



Recommendations for identifying and managing diversity in cultural heritage data

Revision	1.0
Last updated	30.08.2022
Author(s)	<i>Kerstin Herlt, Kristina Rose (both DFF /EFG), Kerstin Arnold (APE)</i>
Contributor(s)	<i>Valentina Bachi (Photoconsortium), Maria Teresa Natale, Marco Fiore (both MUSEU), Gerda Koch (Open-Up), Nicole Emmenegger, Chiara Latronico (both NISV/EUScreen), Tom Miles (Europeana Sounds), Adina Ciocoiu, Marijke Everts, Ariadna Matas, Henning Scholz (all EF)</i>
Dissemination Level	<i>For internal distribution only</i>

Table of content

	29
1. Recommendations for identifying and managing diversity, inclusion and equity in cultural heritage data	32
R1 Recommendations on raising awareness and creating knowledge	32
R2 Recommendations on creating more inclusive catalogue records	33
R3 Recommendations about notifications, content warnings and AI supported search tools	35
R4 Recommendations for collaboration and co-creation	36
2. Resources	37

1. Recommendations for identifying and managing diversity, inclusion and equity in cultural heritage data

The aim of the case studies was to explore possibilities for dealing with potentially offensive, inappropriate or outdated content and metadata by creating different scenarios from different user perspectives. The model of the consequence decision tree enabled the WG to evaluate and assess the problem very precisely by using concrete examples either from Europeana collections or examples from content providers and aggregators that are not yet published on Europeana.

The recommendations of the WG are based on the findings of these seven user stories. The overall objective for the recommendations is to provide cultural heritage institutions and other data providers as well as aggregators a framework to use when looking into D&I&E in their contexts along with the consequence decision tree as a potential tool that can help them develop their questions and challenges.

While we consider the following recommendations to be applicable to other examples as well, there might be content that requires other approaches (e.g. atrocity images, images from human remains or content that is otherwise deeply disturbing or harmful).

R1 Recommendations on raising awareness and creating knowledge

R 1.1 Raise awareness

- Recognise that some traditional approaches of describing collections are inappropriate because they perpetuate offensive, racist and colonial attitudes (even unintendedly)
- Try to get in the mindset of the different stakeholders, this might help to overcome a “gatekeeper’s” attitude
- Organise training events

R 1.2 Create knowledge and consensus on the language to be used

- Build a cross-team of cataloguers, archivists and volunteers to review how language is used to catalogue and describe collections items
- Consult inclusive language glossaries - useful resources exists already as work in progress - Words Matter, Histories of Colours - Inclusive Glossary Project (see chapter 6)
- Define areas of special interest and try to collaborate with communities on inclusive languages

R2 Recommendations on creating more inclusive catalogue records

Catalogue records have often been created decades - or even centuries - ago and are part of the history of a cultural heritage object. At the same time, they are living documents as they will evolve throughout their lifecycle in accordance with the holding institutions’ contemporary policies. Whether or not inappropriate records should be replaced or enriched and contextualised is influenced by many different factors: cataloguing practices, database structure, the record or media type (image, AV, sound, archival record), the type of descriptive element (captions, synopses, transcripts etc). Lastly, it depends on the resources available to an institution to decide in which way taking actions to rework and rethink cataloguing practices is a feasible task.

R 2.1 Adjust, enrich or replace terminology in the descriptive metadata of cultural heritage

There are various reasons in favour of keeping the original wording such as a) this being historical data and b) erasing the original wording seems like hiding past practices. Should an institution decide to remove or contextualise the

original wording, the bullet points below detail aspects to look out for in the process.

