

Building a Framework of Metadata Change to Support Knowledge Management

Oksana L. Zavalina* and Priya Kizhakkethil†
University of North Texas College of Information
3940 North Elm, Denton, Texas 76203, United States
*oksana.zavalina@unt.edu
†priya.kizhakkethil@my.unt.edu

Daniel Gelaw Alemneh‡, Mark E. Phillips§ and Hannah Tarver¶
University of North Texas Libraries
1155 Union Circle, Denton, Texas 76203, United States
‡daniel.alemneh@unt.edu
§mark.phillips@unt.edu
¶hannah.tarver@unt.edu

Published 17 April 2015

Abstract. Evolving user needs and relevance require continuous change and reform. A good digital collection has mechanisms to accommodate the differing uses being made of the digital library system. In a metadata management context, change could mean to transform, substitute, or make the content of a metadata record different from what it is or from what it would be if left alone. In light of the evolving compliance requirements, this paper analyses the three most common types of change within metadata records as well as their subcategories and discusses the possible implications of such changes within and beyond the metadata records.

Keywords: Digital libraries; metadata change; metadata evaluation; metadata quality; digital curation; lifecycle management.

1. Introduction

As more and more cultural heritage institutions are managing digital resources, the extent and variety of digital content being managed is growing rapidly. In light of the growth and diversity of digital content and the ever increasing interest by users in access to digital content, many stakeholders try to prioritise digital contents' life cycle management improvement.

The goal of knowledge management is to support knowledge creation, assimilation, dissemination, and application. One of the central building blocks of digital collections, which serve as repositories of knowledge is

descriptive metadata that describes information objects to ensure that knowledge contained in them is discoverable by the potential users of this knowledge. A good digital collection is sustainable over time, which is to say its individual items are curated and actively managed during their entire lifecycle in both a trusted and cost effective manner.

As many digital repository systems develop, various metadata is needed to support wider requirements at various levels. Maintaining high quality metadata for digital objects requires a framework that provides the appropriate capability to modify and update existing metadata. To enable and guide effective metadata management, it is essential to understand patterns of change in metadata records for a particular collection or an item.

2. Review of Relevant Literature

A major requirement that informs digital repository and digital library development is the need for systems to enable the evolution of metadata and digital objects standards over time. With modern technology changing at an unprecedented rate, new requirements and expectations of users have highlighted the importance and necessity of a robust metadata environment in digital library. This

environment is important as digital libraries and repositories strive to fill a need in the modern information landscape.

There have been many rigorous investigations into the quality of various bibliographic records (e.g. Mason, 2007), cataloging records (e.g. Hider and Tan, 2008), database quality (e.g. Hill, 2008), and information quality (e.g. Adams, 2003). Additionally, many researchers have articulated the benefits of metadata quality (e.g. Alemneh, 2009; Chen et al., 2011; Ochoa and Duval, 2006, etc.). Metadata quality evaluation frameworks were developed and tested (e.g. Bruce and Hillmann, 2004; Moen et al., 1998). The basic value proposition of metadata and its quality is a non-contentious issue, because most of the above mentioned researchers believe that the quality of metadata can have significant impact in facilitating access, use, and long-term preservation to digital resources.

Metadata in large databases developed over time is influenced by various changes in the surrounding context. These environmental changes (Thornburg and Oskins, 2007) include the changes that national and international standards for record creation and controlled vocabularies are undergoing, the revisions to goals and metadata creation approaches of hosting and contributing institutions, growth in certain types/formats or subject matter of materials in repositories, and changes in the content and location of fluid materials such as websites and other electronic resources described by metadata records. As a result, in order to keep up with environmental changes, and to maintain or improve metadata quality, metadata records change over time.

In the library community, metadata change has been an important aspect of cooperative cataloging for many years. In particular, catalogers have edited existing bibliographic records obtained in bibliographic utilities such as Online Computer Library Center (OCLC), Research Library Network (RLIN), etc., and those created by other cataloging agencies — libraries, vendors of books and other information resources, etc. — for local use in order to better meet users' needs. In addition, catalogers have updated and improved the quality of existing metadata records to be shared between libraries by either editing the records themselves using the Minimal Level Upgrading OCLC Program, or Enhance OCLC Program, or submitting change requests to the staff of quality control units of the bibliographic utilities. Changes to metadata records to improve their quality are encouraged by the bibliographic utilities. This includes both bibliographic records that describe information objects and authority records that describe the entities that play role a in creation of information objects — persons, organisations,

meetings etc. — and topical or geographical subjects of these information objects. For example, participating libraries receive credits from OCLC based on the quantity and level of improvements made to the records. The staff of cataloging quality control units of bibliographic utilities also routinely edits records (e.g. by merging duplicate records for the same manifestation of information object) (OCLC, 2014a,b).

Some researchers (Stvilia et al., 2004; Stvilia and Gasser, 2008) pointed out the link between metadata quality and metadata change. They highlighted the importance of quantifying the change to metadata records to assist in optimising quality assurance processes and suggested that metadata change should be justified based on changes of value and cost of metadata to provide justification for spent resources. However, the analysis of literature demonstrates virtual absence of research into metadata change. A number of information science studies relied on Wikipedia's so-called "revision metadata" that documents who made a particular revision to the Wikipedia article and when as well as "rollbacks" — the process of restoring a database or program to a previously defined state — to detect vandalism (e.g. Alfonseca et al., 2013; West et al., 2010). Similarly, Yan and McLane (2012) discussed metadata management process for "revision metadata", including the edits, history, and tracking, made to spatial data and Geographic Information System (GIS) map figures. While using administrative metadata that documents revisions as a tool to answer other research questions, none of these studies focused on the changes made to metadata *per se* as opposed to information objects (e.g. Wikipedia articles) described by metadata.

Until 2014, none of the published information science studies measured metadata change, with an exception of a small-scale component in the broader study of collection-level metadata quality in the IMLS DCC aggregation (Zavalina et al., 2008). As part of this study, researchers conducted analysis of the changes made over a period of two years by digital collection developers to metadata records describing their collections in the IMLS DCC. The authors of that study which selected a metadata field as a unit of analysis found that *Subject*, *Audience*, *Size*, *Spatial Coverage* and *Temporal Coverage* metadata was changed the most frequently. However, IMLS DCC research team did not follow up on that study nor did they suggest a framework for categorising and evaluating metadata change. Another team of researchers (Tarver et al., 2014) recently explored characteristics of metadata change over a five-year period in the large database (UNT Digital Collections) that supports versioning of metadata records. Unlike the IMLS DCC study, this study selected entire

metadata record as a unit of analysis and focused on general aggregated quantitative characteristics of metadata change in the database consisting of almost 700,000 metadata records. This characteristics include comparative overall distribution of the frequency of metadata revisions (measured as the number of editing events per record) at different points of time, the comparative distribution of the number of metadata editors involved in record revisions, the changes in metadata record size (measured as the total number of characters) and access status (hidden or displayed to the users), and the changes in the quantitative indicator (0–1) of metadata record completeness measured as the degree to which metadata record includes at least one instance of each of the elements that are required in the local metadata scheme. The authors of that study found that almost 40% of metadata records in the system underwent one or more editing events over time, with substantial number of metadata revisions resulting in increased completeness of metadata records, and that while most of the edited metadata records increased in length, some of the editing events resulted in decreased metadata record length.

Outside of information science field, in general, and metadata quality area in particular, one can see discussion of change in relation to comparison of texts, strings, files, scripts, etc. For example, in computer science, file comparison tools are used for isolating differences between files, programs and applications, including different versions of the same file (Heckel, 1978). One of these tools, DIFF outputs the differences between earlier and later versions of the same file with the help of two commands — “patch” that displays the changes made per line for text files and “diff” that represents the changes required to transform the original file into the new file. A similar Unix-based utility COMM is considered less powerful than DIFF but best suited for use in scripts (Horwitz, 1990). The COMM tool allows comparing two files for both commonalities and distinctions; it outputs one file consisting of three columns: the first two columns contain lines that are unique to the first and second input file respectively, and the last column contains lines common to both input files. Unlike DIFF, COMM discards information about the order of the lines prior to sorting them. Another file comparison tool — the PRETTY DIFF JavaScript application — is a means to algorithmically differentiate between two files (Cheney, 2010).

