

User Feedback and Cost/Value Analysis of Metadata Creation

**Joyce Celeste Chapman
Metadata & Cataloging / Digital Library Initiatives
North Carolina State University Libraries**

**Description Section meeting
SAA 2010 Annual (August 13, 2010)**

Cost and value of metadata

- We assume there to be inherent value in the work we do with metadata
- Libraries are lacking metrics for measuring cost and value of metadata for respective stakeholder communities
- Problem: unlike for-profits, we cannot model on cost versus sales

While the library community assumes there to be inherent value in the work we do with metadata, libraries and archives are almost completely lacking in metrics for measuring cost, benefit, and value of metadata for various stakeholder communities. One of the biggest problems in initiating such analysis is the fact that unlike for profits, Libraries cannot measure “cost” against “sales” – the traditional definition of value -- and must therefore create alternative definitions of value with which to work.

NCSU is interested in initiating a conversation in the field about measuring cost/value for metadata and metadata-related activities. In the past year, our Metadata & Cataloging department has begun a series of research initiatives into this area – this is not research limited to archival metadata -- and my hope is that talking to you about this work will encourage you to at least think more critically about these issues, possibly to join us in our efforts, or to undertake your own studies to evaluate cost and value of archival metadata.

Cost and value of metadata

6/2010 report by the ALA Task Force on Cost/
Value Assessment of Bibliographic Control
<http://connect.ala.org/node/106017>

“Our ability to make sound decisions and mindful changes around bibliographic control is hindered by our lack of operational definitions of value and methodologies for assessing value within our institutions.”

Last year, the Heads of Technical Services in Large Research Libraries Interest Group of ALA charged a taskforce with identifying measures of cost, benefit, and value of bibliographic control for key stakeholder communities and developing a plan for implementing these measures. What the Task Force uncovered when they began their research, however, was that very little has been done in this field and few people are working to answer these questions now. The Task Force’s final report (issued in June 2010) states that Libraries’ “ability to make sound decisions and mindful changes around bibliographic control is hindered by our lack of operational definitions of value and methodologies for assessing value within our institutions.”

Multi-faceted nature of “value” for metadata

As identified in the cost/value report:

- Use (page views, clicks)
- Discovery success
- Interoperability / ability to operate on the open Web
- Display understanding
- Ability to support the FRBR user tasks
- Timeliness / throughput
- Supporting administrative activities and management

The report suggests 7 operational definitions of value for metadata (*read from slide*) that could be used in assessment efforts.

What becomes clear as we look at these many operational definitions is that evaluating “value” will not be possible with a single study: is a multifaceted problem, and will require a community effort and multiple studies that analyze different facets of the problem.

For the study I’m going to talk to you about today, we decided to embark upon our cost/value analysis of archival metadata using the operational definition of “discovery success.” Even within this single operation definition, there are multiple aspects of use, so more specifically this study targeted the FRBR user task *identify*: or the process by which a user confirms that the item described corresponds to what they were searching for (*FRBR: Functional Requirements for Bibliographic Records: a conceptual entity-relationship model developed by the International Federation of Library Associations and Institutions*).

Users, archival metadata, and value through discovery success

- Time study (“cost”): how long do processors spend creating specific metadata elements?
- User study (“value”): how frequently do users use these same elements in completing the FRBR user task *identify*?
- What lessons can we learn by comparing cost and value as defined above?

The goals of this study were as follows: first, to collect timing data on how long processors spend on metadata creation for certain elements; second, to collect use data by observing researcher behavior while attempting to verify the relevancy of finding aids to their searches and third, to see what lessons can be learned by comparing cost and value defined as such.

Which elements were studied?

1. Abstract
2. Biographical / Historical Note
3. Scope and Content Note (collection-level)
4. Subject Headings
5. Collection Inventory
6. Other*

* Catch-all category for all other elements. Required to analyze ratios.

“Collection Inventory” is the entire output of the <dsc> (the Box List, Container List, etc.) This includes series-level Scope and Content Notes.

Methodology

Part I: User study

- 10 advanced archival researchers
- 5 subjective information discovery tasks
- Interview with researchers about how useful they perceive each metadata element to be and their perceptions of how frequently and under what circumstances they use different elements

• This was an IRB approved study, in which 10 advanced archival researchers participated (definition of advanced researcher: people who are familiar with finding aids and have conducted research with them previously, as opposed to “novices,” who would be people who have never used a finding aid and may not understand archives, archival material, or what a finding aid is. The advanced researchers tested in this study were all academic researchers, however, academic associations have nothing to do with the definition of “advanced”).

• Each was asked to perform 5 subjective information discovery tasks

* This was followed by an interview with researchers about how useful they perceive each metadata element to be and their perceptions of how frequently and under what circumstances they use different elements.

User study and interviews

1. Use the collection guide for the [Charlotte Vestal Brown Papers \(MC00219\)](#) collection to complete this task.

You're a historian who has been asked to give a talk at the annual meeting of the North Carolina Chapter of the American Institute of Architects (AIA). You'd like to talk about the history of the association, so you've come to the archives to conduct some research. Examine the Charlotte Vestal Brown Papers. Would they be useful in your research?

