

Core Metadata Elements: Guidelines to Promote Consistency and Access at TAMU

Charity K.M. Stokes, MLS, MA Texas A&M University Libraries
Jeannette Ho, MLS Texas A&M University Libraries



BACKGROUND OF PROJECT

In December of 2017, a metadata working group was formed in response to answer a two-part problem. One was to determine the metadata workflow for digital collections being placed into the TAMU's digital repository. The other was to recommend metadata schema for the DAME, an ecosystem-based approach to digital asset management involving multiple computer programs, services, and databases. Although much of the focus was on how to handle workflow for digital projects and suggest metadata schema for two specific platforms (D-Space and Fedora), it soon became obvious to the task group that all metadata must be compatible with metadata in other systems within the DAME. Thus, we developed guidelines to encourage the consistent application of metadata and standards in order to maximize discovery and accessibility to our users, while also trying to be efficient with library resources.

ELEMENTS

The Guidelines are organized into two main sections: "Core" metadata elements and "Non-core" elements.

Core elements: These elements are considered to be the most essential within the TAMU repository. They include elements that are designated as "Mandatory" in all circumstances, as well as those that are designated "Mandatory if applicable and available."

- Mandatory (M):** An element labeled "M" for Mandatory" must always be recorded. For example, "Title" is required for all resources, regardless of whether a resource has been given one by its creator. If a resource lacks a title, metadata providers are instructed to supply one themselves or record "Untitled" if this cannot reasonably be done.

- Mandatory if applicable and available (MA):** An element labeled "MA" for "Mandatory if applicable and available" must be recorded if 1) it applies (i.e., is relevant) to a particular resource, and 2) If this information is known or can be easily obtained. A metadata provider is not required to spend more than a minimum amount of effort to look for such information. If it cannot readily be found, then the element may be skipped. In other cases, a particular element may simply not relevant or applicable to a particular resource. For instance, the elements "creator" or "language" would not be recorded for a collection consisting of recorded sounds of nature, such as birdsongs.

Non-core elements: This refers to metadata that are not required, but may be recorded to provide "extra" information if the metadata provider judges them to be useful. They include elements that are "recommended" and "optional."

- Recommended:** An element with this label is not required for all metadata, but may be particularly useful for certain types of collections. The metadata provider is encouraged to apply them if he or she considers them to be helpful in making resources more discoverable. An example of a "recommended" element would be a genre term that designates what a digital resource IS (i.e., a photograph) than what it is ABOUT.

- Optional:** An element with this label is neither required nor recommended, but may optionally be recorded if it fulfills a special need of a collection (e.g., specifying the audience level for a juvenile video), or if it can enhance the use of a resource by providing context or additional information (e.g., specifying that a resource is a chapter of a particular book).

How Much Metadata to Provide? : Three Options for "Completeness"

- Minimal Level:** To be acceptable within the repository, a metadata record must at the very least, contain the elements designated as "Mandatory" in all situations. A record containing only these elements is considered to meet the bare minimum standard of completeness. Although fuller metadata is encouraged, this minimal approach may be taken if metadata providers lack the time or expertise to provide more complete information.
- Standard level:** A metadata record consisting all "Core" elements ("Mandatory" AND "Mandatory if applicable and available elements) is considered "standard," in that it should provide enough information to enable discovery of a digital resource for most users. Metadata providers are encouraged to strive toward this level.
- Fuller level:** A metadata record consisting of all "Core" elements and all recommended and/or optional elements that are relevant to a particular resource would be considered as "full" as possible. While a "standard" level record should provide sufficient information, metadata providers may choose to make the record as full as needed if this would improve discovery for a specialized type of resource or enhance its usefulness for its target audience.

Requirements for Minimum, Full and Standard Records

Minimum

To be acceptable within the repository, a metadata record must at the very least, contain the elements designated as "Mandatory" in all situations plus at least ONE subject term/heading when applicable. A record containing only these elements is considered to meet the bare minimum standard of completeness. Although fuller metadata is encouraged, this minimal approach may be taken if metadata providers lack the time or expertise to provide more complete information.

Title (M, non-repeatable)
Definition: The primary name given to a resource.

Usage guidelines: Typically, a title will be a name by which the resource is formally known. It is usually assigned by the creator of the resource and is found on the resource itself. Transcribe the title as accurately as possible. If no title is found on the resource, look for one on materials that supplement the resource. If no title can be found for a resource, supply one when feasible.

Content Type (M, repeatable)
Definition: The nature or the genre of the resource. It specifies the characteristics and general type of content of the resource.
Usage guidelines: Use DCMI Type vocabulary (<http://dublincore.org/documents/dcmi-type-vocabulary/>). Choose the most specific term that is available.

Digital Publisher (M, repeatable)
Definition: An entity responsible for making the resource DIGITALLY available. For example, a resource (e.g., a technical report) may have been previously published by the Texas A&M Transportation Institute, but the Libraries would be the entity digitizing it and making it available online in our repository.
Usage guidelines: Since the Texas A&M University Libraries is responsible for making the described resources available in their current form, always consider it to be the digital publisher.

