Inclusive Metadata Toolkit - Draft for DLF Review

Request for Feedback

Welcome, Reviewers! We are very grateful for the time and bandwidth you are contributing to reviewing and providing feedback for improving the Cultural Assessment Working Group's Inclusive Metadata Toolkit.

We are requesting feedback from fellow DLF groups and leaders via this form: Inclusive Metadata Toolkit Draft Feedback, and/or comments or suggestions directly in this Google Doc.

The current Review phase for the IMTF Toolkit will run September 03 - September 20.

If possible, please provide **feedback by Friday, September 20, 2024**.

The feedback form asks for your evaluation of the Toolkit:

- structure, and organization/ease of navigation
- any gaps in coverage of inclusive/reparative metadata topics?
- other key resources that should be included? (if so, there's a section in the Reviewers' feedback form, to submit your suggestion!)
- Please also feel free to point out any unclear or inaccurate language, or areas of the draft that seem confusing or unhelpful.

We also welcome other concerns, questions, and suggestions for improvements that Reviewers may identify in this draft.

Thank you again for taking time for this!

About this Draft

This Google doc contains a final draft of the "Toolkit document", with links to the accompanying "Toolkit Resources List", which is currently a shared spreadsheet in Google Docs.

The publicly shared version of the overall Inclusive Metadata Toolkit will have these same two components in a finalized form:

1. The Toolkit document will be a formatted, accessible PDF with features like callout/info boxes, tables, and maybe graphics, available online in an OSF repository.

2. The Toolkit Resources List will be a sortable, filterable publicly-viewable spreadsheet, with an accompanying submissions form for folks to submit ideas for candidate Resources to add, so that the Toolkit overall can grow over time.

Table of Contents

1. Introduction

There has been a rising interest in inclusive metadata within libraries and archives communities because of the increasing recognition of the ways that traditional descriptive practices have contributed to an inequitable knowledge landscape and less effective discovery and usage of resources. However, despite ample academic literature and popular discussions on the topic, it can still be difficult for metadata professionals to find concrete ways to incorporate inclusive metadata principles into their daily work practices.

This Toolkit is a collection of resources, tools, and [anything else? how to describe what is here?] gathered by the Digital Library Federation (DLF) Assessment Interest Group Cultural Assessment Working Group's Inclusive Metadata Task Force between 2021-2024. Our goal is for the Toolkit to provide support for folks who want to promote and implement inclusive metadata practices in their work/at their organizations, by outlining overall considerations, and pointing to useful resources for guidelines and developing strategies that are relevant to your context.

The overall IMTF Toolkit has two components:

1. The complete Toolkit Resources List, which is sortable/filterable by a set of parameters such as topic area, metadata type, [decide/include other parameters]:
2. This Toolkit document, which seeks to provide context for the Resources, by outlining challenges and suggestions within the broad topical areas that we used in selecting Resources.

We hope that the Toolkit Resources List will be an ongoing collaborative resource that practitioners will add to over time. To this end, we encourage Toolkit readers to propose additions via this form [link added after review period to avoid form confusion]. We are especially interested in collecting examples of case studies from practical scenarios, to share information about how various organizations are approaching their own circumstances. We hope the Resources List can support ongoing sharing of information around common challenges.

This Toolkit document is expected to be static, or at least to change less frequently. This version is meant to provide a roadmap/guide to the larger group of Resources in the spreadsheet. If there is interest from a future task force, this document could be updated/improved in a future version.

Inclusive and Reparative Metadata

Inclusive metadata describes an approach to descriptive metadata that is grounded in empathy, care, and respect, and aware of the historical and ongoing inequalities that shape the knowledge landscape. Inclusive metadata work seeks to employ a wider range of perspectives in order to more accurately recognize, represent, and respect the breadth of human experiences. An example of an inclusive metadata practice is choosing a more accurate and preferred term used by a community to describe itself rather than an established authority term when describing archival materials representing that community.

Reparative metadata addresses the impact of oppressive and harmful descriptive norms and practices by implementing inclusive metadata principles in the assessment and remediation of

existing and legacy metadata. An example of a reparative metadata project is identifying materials in an anthropological collection labeled with an offensive or derogatory name for a community, and adding the current preferred term for that community to those materials and a contextualizing statement to the collection about why harmful language was retained in the collection as evidence of the archive's role in the colonial project.

Resources about Defining Terms:

[Reparative Description](), Dictionary of Archives Terminology, Society of American Archivists

[Inclusive Description](), The DEI Metadata Handbook, Iowa State University Digital Press

1.1.   Who is this Toolkit for?

The IMTF Toolkit is for those who engage in this work may come from libraries, archives, museums, and they might be working with archival, bibliographic, or digital collections-specific metadata schemas. The principles and techniques of inclusive metadata might be applied to existing description (which entails more of a "reparative" focus) or guide new metadata creation. This toolkit is broadly applicable to many of these situations, with a particular focus on digital collections metadata.

1.2.   Purpose of this Toolkit

Inclusive metadata work can be a rewarding, collaborative, and meaningful way to connect both with communities (of users and of those represented in metadata descriptions) and with other metadata practitioners. At the same time, it can be challenging. Much of this work is cyclical and iterative, and will need to be done repeatedly as it would be impossible to identify and fix all the issues in one go.

