

Better Guidelines, Better Functionality: How Metadata Supports the Cycle of System Improvement at UNT

Hannah Tarver
University of North Texas
Libraries Digital Projects
Unit, USA
hannah.tarver@unt.edu

Abstract

The University of North Texas (UNT) Libraries recently revised their Metadata Input Guidelines in order to improve usability and accessibility for metadata writers, and to enhance the quality of metadata that drives new features in their digital systems. This paper describes important considerations in the revision process and also demonstrates the relationship between quality metadata and system functionality that ultimately benefits both metadata creators and system end-users.

Keywords: metadata; input guidelines; schemas; system functionality; quality control; faceted searching

1. Introduction

As part of the University of North Texas (UNT) Libraries' mission to support virtual collections for research and scholarship, the Digital Projects Unit maintains the digital collections which include both the UNT Digital Library and The Portal to Texas History. The Digital Library functions primarily as a repository of university scholarly and creative works while the Portal seeks to draw in items from across Texas to serve as a virtual resource for the rich history of the state. Although both systems started modestly, they currently contain more than 130,000 objects combined. To meet the requirements of the increased volume and the needs of our users, the system infrastructure has had to change radically to support the growth of the collections.

In 2009, the University of North Texas (UNT) Libraries launched Aubrey, a new system that is now the framework for all of our digital collections. The system, which continues to evolve, contains many features that rely directly on the quality of the metadata records to offer users options for browsing and narrowing search results. To support this aspect of the new system, we decided to review and overhaul our guidelines for metadata creation (UNT Libraries, 2009). This paper looks at the process and aspects that we found most important in creating an accurate set of guidelines, with examples of functionality in the system that were made possible by having consistent metadata.

2. Quality Guidelines Allow for Quality Metadata

When the original metadata schema was written in 2004, using Dublin Core guidelines as a basis, it served well as a platform for describing fields within UNT Libraries metadata records. However, it did not contain enough information to clarify the various complex problems that metadata creators encounter. The instructions were adequate for generic, straightforward objects, but names and other attributes that did not easily fit within the framework had no specific guidance. Additionally, the number of partner departments and institutions contributing objects to the collections has increased exponentially in the last six years and introduced a greater variety of resource types. The original guidelines addressed neither the issues specific to formats that were not previously part of our digital collections, nor the complex situations and difficult formatting issues that resulted. Good metadata does not come about merely by assertion; consistency and quality in metadata records require reference materials, tools, and other support.

We chose to make our metadata guidelines a priority to facilitate better understanding amongst metadata creators and ensure consistency in metadata records.

We used the original schema as a template to draft our new guidelines, augmented by group suggestions and experiences from writing records within the system. However, it soon became clear that the format needed to change drastically. Although it contained the necessary elements, the new document did not have the clarity needed for the diverse user group that it is meant to serve. Our digital collections have more than one hundred partners that contribute objects, including additional UNT library and university departments, other universities, and outside organizations, all with diverse staffs of varied levels of cataloging and metadata experience (see Figure 1). Since our goal was to facilitate improved consistency and quality in metadata records throughout the system, the language and format of the guidelines had to become more accessible and usable for both new metadata creators looking for basic information and more seasoned metadata writers looking for answers to unusual problems.