Rights/Access (M, repeatable)
Definition: Information about rights held in and over the resource and terms of its use and reproduction.
Usage guidelines: Use a controlled copyright statement from <https://rightsstatements.org/en/> to ensure consistency of the rights statement across all items in the institutional repository

Reformatting (M, repeatable)
Definition: Whether a resource was digitized from another physical format or was originally created in online form.
Usage guidelines: Use phrases recommended by DLF/Aquifer Guidelines

Filenames (M)
Definition: Filename used to identify a digital resource.
Usage guidelines: Filenames are mandatory, since it is essential to be able to connect the metadata with the particular digital image that is being described.

Subject (MA, repeatable)—Only one subject heading or keyword is required for a minimal-level record if applicable
Definition: The topic of a resource
Usage guidelines: May be controlled or uncontrolled. It is strongly preferred that controlled subject terms be used from your choice of thesauri.

Example

Title	Texas beyond the periphery: an archaeological study of the Spanish missions during the 18 th century
Content type	text
Digital publisher	Texas A&M University, Libraries
Rights/Access	In copyright
Rights/Access	For more information about rights and restrictions relating to this resource, go to: http://rightsstatements.org/vocab/InC-EDU/1.0/
Filename	1556635.pdf
Reformatting	reformatted digital
Subject	Spanish mission buildings—Texas
Subject	Anthropology

Standard

A metadata record consisting all "Core" elements ("Mandatory" AND "Mandatory if applicable and available elements) is considered "standard," in that it should provide enough information to enable discovery of a digital resource for most users. Metadata providers are encouraged to strive toward this level.

Creator (MA, repeatable)
Definition: A person, organization, or service primarily responsible for the intellectual content of the item.
Usage guidelines: Enter the components of a name in the following order: {Last name}, {First name or initial}, {Middle name or initial, if known}.

Date published (MA, non-repeatable)
Definition: Date of original publication, presentation or distribution of a resource prior to digitization.
Usage guidelines: Always record date of issuance if resource was previously published prior to digitization.

Date created (MA, non-repeatable)
Definition: Date of the creation of the resource in its original form.
Usage guidelines: This element is mandatory if applicable and available if Date published is not available.

Summary/Abstract (MA, repeatable)
Definition: A summary or abstract of the resource.
Usage guidelines: This is a free-text field where the resource is described. It is recommended to copy and paste an existing summary or abstract if one lacks time or the resources needed to compose one. One may edit an existing one, if higher quality is desired.

Language (MA, repeatable)
Definition: A language of the resource.

Institution/Department (MA, repeatable)
Definition: An entity responsible for contributing a collection to the repository.
Usage guidelines: Record this element when the contributor happens to be a different institution or a more specific unit, such as a library, college or department.

Local Digital Identifier (MA, repeatable)
Definition: Local identifier (not including filename used to identify a digital resource)
Usage guidelines: Locally assigned identifiers other than filenames are mandatory if they are available for a given digital resource.

Standard Identifier (MA, repeatable)
Definition: An unambiguous reference to the digital resource within a given context that adheres to an international standard (e.g. ISBN or DOI)
Usage guidelines: Record this element if applicable. The standard identifier must refer to the digital form of the resource, and not the physical form of the resource before it got digitized.

Edition/Revision Information (MA, repeatable)
Definition: Information identifying a particular edition or version of the resource.
Usage guidelines: Record if applicable and available for a given resource. This element may apply to some resources that may have multiple versions.

Example

Title	Texas beyond the periphery: an archaeological study of the Spanish missions during the 18 th century
Content type	text
Date issued	1994
Abstract/Summary	The majority of Spanish settlement in Texas took place during the 18 th century when Spain was recovering from a serious economic decline that began ca. 1500 and continued until 1760. Spain's failure to maintain her settlements in Texas has been attributed to this economic decline—frequently referred to as "the century of depression."
Creator	Carlson, Shawn Bonath
Publisher	Texas A&M University, Libraries
Rights/Access	In copyright—educational use permitted
Rights/Access	For more information about rights and restrictions relating to this resource, go to: http://rightsstatements.org/vocab/InC-EDU/1.0/
Filename	1556635.pdf
Reformatting	reformatted digital
Language	eng
Subject	Spanish mission buildings—Texas
Subject	Anthropology
Subject	Spaniards—Texas—Antiquities
Subject	Texas—Antiquities

Full

A metadata record consisting of all "Core" elements and all recommended and/or optional elements that are relevant to a particular resource would be considered as "full" as possible. While a "standard" level record should provide sufficient information, metadata providers may choose to make the record as full as needed if this would improve discovery for a specialized type of resource or enhance its usefulness for its target audience. Below are a few of the recommended and optional elements.

Alternative Title (Recommended, repeatable)
Definition: An alternative name for the resource.

Genre (Recommended, repeatable)
Definition: The form or genre of the resource.