Some of the possible challenges this toolkit is intended to address include:

- Finding problematic terminology and assessing unstructured description, and working with different types of metadata found in, for instance, library catalogs, finding aids, digital collections)
- Reviewing collections' descriptive metadata (both controlled and uncontrolled fields) will vary based on schemas and formats of resources (finding aids, MARC, Dublin Core, MODS and other non-MARC schemas) and the technology used to store and display metadata
- Reciprocal relationship building
- It is recommended that those involved in reparative metadata work consult with others who work with these materials intimately (archivists, metadata practitioners, etc.) or subject experts on content in the collections to gather their feedback on problematic and sensitive areas within the collections. Reciprocal relationship building takes time, but is a very necessary element.
- Prioritizing collections or projects where work will be invested
- How to decide how to allocate time and resources, and which collections/projects will be prioritized
- Maintenance and ongoing effort

- Maintenance is a lasting commitment. It can require reassessment and continually reviewing to update metadata, which can be a challenge

1.3.   How to use this Toolkit

1.3.1.   Toolkit Sections

Sections in this document:

- The **Getting Started: Learning about Inclusive Metadata Work** section presents a selection of resources aimed at practitioners at different levels of knowledge and engagement with inclusive metadata concepts, and can serve as a personal jumping off point or the seeds for a reading group.
- The **Getting Started: Considerations for Advocating Change** section outlines considerations for advocacy and activism around inclusive metadata, key questions to ask, and concrete strategies for action at various levels: institutional, local community, and higher-level/cross-institutional.
- The **Take Action** sections dive deeper into some of what is covered in the Getting Started sections by discussing particular frameworks, highlighted resources, and considerations for topics like **Assessment**, **Building and Maintaining Workflows**, **Controlled Vocabularies**, and **Statements on Metadata and Archival Collections**

1.3.2.   What section to start with?

This toolkit is intended to be flexible and used however will best benefit the individual and institution embarking on this work. For example, if your institution is not committed yet to undertaking this work, the "Getting Started: Considerations for Advocating Change" section might be the spot for you to start. If you are looking to learn about new tools, the Building and Maintaining Workflows section might be for you.

Each section introduces its topic, offers ideas of where to start, and highlights a few selected resources. A sortable and filterable list of many more resources is available in the Toolkit Resources List.

1.4.   Maintenance plan

The Toolkit document will be saved as a PDF and uploaded to the DLF's Open Science Framework (OSF) repository. Working drafts from the development of the Toolkit are in the [DLF/CAWG-AIG google drive location], maintained by DLF.

The Resources Appendix and Resource List spreadsheet will be available as a dynamic resource. To submit additions to the Resource List, please do so via this form [link added after review period to avoid form confusion], through the end of December 2026. The expected sunset date for the Resources Appendix is January 2027. At that time the spreadsheet data will be saved as csv and added to the OSF deposit. The static site files will also be added to the OSF deposit.

Maintainers will be responsible for:

- maintaining the Resource List spreadsheet in Google Sheets
- monitoring form for suggested additions, adding resources when approved by [who approves these?]
- at sunset of project, exporting data to CSV files and adding to OSF
- [possible communications/updating IMTF page in DLF wiki with links to sites as relevant?]

Maintainers for this project from September 2024-December 2026 will be: ==Morgan McKeehan, CAWG co-facilitators.==

1.5.   Toolkit contributors

This toolkit was developed by the Digital Library Federation's Cultural Assessment Working Group (CAWG) Inclusive Metadata Task Force. The work on the toolkit took place from 2020-2024 and was co-led by Morgan McKeehan and Rachel Jane Wittmann, with support from CAWG co-leaders over the years (Hannah Skates-Kettler, Jenny Bradshaw, Alexandra Provo, and Jackson Huang). Toolkit contributors include: Chris Day, Emma Beck, Morgan McKeehan, Rachel Wittmann, Sarah Lynn Fisher, Alexandra Provo, Jenny Bradshaw, Sharon Mizota, Rachel Jaffe, Madison Chartier, Allison Bailund, Emily Crawford, Elyse Fox, Jennifer Roper, Andrea Elizabeth Schuba, Lara Friedman-Shedlov, Staci Ross, Kristen Reid Boldridge, Michael Stewart, Christine Kim ==[more toolkit contributors will be added as we receive permission to credit them for their work]==

2.   Getting Started: Learning about Inclusive Metadata Work

The "Learning" group of Resources in the IMTF Toolkit have been selected with the aim of providing an introduction for anyone who would like to learn more about the primary issues within inclusive metadata work. These resources were selected to provide an overview of broad concepts and challenges. We have also organized this group according to levels and topical themes, with the intention of making it easier for readers to choose an area to focus on.

All of the Learning resources and their citations ==are available in the **[Toolkit Resources List: Learning](#)**==.

Levels
Resources in the Learning group section are organized into three ==levels: 'foundational', 'intermediate', and '==advanced'. These level tags are meant to give potential readers a sense of what to expect from a resource, and may be helpful for prioritizing where to start.

- **foundational**: If you are looking for a good place to start with a topic, these resources provide an approachable orientation that introduces terms and concepts, and outline general guidelines for a broad area.

- **intermediate**: Focused explorations of theories and tools within a particular topic area. These build on basic concepts to propose new approaches to common challenges.

- **advanced**: These tend to be more specialized, practical, in-depth examinations such as case studies. The aim for these resources is to provide examples of "how-to do this" for a specific topic. If you are looking for a deeper dive, peruse the "advanced" resources for a given topic.

- **"spark"**: A shorter article with a broad overview of a topic intended for either a discussion group or sharing with stakeholders when pitching a project. They are meant to be conversation starters and can be found across learning levels.