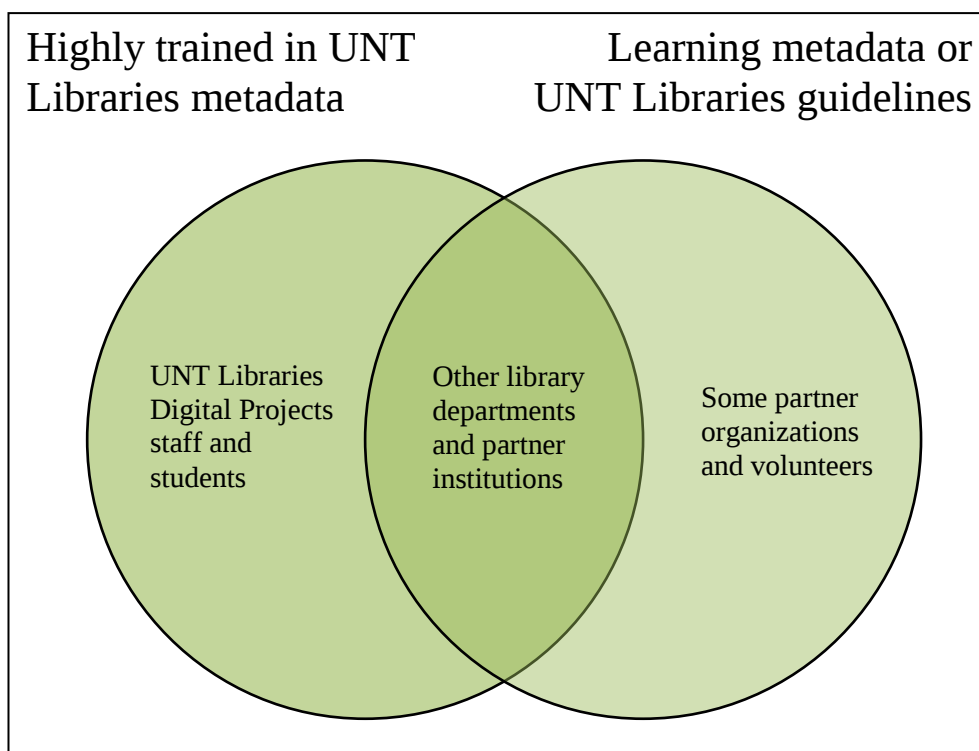


FIG. 1. Range of metadata creators entering information using UNT Libraries standards.

2.1. Clarity and Usability

One of the most important goals of creating revised metadata guidelines for the UNT Libraries was to make it more useful and usable by our entire spectrum of metadata creators. To facilitate usability, the formatting and instructions were both reworked. In the original schema, each metadata field had a separate page containing the same elements. We chose to keep that overall formatting, but changed the sections that we included and labeled them without jargon. For example, one of the first sections, “Where Can the [Element] Information be Found?” describes what library jargon would refer to as “the chief source of information” for the element, depending on the kind of object a metadata creator is describing. Eliminating jargon and using leading questions as headings help users skimming the guidelines to more easily understand where to look for the instructions that they need (Redish, 2007). This need for clarity and reduction of jargon was also a factor in our decision to change the name from “Metadata Schema” to “Metadata Input Guidelines.”

Sections explaining the formulation of field values needed further simplification to make it easier for users to find only the relevant guidelines, examples, and exceptions. The original schema included some tables to outline technical requirements, but the guidelines for field values were written in paragraph form or text-heavy lists that were difficult to read (see Figure 2). Additionally, the instructions on field values were interspersed with commentary about the relationships of the field to other metadata elements, MARC (MACHine Readable Cataloging) fields, ISO (International Organization for Standardization) standards, and other usage or conformance notes that made it difficult to focus only on the relevant guidelines.

Input guidelines:

Input guidelines for Content Description.

- Enter descriptive text about the resource. Free text descriptive information can be taken from the item itself or formulated by the metadata cataloger. Possible sources of content description include text, abstract or other structured description, container or cover notes, video contents, notes written on the back of the photograph, and careful observation by the metadata cataloger.
- The content description should be specialized information not included in other elements, concise and precise, while keeping in mind that the exact form and completeness of the content description is flexible and should be appropriate for the nature and importance of the resource.
- As a rule of thumb, describe the item using only the information in hand without doing any additional research--unless it is believed that additional research and description is required for a patron to locate the resource or to understand the intellectual content of the resource.
 - Only if necessary, an encyclopedia or other reference source can be used. For example, if the resource is an image of a country's President, you may want to add the dates that the person held that office. If specific information that would not be considered common knowledge is included from a reference source, cite the reference source.
- Avoid any commentary on or interpretation of the item being described.
- Since the description field is a potentially rich source of indexable vocabulary, care should be taken in describing content. For instance, in the first one or two sentences, each non-textual item should be described so that a user who is visually disabled will understand its intellectual content if this is not already clear from the title.

FIG. 2. The original guidelines used large blocks of text to describe input rules.

