High-Quality Access for Electronic Content: A Proposal for a Metadata Quality Assessment for Digital Collections

Dr. Heather Lea Moulaison
The iSchool at the University of Missouri
November 5, 2013
ASIST Annual Meeting
Montreal, Canada
This Conference: Metadata Quality

• Theme of concern for “metadata quality”
• How much metadata is necessary?
  – Who should decide on the metadata to be included in a description/collection/domain?
• Is it important who or what creates the metadata?
• How do changes in standards (for encoding, for example) affect access, ultimately?
The Nature of Electronic Content

• Digital content is fragile

• Poor quality metadata has always had the potential to hide materials from online discovery
  – Yet even if labeled “Untied States” online, physical materials could still be discovered in open stacks collections

• Digital content requires high-quality metadata to be effectively organized, accessed, used, and re-used in knowledge organization systems (KOSs)
Metadata Standards

• Standards used for
  – Describing contents
  – Providing intellectual access
  – Encoding information about electronic files/formats
  – Designating relationships to other objects in the collections,
  – Indicating rights information and use
  – And encoding metadata for storage in the system

• Standards contribute to quality by promoting interoperability, authenticity, preservation, etc. if applied correctly
  – Identifying issues before it’s too late is key
Metadata Creation *In Situ*

- Professionally-created metadata is an expensive undertaking
  - And it is not perfect. Humans make mistakes.
- Whenever possible, metadata is:
  - Automatically generated
  - Supplied by the content provider
  - Possibly crowd-sourced
- Metadata is also supplied by para-professionals or trained repository workers (students and volunteers)
  - We assume that this metadata is more imperfect than professionally created.

What overarching criteria for *quality* metadata can be applied in these diverse circumstances, where content in a single record may be provided by any number of these metadata creation methods?
Quality in Metadata

...Is difficult to define and even more difficult to assess.

Subjective Nature of Quality

• “Quality judgments are by definition subjective and incomplete” (Conway, 2011, p. 299).

Guiding questions:
How can we establish flexible parameters for quality in the discussion of metadata for electronic content and how can it be documented in working systems?
Models for Metadata Quality

• Broadly speaking, models for metadata quality
  – Have as a goal the access and use of contents by the targeted community of users
  – Promote improvements in metadata used in systems
Mapping between the Bruce and Hillman Framework and the Gasser and Stvilia Framework (Park, 2009)
Metrics for Metadata Quality

• Metrics can be considered to be completeness, accuracy, and consistency (Park, 2009; Zeng & Qin, 2008)

• Quality can be measured across – fields, records, collections

• Yet assessing existing metadata across a collection is incredibly time-consuming and expensive.
Preservation and Access Initiatives for Electronic Content

• Ultimately, e-resources and their metadata need to be stored and made available over the long-term

• Measures are being taken to ensure long-term access to curated content. Some examples:
  – Internet Archive
  – Portico
  – HathiTrust
  – Dryad
Self-Assessment Initiative for Digital Repositories: DSA

- Repositories housing content/metadata are an important part of initiatives promoting access.
- The Data Seal of Approval [http://datasealofapproval.org/](http://datasealofapproval.org/) requires:
  - Documentation of repository procedures
  - Self-study and self-assessment based on community-agreed upon framework
  - Openness for documentation and assessment
  - Peer-review component
  - End result: a seal that should inspire the trust of users.
- Individual repositories can undergo the process to earn the DSA.
Proposal: Metadata Quality Assessment (MQA)

• After further analysis and study of metadata practices and needs, a set of core standards for metadata creation should be established, with community input for the levels of work
  – The MQA is meant to inspire confidence not in users, per se, but in information professionals

• As part of the application process, repositories would submit and make openly available information at the collection level.
Potential Categories for Self-Study

Each, possibly with 1-5 stars that can be self-assigned according to information community-based criteria:

• A listing of metadata standards used in the collection
• Documentation for the application of the standards
  – best practices, training materials, application profiles, etc.
  – Documentation about the process for applying the standards
• Strategies for periodic analysis/ evaluation of the metadata
• Criteria based on preservation components
  – Authenticity, sustainability, future use
• [information about digital content in the collection:
  – cultural importance of the contents being described
  – potential for use in other systems due to open access rights]

• Peer reviewers from the information community would then evaluate the metadata creation and evaluation process brought forth in the self-study
The Proposed MQA

Would:

• Clarify and quantify essential characteristics of the collection’s metadata
  – to facilitate sharing and reuse

• Be available to individual collections
  – With different levels available according to metadata creation practices and preservation plans

• Propose essentially collection-level meta-metadata to be used by information professionals
End-goals of MQA

- Allow for the subjective assessment of metadata quality by other information professionals as a way of promoting collaboration, access, and reuse.
- Help information communities to assess and target collections, e.g.:
  - At-risk collections for collaborative assistance,
  - High-quality collections for partner projects, etc.
Conclusions

• Ultimately, *quality metadata* allows users to find, access, and use content
  – Some metadata is not high quality
    • shouldn’t be taken as high-quality metadata
  – Repositories should understand what others are doing to provide access and make sure their efforts are indeed promoting access, use, and re-use

• Research can and should guide practice in the establishment of an agreed-upon set of criteria for self-study and community-based certification