The so-called edit distance — the number of transformations needed to arrive from one version of the code to another — is often measured as a means of calculating the degree of similarity between the versions. Several algorithms compute the edit distance for various entities such

as strings, ordered and unordered trees and graphs, and consider basic change operations (Bille, 2005). According to Maynard *et al.* (2007), the ontology change typology was created for the use of ontology developers and researchers; this change typology includes basic change operations such as Add or Delete (e.g. Delete superclass) and complex change operations (e.g. Add an entire subtree). Klein and Noy (2003, p. 5) list among the 80 basic change operations the Add, Remove, and Modify operations, specifying that Modify operation means that “an old value is replaced by a new value”. Ontologists have also developed a PROMPTDIFF tool for differentiating ontologies; this tool first reconstructs the basic change operations and then applies a set of rules to infer complex change operations (Noy *et al.*, 2004). The PROMPTDIFF allows for the detection of high-level changes, which provide richer semantics. For comparing the hierarchically-structured XML files — the syntax in which metadata, including digital library metadata describing information objects, is stored, transmitted, and exchanged — the DIFF XML tool was developed (e.g. Kher, 2004; Barton, 2010). The XSD DIFF tool was also recently proposed to identify differences in XML metadata (Gautam and Parimala, 2012).

Extensive work on metadata versioning done in computer science domain (e.g. the works reviewed by Soules *et al.*, 2003) contributes to creating more efficient ways of storing file system metadata versions in the systems of various kinds, including content management systems utilised by digital libraries and repositories. However, it does not offer the ways to meaningfully analyse the nature of change in metadata which describes information objects to support information discovery in digital libraries and repositories. As the importance of descriptive metadata in digital repositories and its underlying quality becomes more widely known and accepted, it is important for the terminology around the concepts of metadata change to be articulated in order to facilitate research and discussion in the area of metadata change.

3. Research Design

The project reported in this paper sought to develop a terminology for discussing and reporting metadata change. This was achieved through two steps. The first step involved drafting of the general hierarchical framework of metadata change categories that would be context-independent and that would apply to any digital repository or full-text digital library regardless of the type of information objects it contains (e.g. images, research papers, dissertations, historical letters, government publications, etc.), the metadata scheme used (e.g. Dublin

Core, Machine Readable Cataloging MARC, Metadata Encoding and Transmission Standard METS, Metadata Object Description Schema MODS, Visual Resource Association (VRA) Core, etc.), and/or content management platform used (e.g. DSpace, ContentDM, Fedora, Islandora, etc.). The second step involved testing the developed framework in the analysis of metadata change in the digital repository that includes metadata records versioning capability.

3.1. Drafting the general framework of metadata change

The authors of this paper identified three broad main categories (Fig. 1) that illustrate continuous change of metadata: (1) Add, (2) Delete, and (3) Modify. Both these three top categories and the subcategories of metadata change represented in Fig. 1 emerged from brainstorming by this paper’s authors and were mostly based on anecdotal evidence drawn from experience creating and maintaining library metadata in various systems, as well as on logical assumptions. The first two of these categories — Add and Delete — correlate with the basic change operations considered in calculating edit distance in computer science: insert node and delete node (e.g. Bille, 2005; Gao et al., 2009) as well as with major ontology change types: basic change operators of add and remove/delete (Klein and Noy, 2003; Maynard et al., 2007). In addition, they also incorporate more specific subcategories that are applicable for digital library/repository metadata. The third major category of metadata change in our initial framework of

metadata change, Modify, overlaps, but only to some extent, with another basic change operation discussed in computer science literature: the one referred to as modify in relation to change in ontologies (e.g. Klein and Noy, 2003) or as substitute node (Bille, 2005; Gao et al., 2009) in the context of change in other types of files.

The initial framework was created with two levels of metadata change analysis in mind: the level where the unit of analysis is the entire metadata record and the level where the unit of analysis is an individual metadata element and/or a data value associated with it. We assumed that there were three major entities in metadata record that could be affected by change:

- Metadata elements represented as fields in the database (e.g. Description or — if expressed in XML syntax — `<description>`),
- Qualifiers or attributes of these fields (e.g. `<description qualifier = “physical”>`, `<description display = “hidden”>` etc.), and
- Data values that are associated with these elements/fields and constitute the actual representation of an information object’s characteristics or content (e.g. “10 × 12 in.”, “includes color map”).

Based on these three dimensions, we initially identified three subcategories under Add and Delete change categories. The first subcategory concerned metadata elements and included (1) adding a metadata element that was not previously part of the metadata record or deleting a metadata element that was previously part of the metadata record. The second subcategory concerned qualifiers or attributes of a metadata element and included (2) add or delete qualifier or attribute of a metadata and included element. The third subcategory concerned data values and included (3) add or delete the data value associated with an element or an instance of a field containing this data value. This subcategory took into account that depending on the information retrieval system and/or metadata scheme in use, multiple data values associated with the same element/field in the same metadata record could be

- Either enumerated together in a single instance of a metadata element/field (as in “Physical Description: 10 × 12 in.; includes color maps”) or
- Housed in separate instances of the same element/field (as in “`<description qualifier = “physical”>` 10 × 12 in. `</description>`; `<description qualifier = “physical”>` includes color maps `</description>`”).

In our initial general repository-independent framework of metadata change, the Modify top change category included three major varieties of modification — (1)

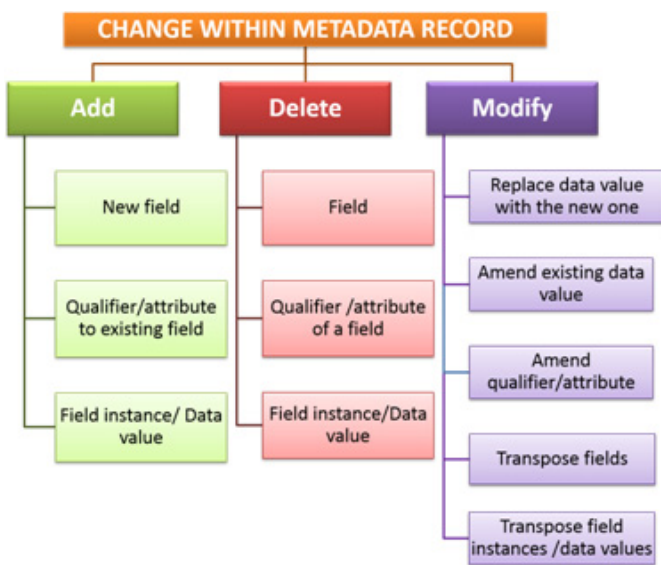


Fig. 1. The most common changes within metadata records: Initial general framework.

Replacement, (2) Amendment, and (3) Transposition — further explained below. Authors’ metadata creation and/or management experiences suggested that the Replacement category would be applicable to data values, as in the modification from `<description qualifier = “physical”>` includes color maps `</description>` to `<description qualifier = “physical”>` in blue cardboard container `</description>`. We also expected that amendments would apply to both data values and qualifiers or attributes, as in the modification from `<description qualifier = “physical” display = “hidden”>` includes color maps `</description>` to `<description qualifier = “physical” display = “unhidden”>` includes 3 folded color maps `</description>`, where both the attribute value of element attribute “display” and the data value are amended. We also expected the Transposition subcategory of Modify to apply to two dimensions of metadata record: (1) data values or element instances (e.g. a modification from `<description qualifier = “physical” display = “unhidden”>` includes color maps `</description>` to `<description display = “unhidden” qualifier = “physical”>` includes color maps `</description>` or a modification from “Physical Description: 10 × 12 in.; includes color maps” to “Physical Description: Includes color maps; 10 × 12 in.”) and (2) the elements themselves when the order of elements in the record changes (e.g. when in the earlier version of the metadata record the Physical Description field is immediately preceded by the Genre field and in the later version of the same record the Physical Description field appears before the Genre field. This resulted in five subcategories of Modify change category in our initial framework of metadata change (Fig. 1).

We assumed that Modify metadata change category would include both manual and automated processes and also the use of local or external sources such as controlled vocabularies. We also expected that Amendment subcategory would be the most commonly occurring type of modification that would include, among other changes, corrections of misspellings, capitalisation and punctuation changes, and change from natural language keywords to their controlled-vocabulary equivalents.