Guide to the Charlotte Vestal Brown Papers 1767-1990, MC 00219

- [Abstract](#)
- [Biographical/Historical Note](#)
- [Scope and Content Note](#)
- [Subject Headings](#)
- [Arrangement](#)
- [Collection Inventory](#)

- Users had to determine the relevancy of the finding aid to their search
- The interface to the finding aids forced a conscious choice between metadata elements

For the user tasks, participants were presented with an institution-neutral interface in which they were not allowed to use find-in-page. They weren't allowed to use find-in-page because FIP essentially negates the need for decision-making on the part of the user about where to go within the finding aid, they aren't performing value judgments about metadata elements, which is what we were quantifying. Each task presented the participant with an information need and asked them to explore a particular finding aid to determine whether or not it met their need. The interface to the finding aid forced a conscious decision on the part of the user about which metadata element they were going to navigate to: metadata is presented in a list, and these links are the only way to see any part of the finding aid. After exploring one element, a user returned to the list of elements and chose another (or ended the task, depending on whether they felt they had enough information to answer the question).

User study: behavior

- Answers were planted in every metadata element for every task.
- At least once during the test, 9 of 10 participants read a metadata element and failed to see the relevant information. 55% repeated the behavior.
- Reading styles relied heavily on skimming, scanning for keywords, and reading only the first part of paragraphs. 66% of participants said that instead of reading a metadata element fully they were scanning for keywords.

Before I start talking about usability findings for individual elements, there was some general behavior that presented a really interesting view of how researchers interact with finding aids overall at this point in their research process.

Data related to the search query was embedded in **each** of the five metadata elements in every task (failure should have been impossible).

However, at least once during the test, 9 of 10 participants “read” a metadata element and completely failed to see the relevant information. 55% repeated this behavior.

Reading styles relied heavily on skimming, scanning for keywords, and reading only the first part of a paragraph before skipping to another metadata section or another paragraph. 66% of participants said that instead of reading a metadata element fully they were scanning for a set of keywords they had in their head related to their search.

User study: behavior

- 20% of participants said the longer a metadata element the more likely they were to leave without reading much or anything at all.
- 55% of participants quit reading even a single paragraph part way through at least once during the study. Some abandoned a metadata element after as little as 5-6 seconds, declaring that it did not contain relevant information (it did.)
- Only 1/3 of participants showed clear evidence of reading metadata carefully at any point during the test.

20% of participants said the longer a metadata element is, the more likely they are to leave without reading much or anything at all at this point in the research process.

55% of participants quit reading even a metadata element composed of a single paragraph part way through at least once during the study. It did not appear that participants were doing this consciously. In some cases, researchers stop reading a metadata element after as little as 5-6 seconds, declaring (incorrectly) that there was nothing there related to their search and moving to a new element.

Only a third of participants showed clear evidence that they were reading a metadata element in full and carefully at any point during the test.

While it is likely that the testing situation influenced behavior to some degree (the test doesn't replicate a real research situation, the outcome isn't personally important to them), I believe this behavior probably represents a somewhat enhanced version of what they do during the information relevancy verification process.

User study: behavior

- On average, participants were able to locate information 42% faster when it was presented in a bulleted list rather than a narrative paragraph.
- *“To my mind, if this were a set of lists of bulleted points...it would be so much easier to scan...just, typographically [these paragraphs are] very hard to read, and to read in the kind of scanning way that a researcher would want to find stuff.”*

Several tasks tested the participants against each other in two groups, using the same metadata but presenting it in narrative paragraphs to one group and as bulleted lists to the other.

On average, participants were able to identify the content of the collection to be relevant to their search 42% faster when using finding aids that employed bulleted list in the scope and content note rather than a narrative paragraph. The quote below summarizes participants general feelings on the subject.

User study and interviews

- **Behavior score:** based on metadata choices made by users while performing tasks.
- **Rank score:** participants were asked to rank metadata elements against each other in terms of overall usefulness.
- **Frequency score:** participants were asked to rate how frequently they use each element and to discuss for what purposes they use each element.

There were three different use scores that came out of this study: one was for the participant behavior during the user study and that was the “behavioral score.” The others emerged from the interview with the participant in which each was asked to rank the metadata elements against each other in terms of overall usefulness (“rank score”) and to rate how frequently they think that they use each element (“frequency score”).

Analyzing usability findings

The three types of analysis produced identically ranked results(!)