One of the first steps was to more clearly delineate the information that we wanted to include in the guidelines. Information about formatting field values was moved into one section near the start of each element page; notes about standards compliance, compatibility, and appropriate usage were moved to a “comments” section near the end of each page so that they would remain accessible to those who need or want a more technical understanding, but would not hinder the majority of users who need to see only the primary guidelines. This change also helped to clearly establish which aspects of metadata creation we wanted to address and how we wanted to approach it.

For the value-formulation guidelines, we eliminated the paragraphs and dense blocks of text. Instead, we tried to apply editing techniques that would make the information more “skimmable” (Bowles & Borden, 2004). General guidelines that applied to an entire element or sub-element were broken into shortly-phrased bulleted lists (see Figure 3). Listing information makes it easier for readers to process each point and find the most relevant information for the situation; using lists also makes it more likely that readers will look at every point instead of skipping to the next section without assimilating all of the important aspects of the previous section (Redish, 2007).

How Should the Content Description be Filled in?

Guideline		Example	
- Describe the subject matter of the item - Use complete sentences with proper grammar and punctuation - Avoid any commentary on or interpretation of the item being described			
It is recommended that the description start with a statement of the item type		Photograph of... Yearbook for Central University..	
Be descriptive but only include detail that would be helpful to users:			
For photographs and artwork, be as descriptive as possible about what the image shows		Photograph of a cowboy riding a brown bull in an arena. A rodeo clown stands off to the right. Behind him, people watch from the other side of a red fence.	
For texts, give a brief overview of the item		Biography of Alexander Gregg includes background information about his family as well as a sketch of his life and involvement in the church throughout his career until his death in 1893.	
For letters, be sure to state who the letter is to and from, giving a brief overview of the content		Letter from James E. Sutherlin to his family describing his plans for saving money to buy a new car once he gets home.	
For postcards, describe them as photographs (or artwork) if there is nothing written on the back; if the postcard has a message written on the back, treat it as a letter		Postcard showing tents of the New Hampshire camp at Ft. McIntosh, Laredo, Texas. Postcard of the Custom House building in Nuevo Laredo, Mexico. The back of the postcard includes a thank you note addressed to Capt. Elmer C. Croom from L. R. de la Peña.	
- For serials, a generic description of the content in the series can be used in every record - Make relevant changes regarding topics of specific volumes or indexes		Quarterly publication containing genealogical information about families in East Texas including fifth generation charts, family histories, and lists of records (births, deaths, etc.). This issue focuses on "Dallas Goes to War: Life on the Homefront"	

FIG. 3. A portion of the same content description guidelines with the new formatting.

More specific guidelines about how to format field values for elements or sub-elements were also broken into lists, however, we also chose to place instructions into a table side-by-side with examples (see Figures 3 and 4). By creating explicit sets of guidelines that are immediately illustrated with examples, it is easier for users to understand how to apply our instructions; if we only included examples in a later section, the practical applications would be disconnected from the guidelines and someone new to the system might need to navigate back and forth to fully understand the guidelines and how to apply them. Additionally, this format helps to mitigate a reader's impulse to jump straight to examples without fully reading and comprehending all of the relevant instructions.

To make instructions even clearer, we included shading in our guideline tables to set off particular sets of information and to group similar instructions together (see Figure 4). Not only does this formatting break information into more accessible sets of guidelines, the shading helps

to focus readers on how the two columns (guideline and example) in each row relate to one another (Redish, 2007).

Finally, to improve navigation, we included a table of contents at the top of each page with links to each major section. This helps new users understand which aspects of metadata creation are addressed in the page while allowing other users to jump directly to the relevant section to find the answer to a difficult problem, links to a specific tool, or other information (see the Appendix for an example of complete guidelines for a field).