3.2. Framework testing

Our next step was to test the framework in an analysis of a random stratified sample of metadata records from a digital library system that archives versions of metadata records. We selected for our testing 157 metadata records from multiple digital collections in a centralized digital library hosted by the University of North Texas (UNT) Libraries — the UNT Digital Collections. The UNT Digital Collections include the UNT Digital Library, containing items owned by UNT and the output of the University’s

research, creative, and scholarly activities; The Portal to Texas History, containing historical materials owned by over 200 partner institutions across the state of Texas; and the Gateway to Oklahoma History, containing primarily newspapers and photographs through partnership with the Oklahoma Historical Society. The UNT Digital Collections incorporate different types of materials including theses, dissertations, artwork, performances, musical scores, journals, government documents, rare books and manuscripts, newspapers, photographs, and historical posters. All items in the UNT Digital Collections are described using a locally-modified Dublin Core metadata scheme.

The metadata records initially created and made visible to the end users between October 2009 and December 2012 and last modified — with retaining the same record access status of “unhidden” — between January and April of 2014 were selected for analysis. For the purposes of this exploratory study, we focused on comparing initial and latest (as of April 2014) versions of these metadata records and ignored the intermediate versions. Two researchers analysed the records in their native XML format. The coding was based on the initial framework of metadata change (Fig. 1). In the process of analysis, additional, more granular metadata change subcategories emerged.

Some of the elements/fields in the UNT Digital Collections metadata records include subelements/subfield (for example, name type, affiliation, and name subfields for Creator and Contributor fields). This, however, might not be the case in some other digital repositories, so inclusion of changes in subfields as separate subcategories of Add, or Modify in the repository-independent framework of metadata change is not justified. Therefore, we interpreted the addition, deletion, or modification of subfields in the same way as addition, deletion or modification of fields.

The metadata records in UNT Digital Collections included up to twenty one descriptive metadata elements which describe the information object and its various attributes, and several administrative metadata elements which describe the metadata record itself (e.g. the date and time of creation and revision, the names of metadata specialists who created and revised the record, etc.) and are for the most part generated automatically. For the purposes of this study, which focuses on change of descriptive metadata, the administrative metadata fields that appear at the end of each record were excluded from analysis.

4. Findings and Discussion

Figure 2 shows twenty one descriptive metadata elements in the order in which they appear in the UNT Digital

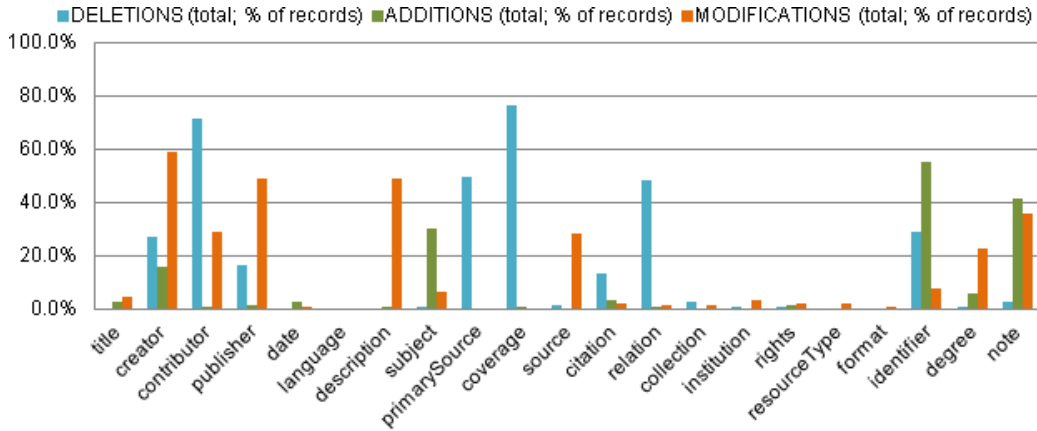


Fig. 2. Relative frequency of distribution of major metadata change types, % of records ($n = 157$).

Collections metadata records. Our analysis (Fig. 2) revealed that almost a quarter of descriptive metadata elements — Coverage, Contributor, Identifier, Creator, and Publisher — changed in more than a half of analysed metadata records. On the other hand, over a third (8) of descriptive metadata elements — Title, Date, Collection, Institution, Rights, Resource Type, Format, and Language — contained changes in only 10% or less of the records in the sample. At the same time, a number of records in the sample exhibited multiple changes. For eleven metadata elements, we observed more than one change type or subcategory in the same record. For example, Creator and Note metadata elements contained multiple change types in more than a third of records; Contributor and Identifier elements contained multiple change types and subcategories in approximately a quarter of metadata records each.

Of the three major metadata change types, modifications occurred the most often, in 16 descriptive metadata elements out of 21 (Fig. 2), followed by deletions (11 elements), and additions (10 elements). Three metadata elements — Creator, Publisher, and Description — were the most modified elements, with modifications in these elements found in more than 40% of records. Four metadata elements — Coverage, Contributor, Primary Source, and Relation — underwent the most deletions (more than 40% of records each). Additions most often occurred in Identifier (55.41% of records), Note (41.4%), and Subject (29.94%) metadata elements.

We have also observed that records in the sample exhibited the change in the number of instances of one or more descriptive metadata elements: in most cases through addition of new instances, but some deletions were also observed. Almost 55% of records in the sample exhibited the change between the initial and latest edited record version in the number of instances of Identifier element,

over 40% in the number of instances of Note element, almost 32% in the number of instances of Subject, and almost 12% in the number of instances of Citation element.

4.1. Addition

In the UNT Digital Collections, addition occurred the least often among the three major types of metadata change. In our analysis of the records, we identified the same three subcategories of addition that were initially included in the draft framework of metadata change: addition of a data value, addition of an element qualifier or attribute, and addition of a new field or subfield that had not been included in the initial version of the metadata record. Figure 3 shows the frequency of occurrence of each of the subcategories of addition for each of the elements in the records. Overall, only two metadata elements out of twenty one — Identifier and Note — exhibited all three subcategories of addition in the analysed dataset.

Addition of a new field or subfield not present in the original version of metadata record was observed in nine descriptive metadata elements. However, only one metadata element — Format — exhibited this kind of metadata change in more than 20% of analysed metadata records. An example of this type of addition can be seen in the Date field in the metadata record for a photograph (Fig. 4). The initial version of the record in the left half of this figure has only one Date element — the Date of Digitisation — while the latest version of the same record in the right half of the figure has both the Date of Digitisation (interestingly, repeated twice, which creates unnecessary redundancy) and the Date of Creation. Similarly, another example of a metadata record — for a conference paper (Fig. 5) — has only one Title element (the Official Title, as can be seen from field qualifier) while the latest version of the same record has both the Official Title and the Series

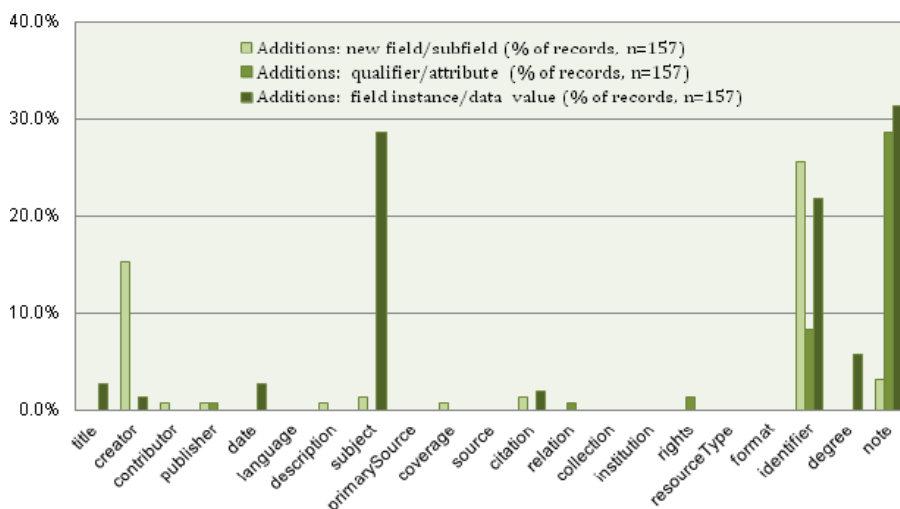


Fig. 3. Addition subcategories, % of records (n = 157).