Topics/Themes
"Learning" resources were selected with the intention of covering the range of topical themes that inclusive metadata work often aims to address.

- Disability
- Gender
- Indigenous sovereignty
- Privacy
- Race and ethnicity
- Religion/spiritual practices
- User-facing communications about materials and description (such as content warnings/about archival description)

3.   Getting Started: Considerations when Advocating Change

This section provides recommendations on strategies for how to get started with inclusive metadata work while balancing advocacy and operational considerations as active parts of metadata practitioners' work.

All of the resources and their citations are available in the **Toolkit Resources List : Advocating Change**.

3.1.   Advocacy, Activism, and Assessing Risk

Advocacy can ensure that our work supports broader change throughout the galleries, libraries, archives, and museums (GLAM) field rather than isolated implementations. The goal is to increase visibility, support challenging conversations, and provide a roadmap for others.

However, advocacy is also a challenging topic. Particularly during the period of time when our group has developed this Toolkit, within the United States there has been drastic change in conversations around inclusivity. For example, in certain parts of the country, practitioners are no longer able to use words like "inclusive" or "diverse" without risking censorship, backlash, or worse. In such cases, it may

be preferable to use "reparative." Accordingly, this section addresses essential considerations for advocacy that we hope can help practitioners evaluate potential levers for change in their contexts, even when movement seems limited.

Advocacy and activism activities can have differing levels of cost and associated risk. It's important to consider which type of activity you will engage in.

Advocacy generally involves getting others to support ideas by working within a system, with or in support of organizations. Activism, on the other hand, often challenges ideas by working outside the system to influence it, sometimes against powerful organizations or individuals.

Whether aiming for advocacy or activism, deciding your approach involves assessing perceived risks, both personally and professionally. Opening up to feedback can evoke vulnerability, like worries about capacity to respond or inability to make changes.

Which approaches are sustainable will depend on the current level of institutional support for inclusive metadata activities, as well as broader local, state, and national contexts. For example, there may be laws in your state regarding funding or support for diversity, equity, and inclusion work. Also consider the status of who will lead such projects and if there are protections for this work due to academic freedom.

3.2.   Checklist: Advancing Inclusive Metadata Initiatives

Some questions to ask as you begin to advocate for inclusive metadata and reparative description and develop feasible strategies at your organization:

- Why is this work a priority?
  - Possible answers: user-focused approach, improving access, aligning with best practices in the field
- How does this work connect or align with larger institutional mission, priorities, and/or strategic areas?
- What risks are involved for practitioners engaged in inclusive metadata work?
- How does the institution perceive itself? What does the institution value; conversely, what does it consider risky?
- What language will resonate with the people you are trying to convince of the value of this work?
  - For example, if changing LCSH is perceived as daunting or too political, what about taking action because of the institution's self-identity as "innovative"/"flagship" etc?
- What staffing is needed to engage in this work?

3.3.   Levels of Action: Considerations and Strategies

There are a range of levels at which metadata advocacy work can take place. Consider what level of inclusive metadata advocacy is most relevant and feasible for your goals: within your own institution, relative to a local community, or cross-institutional – which might be at a state, national, or international level.

3.3.1. Institutional

Within the scope of inclusive metadata and reparative description practices, consider if there may be actionable steps that might not need administrative approval or buy-in. For example, updating recently changed Library of Congress Subject Headings and corresponding descriptive metadata to match updated and preferred terminology (e.g. "Enslaved persons" updated from "Slaves") and adding community-based controlled vocabulary terms (e.g. Homosaurus) to subjects, may fit in the purview of a metadata practitioner's role. Internal advocacy to gain administrative approval and technology support is most likely needed to implement a Harmful Language Statement or override problematic LCSH terms.

Within an organization, creating policies may be challenging, while establishing internal guidelines, style guides, or best practices may be more achievable and can establish benchmarks for assessing inclusive metadata initiatives. While reparative metadata remediation projects are necessary, it is critical to create internal policies or guidelines to develop consistent description practices and prevent continual harmful practices resulting in more remediation projects. If implementing policies at your organization is difficult, internal guidelines are a good step towards progress in this area.

Strategies for action

- Create a local controlled vocabulary (see the "Controlled Vocabularies" section of this Toolkit).
  - Example: "Improving Subject Headings for Iowa Indigenous Peoples" *Library Resources & Technologies* https://journals.ala.org/index.php/lrts/article/view/7667/10780
- Identify harmful terminology and locate it in your metadata (see Take Action sections on Assessment and Building and Maintaining Workflows)
  - Example: Make a list of problematic terms relevant to your collections or draw from a list like Problem LCSH and search on keywords known to be offensive in your collections
- Host webinar or film screenings or a discussion session
  - Examples: Change the Subject; Incorporating Critical Cataloging into Your Work by We Here
- Create an interest or working group to discuss how to integrate these concerns/awareness into existing metadata workflows, processes, etc.
- Start a reading group. Suggest readings from the Inclusive Metadata Toolkit Learning/Resource List
- Apply shared guidelines on replacing harmful or outdated terms in descriptive metadata, for example updating controlled vocabulary terms, using alternative controlled vocabularies, or revising aggrandizing language (see Take Action sections on Controlled Vocabularies and Building Workflows: Reparative Metadata Tools)
  - Examples: Archives for Black Lives in Philadelphia (A4BLiP) Anti-Racist Description Resources, Metadata Best Practices for Trans and Gender Diverse Resources
- Create institutional guidelines, style guides, or internal documentation for metadata best practices that include inclusive metadata strategies.
  - Examples: Yale's Reparative Archival Description Working Group: Yale records on Japanese American incarceration during World War II, https://guides.library.yale.edu/c.php?g=1140330&p=8319099; Harvard Joint Processing