Guideline	Example
Individual Names:	
- Enter as much of the names as are known using proper formatting: - Invert names (Last, First Middle) - Use initials if the full names are not known - Use spaces between initials - Put additional middle names after the first name - Consider both parts of a hyphenated name the 'last name' - Consider multiple parts (von, de la, etc.) as part of the last name	Hébert, Rachel Bluntzer Reid, Samuel C. Blackburn, J. K. P. Briscoe, Mary Jane Harris Donahue-Smith, James A. de la Peña, L. R.
If it is unclear which part of the name is the surname, enter the name as it appears on the item	
- Use appropriate abbreviations: - Only include known titles (Dr., Rev., Capt., etc.) before the first name if: - The title is necessary for clarification of the name - The title is the only part of the name known (aside from a surname) - Include suffixes that are a part of the name (Jr., Sr., etc.) at the end of the name after a second comma	Morris, Mrs. Harry Joseph Jones, Dr. Roberts, Frank H. H., Jr.
- Do not include: - nicknames - abbreviations - titles that do not fit the criteria above - job or educational qualifiers (C.E., D.D.S., Ph.D., etc.) - These can all go in the information section of the field.	
If known, use the authorized form of the name from the Library of Congress Authorities	Kittrell, Norman G. (Norman Goree), 1849-1927

FIG. 4. Guidelines were rewritten side-by-side with examples for clarity.

2.2. Completeness

Although clarity was a key aspect of the new metadata guidelines, we also wanted to describe our standards thoroughly. The original schema had become outdated, in part, because it included only the most generic rules and examples. For example, it stated that a creator or contributor name should not include titles or suffixes to avoid entries such as “Smith, Mr. John L.” and to maintain more standardized names. However, that rule required names such as “Mrs. Harry Joseph Morris” and “Frank H. H. Roberts Jr.” (these examples are in the table from Figure 4) to be entered as “Morris, Harry Joseph” and “Roberts, Frank H. H.” respectively. Neither of those names represents the persons who were actually responsible for the works, which makes it difficult to maintain the integrity of the information while also using appropriate formats. Therefore, one of the primary goals in rewriting the schema was to address as many unusual situations as possible and to record current precedents.

Several authors have written about writing metadata guidelines and the topics that need to be covered. Chan and Zeng (2006) comment that

For each element defined, a metadata standard usually provides content rules for how content should be included (for example, how to identify the main title), representation rules for content (for example, capitalization rules or standards for representing time), and allowable content values (for example, whether values must be taken from a specified controlled vocabulary or can be author-supplied, derived from text, or added by metadata creators working without a controlled term list) (Metadata Schema section, para. 3).

To make our instructions more complete, we included the following information for each element: name, definition, where the information can be found (chief sources of information), how the field functions in the system (field parts, repeatability, etc.), guidelines for creating field values (with examples, as shown in Figure 4), additional examples from a variety of situations, and, when applicable, comments (information about related fields, relevant non-UNT Libraries standards, etc.) and resources (links to controlled vocabularies and external tools). A complete element page is included in the Appendix as an example.

In particular, we spent time choosing a diverse set of examples from The Portal to Texas History as well as the UNT Digital Library to cover as many situations as possible for each element. It seemed important not only to clearly state our guidelines and expectations for formulating field values, but also to show numerous use-cases as a way of illustrating our rules. As a whole, incorporating so many clearly-delineated sections and examples gives a well-rounded view of the way that each element functions within our metadata system.

2.3. The Syntax-Semantics Dichotomy

The original schema included information about technical specifications for each element, however, one goal of the UNT Libraries was to move away from pairing system requirements and field formatting. This distinction between syntax and semantics has been noted by several authors (Duval, Hodgins, Sutton, & Weibel, 2002; Chen & Zeng, 2006), some of whom even note the desirability that the two be separated (Duval et al., 2002). The UNT Libraries intended to have one document (the UNT Libraries Descriptive Metadata Definition: Version 3.0) that explicitly states technical information about each element - including whether or not it is required or repeatable, has qualifiers, and the data type that is entered into the field - and a second document (the Metadata Input Guidelines) that gives details about how to format values. In theory, it makes sense to separate these two aspects of the standards; in practice, the line between syntax and semantics is not always simple.

