Metadata Change:
Meeting the Evolving Requirements

Oksana L. Zavalina  
Priya Kizhakkethil  
(UNT iSchool)

Daniel Gelaw Alemneh  
Mark E. Phillips  
Hannah Tarver  
(UNT Libraries)

ICKM 2014, Antalya, Turkey
Metadata in knowledge management

- **Data**: Unorganized, unprocessed facts / observations
- **Information**: Data organized and set into a context
- **Knowledge**: Fully internalized information
- **Wisdom**: Capacity to take appropriate action on basis of knowledge

Described/represented with metadata

(Rowland, 2007)
Metadata in data curation lifecycle

Metadata for discovery and reuse

(Higgins, 2008)
Background of the study

1. Change in metadata records over time to keep up with “environmental” changes:
   • Growth in certain types/formats and subject matter of materials in repositories
   • Changes in the content & location of fluid materials (e.g., linkrot in websites)
   • Goals of hosting & contributing institutions
   • KOS: classification systems & controlled vocabularies
   • National & international standards for record creation.
Background of the study

1. Change in metadata records over time to keep up with “environmental” changes (Thornburg & Oskins, 2007):
   • Growth in certain types/formats and subject matter of materials in repositories
   • Changes in the content & location of fluid materials (e.g., websites)
   • Goals of hosting & contributing institutions
   • KOS: classification systems & controlled vocabularies
   • National & international standards for record creation.

2. Metadata change to improve quality
   • encouraged by bibliographic utilities (OCLC, RLIN, etc.)
Background of the study (2)

Metadata quality criteria suggested (over 20):

- Access
- Accuracy
- Availability
- Compactness
- Compatibility
- Completeness
- Comprehensiveness
- Content
- Consistency
- Cost
- Data structure
- Ease of creation
- Ease of Use
- Economy
- Flexibility
- Fitness for Use
- Informativeness
- Protocols
- Quantity
- Reliability
- Standard
- Timeliness
- Transfer
- Usability
Background of the study (2)

Most important metadata quality criteria (Park & Tosaka, 2010):

- Access
- **Accuracy**
- Availability
- Compactness
- **Completeness**
- Comprehensiveness
- Content
- **Consistency**
- Cost
- Data structure
- Ease of creation
- Ease of Use
- Economy
- Flexibility
- Fitness for Use
- Informativeness
- Protocols
- Quantity
- Reliability
- Standard
- Timeliness
- Transfer
- Usability
Background of the study (2)

Metadata quality criteria (3 most important):

**Accuracy**
- Format and formatting errors
- Spelling and typographical errors
- Have accepted methods been used for creation or extraction of metadata?
- What has been done to ensure valid values and structure?
- Are default values appropriate, and have they been appropriately used?

**Completeness**
- Number of elements per record
- Practice of presenting “blank” elements
- Utilization and selected characteristics of “mandatory” and “optional” elements
- Does the element set completely describe the objects?
- Are all relevant elements used for each object?

**Logical Consistency/Coherence**
- Are values in metadata elements consistent throughout?
- How does it compare with other data within the community?

(Moen, Stuart, & McClure, 1998; Bruce & Hillmann, 2004)
(Bruce & Hillmann, 2004)
Background of the study (3)

- Link between metadata change and metadata quality
- Planning for metadata change to improve metadata quality and facilitate data/information discovery and reuse
  - need for common terminology
  - no metadata change categorization available
  - CS literature offers categorization of change in graphs, trees, ontologies etc. (*Bille, 2005; Gao et al., 2009; Klein & Noy, 2003; Maynard et al., 2007*)
  - 3 major change types: Add, Delete, Modify
Our Initial Framework of Metadata Change

- Anecdotal evidence
- Brainstorming
Testing the framework

• Digital repository that archives versions of metadata records: UNT Digital Collections
  • UNT Digital Library (incl. institutional repository)
  • Portal to Texas History
  • Gateway to Oklahoma History

• Random stratified sample of metadata records
  • Different types of materials from different collections
  • 157 records, 2 versions of each: initial and latest (April 2014)
    • Locally-modified Dublin Core metadata scheme
    • Native XML format

• In-depth manual comparative content analysis
Distribution of Change Categories in 21 Metadata Elements

- DELETIONS (total; % of records)
- ADDITIONS (total; % of records)
- MODIFICATIONS (total; % of records)
Metadata Change: Record Example 2

Empty field / subfield deletion

Deletion of field instance
Metadata Change: Record Example 2

Empty field / subfield deletion

Field instance/ data value addition

Deletion of field instance
Metadata Change: Record Example 2

- Transposition of subfields (modification)
- Empty field / subfield deletion
- Field instance/data value addition
- Data value modification: amendment
- Deletion of field instance
Metadata Change: Record Example 3

Empty subfield deletion

Field instance / data value addition (KOS)
Metadata Change: Record Example 3

Data value modification: replacement

Empty subfield deletion

Field instance / data value addition (KOS)

Empty field deletion
Metadata Change: Record Example 3

Data value modification: replacement

Data value modification: transposition

Empty subfield deletion

Field instance / data value addition (KOS)
Metadata Change: Record Example 4

Populated field deletion

Empty field deletion
Metadata Change: Record Example 4

Populated field deletion
Subfield addition
Empty field deletion
Field instance / data value addition
Metadata Change: Record Example 4

- Populated field deletion
- Subfield addition
- Empty field deletion
- Field qualifier addition
- Field instance / data value addition
Metadata Change: Record Example 4

- Populated field deletion
- Subfield addition
- Empty field deletion
- Field qualifier addition
- Populating empty field
- Field instance / data value addition
Metadata Change Framework (Revised)
Metadata Change & Metadata Quality

- Completeness ?
- Accuracy ?
- Completeness ?
- Accuracy ?
- Consistency ?

Addition
- New field / subfield
- Qualifier/attribute to existing field
- Data value / Field instance

Deletion
- Field
- Field qualifier / attribute
- Data value / Field instance

Modification
- Population of empty field
- Replacement
- Amendment
- Transposition
Summary

• A good digital collection is sustainable over time.

• Maintaining high quality metadata about every digital object requires a framework to carry out quality assurance measures during entire lifecycle.

• Ensuring metadata quality -- interoperability, completeness, consistency, or accuracy -- necessitates some sort of changes.

• Understanding the changes and managing high quality metadata requires a cyclical process that balances the evolving and sometimes conflicting needs and requirements.
Our view:

(adapted from Higgins, 2008)
Future Research

Test the framework in various contexts using a variety of metadata schemes
Future Research

Test the framework in various contexts

- digital libraries
- digital repositories:
  - research products (publications, presentations etc.)
  - research data
- bibliographic databases (e.g., WorldCat)

using a variety of metadata schemes
Future Research (2)

Refine mapping of metadata change categories and subcategories to 3 major metadata quality criteria

Map metadata change categories and subcategories to a wider set of (more specific) metadata quality criteria
Cited Works


Questions?
Comments? Ideas?

Oksana L. Zavalina: Oksana.Zavaina@unt.edu