

This survey is a product of the Digital Library Federation Assessment Interest Group Metadata Working Group.

 DIGITAL LIBRARY



Survey of Benchmarks in Metadata Quality

CONSENT INFORMATION

BACKGROUND

You are being invited to participate in a research project conducted by the Digital Library Federation Assessment Interest Group Metadata Working Group and led by Rachel Wittmann, Digital Curation Librarian, University of Utah and Andrea Payant, Metadata Librarian, Utah State University. This survey investigates metadata quality and how it is measured in libraries, archives, and museums. The goal is to gain insight into

participant's methods and criteria for evaluating metadata quality and to identify potential gaps in present knowledge and practices. The survey will be open through June 30, 2019. This survey is open to anyone self-identifying as a metadata professional who are currently employed in a library, archive, museum or other cultural heritage institution. Specifically, eligible participants for this study will be individuals with some responsibility for metadata management and/or creation for a digital repository.

STUDY PROCEDURE

This is a virtual multi-location study. This survey will ask respondents about current metadata practices at their organizations. Participants will answer a series of questions related to current metadata requirements and evaluation practices. The survey should take approximately 15-30 minutes to complete. Respondents should only take the survey once, although there is a repeatable section for a respondent to input data for multiple repositories that they manage if desired. It is encouraged to have one representative per organization complete this survey.

RISKS

While there may be a loss of confidentiality with participating in any research study, the research team conducting this survey will minimize this risk by making the survey anonymous. Participants interested in maintaining anonymity should take this survey in a

private location and not self-identify repository or contact information.

BENEFITS

While you will not directly benefit from participation, your participation may help understand current metadata practices in establishing and measuring metadata quality in the field.

CONFIDENTIALITY

The results of this study may be published in professional journals. It may also be used for educational purposes or for professional presentations. However, no individual respondent will be identified. Your participation in this project is anonymous. We will not be collecting any personally identifying information.

PERSONS TO CONTACT

If you have any questions, you may contact Rachel Wittmann at rachel.wittmann@utah.edu or (801) 581-8294. You may also contact Andrea Payant at andrea.payant@usu.edu or (435) 797-4343.

The University of Utah Institutional Review Board has approved this study. Contact the University of Utah Institutional Review Board (IRB) if you have any questions regarding your rights as a research participant. Also, contact the IRB if you have any questions, complaints or concerns which you do not feel you can discuss with the investigator. The University of Utah IRB may be

reached by phone at (801) 581-3655 or by email at irb@hsc.utah.edu.

You may also contact the Research Participant Advocate (RPA) by phone at (801) 581-3803 or by email at participant.advocate@hs.utah.edu.

The Institutional Review Board for the protection of human participants at Utah State University has approved this research study. If you have any questions or concerns about your rights or a research-related injury and would like to contact someone other than the research team, you may contact the IRB Administrator at (435) 797-0567 or by email at irb@usus.edu to obtain information or to offer input.

VOLUNTARY PARTICIPATION

Your participation is voluntary and you may refuse to participate or withdraw at any time. You may also refuse to answer any question without any penalty.

COSTS AND COMPENSATION TO PARTICIPANTS

There are no costs or compensation for participating in this study.

CONSENT:

By selecting “Yes, I agree”, I confirm that I am 18 years or older, have read the information in this consent form, and have had the opportunity as questions. I voluntarily agree to take part in this study.

Q1

- ☐ Yes, I agree
- ☐ No, I do not agree

Default Question Block

Survey Introduction

PURPOSE

This survey investigates metadata quality and how it is measured in libraries, archives, and museums. The goal is to gain insight into participant's methods and criteria for evaluating metadata quality and to identify potential gaps in present knowledge and practices. Responses will be analyzed by the Digital Library Federation Assessment Interest Group Metadata Working Group, and anonymized results will be shared with the community.

DEFINITIONS

Aggregation Project: Organization or other body that collects metadata, or metadata and digital objects, from multiple sources and other organizations for the sake of unified presentation, discovery, and collocation of the described resources. Aggregators could include organizations

that regularly or in an automated fashion retrieve metadata from known sources, or they could include organizations that collect and expose contributed metadata from partners (or any mix therein). Example(s): DPLA, (<https://dp.la/>); SHARE (<http://www.share-research.org/>).

Digital Collections: A collection of digital resources collected and managed by an organization with metadata about those resources made accessible via an online interface. Resources are typically digitized cultural heritage, primary-source material that would otherwise be inaccessible without physically visiting the local organization, or unique born-digital resources.