While writing the guidelines, it became clear that the distinction between technical system requirements and style could only be maintained to a certain point without sacrificing clarity. For example, whether or not the creator field can be repeated affects how many creators a metadata writer can include. Similarly, the type of creator (person or organization) and role of the creator (author, photographer, draftsman, etc.) are required but use controlled vocabularies; how a metadata creator enters these values is dependent on whether the system uses a drop-down menu to display options or has a text field that must be filled in after the data enterer looks up the appropriate codes.

In addition to the practical requirements, expecting partners with little or no metadata experience to look at multiple documents, one of which is highly technical and jargon-based, in order to write a metadata record would be counter-productive toward our goals of quality and consistency in records and clarity of our guidelines. To bridge this gap, we included a section titled "How [Element] Works in the Metadata Form" to describe whether each element is required or repeatable, and to specify the kind of input needed (text, controlled terms from a drop-down list, etc.). Usage instructions also reference some technical aspects, when appropriate, to ensure that metadata creators are clear about how to fill in each section of an element.

2.4. Accessibility

In terms of consistency, guidelines can only be effective if every user creating or editing metadata has access to the established standards. In the case of UNT Libraries, it was known from the start that the metadata guidelines would have to be available via the Internet so that they would be as accessible to the many partners scattered across the state as they are to departments on campus. Additionally, in the UNT Libraries system, the metadata guidelines are accessible to editors directly from the metadata entry form (see Figure 5).



FIG. 5. The UNT Libraries editing system connects to guidelines from each field using a link in the upper-right corner.

Although Internet access was primarily a consideration for our digital collections partners, it was also our intention to make our standards freely available to the wider community. Having our documentation available promotes dialog as well as access to resources that some have previously observed are lacking in the digital field (Bruce & Hillmann, 2004).

2.5. Flexibility

Another consideration that the UNT Libraries had when choosing a means of access was the flexibility to change the guidelines as necessary without difficulty. Realistically, guidelines will not stay static forever unless the collection remains highly specialized in a static environment. Currently, we update the UNT Libraries' guidelines as often as necessary to maintain currency. The guidelines are updated if instructions need additional clarification, if new precedents are set, if information is missing, or if links to external pages change. The core metadata standards do not change, but changing system requirements, the acquisition of new kinds of items, or other situations may require guidelines to be updated. Even barring major changes, guidelines will need to be updated for clarification purposes any time that objects do not fit the current guidelines exactly; not changing the guidelines means that precedents set for one object may not be accurately recalled when another object with a similar exception enters the collection. In this way, flexibility of guidelines is even more important for consistency than a set of permanent guidelines that never changes.

3. Quality Metadata Allows for More Functionality

Maintaining quality metadata could be considered an end in itself for the preservation and diffusion of accurate information. However, as an added incentive, in the UNT Libraries system there is a direct correlation between the quality of the metadata and the functionality of the user interface. Some authors (including Qin, Liu, Lin, & Chen, 2009) have noted that many digital systems do not allow for sufficient options so that end users can easily narrow search results to a relevant subset. One of the most obvious changes in the UNT Libraries user interface is the recently-implemented option for faceted searching, which was not previously available. Now, when a user searches for a term or browses a segment of the collection, menus on the left side of the screen allow him or her to narrow search results by various criteria which may include the holding partner, collection, U.S. county, decade, series or serial title, language, resource type, degree, discipline, country, and access (see Figure 6). Although all of the criteria are available within the system, the faceting options presented to users will vary based on the contents of the

search results list (a facet menu will not be presented for any field that has the same value for all of the search results). These options are made possible only by ensuring quality metadata and taking advantage of the fields built into our records since the system draws on field values with structured formatting to generate faceted results. In fact, when the system was first switched over, it was blatantly obvious just how many records had too little information, incorrect formatting, or other errors that affected retrieval of faceted information.

