



Co-funded by  
the European Union

# DEBIAS

## Representing diversity in metadata

Recommendations and implementation strategy for  
cultural heritage institutions and professionals

Image: [Worldwide Wall](#) - Associazione Culturale GoTellGo, Italy - CC BY



<https://creativecommons.org/licenses/by-sa/4.0/>



## Document information

<b>Current version</b>	1.0	
<b>Contributors</b>	Sofie Taes, Fred Truyen (KU Leuven), Cristina Da Milano (ECCOM), Kerstin Herlt, Kristina Rose, Julia Welter (DFF)	
History of changes		
<b>Version</b>	<b>Change(s) applied</b>	<b>Contributors</b>
1.0	Copied from DE-BIAS deliverable D5.2 "Recommendations to represent diversity in metadata"	Kerstin Arnold (APEF)

### Target audience

These recommendations are meant for colleagues working at cultural heritage institutions and with cultural heritage collections. This includes decision makers in higher and senior management as well as staff being responsible for preserving and describing cultural heritage collections, ensuring their accessibility to the public, and publishing their descriptions either in a digital form, e.g. in cataloguing systems available in the reading rooms or shared online, or in analogue form, e.g. in the context of a physical exhibition as labels next to the works and as part of print publications.

### Learning goals

The recommendations in this document aim at raising awareness of the various aspects to be taken into account when considering debiasing cultural heritage collections. Not each of these aspects will apply in every context and to every use case. Hence these recommendations are intended to help the readers understand what it is that they need to look out for and how different tools, methods, and approaches can complement each other. With the [Implementation strategy](#) introduced at the end of this document, readers will also have a tool to help them in applying the recommendations in a step-by-step approach.



## Transversal themes

### Investing in a holistic approach

Promoting not only access but also participation in activities related to heritage and community representation requires a serious investment in community engagement processes. These processes should be viewed as strategic assets that enable cultural organisations to place the public—not just as visitors, but as individuals and communities—at the centre of their actions. Cultural heritage institutions need to be empowered to invest in opening up participation in decision-making processes and fostering representation. This approach aligns with the vision of the Faro Convention, which emphasises the role of communities in shaping the values of heritage and views heritage and its connected values as a dynamic process rather than something fixed and immutable.<sup>1</sup>

### Use of technology in the cultural heritage sector

DE-BIAS raises the question of the role technology can play in including communities at risk of sociocultural exclusion. The project has demonstrated that technological innovation is highly effective when it serves engagement and inclusion. By integrating advanced technologies into the cultural heritage sector within a structured framework that promotes cooperation between technology and culture and stimulates co-creation with the concerned communities, DE-BIAS has initiated a pioneering reflection on technology's role in participation and engagement. For technological innovation to be meaningful in promoting inclusion, it must be designed and implemented with input from both tech and socio-cultural professionals. Additionally, it requires awareness-raising, support, and evaluation from policymakers.

### Capacity building

It is essential to strengthen the capacity of those involved in promoting access, participation, and inclusion through cultural heritage. This involves activating capacity-building processes that develop not only hard skills (technical and theoretical knowledge) but also soft skills, such as creativity and empathy, which are crucial for fostering community engagement — central to the DE-BIAS project. Good practices from other sectors, particularly the social sector, can inspire our efforts, always keeping in mind the principles of the right to cultural participation.

---

<sup>1</sup> Council of Europe Framework Convention on the Value of Cultural Heritage for Society (FARO Convention), 27 October 2005, <https://rm.coe.int/1680083746>