Guidelines: Reparative Description:
https://sites.harvard.edu/joint-processing-guidelines/description/reparative-description/ ;
A Guide to Conscious Editing at Wilson Special Collections Library
https://library.unc.edu/project/conscious-editing-initiative/

- Write a statement or policy on harmful language in your collections (See Section 6).
  - Examples:
    https://cataloginglab.org/list-of-statements-on-bias-in-library-and-archives-description/
  - Peltzman, S., & Besser, K. (2022). Toward Ethical and Inclusive Descriptive Practices. Journal of Critical Digital Librarianship, 2(1). https://doi.org/10.31390/jcdl.2.1.03
    - *article about crafting a descriptive practices statement*
- Create a feedback form so harmful language or other inclusive metadata issues can be surfaced
  - Example: NYU Libraries' Changing the Subject project and harmful language reporting form Changing the Subject project (Google Doc)

3.3.2. Local communities

Engaging directly with a local community is a way to build trust and ensure your efforts are in line with the wishes of community members represented in or impacted by your institution's metadata. These kinds of relationships are a commitment. It's important to consider things like continuity in the face of staff turnover or change, compensation for community members' participation, and accountability measures to ensure there is follow-up and communication about any actions taken.

Strategies for action

- Facilitate a conversation, focus group, user study, or other collaborative project with partners and community members represented in, or impacted by, your metadata to learn more about their values, use cases, concerns, problematic language, or priorities.
- Listen to the perspectives and voices of those represented in metadata
  - Example: Koford, Amelia. "Engaging an Author in a Critical Reading of Subject Headings." Journal of Critical Library and Information Studies 1, no. 1 (2017): 1–2.
- Host webinar or film screenings or a discussion session open to the public or a specific community
  - Examples: Change the Subject; Incorporating Critical Cataloging into Your Work by We Here

3.3.3. High-level / cross-institutional

When working cross-institutionally or at higher levels such as state, national, or international, there are advantages and disadvantages. The advantage of working at higher levels is that these can provide authority/sway for shaping local levels. For example, working at the national level, advocates might seek to change shared controlled vocabularies or standards, such as Library of Congress Subject Headings or cataloging standards, or cooperative guidelines and policies, such as Program for Cooperative Cataloging documentation. Disadvantages include slower rate of change, more partners in the room, and potential mismatch between higher-level and local needs. As an example, Indigenous

communities in different parts of the world have different preferred terminology (for example, First Nations vs. Native American).

Strategies for action

- Create an interest or working group to discuss how to integrate these concerns/awareness into existing metadata workflows, processes, etc.
- Publish an article, white paper, blog post, etc. about your work
- Propose changes to national or international controlled vocabularies or cataloging guidelines
- Participate in editorial groups for shared metadata standards or best practice guidelines

---

3.4. Highlighted Resource: Evaluating your Capacity for Advocacy

The IFLA Advocacy Capacities Grid is a framework that can be used to support organizations in evaluating their current advocacy capabilities across core activity areas. The tool's purpose is to help organizations use time and resources effectively by identifying strengths, gaps, and priorities for focused effort.

The Capacities Grid provides a matrix for self-assessment across these key areas:

- Understanding and engaging current policies, laws, and lawmakers
- Crafting effective communications strategies
- Developing strategic partnerships for leverage
- Assessing and demonstrating impact

To learn more about using this framework, see IFLA's Library Policy and Advocacy Blog post:"What advocacy activities are libraries undertaking? Analyzing the Results of the IFLA Regional Advocacy Priorities Study (Part 2)"

---

4. Take Action: Assessment

In this section, we've suggested how generally accepted metadata assessment categories can be applied/extended to improve metadata inclusivity in a contextually relevant way for each institution. While recommendations for assessing metadata quality are well-established, the concept of "inclusivity" has not been explicitly included in these frameworks. However, established categories and criteria for metadata assessment are directly relevant to improving an institution's metadata inclusivity. Applying the lens of inclusivity to existing metadata assessment frameworks provides a way to address challenges in identifying and replacing harmful, outdated descriptive language.

Definition

We define "assessment" for inclusivity as:

> a set of principles and practices that extend established general categories and criteria for metadata assessment to the specific work of representing individuals and groups within collections in a way that aligns with inclusive language principles: using descriptive terms that are respectful of the humanity of the people they represent, as well as seeking to increase the depth and breadth of who is represented and in collections.

4.1.   Assessing Inclusivity of Metadata: Suggested Criteria

The table below explores the characteristics of one metadata quality assessment framework as defined in the chapter "The Continuum of Metadata Quality: Defining, Expressing, Exploiting" (Bruce and Hillmann 2004) of *Metadata in Practice*. These characteristics (completeness, accuracy, conformance to expectations, consistency, and timeliness) are widely used in the GLAM field to assess metadata quality, however, as noted above, inclusivity is not included as a discrete category in Bruce and Hillmann's criteria or in most other assessment frameworks. The table below explores the definition of each criteria as it relates to general metadata quality and how each definition relates to inclusive metadata practices.

