

Summary Benchmarks

- The purpose of these criteria is to suggest an “industry standard” for absolute minimum requirements and a “gold standard” for the quality of a descriptive metadata record
- “Suggested” criteria do not rise to the level of “absolute minimum,” but are considered likely priorities based on our research and experience
- Usage:
 - Each criterion is intended to be “system agnostic” but some may not apply to every situation (e.g., local field requirements)
 - Criteria are binary -- i.e., the set being evaluated must meet all points or it does not meet the benchmarking standard
 - These benchmarks focus solely on the quality of metadata entry, not the quality of information (i.e., *available* information is all entered correctly, although we might wish that additional information is known about an item to improve the record)
 - This framework is intended to be scalable (it is written in the context of 1 record, but could apply across a collection, resource type, or an entire system)

	Minimal	Suggested	Ideal
Content relevance			
These criteria are about the relationship between the record and the item	• The record is specific/scoped correctly (e.g., <i>records for single manuscripts are not described as whole collections of materials or vice versa</i>)	• The record describes the item that it is attached to (i.e., <i>there is not a mismatch between an item and a record describing a different item</i>)	• All metadata values align with expectations for the material type (e.g., <i>language use for text vs. images; theses/dissertations have creator or author values</i>) • When applicable, relationships between items and parent collections are clearly represented • All information not specific to the item has been removed • Administrative/acquisition information is appropriately noted and/or hidden from the public, if applicable



This work is licensed under a Creative Commons Attribution 4.0 International License
To view a copy of the license, visit <https://creativecommons.org/licenses/by/4.0/>.

DLF-AIG-MWG Metadata Quality Benchmarks Sub-Group
Last revised: 2025-09-26

	Minimal	Suggested	Ideal
Existence of values			
These criteria are related to when fields should have values or not (regardless of value formatting or correctness)	Every record has a title	- All locally-required fields have values - All conditionally-required fields have values - Fields that require multiple parts or qualifiers have all components - Records have some type of subject value (<i>e.g., subject, keyword, genre, etc.</i>) - A rights statement is present (standardized statements preferred)	- Relevant recommended/optional fields have values - All relevant information about the item is included - Non-required qualifiers or field parts are added to provide enhanced information or functionality "Null" values are used consistently, according to local guidelines (<i>e.g., N/A, unknown, leave blank, etc.</i>) - Fields/subfields that cannot be repeated occur only once
Content of values			
These criteria are focused on the field values specifically – <i>e.g., string values, formatting, etc.</i>	- Value content matches the field type (<i>e.g., "date" fields do not contain text strings</i>) - No values exceed applicable system character limits - There is no text encoding that "breaks" records (<i>i.e., records are not prevented from being displayed</i>)	- Stray character encoding has been removed (<i>e.g., MARC subfields that may have been pasted in</i>) - All "placeholder" values have been replaced/ removed and are not present in the publicly accessible record (<i>e.g., {{{name}}}</i> or <i>YYYY-MM</i> or <i>notes intended to be removed by editors</i>)	- All values are appropriate lengths for their fields - All values that ought to align with standards conform to applicable vocabularies or rules (<i>e.g., names, dates, locations, subjects, etc.</i>) - All values are spelled correctly - Text fields use appropriate punctuation, grammar, abbreviations, etc.

	Minimal	Suggested	Ideal
Community/audience			
These criteria describe how metadata records reflect the needs of users		Extremely problematic/offensive terms have been removed or handled appropriately	- Reading level & language use is appropriate for all (relevant) communities or audiences <i>(e.g., collections intended for students match their reading level; collections intended for scientific research use technical terminology)</i> - Vocabulary usage aligns with the needs of the audience and material type <i>(e.g., technical vocabularies are used for scientific materials, but not for collections intended for amateurs)</i> - Values connected to interface functionality are correct <i>(e.g., field values used for browsing or filtering search results)</i> - Record language has been evaluated/updated to align with best practices related to reparative metadata, inclusive language, etc.