Fig. 4. Metadata record example 1: Initial version (left) and latest edited version (right).

Title. One can also see an example of a subfield addition in Fig. 5: in the latest version of the metadata record, all three instances of Creator field include a new Info subfield which holds information about creator’s affiliation and which was not included for Creator element in the initial version of the same metadata record.

Addition to data values was observed the most frequently among all subcategories of addition, in eight

metadata elements: Title, Creator, Date, Subject, Citation, Identifier, Degree, and Note. Three metadata elements — Note, Identifier, and Subject — exhibited addition in the data value in more than 20% of the records each. In the example of metadata record for conference proceedings paper (Fig. 5), in Degree field which holds information about submitting units of the university, the third field instance containing the third data value — “Digital

<pre> 3 <title qualifier="officialtitle">The DataRes Research Project on Data Management</title> 4 <creator qualifier="aut"> 5 <typeper</type> 6 <name>Halbert, Martin </name> 7 </creator> 8 <creator qualifier="aut"> 9 <typeper</type> 10 <name>Moen, William E.</name> 11 </creator> 12 <creator qualifier="aut"> 13 <typeper</type> 14 <name>Keralis, Spencer</name> 15 </creator> 16 <contributor qualifier="orm"> 17 <info>Faculty of Information</info> 18 <typeorg</type> 19 <name>University of Toronto</name> 20 </contributor> 21 <publisher> 22 <location>[New York, New York]</location> 23 <name>Association for Computing Machinery (ACM)</name> 24 </publisher> 25 <date qualifier="creation">2012-02</date> 26 <language>eng</language> 27 <description qualifier="content">This article discusses data management. The University of North Texas Library and Information Resources, have received \$226,786 from the Institute of Museum and Library Ser project to investigate how the library and information science profession can best respond to emerging management in universities. This project will address broad new issues concerning the emerging roles, arising from requirements announced by NIH, NSF, IMLS and other funding agencies for data management p description 28 <description qualifier="physical">3 p.</description> 29 <subject qualifier="KWD">research </subject> 30 <subject qualifier="KWD">data management</subject> 31 <subject qualifier="KWD">policies</subject> 32 <subject qualifier="KWD">management</subject> 33 <subject qualifier="KWD">documentation</subject> 34 <subject qualifier="KWD">security</subject> 35 <subject qualifier="KWD">legal aspects</subject> 36 <subject qualifier="KWD">verifications</subject> 37 <primarySource/> 38 <coverage/> 39 <source qualifier="conference">Proceedings of the iconference, 2012, Toronto, ontario, Canada</source> 40 <citation/> 41 <relation/> 42 <collection>UNTSW</collection> 43 <institution>UNT</institution> 44 <rights qualifier="access">public</rights> 45 <resourceType>text_article</resourceType> 46 <format>text</format> 47 <identifier qualifier="DOI">10.1145/2132176.2132300</identifier> 48 <degree qualifier="department">Libraries</degree> 49 <degree qualifier="department">Library and Information Science</degree> 50 <note/> </pre>	<pre> 3 <title qualifier="officialtitle">The DataRes Research Project on Data Management/< 4 <title/> 5 <title qualifier="seriestitle">The DataRes Project</title> 6 <creator qualifier="aut"> 7 <info>University of North Texas</info> 8 <typeper</type> 9 <name>Halbert, Martin</name> 10 </creator> 11 <creator qualifier="aut"> 12 <info>University of North Texas</info> 13 <typeper</type> 14 <name>Moen, William E.</name> 15 </creator> 16 <creator qualifier="aut"> 17 <info>University of North Texas</info> 18 <typeper</type> 19 <name>Keralis, Spencer D. C.</name> 20 </creator> 21 <contributor qualifier="orm"> 22 <typeorg</type> 23 <name>University of Toronto. Faculty of Information Studies.</name> 24 </contributor> 25 <publisher> 26 <location>[New York, New York]</location> 27 <name>Association for Computing Machinery</name> 28 </publisher> 29 <date qualifier="creation">2012-02</date> 30 <language>eng</language> 31 <description qualifier="content">Paper for the 2012 International iconference. This presentation discusses data management and the DataRes Project. </description> 32 <description qualifier="physical">3 p.</description> 33 <subject qualifier="KWD">research data management</subject> 34 <subject qualifier="KWD">policies</subject> 35 <subject qualifier="KWD">documentation</subject> 36 <subject qualifier="KWD">security</subject> 37 <subject qualifier="KWD">legal aspects</subject> 38 <subject qualifier="KWD">verification</subject> 39 <source qualifier="conference">iconference, 2012, Toronto, Ontario, Canada</source> 40 <relation qualifier="referenced-by">The DataRes Research Project on Data Management: ark:/67531/metadc77152</relation> 41 <collection>UNTSW</collection> 42 <institution>UNT</institution> 43 <rights qualifier="access">public</rights> 44 <resourceType>text_paper</resourceType> 45 <format>text</format> 46 <identifier qualifier="DOI">10.1145/2132176.2132300</identifier> 47 <degree qualifier="department">Libraries</degree> 48 <degree qualifier="department">Library and Information Science</degree> 49 <degree qualifier="department">Digital Scholarship Co-Operative</degree> 50 <note qualifier="display">Abstract: The University of North Texas together with the Council on Library and Information Resources, have received \$226,786 from the Institute of Museum and Library Services for a two year research project to investigate how the library and information science profession can best respond to </pre>
---	--

Fig. 5. Metadata record example 2: Initial version (left) and latest edited version (right).

Scholarship Co-Operative” — was added to the two data values initially provided in two instances of this field. Similarly, in the example of metadata record for a dissertation (Fig. 12), several field instances of a Subject field, containing the terms from the Library of Congress Subject Headings controlled vocabulary, were added.

Addition of a qualifier or an attribute to a field that had not originally included any qualifiers or attributes was observed less often than the two other types of addition, in five descriptive metadata elements: Publisher, Relation, Rights, Identifier, and Note. Only one descriptive metadata element, Note, exhibited addition of a new qualifier or attribute in more than 20% of records. In the metadata record shown in Fig. 5, the Relation field which had no qualifiers in initial version of the record received a qualifier Referenced By in the latest version of the record.

4.2. Deletion

Deletion, the second major type of metadata change, occurred somewhat more often overall than addition. In the process of analysis, we identified three subcategories of deletion: deletion of a field or subfield, deletion of an element’s qualifier or attribute, and deletion of a data value. For the first subcategory of deletion, our analysis revealed

two kinds: deletion of a previously empty field or subfield and deletion of a previously populated field or subfield.

As shown in Fig. 6, deletion of a field or subfield occurred in a variety of elements (a total of 12) and in high proportion of records overall. Most of the field or subfield deletions happened when the field or subfield that had been empty in the initial version of metadata record was removed from the edited version of metadata record. In most of the records analysed in this study, Coverage, Contributor, Primary Source, and Relation metadata elements had been empty fields in initial versions of the records, and these four elements also were the ones with the highest number of occurrences of deleted empty field/subfield (48.4–76.4% of metadata records each). Both examples of metadata records presented above (Figs. 4 and 5) underwent deletions of initially empty or unpopulated fields. For instance, in the record for a photograph (Fig. 4), the initially unpopulated Primary Source field and two out of three instances of a Coverage field — the ones with sDate and eDate qualifiers for temporal coverage — were removed in the process of metadata record revision and are not present in the latest version of the record. In the record for a conference paper (Fig. 5), three fields that had been empty in the initial version — Primary Source, Coverage, and Citation — were deleted. Similarly, in the record for

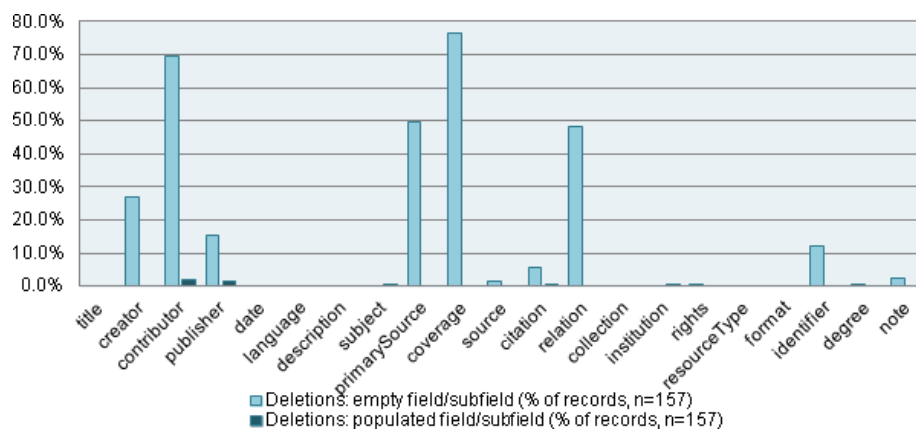


Fig. 6. Deletion of field/subfield, % of records ($n = 157$).

a dissertation (Fig. 12), the empty Coverage field was removed.