- *Descriptive texts*: Captions (see case study 4.2), object descriptions and synopses which are inappropriate or reflect a stereotypical world view could be replaced by more inclusive ones. For the purpose of versioning, the former descriptions could still be kept in the database with contextualising information such as a date and a source if available. An example for this could be being specific about the people who are represented. The homogenising colonial discourse refers to “Africans” or “indigenous people”. To counteract this homogenisation, use the name of the community or ethnicity (see case study 4.1 where people have been identified as Kunama people, or case study 4.4 that distinguishes between Djur, Bongo or Niam-Niam).
 - *Synopses*: Some data providers distinguish between fictional and documentary material. In case of documentary films, for example, inappropriate or outdated terminology could be replaced by more inclusive terms. In the case of fictional films replacing “Gypsy” by “Roma people” or “Indians” by “First Nations people” would be contrary to reality insofar reference to an existing fictional stereotype in cinema of that time would be lost.¹ In that case, it is recommended to use original sources (with authorship attributions) and to add / import information from various sources (with citation) to have richer and more up to date descriptions.
- *Titles*: Description policies often ask for original titles not to be changed (see case study 4.1). In that case, an alternative / descriptive title could be added. If the title is given by the data provider themselves, it can be replaced. We would recommend keeping the former title in the database, e.g. by indicating “Former [data provider’s name] title”.

R 2.2 Review or add keywords and subjects

Possible points of action depend on whether an institution reviews their existing vocabularies used for indexing or gets newly started with indexing (see case study 4.1). Similarly, different approaches are needed when an institution uses its own vocabulary in contrast to the usage of an external vocabulary managed by a different institution or community (see case study 4.5).

¹ Recommendations of the Catalogue Project Group - EYE Filmmuseum, 20th July 2021. Unpublished Paper.

- When reviewing existing keywords from a vocabulary managed by the CHI itself, consider replacing inappropriate terms or adding new, inclusive keywords to existing ones in the database records. Inclusive glossaries are already available as work in progress, e.g. Words Matters or the Inclusive Glossary from the History of Colours project (see chapter 6).
- When reviewing existing keywords from a vocabulary managed by the CHI itself or when creating such a vocabulary anew, use thesaurus functionalities like preferred names to install a reference between outdated terms, that are in use, and their appropriate counterpart(s) - e.g. "Roma" would be the preferred term that is linked to "Gypsy" (fictional stereotype of Roma).
- When reviewing existing - or considering new - keywords from vocabularies that are managed by external institutions and communities, get in touch with the respective managing body (see case study 4.5) when concerns about specific terms arise.
- When getting started with indexing during cataloguing of a collection, make use of inclusive language. Consider the inclusivity to be part of the criteria to apply when choosing a vocabulary or building a vocabulary for one's institution.

As a general recommendation, it should be noted that vocabularies that support multilinguality and/or are available as linked open data (LOD) are especially valuable to connect one's metadata with that of other institutions. Furthermore, using Unique Resource Identifiers (URIs) instead of a literal keyword is helpful to distinguish between concepts that have the same names. "Roma" or "Romani" could for example mean the city in Italy (<https://www.geonames.org/3169070>), the people (<http://www.wikidata.org/entity/Q8060>) or the language (<http://vocab.getty.edu/aat/300389150>). Identifying them by URIs not only distinguishes the concepts, but provides the opportunity to gather more information from the vocabularies that provide the URI.

All the points mentioned before in R 2.2 apply to LOD vocabularies as well.

R3 Recommendations about notifications, content warnings and AI supported search tools

- When engaging with D&I&E, consider publishing notifications about potentially harmful content in the collections your website or service gives access to (see case study 4.7).²
- Consider inviting users to give feedback when encountering content or terminology they consider harmful. This type of feedback is extremely important to content providers and aggregators in order to concretely improve collections and take an inclusive approach regarding minoritised communities of which they preserve the heritage. Based on the feedback received, it will be the responsibility of the individual data provider to decide whether or not they will remove or contextualise the content, while the role of the aggregators is more in the context of helping to raise awareness about sensitive content, sharing best practices and offering training.
- A content warning pop-up is a useful tool to warn users about the display of harmful content such as atrocity images, images of war crimes, disturbing medical materials etc. If possible with regard to technical possibilities and resources available, consider implementing content warning pop-ups to allow users to decide whether or not they want to see this kind of content.
- Consider the joint development of AI supported tools to detect potentially offensive language: many CHIs cannot update legacy metadata in their databases due to the sheer size and scope of their collections and the complexity of the endeavour.³

R4 Recommendations for collaboration and co-creation

- Start small: define a specific topic or theme you want to address, such as colonialism, the mis- or underrepresentation of specific minoritised communities in your collections, etc.