Metadata Quality Characteristics - General & Inclusive		
Criteria	**General Assessment**	**Assessing for Inclusivity**
Completeness	Is the metadata element, property, and/or attribute present?	To what extent does collections metadata represent the people within the communities the collections are from? Who is represented, who is missing, what are the areas of gaps?
Accuracy	Are metadata values semantically and syntactically correct?	Are correct spelling, diacritical markers, and

Metadata Quality Characteristics - General & Inclusive		
		punctuation used for names and other terms?
Conformance to Expectations	Do the metadata values adhere to the expectations of your defined user communities (internal and external)?	Are the terms preferred by the communities represented used in the descriptive metadata? Are the most current subject headings used?
Consistency	Are semantic and structural values and elements represented in a consistent manner across records? Are values consistent within your domain?	Once identified, are updated or preferred terms and concepts used consistently in the metadata?
Timeliness	Do you assess metadata from legacy projects or systems before reusing it?	Language changes over time and terms that were once preferred or commonly used can become harmful. Are there known legacy collections that need remediation work? Is there a schedule for assessing metadata (e.g., every 10-20 years)?

Figure 1. Table translating general metadata quality characteristics to inclusive quality characteristics

4.2. Assessing Impacts of Inclusive and Reparative Work

In addition to assessing metadata quality, the social, community, and organizational impacts of inclusive metadata work can be a site for reflective assessment. Ricky Punzalan has written about "Areas of Impact" for reparative archival work in his talk, "Words Matter: Reparative Description as Decolonial Action." These areas could also be applied to inclusive metadata projects. Below is a brief summary of each area:

- **Knowledge**: New knowledge about collections is generated and collections are used in meaningful ways across diverse settings.

- **Attitudes**: Staff attitudes and practices have shifted and community members feel more welcome interacting with collections and staff.

- **Professional discourse**: Staff have a new understanding of responsibilities and "ownership" of collections.

- **Institutional capacity**: Guidelines for reparative/inclusive practices and new ways of managing collections have been developed.

- **Policy**: New policies around collections care and representation are in progress or in place.

- **Relationships**: Reciprocal relationships between institutions and community members are forming or in place.

4.3.   Highlighted Resources: Assessment

Here are two frameworks for metadata audits and analysis. In combination with some of the tools listed in the "Building Workflows" section, they can help get you started in developing a reparative metadata remediation project:

- Westbrook, R., Johnson, D., Carter, K. Lockwood (2012) "Metadata Clean Sweep: A Digital Library Audit Project" *D-Lib Magazine*, http://www.dlib.org/dlib/may12/westbrook/05westbrook.html
- Traill, S. and Patrick, M. (2022) "Core Concepts and Techniques for Library Metadata Analysis" *Code4Lib Journal,* https://journal.code4lib.org/articles/16078

Other Assessment Frameworks and documents

- DLF Assessment Interest Group's Environmental scan: https://dlfmetadataassessment.github.io/projects/environmental-scan/
- *Survey of Benchmarks in Metadata Quality: Initial Finding*s details the response data gathered from the Survey of Benchmarks in Metadata Quality
- ALA Core Metadata Standards Committee's Metadata Schema Assessment Framework: 🖹 Metadata Schema Assessment Framework_PublicReview

- DLF Assessment Interest Group's sample metadata quality metrics: https://dlfmetadataassessment.github.io/projects/framework/
- DLF Survey of Benchmarks in Metadata Quality: Initial Findings (2020), https://dlfmetadataassessment.github.io/assets/pdf/2020-dlf-mawg-mqb-white-paper.pdf
- Kumar, V., Chandrappa, & Harinarayana, N. (2024). Exploring dimensions of metadata quality assessment: A scoping review. Journal of Librarianship and Information Science, 0(0). https://doi.org/10.1177/09610006241239080 (Literature review).

5.   Take Action: Tools and Workflows

Reparative work on legacy descriptions in finding aids or digital collection metadata can be overwhelming. This section provides suggestions for low-overhead software tools to assist with:

- project management for individual and team-based work
- metadata audits and assessments, including identifying and remediating problematic language
- data-wrangling

The profession increasingly recognizes that remediation, like description creation itself, is an iterative process. Additionally, with capability for nuanced description comes accountability for auditing and maintaining metadata integrity. Keeping track of individual changes and which ones to track can be complicated and unruly if not handled systematically.

Along with the tools noted, the resources, and examples in this section suggest ways to approach metadata audits and assessments systematically, reproducibly, and manageably in terms of workload. It addresses challenges such as:

- Auditing metadata to meet professional standards and ethical obligations
- Leveraging regular assessments to inform active (rather than passive or reactive) remediation projects

Areas potentially needing regular maintenance include, but are not limited to:

- Routine changes to vocabularies for subject headings, demographic terms, and/or keywords (coming from the vocabulary's maintainer)
- Changes in form of name of creators, contributors, or human subjects per preferred authority files
- Changes to geographic location names, especially derogatory name changes
- Changes to classification schedules used by the collections
- Local changes made (if applicable: is the change still valid, does the reason for the local change exist) or local systems created to supplement controlled data
- Uncontrolled descriptive metadata (esp. Digital Collections: Title, description; Archives: scope and content notes, historical notes, etc.)

5.1.   Tools for Data Manipulation and Project Management

General Metadata Tools for Data Manipulation

There are a number of software and tools that can aid in reparative descriptive work, particularly for finding, identifying, and remediating metadata and description. While these tools make large-scale data manipulation easier, some do come with a learning curve, particularly for those who may not be familiar with Python, SQL, and other programming languages.

All of the resources and their citations are available in the **Toolkit Resources List : Tools for Data Manipulation and Project Management**.