Deletions of initially populated fields or subfields occurred infrequently and were only observed for five metadata elements: Contributor, Publisher, Subject, Citation, and Institution. One can see examples of such deletion in the record for a conference paper in Fig. 5. The second instance of a Contributor field in the initial version of this record — the one holding the data value for Faculty of Information at University of Toronto — was deleted and is not present in the latest version of the same record.

As shown in Fig. 7, deletion of a data value was found in only two descriptive metadata elements and in a moderate proportion of metadata records in the sample: Citation (7%) and Collection (2.5%). In the metadata

record in Fig. 4, the first of the two data values in Collection field — UNTCVA — included in the original version of the record was subsequently deleted.

Deletion of an element’s qualifier/attribute (Fig. 8) was observed more often overall (16.6% of records in the sample) but only in one metadata element: Identifier. In the third example of a revised metadata record, the record for a standard (Fig. 9), the first instance of Identifier field had had a qualifier LOCAL-CAT-Number in the initial version of the record which was removed in the process of revision.

As can be seen in Figs. 6–8, none of the individual metadata elements exhibited all three subcategories of deletion, and only two descriptive metadata elements — Citation and Identifier — exhibited two deletion subcategories. In Citation element, data value deletion and field/subfield deletion

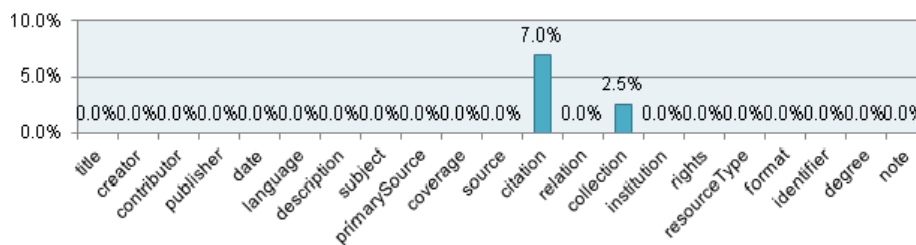


Fig. 7. Deletion of data value, % of records ($n = 157$).

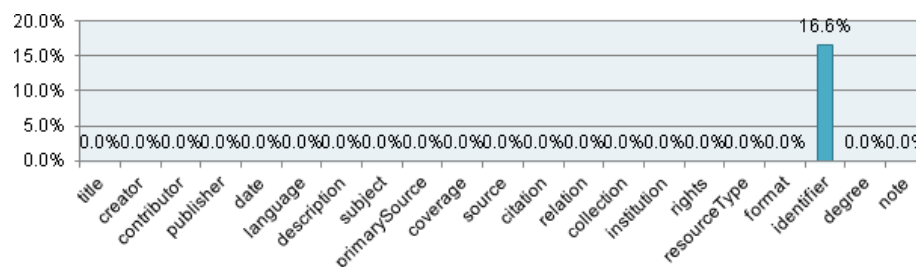


Fig. 8. Deletion of element qualifier/attribute, % of records ($n = 157$).

Fig. 9. Metadata record example 3: Initial version (left) and latest edited version (right).

were observed; in Identifier element, element qualifier deletion and field/subfield deletion were observed.

4.3. Modification

Most of the records in the sample underwent modifications in one or more metadata element. Our analysis revealed four major subcategories of modification: populating previously empty field or subfield (not on the draft framework of metadata change), amendment, replacement, and transposition. Figures 10, 11, 13 and 14 show the frequency of occurrence of each of the subcategories of modification for each of the descriptive metadata elements in the records.

Some of the empty fields and/or subfields received their data values in the process of editing over time (Fig. 10). The Note was the element where this kind of metadata

change occurred the most often (in 28.7% of all analysed records); the Identifier empty field was also supplied with the data value quite frequently (7.6% of the records). For ten additional metadata elements, this kind of metadata change was observed in a small proportion of records (between 0.6% and 1.3%). The example of metadata record for a conference paper (Fig. 5) illustrates this subcategory of a metadata change: the Relation and Note fields which were empty in the initial version of the record, received data values in the final version of the same record.

4.3.1. Replacements

Complete replacement of a data value, field's qualifier or attribute with the new one is a subcategory of metadata change proposed in our framework that represents a

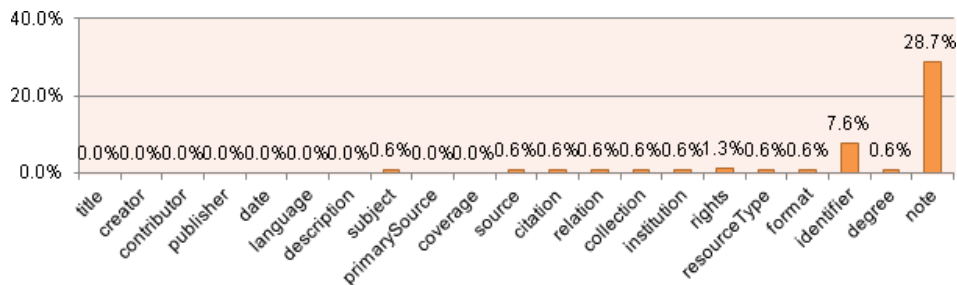


Fig. 10. Populating previously empty field/subfield, % of records ($n = 157$).

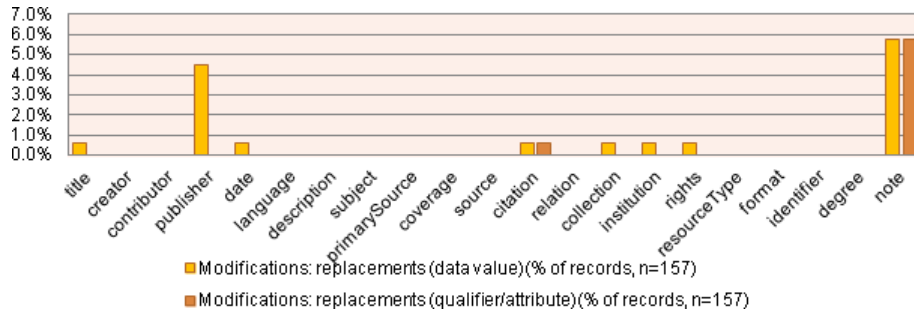


Fig. 11. Replacements, % of records (n = 157).