Digital Content: All digital materials produced or collected by an organization, including both primary source material and scholarly research.

Digital Library Consortium: A group of digital libraries, typically from a particular geographic region (state or multi-state), with digital content which is harvested from local DAMS to a consolidated discovery system. For example, the Mountain West Digital Library, Digital Library of Georgia.

Element: A designated part, aspect, or attribute of a metadata record or more generic dataset. Examples: Dublin Core Metadata set (dc:title) ; CSV header of your local spreadsheet.

Institutional Repository: A regularly maintained online platform where digital resources and their metadata are made accessible via an online interface. Resources are typically produced by the organization itself and/or created by its constituents or members. Typically these repositories include, but are not limited to, scholarly works and related material, e.g. articles, datasets, monographs, and theses.

Metadata Benchmarks: Criteria to measure metadata quality in comparison with a standard of metadata quality set by a collective of organizations.

Metadata Application Profile: A set of metadata elements, properties, and attributes with definitions and requirements (e.g., obligation, range, cardinality) that provides an overview of usage for a particular user community or domain. MAPs could be machine-readable (as in the DCMI usage of the term) or they could be human readable documentation (examples being a lot of the PCDM profiles, DPLA MAP, or other).

Metadata Schema: Technical or semantic metadata guidelines for a particular system or domain.

Metadata Standard: A set of metadata elements and definitions agreed upon by a particular domain or community for a particular purpose.

Non-MARC Metadata: Metadata schema that are commonly used for digital content. Examples include: Dublin Core, MODS, Visual Resource Administration (VRA) Core, etc.

Repository: Collections of items managed by a single metadata application profile (MAP) or schema. Includes digital collections, institutional repositories, and aggregation projects.

SURVEY OUTLINE

Part 1: Respondent Profile

Part 2: Metadata Basics

Part 3: Metadata Inclusion and Metadata Evaluation Grids

Part 4: Metadata Quality Assessment

This survey should take about 20 minutes to complete.

Part 1: Respondent Profile

This section will gather information about the respondent's organization and their experience with metadata.

Q2-2a What kind of organization do you work for?

- ☐ Libraries - Academic
- ☐ Libraries - Public

- ☐ Libraries - Special
- ☐ Archives
- ☐ Museum
- ☐ Consortium
- ☐ Aggregation Project
- ☐ Other (please specify)

Q3 If you are willing, please fill in the name of the organization, consortium, or aggregation project you are representing:

Q4 How many total employees work for your organization?

- ☐ 1-10
- ☐ 11-50
- ☐ 51-100
- ☐ 101 +

Q5 How many full time employees in your organization work with non-MARC metadata? If half-time, indicate .5, etc.

Q6 How long have you been working with non-MARC metadata?

- ☐ Never
- ☐ 0 to 4 years
- ☐ 5 to 9 years
- ☐ 10 + years

Q7 How long have you worked with MARC metadata?

- ☐ Never
- ☐ 0 to 4 years
- ☐ 5 to 9 years
- ☐ 10 + years

Q8-8a What tasks are your responsibility when working with metadata?
Select all that apply.

- ☐ Creating descriptive metadata
- ☐ Setting guidelines and best practices
- ☐ Supervising metadata creators
- ☐ Quality control checks
- ☐ Managing existing metadata (migration, remediation, enhancements)
- ☐ Other (please specify)

Part 2: Metadata Basics

This section will gather information about metadata practices and technology used for a particular repository. A repository, as defined in the beginning of this survey, is collections of items managed by a single metadata application profile (MAP) or schema. This includes digital collections, institutional repositories, and aggregation projects.

Q9-9a How many repositories does your organization manage?

- ☐ 1-2
- ☐ 3-4
- ☐ 5-6
- ☐ Other (please specify)

The next set of questions are to gather information about a single repository and its metadata guidelines. **Please note, this section of the survey is repeatable and should take about 10-15 minutes to complete.** If you manage more than one repository, please feel free to fill out the following section once for each MAP or schema. If you have several legacy projects that are similar, please select the one that is most representative of your current activities.

Q10 For how many repositories would you like to fill out the following section?

Q11 Does this repository serve as an institutional repository, a platform for digital collections, or both?

- ☐ Digital Collections
- ☐ Institutional Repositories
- ☐ Both

(Optional) Name of Repository

Q12
-12a What type of system is being used?