Three tools that might be good options are:

- OpenRefine (free, open-source data analysis and manipulation tool)

- MarcEdit (free desktop tool to batch edit MARC data)

- LibreOffice Calc (free spreadsheet tool that can find and resolve character encoding issues and helpful for format conversion)

5.2.   Tools for Reparative Metadata Work and Assessment

The following tools and resources can help you determine what terms to target for reparative work, find harmful or problematic terms in metadata files, or remap terminology in your system.

All of the resources and their citations are available in the **Toolkit Resources List : Tools for Reparative Metadata Work and Assessment**.

For coming up with a list of problematic subject terms:

- Problem LCSH (https://cataloginglab.org/problem-lcsh/)

  - This is a crowdsourced compilation of Library of Congress Subject Headings (LCSH) hosted by the Cataloging Lab that some consider problematic (for a variety of reasons). Opinions will vary about what constitutes a "problem heading"; this may include offensive terminology, headings that aren't clear, headings that inappropriately conflate multiple ideas, or terminology that's not commonly used/understood.

For analyzing Dublin Core or other metadata schemas in CSV format:

- Marriott Reparative Metadata Assessment Tool (MaRMAT) https://github.com/marriott-library/MaRMAT.

  - This tool provides a graphic user interface for generating assessment reports on CSV-based metadata. The user can select from included lexicons based on harmful language and outdated/problematic LCSH terms, or can run the tool against a user created lexicon.

- mrs (https://github.com/ngeraci/mrs)

- This is an experiment in creating a Python library designed to detect names of people with the structure "Mrs. [male first name] [last name]," such as "Mrs. Ralph Mayer" or "Mrs. Tomás Rivera." It uses spaCy and gender-guesser.

For auditing EAD finding aids and MARC XML:

- Duke University Libraries - Description Audit Tool
  https://github.com/duke-libraries/description-audit.

  - This tool is designed to query descriptive metadata in EAD and MARCXML files against a pre-configured set of harmful terms. Tool includes binaries, documentation, and source code for public use. Repository also includes a lexicon which lists possible offensive keywords.

- XQuery for searching XML files

  - Inclusive metadata example:
    https://github.com/kellybolding/scripts/blob/master/terms_of_aggrandizement.xquery

For remapping MARC subject heading data:

- MARC Subject Headings Remapping through OCLC ("Create locally preferred subjects for display and search expansion." *OCLC*. Last updated 15 June 2023.
  https://help.oclc.org/Discovery_and_Reference/WorldCat_Discovery/Display_local_data/Create_locally_preferred_subjects_for_display_and_search_expansion?sl=en

  - This web page gives advice on how to remap harmful subject headings so that less problematic headings display in their place in your online public access catalog. This remapping project does not change the MARC record.

6. Take Action: Controlled Vocabularies

While controlled vocabularies are harder to change than free-text descriptions since they often require proposals and vetting, editing or proposing new terms has been a key (and sometimes quite visible) activity in inclusive metadata practice. Section 5 provides resources for assessing outdated or harmful terms used. This section will provide resources for advocating for changes to Library of Congress vocabularies and ideas for identifying alternative controlled vocabularies beyond the Library of Congress Subject Headings (LCSH).

All of the resources and their citations are available in the **Toolkit Resources List : Controlled Vocabularies**.

6.1. Library of Congress Controlled Vocabularies: Proposing changes

The Library of Congress Subject Headings (LCSH) is a widely-used controlled vocabulary in both MARC-based systems and in digital collections, and its terms are often a focus of intervention. Changing LCSH involves a formal proposal process, typically conducted via the Program for Cooperative Cataloging. Even if your organization is not a PCC member, there are several pathways to get involved in submitting proposals.

6.1.1. SACO Proposals and Funnels

The Program for Cooperative Cataloging (PCC)'s Subject Authority Cooperative Program (SACO) is a membership organization that provides training for the creation of Library of Congress subject and classification proposals. Members can submit proposals for adjustments or additions to Library of Congress Subject Headings (LCSH), Library of Congress Children's Subject Headings (CYAC), LC Genre/Form Terms (LCGFT), LC Demographic Group Terms (LCDGT), LC Medium of Performance Thesaurus for Music (LCMPT), and LC Classification (LCC) schedules.

SACO Funnels are groups of catalogers with shared expertise to pool their efforts in developing and submitting proposals. These groups can be joined or contacted directly to suggest subject proposal topics. Funnels coalesce around various categories, including regional, material form, discipline, and identity.

- SACO Funnels website: https://www.loc.gov/aba/pcc/saco/funnelsaco.html

6.1.2. Cataloging Lab

The Cataloging Lab is a collaborative project whose goal is to connect catalogers interested in working together on subject and name authority proposals. One of the aims is to demystify the proposal process, get people with interest in a specific topic connected with those with proposal submission experience, and help people create proposals likely to be approved. Proposals are reviewed by the Cataloging Lab community.

6.1.3. ARLIS/NA's Cataloging Advisory Committee

Some professional societies may have cataloging committees that are charged to work on subject or name authority proposals. For example, ARLIS/NA's Cataloging Advisory Committee maintains a form to solicit proposed changes to problematic art-related subject headings.

- "Proposed changes to problematic art-related subject headings" submission form: https://forms.gle/9jQKG3osUFpuheTr7

6.1.4. Highlighted Resources: Proposing changes to LC Controlled Vocabularies

- Creating subject headings for Indigenous topics : a culturally respectful guide, https://ir.library.oregonstate.edu/concern/defaults/8c97kz40m
  - This manual provides guidance on creating subject heading proposals for Indigenous topics that can be submitted via 6the Subject Authority Cooperative Program (SACO) for addition to the Library of Congress Subject Headings.