metadcl1003.untl.xml.0 x	metadcl1003.untl.xml x
2 <metadata>	2 <metadata>
3 <title qualifier="officialtitle">The Assimilation of Baroque Elements in Ferruccio Busoni's Compositions as Exemplified by the Fantasia nach Bach and the Toccata</title>	3 <title qualifier="officialtitle">FPGA Implementation of Low Density Parity Check Codes Decoder</title>
4 <creator qualifier="aut">	4 <creator qualifier="aut">
5 <name>Vijayakumar, Suresh</name>	5 <type>per</type>
6 <info/>	6 <name>Vijayakumar, Suresh</name>
7 </creator>	7 </creator>
8 <contributor qualifier="cha">	8 <contributor qualifier="cha">
9 <name>Mikler, Armin R.</name>	9 <info>Major Professor</info>
10 <info>Major Professor</info>	10 <type>per</type>
11 <type>per</type>	11 </contributor>
12 </contributor>	12 <contributor qualifier="cha">
13 <contributor qualifier="cha">	13 <info>Major Professor</info>
14 <name>Fu, Shengli</name>	14 <type>per</type>
15 <info>Major Professor</info>	15 </contributor>
16 <type>per</type>	16 <contributor qualifier="cmr">
17 </contributor>	17 <type>per</type>
18 <contributor qualifier="cmr">	18 <name>Sweany, Philip</name>
19 <name>Sweany, Philip</name>	19 <info/>
20 <info/>	20 <type>per</type>
21 <type>per</type>	21 </contributor>
22 </contributor>	22 <publisher>
23 <publisher>	23 <location>Denton, Texas</location>
24 <name>University of North Texas</name>	24 <name>University of North Texas</name>
25 <info/>	25 </publisher>
26 <location>Denton, Texas</location>	26 <date qualifier="creation">2009-08</date>
27 </publisher>	27 <date qualifier="digitized">2009-10-16</date>
28 <date qualifier="creation">2009-08</date>	28 <language>eng</language>
29 <date qualifier="digitized">2009-10-16</date>	29 <description qualifier="content">Reliable communication over the noisy channel has become one of the major concerns in the field of digital wireless communications. The low density parity check codes (LDPC) has gained lot of attention recently because of their excellent error-correcting capacity. It was first proposed by Robert G. Gallager in 1960. LDPC codes belong to the class of linear block codes. Near capacity performance is achievable on a large collection of data transmission and storage. In my thesis I have focused on hardware implementation of (3, 6) - regular LDPC codes. A fully parallel decoder will require too high complexity of hardware realization. Partly parallel decoder has the advantage of effective compromise between decoding throughput and high hardware complexity. The decoding of the codeword follows the belief propagation alias probabilistic propagation algorithm in log domain. A 9216 bit, (3, 6) regular LDPC code with code rate 1/89; was implemented on FPGA targeting Xilinx Virtex 4 XC4VLX80 device with package FF1148. This decoder achieves a maximum throughput of 82 Mbps. The entire model was designed in VHDL in the Xilinx ISE 9.2 environment.</description>
30 <language>eng</language>	30 <subject qualifier="KWD">belief propagation</subject>
31 <description qualifier="content">Reliable communication over the noisy channel has become one of the major concerns in the field of digital wireless communications. The low density parity check codes (LDPC) has gained lot of attention recently because of their excellent error-correcting capacity. It was first proposed by Robert G. Gallager in 1960. LDPC codes belong to the class of linear block codes. Near capacity performance is achievable on a large collection of data transmission and storage. In my thesis I have focused on hardware implementation of (3, 6) - regular LDPC codes. A fully parallel decoder will require too high complexity of hardware realization. Partly parallel decoder has the advantage of effective compromise between decoding throughput and high hardware complexity. The decoding of the codeword follows the belief propagation alias probabilistic propagation algorithm in log domain. A 9216 bit, (3, 6) regular LDPC code with code rate 1/89; was implemented on FPGA targeting Xilinx Virtex 4 XC4VLX80 device with package FF1148. This decoder achieves a maximum throughput of 82 Mbps. The entire model was designed in VHDL in the Xilinx ISE 9.2 environment.</description>	31 <subject qualifier="KWD">LDPC codes</subject>
32 <subject qualifier="KWD">belief propagation</subject>	32 <subject qualifier="KWD">FPGA</subject>
33 <subject qualifier="KWD">LDPC codes</subject>	33 <subject qualifier="LCSH">wireless communication systems.</subject>
34 <subject qualifier="KWD">FPGA</subject>	34 <subject qualifier="LCSH">telecommunication systems.</subject>
35 <coverage qualifier="sDate"/>	35 <subject qualifier="LCSH">field programmable gate arrays.</subject>
36 <coverage qualifier="eDate"/>	36 <subject qualifier="LCSH">Decoders (Electronics)</subject>
	37 <collection>UNTETD</collection>

Fig. 12. Metadata record example 4: Initial version (left) and latest edited version (right).

semantic match to basic change operators Modify and Substitute node discussed in computer science literature (Bille, 2005; Gao *et al.*, 2009; Klein and Noy, 2003). This type of metadata change was observed in eight descriptive metadata elements in our study: Title, Contributor, Publisher, Date, Citation, Collection, Institution, Rights, and Note (Fig. 11). Replacements of qualifiers or attributes were observed in only two out of these eight descriptive metadata elements: Note and Citation. The qualifiers or attributes of a Note field underwent replacements the most frequently (over 5% of all analysed records). The data values of Note and Publisher fields

underwent replacements the most frequently: in over 5% of all analysed records for Note, and in over 4% of all analysed records for Publisher. In the example of a metadata record for a dissertation (Fig. 12), the initial erroneous data value in the Title field, “The Assimilation of Baroque Elements in Ferruccio Busoni’s Compositions as Exemplified by the Fantasia nach Bach and the Toccata” was replaced with a completely new correct one, “FPGA Implementation of Low Density Parity Check Codes Decoder”. The example of a metadata record for a conference paper above (Fig. 5) provides another illustration for a common case of replacement: the initial

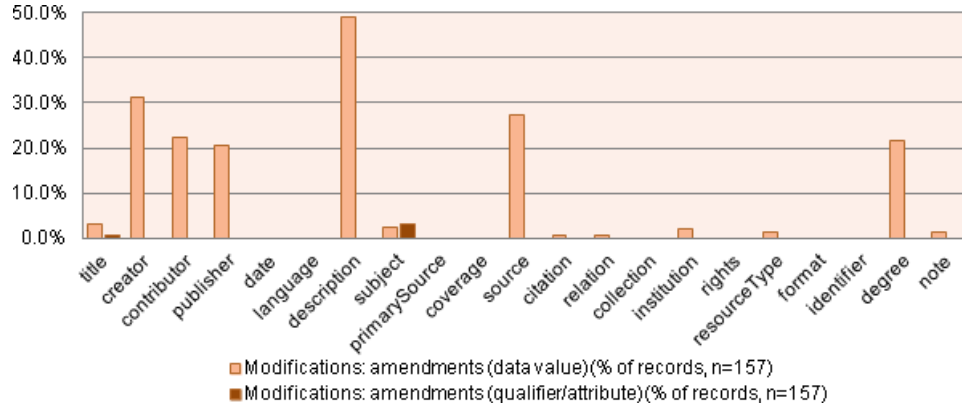


Fig. 13. Amendments, % of records ($n = 157$).

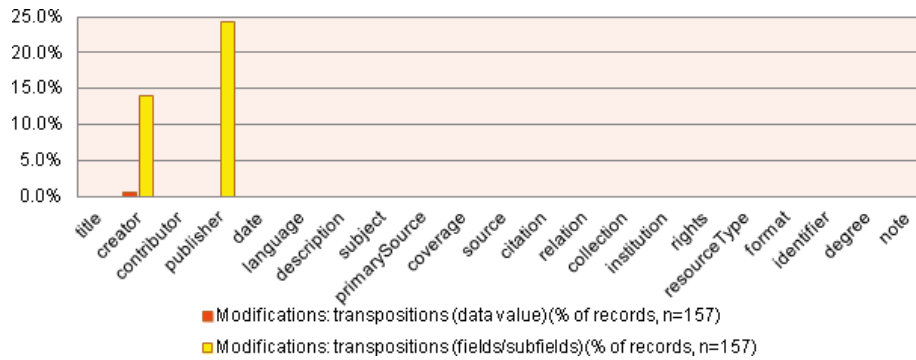


Fig. 14. Transpositions, % of records ($n = 157$).

detailed data value in Description field was moved to another field (Note), and the new, brief data value replaced it.

4.3.2. Amendments

Amendments — modifications that are less drastic than complete replacements and that do not have a direct semantic match among basic change operators discussed in computer science literature — were observed substantially more often than two subcategories of modifications discussed above: Populating of previously empty fields and replacements. They were also observed more often than the transpositions discussed in the next subsection, thus supporting authors’ assumption that amendment is the most common type of metadata modification. We identified two kinds of amendments: amendments to the data value and amendments to the element’s qualifier or attribute. Amendments to data value were observed (Fig. 13) in a total of twelve metadata elements out of 21. The Description field demonstrated this kind of amendment the most frequently, in almost 50% of all analysed records. The data values of five more elements — Creator,

Contributor, Publisher, Source, and Degree — were amended in at least 20% of all analysed records.