# Recommendations for CHIs

## Metadata

1. **Perform metadata audits:** Adopt an institutional practice of conducting regular reviews and updates of cultural heritage metadata to ensure that it aligns with current social values and is free from bias. These audits should emphasise the application of inclusive and accurate descriptions for cultural assets while ensuring that outdated or harmful terms are properly contextualised rather than removed. Special attention should be given to collections that reflect the heritage of marginalised and historically disadvantaged communities.
2. **Examine your collections' metadata** by considering three aspects: the images, the metadata itself, and how they relate to each other. Think about this from two angles: first, look for terminological bias, which means using outdated or harmful language to describe things. Second, consider nominative bias, which is when the metadata misrepresents the objects. This can happen when individuals or items are not properly named, are generalised, or are not mentioned at all.
3. In the same vein, conduct your analysis with the understanding that **bias** often arises from a **combination of words rather than from isolated words**. It's important to explore whether the tools and workflows you use consider expressions as a whole, rather than focusing solely on individual words. For instance, look into how they handle phrases like "medicine man" and compound words such as "multiracial" and "Judenschule."
4. **Be language- and time-sensitive when assessing your metadata.** Like socio-political realities, ethics, and conventions, language is constantly evolving. Expressions deemed harmless and neutral in one era may later be viewed as inappropriate, reappropriated, or redefined — and vice versa. Moreover, instances of bias often do not translate directly across different languages.
5. **The importance of disambiguation cannot be emphasised enough.** When a term that is typically considered offensive is part of a name, organisation, or place, it can lead to false positives in automated analyses. Additionally, many terms can be perceived as either offensive or neutral depending on their context (for example, "ape" as an animal versus an insult). While the use of Named Entity Recognition (NER) and Large Language Models (LLMs) can help mitigate this risk, it is crucial that existing descriptions provide sufficient context for successful disambiguation.
6. **Implement flexibility in metadata value standards:** CHIs should adopt flexible metadata standards that facilitate the continual updating of terms and descriptions as new biases are identified or societal norms evolve. This approach will help institutions maintain the relevance of their collections and eliminate harmful language. While citing controlled vocabularies is generally considered best practice, it does not eliminate the presence of biased language. Therefore, it is essential for the governing bodies of these vocabularies to conduct careful and ongoing reviews.
7. **Adopt an intersectional review process.** Biases are often compounded when metadata overlooks the intersection of multiple identities, such as gender, race, and geographic origin. To address this, use intersectionality as a guiding principle when reviewing terms that may appear



innocuous in isolation but reveal deeper systemic issues when considered in combination with other fields. Establish a **cross-category review system** that assesses metadata not just for individual biases, but for intersecting biases across categories. For instance, terms in the ethnicity field may also carry implicit geographic biases that require attention.

8. **Focus on high-risk metadata fields:** Metadata fields that commonly include free text, such as titles and descriptions, are more likely to contain bias and should be prioritised in audits and revisions.
9. CHIs should develop **quick assessment methods**, such as immediate audits, and provide users with **easy feedback options** to flag problematic content. This approach will help address issues in a timely manner, ensuring more accurate representation.
10. **Incorporate bias detection frameworks:** Utilise frameworks for detecting and analysing bias within cultural heritage metadata. Frameworks such as the DE-BIAS typology offer a structured approach to identifying and mitigating bias, ultimately enhancing the accuracy and inclusivity of cultural narratives.
11. Consider using or creating **standardised, yet domain- and community-specific thesauri** of inclusive terms that are relevant to your collections and the heritage communities they represent.
12. **Contextualisation over replacement of offensive terms:** It is highly recommended to contextualise offensive terms in cultural heritage metadata rather than replacing them outright. This approach ensures historical accuracy while providing the necessary context to understand the terms in their original usage.
13. **Address “Shades of bias”<sup>2</sup>:** Subtler forms of bias, such as euphemisms, generalising terms, or politically charged language, can be pervasive and often go unnoticed. Aim to recognize and address these shades of bias by training staff to identify linguistic patterns, such as the transformation of neutral terms into dysphemisms (e.g., the pejorative use of “foreigner”).
14. **Adopt cataloguing practices and metadata updates that incorporate adjusted thesauri,** including terms that reflect more specific and less generalised identities. For example, replace the broad term “Asian” with specific ethnic or national identifiers.
15. **Dynamic and flexible approaches:** Maintain flexibility in controlled vocabularies to accommodate emerging terms and concepts that reflect the dynamic nature of culture and identity. Encourage experimentation and innovation in metadata practices to more effectively represent diverse communities.

---

<sup>2</sup> “Shades of bias”: Analysis of patterns of bias in cultural heritage collections, particularly on Europeana.eu, revealed varying degrees of misrepresentation at the linguistic level, e.g. misappropriation, stereotypes, derogatory language, diminutive language, euphemism, omission/erasure, fragmentation. See: D2.1 “Report on research into bias types and patterns, including a typology applied to Europeana use cases and a vocabulary co-created with communities” (available for download at <https://pro.europeana.eu/project/de-bias>).