- CAC Launches Form to Propose Changes to Problematic Art Subject Headings, https://catablog.arlisna.hcommons.org/2022/01/20/cac-launches-form-to-propose-changes-to-problematic-art-subject-headings/

  - "Many hands make light work and the Cataloging Advisory Committee aims to provide a way to connect art librarians who have identified headings in need of revision with art catalogers with SACO access. And while we acknowledge the full range of subjects that can crop up in an art library collection, we are focusing our efforts on headings in art, architecture, design, and related fields."

6.2.   Controlled Vocabularies outside of Library of Congress

This section will provide some controlled vocabulary options outside of the Library of Congress.

6.2.1.   Specific subject vocabularies

These options are more specific to certain topics and can be more inclusive, but might not be an exhaustive list.

- African Studies Thesaurus
  - Released by the library staff at African Studies Centre Leiden, provides a structured list of English terms covering African studies with an emphasis on the social sciences and humanities.
- First Nations, Metis and Inuit – Indigenous Ontologies (FNMIIO)
  - Released through the National Indigenous Knowledge and Language Alliance (NIKLA) and represents an initial effort to improve the representation of Indigenous peoples and communities in libraries, archives and other cultural memory institutions.
- Homosaurus
  - An international linked data vocabulary of Lesbian, Gay, Bisexual, Transgender, and Queer (LGBTQ) terms.
- Indigenous Subject Headings in Manitoba Archival Information Network (MAIN)
  - Subject headings altered to be more inclusive and respectful of Manitoba's Indigenous people.
- Subject Heading and Term Source Codes
  - A collected resource with all of the thesauri, lists, and databases that have a MARC code.
- University of British Columbia (BC) Xwi7xwa Library's Indigenous Knowledge Organization
  - This includes the "Names for BC First Nations in BC List" which is updated and expanded as needed.

6.2.2.   Genre vocabularies

Genre terms identifying harmful content can be used on their own or in conjunction with statements on metadata and archival collections. The ALA Association of College and Research Libraries (ACRL) Rare Books and Manuscripts Section (RBMS) maintains the Controlled Vocabulary for Rare Materials Cataloging (CVRMC), a genre/form terms vocabulary that includes a section of genre terms for

[prejudicial materials](#), useful for indexing works that are prejudicial in nature, or that are the byproduct of prejudicial and hateful systems and ideologies. These terms are intended for collocating resources for research on the study of oppression, as well as to help metadata workers surface the essential nature of resources rather than contributing to or masking harm. Beyond the rare materials cataloging community, these terms can be used to identify biases and harmful content. New terms are added to the CVRMC at the request of community members using a [proposal form](#) and are vetted for inclusion by the RBMS Controlled Vocabularies Editorial Group.

6.2.3.   Highlighted Resource: Controlled Vocabularies outside of Library of Congress

[Controlled Vocabularies (CV) Decision Tree](#): The Samvera Metadata Interest Group has created a framework to help provide guidance on evaluating, selecting and using controlled vocabularies. Though it is Samvera-specific, the general principles of these guidelines can be applied by anyone selecting a controlled vocabulary for use in a system.

6.3.   Creating Local Controlled Vocabularies

While it is best practice to use a controlled vocabulary, there are cases where you may need to create a local one. Perhaps the people or groups of people represented in your collections aren't included or are misrepresented in the available vocabularies, or you may have users who need to search for culturally specific or hyperlocal concepts or terms. Local vocabularies enable you to control precisely the terms applied to your records, although it is important to remember that these terms won't necessarily be compatible or meaningful if you share your records with aggregators or other institutions.

6.3.1.   Where does your vocabulary come from?

Local vocabularies may consist of a mixture of terms from multiple outside vocabularies, or they may be completely customized lists. In both cases, it is important to do your research and make sure you choose terminology that is accurate and respectful of the people it represents. Even though your vocabulary may only be used at your institution, it is important to record the sources for the terms you choose. Like existing controlled vocabularies, your local vocabulary will need to be updated as language and cultural values change, and it is important to know where the terms you select came from. If you are mixing terms from existing vocabularies, it's best to record each term's unique identifier and/or URI. If you are using custom terminology, it's helpful to cite a source where each term is defined and in use, such as an encyclopedia entry, scholarly article, oral history, or personal communication.

6.3.2.   Where should you keep your vocabulary?

Depending on the system you use to manage your collections, you may be able to store your local vocabulary alongside the external vocabularies you already use. But if your system doesn't allow this, you can look into setting up an instance of [Wikibase](#)—a free, open source database that can be used

to provide your vocabulary as linked open data. By putting your vocabulary into a Wikibase, you enable it to be used, not only in your system, but in others as well. However, not every institution will have the technological know-how or resources to install Wikibase. A low-cost method for vocabulary storage is simply to use an online spreadsheet that can be shared internally.

You may also want to look into contributing your vocabulary terms to [Wikidata](#). Wikidata is a huge, crowd-sourced, open access repository of terminology that many institutions increasingly use as a source of authority records and standardized terminology. Contributing your local terms to Wikidata means they can be used as linked open data, in your system or by others. Note however, that once a term is in Wikidata, it can be modified and enhanced by anyone, including changing the term itself.