The metadata records for various types of information objects provided illustrations for this subcategory of modification. For instance, in the example record for a conference paper (Fig. 5), the data values in two fields — Creator and Publisher — were amended. The name of the second creator, Spencer Kerulis, was edited to result in what looks like the authorised form of name from a controlled vocabulary such as the Library of Congress Name Authority File (although the source of the heading, or the name of controlled vocabulary, is not identified in the qualifier of a Creator field). In this same metadata record, the data value for Publisher field initially included the acronym in parenthesis after the full name; the acronym was removed in the process of editing and is no longer part of the data value.

4.3.3. Transpositions

Finally, we observed one more subcategory of modification which did not have a counterpart among basic change operators discussed in computer science literature — a

transposition when the data values within the same field or the instances of a field or subfield are rearranged and the order in which they appear is changed between the initial and the subsequent version of the metadata record. This subcategory of metadata change was observed in only two metadata elements: Creator and Publisher (14% and 24% of all the analysed records, respectively). The vast majority of this kind of metadata change occurred with subfields; data values were transposed much less often (Fig. 14).

Examples of transpositions can be seen in the metadata record examples above. For instance, one of the typical transpositions was moving the Type subfield with its value “per” which stands for a personal name in Creator and Contributor fields from the position after the Name subfield to the position before the Name subfield and/or moving up the Info subfield with its data value within the Creator or Contributor field (Fig. 12). We did not observe the transpositions of the top-level fields themselves or changing the order in which the fields appear in the record (e.g. Primary Source field appearing after the Subject field in initial versions of the records and changing its position to appear after the Coverage field in later versions of the records); the order of fields in the records was consistent. We did however observe rearrangements in the order of the multiple creators or contributors for a paper; these represented transposition of data values encoded in multiple field instances of the same field.

We did not encounter any instances of transposition between multiple qualifiers, between multiple attributes, or between a qualifier and an attribute of a metadata element. This can be partially explained by the nature of element qualifiers which are used to further specify the metadata element (e.g. to say that the type of the Date in question in the record for journal article is the Date Submitted; it does not make sense to say that the date in question is both Date Submitted and Date Published). However, the nature of the XML attributes of a metadata element is quite different, and often multiple attributes can be used at the same time, with the same instance of the same metadata element:

- The encoding attribute (allows to specify the controlled vocabulary the term is taken from or — as in case with date — the standard according to which the date is formatted),
- The language attribute (allows to specify the language in which the data value is represented),
- The display attribute (allows to specify whether the element is displayed to the end-user of the metadata record or is only visible to metadata creators)

- The qualifier itself (this attribute is normally called type), and
- Other attributes depending on the metadata scheme used and the specific metadata element in this scheme.

Therefore, although the cases of transposition of XML attributes were not observed in the analysis of a sample of UNT Digital Collections metadata records, which contained only one kind of an attribute — display — this type of change can occur in metadata records from other repositories that use other metadata schemes that make more extensive use of XML attributes (e.g. MODS).

4.4. *Final thoughts*

During this research, the authors identified a number of issues with the analysis of metadata records using the native XML format. The first was different ways of notating and interpreting “empty” elements. This was noted during the analysis of the refined general framework. In the metadata samples from the UNT Digital Collections, a common workflow includes the possibility of metadata editors adjusting metadata records by either a Web form or by adjusting the XML files directly. Because of this, when records are initially written, all possible fields are included but without value. This results in many “missing” values that are often “deleted” during the first edit of the record.

Another issue identified during discussions between the authors was the fact that some of the changes in the structure of the record happen in ways that may not be expected. For example, in the system used for managing these records, there is a standard software module that is responsible for reading and writing metadata records to and from the XML format used in this study. This software module will normalise different representations of “blank” fields and will remove them from the resulting metadata record, as a record is written. The software module has a standard way of ignoring blank fields in the record and thus not writing them back to the XML file saved in the process. Therefore, if a metadata editor was to take an original XML metadata record, open and then save it without changing the values of the metadata, the structure of the resulting metadata record may change slightly.

In addition to this type of automatic deletion of blank fields that happens within the programming module, a number of changes are made to the record in automated ways that are invisible to the end user who views the record in user display, not in its XML behind-the-scenes version. For example, leading and trailing white spaces are

automatically removed from data values, and subjects with consistent delimiters such as Library of Congress Subject Headings strings are normalised to a standard format. One can see visible examples of such change in the XML of the last edited version of the record for a photograph in Fig. 4, which includes six cases of the XML character reference “” in two instances of a Note metadata field which were not part of the initial XML record version. None of these would be visible to the end user looking at the graphic user interface representation of the metadata record.

Programming modules that interact with metadata records often are implemented in a way that has them “changing” the metadata records as these records are opened and saved in systems without input from end users. Often this change is unnoticed to the end user because of the ways metadata is presented in end user interfaces. However, its effect should be noted in future investigations.

5. Conclusion and Future Research

The common ways to ensure metadata quality (be it interoperability, metadata record completeness, consistency, or redundancy) all assume the ability to modify metadata in order to correct gaps in quality, therefore necessitating some sort of change. Information science literature has only begun to explore the significance and implications of metadata quality, in the context of ensuring long term access. In the point of view of the authors of this paper, measuring of metadata change is closely

connected to measuring and improving the metadata quality in digital repositories. An agreed upon vocabulary of characteristics of digital library/repository metadata change can be used to help categorise and analyse metadata change at the granular record-to-record level as well as in broad studies involving hundreds of thousands of records and millions of change events. This paper suggests the categories and subcategories of metadata change that can be used in such analysis.

Based on the results of our analysis of metadata records in the UNT Digital Collections repository, we were able to test and refine the general framework of metadata change. Our data confirms that the three major types of metadata change are addition, deletion, and modification. We identified and empirically confirmed additional subcategories of metadata change. The resulting updated general framework of metadata change is presented in Fig. 15.

Further testing of the general framework of metadata change in different digital repositories, with different metadata schemes will result in fine-tuning of the framework and further discussion and development of an accepted metadata change terminology in various information science communities, including the knowledge management community. Some of the questions worth exploring for further studies include:

- If a data value is removed from one metadata field and added to another (similar to what we observed with Description and Note fields in a number of records in

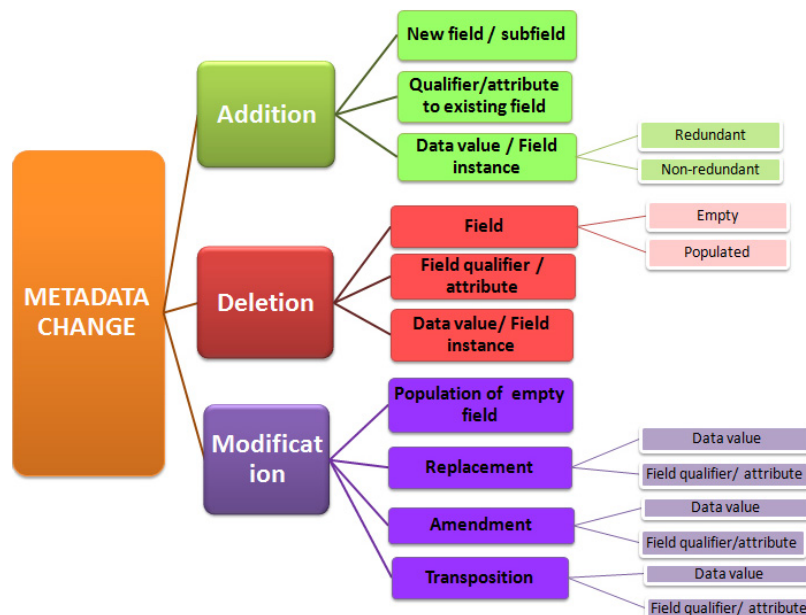


Fig. 15. The metadata change types: Refined general framework.

UNT Digital Collections), how should this change be described? In this case, the total length/text of the record may stay the same, but there could be a distinct change in how the information is represented in the record. The complex change operators discussed in computer science, particularly in relation to ontology change (e.g. Klein and Noy, 2003) might provide some ideas applicable to categorisation of metadata change in digital libraries and repositories.