- ☐ Bepress

- ☐ CollectiveAccess
- ☐ CONTENTdm
- ☐ DSpace
- ☐ Eprints
- ☐ Islandora
- ☐ Omeka
- ☐ Samvera
- ☐ Other (please specify)

Q13 Are you implementing a Metadata Application Profile (MAP)?

- ☐ Yes
- ☐ No

Q14 If you answered yes to implementing a Metadata Application Profile, is it from a governing body (digital library/ consortia) or created specifically for your digital library?

- ☐ Using a MAP created by an external consortia (example, DPLA hub; Mountain West Digital Library)
- ☐ Using a MAP created specifically for local repository

Q15
-15a What metadata schema is being used?

- ☐ Dublin Core
- ☐ EDM

- ☐ MODs
- ☐ PBCore
- ☐ Premis
- ☐ Qualified Dublin Core
- ☐ VRA Core
- ☐ Other (please specify)

Q16
-16a

Which controlled vocabularies are being used? Select all that apply.

- ☐ FAST Subject Headings
- ☐ GeoNames.org
- ☐ Getty Art & Architecture Thesaurus
- ☐ Getty Thesaurus of Geographic Names
- ☐ Getty Union List of Artist Names
- ☐ Library of Congress Genre/Forms
- ☐ Library of Congress Subject Headings
- ☐ Library of Congress Thesaurus of Graphic Materials
- ☐ Medical Subject Headings MeSH
- ☐ Other (please specify)

Q17

Approximately how many descriptive metadata records are in this repository?

Q18
-18a Do you use local or regional controlled vocabularies?

- ☐ Yes (please specify)
- ☐ No

Part 3: Metadata Elements Grids

There are two metadata element grids included in this section: the Metadata Inclusion Grid and the Metadata Quality Grid. This section will gather detailed information on the metadata elements that are included and evaluated for the repository.

Metadata Inclusion Grid

Please indicate if an element is required, optional, or recommended for your repository or project. Select "Required" for elements that must be present in a metadata record in order for a resource to be published online and/or harvested. Select "Recommended" for elements that are strongly encouraged. Select "Optional" for elements that are only included when applicable. **If an element in the list is not relevant to your repository, please do not select any options for that element.** If there are elements missing from this grid that are required,

recommended, or optional for your project, please add these in the free text field below.

Q19a-19z

	Required	Recommended	Optional
Abstract: A summary of the resource.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Alternative Title: An alternative name for the resource.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Collection Title: Name of a group of related resources which the described resource belongs to.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Contributor: An entity responsible for making contributions to the resource.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Coverage: Describes the spatial and temporal characteristics of the resource.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Creator: An entity responsible for making the resource.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Date: The date of the creation of the original resource.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Description: An account of the resource, including item's history, appearance, contents, etc.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Digitization Specifications: Description of process, equipment, and specifications used to convert resource to digital format.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Extent: The size or duration of the resource.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Format: File format of the digital resource.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Genre: Nature of original resource.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Identifier: Unambiguous reference to the resource.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
isPartOf: A related resource(s) in which the described resource is physically or logically included.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Language: Language of the resource.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
physicalLocation: The institution or repository that holds the resource or where it is available.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Publisher: An entity responsible for making the resource available.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

	Required	Recommended	Optional
Relation: A related resource(s).	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Rights: Information about rights held in and over the resource.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Source: A related resource from which the resource is derived.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Spatial: The geographic topic or applicability of the resource.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Subject: Topic that describes what the resource is about.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Table of Contents: A list of subunits of the resource.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Title: A name given to the resource.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Transcription: Transcription or full text of resource.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Type: The nature of the resource (StillImage, MovingImage, Sound, or Text).	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Q19zz

Please name and define any other **required** metadata elements not listed above:

Q19zzz

Please name and define any other **recommended** metadata elements not listed above:

Q19zzzz

Please name and define any other **optional** metadata elements not listed above:

Q20a-20z

Metadata Evaluation Grid

Which metadata elements do you evaluate for quality? Select “Evaluated” for elements that are in your system and evaluated for any measure of quality. Select "Not Evaluated" for elements that are in your system but not measured for quality. **If an element in the list is not relevant to your repository, please do not select any options for that element.** If there are elements missing from this grid that are evaluated for quality, please add these individually in the free text field.