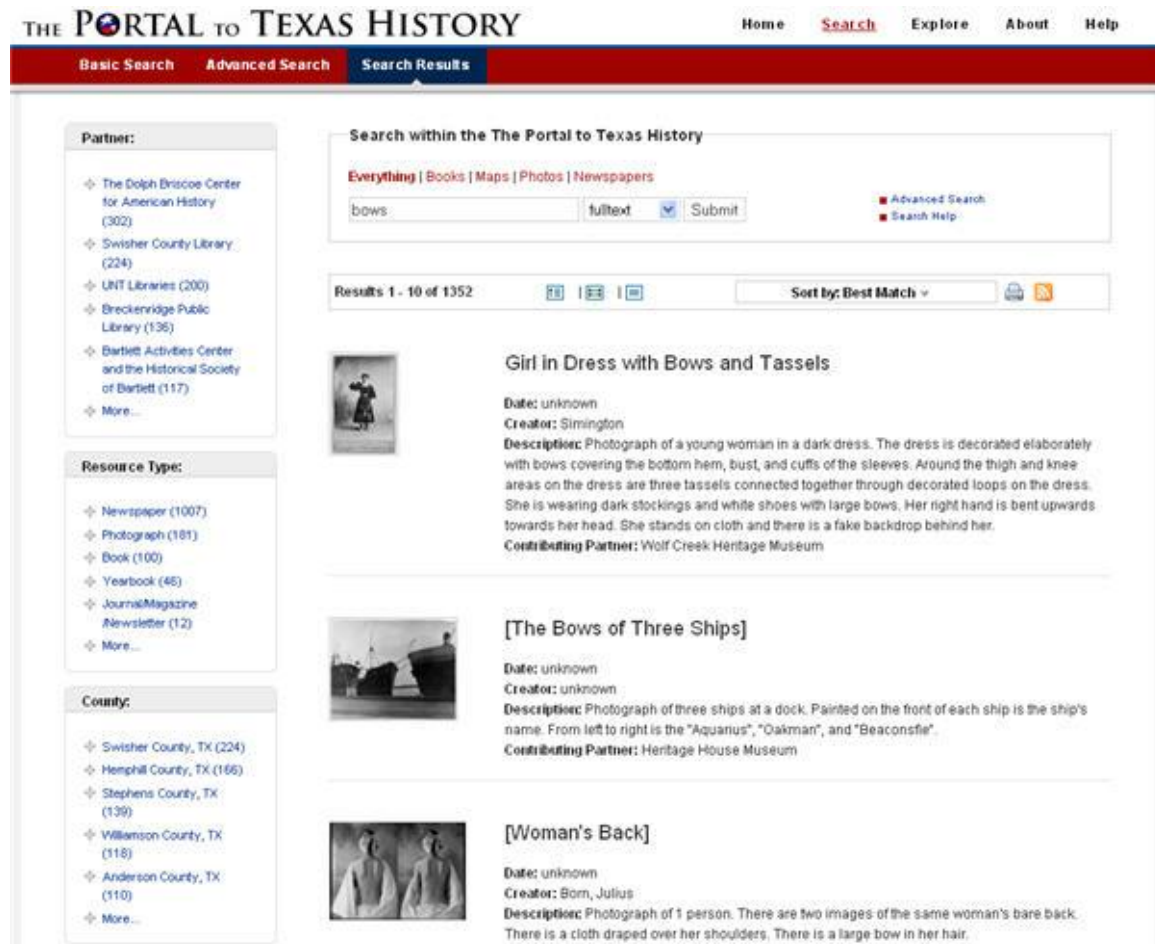


FIG. 6. Faceting search options appear on the left in the user interface of the UNT Libraries digital collections.

Faceted searching in our system has accomplished two things. First, it has created functionality for end users that was not previously available. Someone searching in our digital collections now has tools available to more easily find what he needs or to narrow a group of results to browse. Second, on the other side, creators and editors of metadata have concrete examples regarding the importance of metadata quality and why our guidelines need to be followed. Any of our partners who choose not to include information or to use non-compliant formatting will discover that faceted searching and similar system functionalities do not extend to their collections. Additionally, many editors are more willing to embrace our guidelines when they realize that our instructions were not written arbitrarily and that there is a direct relation to the way that the information is used in the system to meet the needs of information seekers.

