolonisation, restitution, and repatriation*. (n.d.). GRASSI Museum für Völkerkunde zu Leipzig. <https://grassi-voelkerkunde.skd.museum/en/research/decolonisation-restitution-and-repatriation/>
- Okello Abungu, G. (2019). Museums: Geopolitics, Decolonisation, Globalisation and Migration. *Museum International*, 71(1-2), 62-71. <https://doi.org/10.1080/13500775.2019.1638030>

Okello Abungu, G. (2019). Museums: Geopolitics, Decolonisation, Globalisation and Migration. *Museum International*, 71(1-2), 62-71.

<https://doi.org/10.1080/13500775.2019.1638030>

O'Neal, J. (2017, October). *Decolonizing Archives and Museums: What Comes Next?* [Video]. The sustainable heritage network.

<https://sustainableheritagenetwork.org/digital-heritage/decolonizing-archives-and-museums-what-comes-next>

Pitt Rivers Museum. (2020). *Maasai Living Cultures* | *Pitt Rivers Museum*. Pitt Rivers Museum. <https://www.prm.ox.ac.uk/maasai-living-cultures>

Rethinking cataloguing – Collections Trust. (n.d.). Collections Trust.

<https://collectionstrust.org.uk/blog/rethinking-cataloguing-2/>

Revealing the Colonial Past in Museums - Zeitgeist - International Perspectives from Culture and Society. (2021, November). Goethe-Institut.

<https://www.goethe.de/prj/zei/en/pos/22500702.html>

Sandy, H. M., & Dykas, F. (2016). High-Quality Metadata and Repository Staffing: Perceptions of United States–Based OpenDOAR Participants. *Cataloging & Classification Quarterly*, 54(2), 101-116.

<https://doi.org/10.1080/01639374.2015.1116480>

Scholz, H., & Ciocoiu, A. (2022, January 14). *EPF - Publishing guidelines*. Europeana Knowledge Base - Confluence.

<https://europeana.atlassian.net/wiki/spaces/EF/pages/2059763713/EPF+-+Publishing+guidelines>

- Smith, K. (2018, November 7). *Creating metadata for equity, diversity, and inclusion*. Hanging Together. <https://hangingtogether.org/creating-metadata-for-equity-diversity-and-inclusion/>
- Thürlemann, F. (2017, January 25). *Staging the Egyptian Harem for Western Eyes*. Getty Iris. Retrieved August 15, 2022, from <https://blogs.getty.edu/iris/staging-the-egyptian-harem-for-western-eyes/>
- Tropenmuseum. (2018, June 27). *Words matter (English) by NMVW*. Tropenmuseum in Amsterdam - Een museum over wereldculturen. https://issuu.com/tropenmuseum/docs/wordsmatter_english
- Vawda, S. (2019). Museums and the Epistemology of Injustice: From Colonialism to Decoloniality. *Museum International*, 71(1-2), 72-79. <https://doi.org/10.1080/13500775.2019.1638031>
- Vawda, S. (2019, July). Museums and the Epistemology of Injustice: From Colonialism to Decoloniality. *Museum International*, 71(1-2), 72-79. <https://www.tandfonline.com/doi/full/10.1080/13500775.2019.1638031>
- Verloop, M., & Moskalenko, D. (2022, March 28). *What's new on the Europeana website? January - March 2022*. Europeana Pro. Retrieved August 14, 2022, from <https://pro.europeana.eu/post/what-s-new-on-the-europeana-website-january-march-2022>

AFDELING
Straat nr bus 0000
3000 LEUVEN, BELGIË
tel. + 32 16 00 00 00
fax + 32 16 00 00 00
www.kuleuven.be