	Evaluated	Not Evaluated
Abstract: A summary of the resource.	<input type="radio"/>	<input type="radio"/>
Alternative Title: An alternative name for the resource.	<input type="radio"/>	<input type="radio"/>
Collection Title: Name of a group of related resources which the described resource belongs to.	<input type="radio"/>	<input type="radio"/>
Contributor: An entity responsible for making contributions to the resource.	<input type="radio"/>	<input type="radio"/>
Coverage: Describes the spatial and temporal characteristics of the resource.	<input type="radio"/>	<input type="radio"/>

	Evaluated	Not Evaluated
Creator: An entity responsible for making the resource.	<input type="radio"/>	<input type="radio"/>
Date: The date of the creation of the original resource.	<input type="radio"/>	<input type="radio"/>
Description: An account of the resource, including item's history, appearance, contents, etc.	<input type="radio"/>	<input type="radio"/>
Digitization Specifications: Descriptions of process, equipment, and specifications used to convert the resource to a digital format.	<input type="radio"/>	<input type="radio"/>
Extent: The size or duration of the resource.	<input type="radio"/>	<input type="radio"/>
Format: File format of the digital resource.	<input type="radio"/>	<input type="radio"/>
Genre: Nature of the original resource.	<input type="radio"/>	<input type="radio"/>
Identifier: Unambiguous reference to the resource.	<input type="radio"/>	<input type="radio"/>
isPartOf: A related resource(s) in which the described resource is physically or logically included.	<input type="radio"/>	<input type="radio"/>
Language: Language of the resource.	<input type="radio"/>	<input type="radio"/>
physicalLocation: The institution or repository that holds the resource or where it is available.	<input type="radio"/>	<input type="radio"/>
Publisher: An entity responsible for making the resource available.	<input type="radio"/>	<input type="radio"/>
Relation: A related resource(s).	<input type="radio"/>	<input type="radio"/>
Rights: Information about rights held in and over the resource.	<input type="radio"/>	<input type="radio"/>
Source: A related resource from which the resource is derived.	<input type="radio"/>	<input type="radio"/>
Spatial: The geographic topic or applicability of the resource.	<input type="radio"/>	<input type="radio"/>
Subject: Topic that describes what the resource is about.	<input type="radio"/>	<input type="radio"/>
Table of Contents: A list of subunits of the resource.	<input type="radio"/>	<input type="radio"/>

Title: A name given to the resource.

Evaluated ☐

Not Evaluated ☐

Transcription: Transcription or full text of the resource.

☐

☐

Type: The nature of the resource (StillImage, MovingImage, Sound, or Text).

☐

☐

Q20zz

Please name and define any other **evaluated** metadata elements not listed above:

Q20zzz

Please name and define any other **not evaluated** metadata elements not listed above:

Q21

This section will repeat based on the number of repositories you indicated earlier. If you do not want to fill this section out again, please indicate below.

☐ I do not want to repeat this section

Part 4: Metadata Quality Assessment

This section will gather information on current and desired metadata quality evaluation practices.

Q22-22a

Does your organization use any tools for metadata quality assessment? Select all that apply.

- ☐ DPLA OAI Aggregator Tools
- ☐ Gadget
- ☐ LibreCat/Catmandu
- ☐ LODrefine
- ☐ MARCEdit
- ☐ Metadata Quality Control (MDQC) by AVP
- ☐ Metadata Breakers
- ☐ OpenRefine
- ☐ Python pandas
- ☐ Spreadsheet based software (Microsoft Excel, LibreOffice Calc, Google Sheets)
- ☐ Other (please specify)

Q23a-23g

When judging the quality of a metadata record, what aspects are most important by your organization? By dragging and dropping, please rank (1 being most important) the characteristics of quality (per the [DLF AIG Metadata Assessment Working Group Toolkit](#)).

Completeness: The element, property, and/or attribute is present.

Accuracy: Information is correct both semantically and syntactically.

Accessibility: Metadata can be read by both humans and machines.

Conformance to expectations: Values adhere to the expectations of your defined user communities (both internal and external).

Consistency: Semantic and structural values and elements are represented in a consistent manner across records. Values are consistent within your domain.

Timeliness: When the resource changes, the metadata is updated accordingly. When additional metadata becomes available or when metadata standards change, the metadata associated with the resource changes.

Provenance: You have information about the source of the metadata, and you can track metadata transformations back to the original form of the metadata record.

Q24

How do you measure for the characteristics described in the previous question?



Q25

What characteristics would you like to measure but are unable to?



Q26

(Optional) Any thoughts you would like to share?



May we follow up with you if we have further questions? Would you be open to an informational interview? If yes, please leave your name and email address below:



Powered by Qualtrics