Another aspect of functionality is that when metadata is used to a fuller extent, it becomes more apparent how much can actually be accomplished by tailoring system options to metadata that is already in system records. This can create a cycle of improvement (see Figure 7) in which every improvement (either to metadata or to the system that utilizes it) can fuel further

improvements, upgrades, and system functionality. Figure 7 helps to illustrate the way that system functionality is directly dependent on the quality of the metadata records that make up the database. For example, in the UNT Libraries system, faceted searching only works when we maintain consistent formatting and widespread use of the fields that are faceted. In some cases objects will be lost to users who choose to narrow their search with faceting since information about coverage dates and places, for example, are not known or entered in every record. But the potential for these tools is already driving partners to improve their metadata.

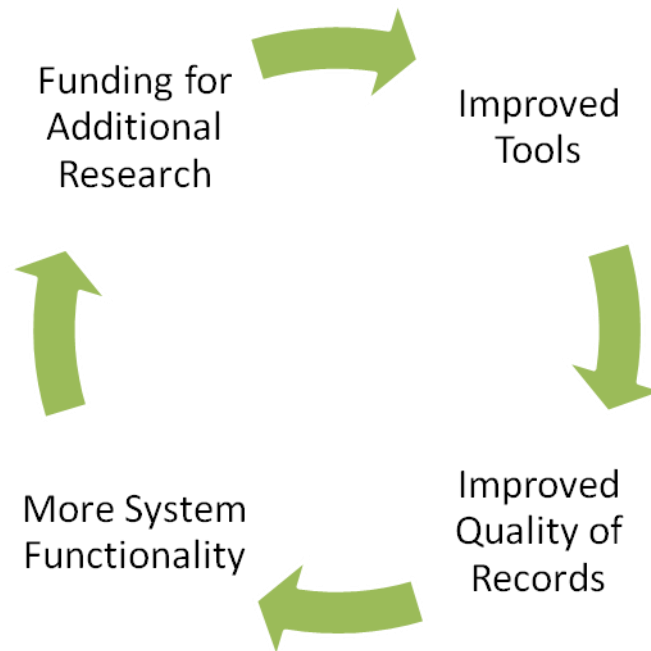


FIG. 7. Cycle showing how good metadata affects system functionality when it is utilized.

4. Conclusion

At the UNT Libraries, we are actively seeking ways to improve the metadata in our systems, starting at the point of creation. Facilitating quality metadata starts with quality resources that support the needs of metadata creators. When we revised our Metadata Input Guidelines, we discovered how important it is to look at every aspect of the available resources, including clarity, completeness, and accessibility of the documentation. Even more essential, however, is the need to have a clear sense of the importance of good metadata as a system component rather than as an end in itself. Although a system may not support particular functionality, having a strictly standardized way of recording information makes development of functionality a possibility for the future. Thus, formatting guidelines are key in digital library systems. Similarly, we have found that it is much easier to gain compliance when we can show metadata creators the direct effect that their formatting has on search results in our system. Quality metadata is powerful because it allows designers to create systems that translate metadata information into usable functionality; promoting the creation of quality metadata and the possibilities that it provides will lead to new ways of finding information and better digital services.

References

- Bowles, Dorothy A. & Diane L Borden. (2004). *Creative editing* (4th ed.). Belmont, CA: Wadsworth.
- Bruce, Thomas R., & Diane I. Hillmann. (2004). The Continuum of metadata quality: Defining, expressing, exploiting. In Diane I. Hillmann & Elaine L. Westbrooks (Eds.), *Metadata in practice* (pp. 238-256). Chicago: American Library Association.
- Chen, Lois Mai, & Marcia Lei Zeng. (2006, June). Metadata interoperability and standardization - A study of methodology part I. *D-Lib Magazine*, 12(6). Retrieved February 21, 2010, from <http://www.dlib.org/dlib/june06/chan/06chan.html>.
- Duval, Erik, Wayne Hodgins, Stuart Sutton, & Stuart L. Weibel. (2002, April) Metadata principles and practicalities. *D-Lib Magazine*, 8(4), [pp.]. Retrieved February 21, 2010, from <http://www.dlib.org/dlib/april02/weibel/04weibel.html>.
- Qin, Jian, Xiaozhong Liu, Xia Lin, and Miao Chen. (2009). Using metadata for query refinement and recommendation. *Proc. Int'l Conf. on Dublin Core and Metadata Applications 2009* (pp.15-20). Seoul, Korea. Retrieved February 8, 2010, from <http://dcpapers.dublincore.org/ojs/pubs/index>.
- Redish, Janice. (2007). *Letting go of the words: Writing web content that works*. Canada: Elsevier Inc.
- UNT Libraries. (2009). Metadata Input Guidelines. Retrieved February 10, 2010, from <http://www.library.unt.edu/digitalprojects/metadata/descriptive-metadata>.