Table of Contents (Recommended, repeatable)
Definition: A list of subunits contained within a resource.

Contributor (Recommended, repeatable)
Definition: Any person or corporate body that makes contributions to a resource but is not its primary creator (e.g., illustrator, editor, etc.).

Related resource (Recommended, Repeatable)
Definition: A resource (either digital or print) that the resource being described is related to

Original publisher (Recommended, repeatable)
Definition: Entity that published a resource before it was digitized.

Physical Extent (Recommended, repeatable)
Definition: A statement of the number and specific material of the units of the resource, and/or physical dimensions of the original item that became digitized.

Sponsor (Recommended, repeatable)
Definition: A person or group that funds or sponsors the development of some aspect of a resource (e.g., funding research, sponsoring an event.).

Source Collection (Optional, Repeatable)
Definition: The original collection that a resource belonged to before it was digitized. For example, this may be a collection of manuscripts in Cushing.

Original resource (Optional, repeatable)
Definition: A related resource from which the described resource is derived either in whole or part.

Example

Title	Texas beyond the periphery: an archaeological study of the Spanish missions during the 18 th century
Alternative Title	An archaeological study of the Spanish missions during the 18 th century
Content type	text
Date issued	1994
Abstract/Summary	The majority of Spanish settlement in Texas took place during the 18 th century when Spain was recovering from a serious economic decline that began ca. 1500 and continued until 1760. Spain's failure to maintain her settlements in Texas has been attributed to this economic decline—frequently referred to as "the century of depression." Introduction – A global approach to Spain's colonization: world-system theory – Historical and cultural geography of New Spain: the interior province of Tejas – The archaeological evidence – Evidence of a world-system in New Spain – Conclusions. Based on a dissertation submitted to Texas A&M University in the field of anthropology in 1990.
Extent	270 pages
Creator	Carlson, Shawn Bonath
Contributor (role-illustrator)	Hugill, Peter J.
Contributor (role-editor)	Shafer, Henry J.
Publisher	Texas A&M University, Libraries
Rights/Access	In copyright—educational use permitted
Rights/Access	For more information about rights and restrictions relating to this resource, go to: http://rightsstatements.org/vocab/InC-EDU/1.0/
Filename	1556635.pdf
Reformatting	reformatted digital
Language	eng
Related resource	Spanish missions in Texas during the 18 th century: a dissertation
Related resource (URL to thesis book was based on)	http://hdl.handle.net/1990/2thesismissionsTexas
Genre	Book
Subject	Spanish mission buildings—Texas
Subject	Anthropology
Subject	Spaniards—Texas—Antiquities
Subject	Texas—Antiquities

Our Process

- To recommend policies for descriptive metadata that would apply to both its DSpace and its (newly implemented) Fedora repositories, the working group (consisting of three cataloging/metadata librarians considered:
 - The literature (Especially a study by Yang, L. (2016) that found that title, subject and description were most useful for keyword searches on the Internet to retrieve items in a repository)
 - Practices at other institutions (Especially an environmental scan study conducted at Emory University in 2017. It is available online at: <http://metadata.emory.edu/about/research.html>)
 - Existing standards:
 - Requirements for Dublin Core and MODS (DLF/Aquifer Guidelines)
 - Recommendations made by another working group in 2017 that examined "core" elements in Resource Description and Access (RDA) and ones recommended as core by the Program for Cooperative Cataloging (PCC)
 - What was practical or realistic in terms of interoperability and expectations for non-expert metadata creators
- After recommendations were approved, we developed:
 - Guidelines describing elements, instructions and examples for recording them, recommended standards (e.g., ISO 8601 for dates) and controlled vocabularies), and mappings to Dublin Core and MODS

Metadata Elements that Do Not Need to be Recorded (i.e., that are System Supplied)

Some metadata elements are automatically supplied by the TAMU Libraries repository when digital resources get uploaded. Thus, there is not a need to record the following types of information, since they would duplicate information that is already generated by the system:

- File format (e.g., image/jpeg, etc.)
- Date of digital publication (e.g., the date a resource became available online in the repository)
- URL at which a resource may be accessed
- Name of a collection
- Duration (i.e., playing time) for audio and video resources
- File size

Next Steps

- Publicizing our metadata guidelines to colleagues within TAMU Libraries who work with metadata
 - Metadata creators (curators, subject selectors, etc.)
 - Digital Initiatives staff
 - Other stakeholders
- Providing education and outreach in the form of a presentation or workshop
- Listening to feedback
- Further testing of guidelines:
 - Within our new Fedora and Avalon-based repositories as these systems continue to get fully implemented
 - To see how well they work with tools/interfaces that we use with Fedora (e.g., Spotlight, our Curators Administrative Platform (CAP), eventually Blacklight?)
- Development of additional core and recommended/optional elements for other types of resources (e.g., serials, digital images, video, etc.)
- Refinement of online self-submission forms for Manakin and Vireo as needed
- Continual adjustment of procedures as necessary—Metadata design and implementation is an iterative process!