## Technology

16. **Use of AI and human oversight for bias detection:** AI-powered tools, such as the DE-BIAS tool, can assist CHIs in automatically detecting and flagging biases within their metadata. However, the use of these tools should be complemented by human oversight to ensure that the nuances of language and cultural context are accurately represented in metadata revisions. Staff must be trained to understand the tool's outputs and to contextualise the flagged terms within the broader narrative of the collections.
17. **Regularly update and expand tools** based on user feedback and evolving language. Consider utilising tools and platforms that allow for dynamic updates — such as vocabulary edits or adaptations of institutional bias-detection rules — without requiring extensive redevelopment.
18. **Consider a phased approach for implementation.** Implementing a bias-detection tool can be resource-intensive. A phased approach enables institutions to gradually integrate the tool while assessing its effectiveness. Begin by applying the tool to a small subset of the collection, focusing on high-risk metadata fields.
19. **There is a need for truly participatory environments** — software and procedures that actively facilitate participation. Provide platforms that promote respectful interaction, fully recognizing contributions and shared responsibilities. We recommend gathering best practices for community engagement software and platforms, shifting from crowdsourcing to genuine co-creation and co-curation activities. Additionally, develop a validated code of conduct and user rights agreement for these platforms as a counterbalance to the one-sided End User Licence Agreements of commercial platforms.
20. **Adapt the technology:** Technological solutions must address the needs of the institutions and communities they serve. Utilise customization features to enhance accuracy, such as selecting collection-specific vocabularies and implementing context-sensitive rules.
21. **Do not forget the user interface.** When developing or deploying AI-supported solutions to enhance diversity and fairness in your metadata, consider the experience you want to create for online visitors to your collections. This could range from content warnings or generic messages to record-specific links that provide context and information. Explore a variety of approaches to ensure a meaningful experience.

## Community engagement

22. **Develop a co-creative community-centred approach.** CHIs should consider participatory processes as a cornerstone of their strategies and operations. These processes can be conducted remotely using collaborative tools to create safe spaces, but they must always be facilitated by social and technological mediators. It is essential to consider the needs of the targeted communities to ensure they are effective partners rather than merely informants for ideas.
23. **Community-centric metadata curation:** Consulting with represented communities is vital for achieving inclusive metadata representation. Their involvement provides valuable insights and helps identify biases that may not be apparent to outsiders. To ensure diverse perspectives are included, engage community representatives in the metadata creation and review process. Organise regular workshops and feedback sessions with community members to discuss and



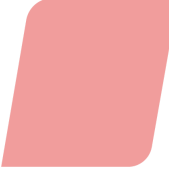
update the language used in metadata. This not only fosters a sense of ownership over their cultural representations but also ensures that their voices are integral to the institution's practices.

24. CHIs should explore strategies to **invite and involve staff members with community backgrounds** in this process, if they wish to participate. It is essential to acknowledge the sensitivities and knowledge that exist within the institution's staff and to include them in these efforts. To enhance community knowledge within its team, a CHI should actively seek to improve diversity and representation among its staff.
25. **Diversified community involvement:** Involve community members from various generations, genders, and professional backgrounds to enhance the critical analysis of sensitive collection metadata. This diversity of perspectives contributes to creating a more inclusive and accurate representation of cultural heritage.
26. **Building trust with community representatives:** Engage with community representatives as allies to build trust and strengthen relationships between the institution and local communities. This engagement involves regular, respectful communication and the establishment of clear protocols for collaboration.
27. **Preparation and knowledge sharing:** Empower community members to become well-prepared and knowledgeable about your collections, operations, and collaboration objectives. This can include demonstrations of search terms that enhance the findability of collections, training opportunities, and providing resources to community representatives so they can effectively contribute to metadata practices.
28. **Sustainable engagement:** Recognise that collaborating with underrepresented groups requires ongoing attention, commitment, and dedication. Engagement should not be viewed as a one-time event; rather, it should inspire a long-term, sustained effort that is embedded in the institution's operations.
29. **Acknowledge contributions:** True engagement with user communities involves recognising contributor rights and effectively navigating and moderating dissenting voices. It must be founded on a mutual agreement based on equality.