6.3.3.   What elements should your vocabulary include?

At a minimum, your local vocabulary should include three elements for each term: a unique identifier, the term itself, and the source of the term, including, if applicable, a URL or URI. Depending on how extensive and complex your vocabulary is, you may also want to include "guide terms," or terms that aren't part of your vocabulary but are used to organize it. For example, "apples" and "watermelons" might both appear under "fruit," which makes them easier to find than if they were simply in an alphabetical list mixed with other types of terms. You may also want to include synonyms, related terms, deprecated terms, or "do not use" terms to provide additional guidance to catalogers.

6.3.4.   Highlighted Resources: Creating Local Controlled Vocabularies

Here are few resources to get you started:

- [Curationist Taxonomy Guidelines](#) Guidelines for using Wikidata as a controlled vocabulary. Specific to Curationist, but provides some guidance as a case study.
- [Diversity, equity and inclusion principles for custom taxonomies](#) This paper discusses the common challenges and considerations associated with the development of custom taxonomies for describing people along the lines of race, gender, sexual orientation, nationality and other identity facets. Most appropriate for a DAMS context, and includes examples.
- [Improving Subject Headings for Iowa Indigenous People](#) Paper on a working group in the Metadata Services department of the Iowa State University Libraries which undertook a project to build, publish, and use a controlled vocabulary of preferred terms for Indigenous communities with ties to land that is now part of the state of Iowa.

7.   Take Action: Statements on Metadata and Archival Collections

In recent years, a wave of institutions have made public statements on inclusive metadata topics such as ethical cataloging, harmful language, and reparative description. These statements may be made at the institutional, collection, or item level. Statements have a range of purposes, from informing people that they might encounter harmful language, sharing what the institution is or is not prepared to do about harmful language, to soliciting feedback/contact from users. In some cases, harmful terminology used cannot be changed. Examples of this happening include the term being used by the original

creator, the field being a transcription field (such as a title), or because the full-text or image contents of the item are harmful. When harmful language may be encountered or potentially harmful messages cannot be changed, consider adding a harmful language statement.

-

All of the resources and their citations are available in the **Toolkit Resources List : Statements on Metadata and Archival Collections**.

7.1.   Institutional statements

Institution-wide harmful language statements or policies are web pages not associated with a specific collection or item. These statements often include institutional background, context of the archives and this work. They might take the form of a policy or a positionality statement.

Goals for your statement (from a 2022 SAA webinar on Reparative Description taught by Stephanie Luke and Sharon Mizota):

- Acknowledge harm; be specific about the types of content in your collections that may cause harm
- Acknowledge bias in your catalog/institution
- Assert that your institution maintains historical accuracy while mitigating harm
- Differentiate between supplied metadata and that created by the institution
- Acknowledge that reparative description is a work in progress
- Provide an avenue for users to report harmful language or content

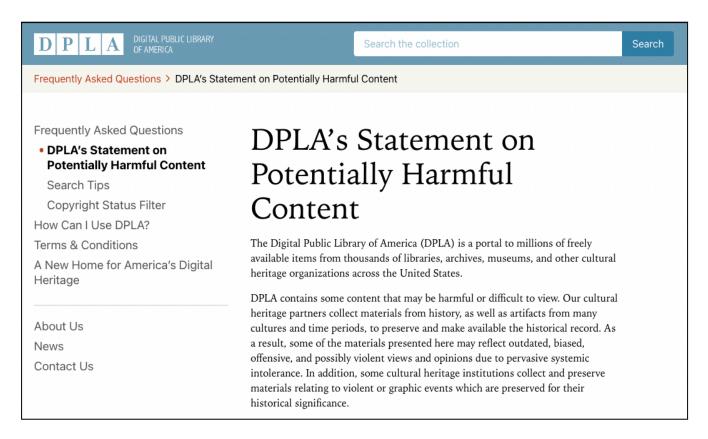Digital Public Library of America (DPLA) encourages others to use or modify their statement.

Figure 2. Screenshot of https://dp.la/about/harmful-language-statement, accessed August 12, 2024.

7.2. Collection statements

Collection statements are used when an entire collection or sub-collection is known to or suspected of having harmful content. These statements can be added to the collection's description field or communicated via a pop up message, depending on the content and system used.

This example is from the Listening to War collection of Recollection Wisconsin. It contains specific information about the types of harmful content that may be encountered, places this content within its historical context, and provides contact information to report offensive or harmful content or language.

Figure 3. Screenshot of https://recollectionwisconsin.org/listeningtowar/harmful-content-statement accessed August 15, 2024.

7.3.   Item statements

Item statements are applied on the individual digital content item level. These are often added to the description field of the record, but exact placement depends on the system.

This example is from University of Texas Arlington Libraries. The harmful content statement for this postcard appears in the lower right corner. It reads:

> **Harmful Content Statement:** This item includes content that may have outdated language or may be graphic or disturbing in nature. Please refer to our Statement of Harmful Language for more information.

Figure 4. Screenshot of https://library.uta.edu/digitalgallery/img/10021366 accessed August 22, 2024.

7.4. Highlighted Resources: Statements on Metadata and Archival Collections

- The Cataloging Lab has compiled a list of examples of both institutional and collection statements.
- Peltzman, S., & Besser, K. (2022). Toward Ethical and Inclusive Descriptive Practices. Journal of Critical Digital Librarianship, 2(1). https://doi.org/10.31390/jcdl.2.1.03
  - *This case study details how the Collection Management unit at UCLA Library Special Collections went about crafting their descriptive practices statement, providing a model for how other institutions can do the same.*

8. Bibliography

The full Toolkit Resources List is at: 🔳 **Toolkit_Resources_All**