² This can be done on data providers' own websites and on aggregators' online collection portals. Europeana is also currently working on a statement and invites users to give feedback on content or terminology they consider inappropriate or harmful. A list of statements created by various organisations can be found here: [List of statements on bias in library and archives description – Cataloging Lab](#).

³ The DE-BIAS project, which has recently been invited to start Grant Agreement negotiations and includes several of the aggregators involved in the WG as well as other members of the Europeana Aggregators Forum, proposes the development of an AI supported tool that automatically detects potentially offensive terms (in multiple languages and co-created with communities and community allies) and suggests alternative terms.

- Seek collaboration with communities affected by your selected topic or theme and with communities' allies (researchers, critical friends, civil society organisations, counter archives, thematic archives) to establish glossaries and vocabularies, to contextualise legacy data and co-curate virtual exhibitions, to co-write editorials etc.
- Open up your institution and share your collections with those communities they refer to.

2. Resources

Glossaries:

- Words Matter | Research Center for Material Culture:
<https://www.materialculture.nl/en/publications/words-matter>
- Carissa Chew. Histories of Colours - Inclusive Glossary Project:
<https://drive.google.com/drive/folders/1JZG0zmzlzPauwqj5JxxUajf5hYkD0ta>

Notifications and Statements:

- List of statements on bias in library and archives description – Cataloging Lab:
<https://cataloginglab.org/list-of-statements-on-bias-in-library-and-archives-description/>
- Archives Hub: <https://archiveshub.jisc.ac.uk/offensivelanguage/>
- National Archives and Records Administration (United States of America):
<https://www.archives.gov/research/reparative-description/harmful-content>
- The National Archives (UK):
<https://cdn.nationalarchives.gov.uk/documents/cataloguing-approach-on-offensive-terminology.pdf>

Best practices, Recommendations

- Museums Association:
<https://ma-production.ams3.digitaloceanspaces.com/app/uploads/2021/11/07135807/Supporting-decolonisation-in-museums-final-version.pdf>
- <https://decolonisingthroughcriticallibrarianship.wordpress.com>

- Natural Science Collections:
<https://museumsandgalleries.leeds.gov.uk/collections/decolonisation-and-natural-science-collections/>
- Historic England:
<https://historicengland.org.uk/whats-new/statements/contested-heritage/>
- Museums Lab Germany <https://themuseumslab.org/>
- Recommendations of the Catalogue Project Group - EYE Filmmuseum, 20th July 2021. Unpublished Paper.
- OCLC Research Library: Equity, Diversity, and Inclusion in Research Library Partnership Institutions Survey (oclc.org)
- OCLC:
<https://www.oclc.org/research/publications/2022/reimagine-descriptive-workflows.html>
- National Archives and Records Administration (United States of America):
<https://www.archives.gov/files/news/archivists-task-force-on-racism-report.pdf>
- SNAC Cooperative: <https://portal.snaccooperative.org/node/545>
<https://portal.snaccooperative.org/node/359>
- Open GLAM: <https://openglam.pubpub.org/pub/decolonization/release/1>
- Biodiversity Heritage Library:
<https://blog.biodiversitylibrary.org/2021/11/understanding-bhl-through-metadata-patterns-of-bio-diverse-knowledge-production.html>

Training / Sources

- Diverse sources database: <https://training.npr.org/sources/>