### Process - workflows

30. **Transparency and accountability:** Implement transparent processes for the public to report offensive or harmful language found in cultural heritage metadata. Establish accountability mechanisms to address and rectify issues promptly.
31. **Document biases and corrective measures:** When biases are detected and corrected, CHIs should document these instances to ensure transparency in their processes. This documentation can also serve as a valuable resource for other institutions seeking to implement similar practices in metadata management.
32. **Ensure transparent reporting:** Metadata bias detection and correction should be an open process with clear reporting mechanisms in place. CHIs should provide public access to information about how biases were detected, the steps taken to correct them, and the measures implemented to prevent future occurrences.



- 
33. **Be ready to participate in innovative capacity building processes:** It is crucial for cultural professionals to understand the complexity of contemporary society and be open to engaging with disciplines beyond traditional sectoral boundaries. Actively participating in lifelong learning activities should be a process embraced by all members of the organisation.
  34. **Practice-driven and -oriented training:** Train staff on the significance of inclusive language and the impact of harmful language on communities. Provide ongoing professional development opportunities that focus on best practices for inclusive metadata management.
  35. **Documentation and dissemination of best practices:** Document successful practices widely within the GLAM sector. Create a repository of case studies and practical guidelines to assist other institutions in implementing similar initiatives. Share the frameworks and guidelines developed within your organisation to help guide the efforts of peer institutions and accelerate sector-wide de-biasing efforts.





## Implementation strategy

This outline presents a roadmap for cultural heritage institutions (CHIs) to implement the recommendations, encompassing short-term, medium-term, and long-term steps.

- **Short-term steps** include immediate actions such as auditing high-risk fields and utilising technological support like the DE-BIAS tool.
- **Medium-term goals** focus on capacity building, ongoing staff training, and establishing transparent processes.
- **Long-term objectives** involve institutionalising regular metadata updates, creating inclusive vocabularies, and fostering sustained community partnerships.

This phased approach can help CHIs manage the process efficiently, particularly when dealing with vast volumes of data and limited resources. The following roll-out is based on priority.

### Phase 1: Immediate focus on high-risk fields

- **Priority:** Collections and metadata fields likely to contain references to **ethnicity, race, geography, and civilisation** should be prioritised for review, as they often hold the most concentrated instances of bias.
- **Action:** Use automated tools like the DE-BIAS tool to scan these fields, identifying high-risk terms. Manual oversight should then be employed for contextual updates and sensitivity checks.
- **Key Focus:**
  - Identify and correct any overtly offensive or outdated terms in these fields.
  - Engage with affected communities to gain insights into preferred terminology and suggestions for contextual notes.

### Phase 2: Training and awareness on intersectionality and bias detection

- **Priority:** Provide staff with the skills needed to detect and address biases that intersect across multiple categories.
- **Action:** Conduct **training focused on intersectionality** to help CHI professionals recognize how categories such as race, ethnicity, and gender intersect and reinforce biases. Be mindful of staff members who may be sensitive to these issues due to their personal experiences.
- **Key Focus:** Train your staff on bias detection, emphasising multi-layered biases that span several fields.

### Phase 3: Subtler forms of bias (shades of bias)

- **Priority:** In the final phase, focus on addressing more subtle biases, including euphemisms, umbrella terms, and terms with politically charged origins.
- **Action:** Conduct a detailed review of thesauri, paying close attention to **shades of bias** such as overgeneralizations and the political framing of terms.
- **Key Focus:** Transition from broad updates to metadata fields to more nuanced corrections, ensuring that even subtle biases are adequately addressed.



## Phase 4: Ongoing auditing and updating process

- **Priority:** Establish long-term protocols for the ongoing auditing and updating of metadata.
- **Action:** Establish **regular auditing schedules** to continuously review metadata for bias. This should include incorporating community feedback loops and adapting to evolving language and societal norms.
- **Key Focus:** Institutionalise a process for regular updates to ensure that metadata remains responsive to changes in community preferences and social values.