- At what point does an amendment to an existing data value constitute a replacement data value? In the exploratory analysis presented in this paper, our team distinguished amendments from replacements based on the operational assumption that amended data value should have at least one meaningful word (i.e. not a preposition, article, or any other kind of a “stop word”) in common with the initial version of the data value. Following this logic, the change of a data value from “data management” to “research data management” which results in a longer data value was interpreted by us as an amendment, just like the change of a data value from “Association for Computing Machinery (ACM)” to “Association for Computing Machinery” which results in a shorter data value. Would the same approach be as applicable to changes in much longer data values in fields like Description, if only one or two of the meaningful words out of dozens and sometimes hundreds are common between the initial and the subsequent version of the data value in the field?
- What significance should be placed on the “amount” or degree of editing to a field or a record? For example, if the only change in a metadata record version is the removal of an extraneous period in a field’s data value, should that be given the same weight as removing or adding entire sentences?

Future studies need to take into account the automatic changes made to the metadata records by the software on opening and closing XML record files. This suggests that future work in this area may benefit from working with a standardised abstraction of the records that would remove some of the effects of these programming modules.

The metadata community needs a common vocabulary to discuss different kinds of metadata change. This will make it easier to compare record versions within a system, and to discuss and compare metadata changes across different systems. Successful digital curation strategies involve mechanisms for both pre- and post-ingest metadata normalisation and quality control. Understanding and managing metadata quality requires a cyclical process

that balances the evolving needs of the users, the requirements of national/international standards, and the local environments of the metadata creators. With this in mind, the ability to accurately describe and communicate change events during the lifecycle of metadata associated with a digital object will be increasingly important.

References

- Adams, S (2003). Information quality, liability, and corrections. *Online*, 27(5), 16–22. Retrieved from <http://search.proquest.com/docview/199956834?accountid=7113>.
- Alemneh, D (2009). Metadata quality assessment: A phased approach to ensuring long-term access to digital resource. *Proceedings of the American Society for Information Science and Technology*, 46(1). Retrieved from <http://digital.library.unt.edu/ark:/67531/metadc29318/>.
- Alfonseca, E, G Garrido, J Delort and A Peñas (2013). WHAD: Wikipedia historical attributes data. *Language Resources and Evaluation*, 47(4), 1163–1190. doi: <http://dx.doi.org/10.1007/s10579-013-9232-5>.
- Barton, L. (2010, September). *What’s The Diff? Using the Metadata Differ to Compare Object Metadata and Correct the Differences: An Oracle white paper*. Retrieved from http://download.oracle.com/otndocs/products/database/enterprise_edition/utilities/pdf/metadiffer11gr2_twp_1009.pdf.
- Bille, P (2005). A survey on tree edit distance and related problems. *Theoretical Computer Science*, 337(1–3), 217–239.
- Bruce, TR and DI Hillmann (2004). The continuum of metadata quality: Defining, expressing, exploiting. In *Metadata in Practice*, D Hillman and L Westbrook (eds.). Chicago: American Library Association, pp. 238–256.
- Chen, Y, Y Wen, H Chen, Y Lin and H Sum (2011). Metrics for metadata quality assurance and their implications for digital libraries. *Proceedings of the 13th International Conference on Asia-Pacific Digital Libraries: For Cultural Heritage, Knowledge Dissemination, and Future Creation*, 24–27 October 2011, Beijing, China.
- Cheney, A (2010). Pretty Diff — Documentation. Retrieved from <http://prettydiff.com/documentation.php>.
- Gao, X, B Xiao, D Tao and X Li (2009). A survey of graph edit distance. *Pattern Analysis and Applications*, 13(1), 113–129.
- Gautam, V and N Parimala (2012). E-Metadata versioning system for data warehouse schema. *International Journal of Metadata, Semantics and Ontologies*, 7(2), 101–113.
- Heckel, P (1978). A technique for isolating differences between files. *Communications of the ACM*, 21(4), 264–268, doi:10.1145/359460.359467.

- Hider, P and K Tan (2008). Constructing record quality measures based on catalog use. *Cataloging & Classification Quarterly*, 46(4), 338–361.
- Hill, JS (2008). Is it worth it? Management issues related to database quality. *Cataloging & Classification Quarterly*, 46(1), 5–26.
- Horwitz, S (1990). Identifying the semantic and textual differences between two versions of a program. *ACM SIGPLAN Notices*, 25(6), 234–245, doi: <http://dx.doi.org/10.1145/93548.93574>.
- Kher, A (2004). *The XML Diff and Patch GUI Tool*. Retrieved from <http://msdn.microsoft.com/en-us/library/aa302295.aspx>.
- Khoo, M and C Hall (2013). Managing metadata: Networks of practice, technological frames, and metadata work in a digital library. *Information and Organization*, 23(2), 81–106.
- Klein, M and N Noy (2003). A component-based framework for ontology evolution. In *Workshop on Ontologies and Distributed Systems*. Retrieved from http://www.researchgate.net/publication/2930642_A_Component-Based_Framework_for_Ontology_Evolution.
- Mason, MK (2007). *Copy Cataloguing: Where is it Taking Us on Our Quest for the Perfect Copy?* Retrieved from <http://www.moyak.com/researcher/resume/papers/clog4mkm.html>.
- Maynard, D, W Peters, M d’Aquin and M Sabou (2007). Change management for metadata evolution. In *Proceedings of the International Workshop on Ontology Dynamics (IWOD-07)*. pp. 27–40. Retrieved from <http://kmi.open.ac.uk/events/iwod/papers/paper-08.pdf>.
- Moen, WE, EL Stewart and CR McClure (1998). *The Role of Content Analysis in Evaluating Metadata for the U.S. Government Information Locator Service (GILS): Results from an Exploratory Study*. Retrieved from: <http://www.unt.edu/wmoen/publications/GILSMD-ContentAnalysis.htm>.
- NISO Framework Working Group (2007). *A Framework of Guidance for Building Good Digital Collections*, 3rd edn. Retrieved from <http://www.niso.org/publications/rp/framework3.pdf>.
- Noy, N, S Kunnatur, M Klein and M Musen (2004). Tracking changes during ontology evolution. *Lecture Notes in Computer Science*, 3298, 259–273.
- Ochoa, X and E Duval (2006). Towards automatic evaluation of learning object metadata quality. In *Proceedings of the 2006 international conference on Advances in Conceptual Modeling: Theory and Practice*, 06–09 November 2006, Tucson, AZ.
- OCLC (2014a). Quality Control. Retrieved from <http://oclc.org/worldcat/catalog/quality-control.en.html>.
- OCLC (2014b). Quality assurance. In *Bibliographic Formats and Standards*, 4th edn., Chapter 5. Retrieved from <https://oclc.org/bibformats/en/quality.html>.
- Soules, C, G Goodson, J Strunk and G Ganger (2003). Metadata efficiency in versioning file systems. In *Proceedings of the Conference on File and Storage Technologies*, pp. 43–58. USENIX Association.
- Stvilia, B and L Gasser (2008). Value based metadata quality assessment. *Library & Information Science Research*, 30(1), 67–74, Retrieved from <http://dx.doi.org/10.1016/j.lisr.2007.06.006>.
- Stvilia, B, L Gasser, M Twidale, S Shreeves and T Cole (2004). Metadata quality for federated collections. *Proceedings of ICIQ*, 04, 111–125.
- Tarver, H, O Zavalina, M Phillips, D Alemneh and S Shakeri (2014). How descriptive metadata changes in the UNT Libraries’ Collections: A case study. *Proceedings of the 22nd Annual Meeting and Conference of the Dublin Core Metadata Initiative (DCMI)*. Retrieved from <http://dcevents.dublincore.org/IntConf/dc-2014/paper/view/235>.
- Thornburg, G and M Oskins (2007). Misinformation and bias in metadata processing: Matching in large databases. *Information Technology and Libraries*, 26(2), 15–26.
- West, AG, S Kannan and I Lee (2010). STiki: An anti-vandalism tool for Wikipedia using spatio-temporal analysis of revision metadata. *Proceedings of the 6th International Symposium on Wikis and Open Collaboration (WikiSym ’10)*, doi:10.1145/1832772.1832814.
- Yan, Y and T McLane (2012). Metadata management and revision history tracking for spatial data and GIS map figures. *Proceedings of the 3rd International Conference on Computing for Geospatial Research and Applications*, doi:10.1145/2345316.2345357.
- Zavalina, OL, CL Palmer, AS Jackson and M-J Han (2008). Evaluating descriptive richness in collection-level metadata. *Journal of Library Metadata*, 8(4), 263–292.