Appendix

This is an example of full guidelines for an element (the language field), which is part of the UNT Metadata Input Guidelines taken from our website.

Language



- [Element Name](#)
- [Definition](#)
- [Where Can the Language Information be Found?](#)
- [How Language Works in the Metadata Form](#)
- [How Should the Language be Filled in?](#)
- [Other Examples](#)
- [Comments](#)
- [Resources](#)

Element Name

Language

Definition

The language(s) of the intellectual content of the resource.

Where Can the Language Information be Found?

Language is determined by examining the item:

Item Types	Information Sources
For text:	■ title page ■ words of the text
For images	■ caption ■ visible words within the image (e.g. a sign in a photograph)
For maps	■ caption or legend ■ place/physical feature labels
For music scores	■ title ■ lyrics
For videos	■ title screen(s) ■ label on disc ■ container cover
For sound files	■ spoken/sung words ■ information on disc/cassette ■ container information
For computer files	■ title page/screen ■ text

How Language Works in the Metadata Form

Parts:

1. Language -- drop-down menu

Repeatable?

Yes - to include multiple languages, click 'Add' to repeat the field

How Should the Language be Filled in?

Guidelines	Examples
■ Choose the appropriate language(s) from the controlled vocabulary	eng - English
■ Include all relevant languages (do not include languages that are merely referenced or only appear as single words in text of another language)	eng - English ger - German
■ For visual images that do not have a textual caption or text shown in the image, choose "no language"	nol - No Language
■ If there are special circumstances or additional information about the language uses of the item, include it in the Note field	Language: eng - English Language: spa - Spanish Note: book is printed in two sections, the first in Spanish and the second in the English translation
■ If the language of the item is not on the controlled vocabulary list (or cannot be determined from available resources), choose "other" and include a note in the record. Notify the metadata administrators if the language is not on the list	Language: oth - Other Note: Language is unknown. Note: Text is in the Kutenai Indian language.
■ If the only text on the item consists of persons' names, "no language" can be chosen and the text can be described in the content description or a display note	Language: nol - No Language Note: The name "Bartimaeus" is carved into the top of the box.

Other Examples

French map

Language: fre - French

Laws of the Republic of Texas

Language: eng - English

Photograph of Berlin showing marquees and shop windows

Language: ger - German

Adding machine with a label saying, "Made in America"

Language: eng - English

Landscape photograph

Language: nol - No Language

Audio recording of testimony in Czech and German

Language: cze - Czech

Language: ger - German

Video in Hebrew with English subtitles

Language: heb - Hebrew

Language: eng - English

Note: Performed in Hebrew with English subtitles.

Article in Amharic with English translations

Language: oth - Other

Language: eng - English

Note: In Amharic with English translations in parallel columns.

Proceedings and abstracts in English and Spanish

Language: eng - English

Language: spa - Spanish

Note: Papers presented in Spanish or English with abstracts in English.

Recording of a music recital sung in Italian and French

Language: fre - French

Language: ita - Italian

Comments

- Because of the global nature of the Internet, use of this field is recommended.
- Preferred usage is to utilize a standard schema for language names as defined by ISO639-2:
 - Three letter language codes, followed optionally by a two-letter country code (taken from the ISO 3166 standard).
 - To simplify data entry, the system does not currently support the addition of a country code to the language code.

Resources

- [UNT Language Controlled Vocabulary](#)

Related content

[Metadata](#)

[Input Guidelines for Descriptive Metadata](#)

This page is maintained by [Hannah Tarver](#) — last modified Tuesday, July 13, 2010, 10:26 AM

Captured from: <http://www.library.unt.edu/digitalprojects/metadata/elements/language-1>