1. Collection Inventory
2. Abstract
3. Subject Headings
4. Scope Content
5. Biographical Note

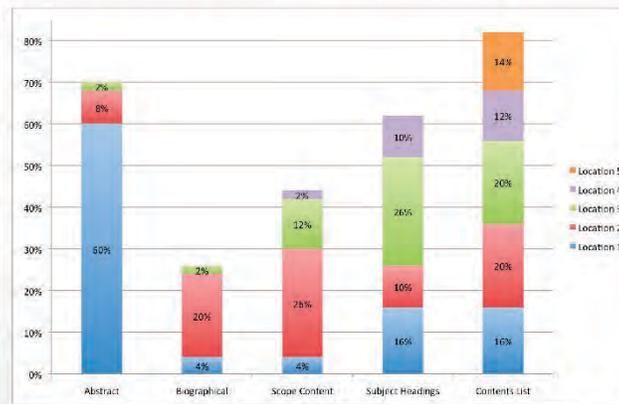
1 = most useful/most used, 5 = least useful/least used

The three types of analysis produced almost identically ranked results (frequency score Scope and Subjects were a tie). Elements were ranked in the following order of importance from most useful to least: inventory, abstract, subjects, scope, biographical.

Items of interest to me here:

- * Did not necessarily expect Inventory to beat out Abstract as #1. Remember that the tasks didn't require users to find particular containers, so there was no need to go to the Inventory.
- * I wouldn't have been surprised if Subjects had come in last, and I am surprised that they beat out Scope Content.
- * For the Rank score participants actually answered this question in terms of overall usefulness, not usefulness specific to the facet of value "discovery success" that was being studied. So it's interesting that they still came back with the same scores.

Behavioral scores by order visited



Another interesting way to look at behavioral scores is order in which participants visited different metadata elements for each task. This chart shows the percentage of total tasks involved a participant going to a certain metadata element first, second, third, etc. for determining information relevancy.

Behavioral scores by order visited



60% of task instances involve a user beginning at the abstract.

NOTE: during the SAA presentation, someone asked if this couldn't have been influenced by the fact that the Abstract was listed first on the list of metadata elements presented to users. I doubt that influenced behavior in the current study to a significant degree because we were studying advanced researchers: or people that had experience with finding aid research, people that already had their own developed patterns of how they interact with a finding aid. They were all supposedly quite familiar with what they would find on the other end of each of these metadata element labels (though it's possible that some weren't). However, if this study were repeated with novice finding aid users, you would definitely want to control for such factors. Another thing to consider is that some of these metadata section labels might make sense to a novice, and others wouldn't. While everyone understands the word "Abstract," not many have any idea what "Scope and Content" means. The advanced users were supposed to understand such differences. There's a lot of info available in published finding aid usability studies about section labels and jargon.

Problematic metadata overlap

- Some participants were confused about the overlap between Abstract and Scope and Content Note
- Out of all the instances in which participants navigated to the Abstract, 64% of the time, they never subsequently looked at the Scope and Content Note

The abstract is interesting because in the post-test Interview, multiple participants discussed having experienced confusion at some point differentiating between the Scope Content and the Abstract due to overlapping content. At NCSU, because we do minimal processing it's not infrequent that our collection level Scopes are a single paragraph long: in such cases, often the entirety is copied into the abstract and I know of other institutions that have similar habits, so it's not surprising that users who see these are confused. While some researchers clearly did know that the Scope was supposed to be a more extensive description, several others didn't and discussed only every using one of these sections and not the other because of overlapping content. This would definitely lead them to miss out on a lot of info if the Scope were the element they didn't look at.

I analyzed the study data to find out what our usage patterns we had and found that 64% of the time in which a user navigated to the Abstract they never subsequently looked at the Scope Content Note (note: this does not mean that all of those people didn't look at the Scope Note for that particular reason. These are two general observations that are being looked at side-by-side, I'm not suggesting total correlation. I'm sure that some of the people who read an Abstract and not a Scope Note did so because they had enough information to answer the question after only looking at one or two elements and just didn't get around to the Scope. Whatever the reason, it's a disturbingly high % of the time that the Scope was not read).

Behavioral scores by order visited



Users are pretty open about where they head second for information relevancy verification (three-way tie).

Behavioral scores by order visited



If the Biographical and Abstract are not checked as the 1st or 2nd location, they are essentially not used to determine relevancy of a collection to a search.

Behavioral scores by order visited



The great popularity of Subject Headings as a third location reflects the tendency I saw in many participants to verify information via Subject Headings, but not to rely on Subject Headings alone as an indicator of relevancy. So they'd look a couple other places first, and then they'd come here last and if info was also found here as a heading, they would feel confident that their answer was correct.

User interviews

Biographical Note

- Only 36% of participants say they use it frequently or always.
- A number of people said they only read it if their research pertains to the creator.
- Some said that they rarely use Biographical Notes during the information discovery process but do use them later in the research process.

User interviews

Subject Headings

- 54% of participants say they use frequently or always.
- Most use them more heavily in the initial phases of information discovery (FRBR *find*) and not for relevancy verification (FRBR *identify*).
- *“The only reason I’d ever use a subject heading is to find other sources.”*

User interviews

Collection Inventory

- 100% of participants say they use frequently or always.
- A number of users said that they use the collection inventory not only to locate/request materials but also as a tool for discovery success.

Methodology

Part II: Timing metadata creation

- Two partner institutions* (both employ minimal processing)
- 9 processors (5 with archives backgrounds, 4 with cataloging backgrounds)
- Data for 14 archival collections
- Time data was tracked to the minute for all aspects of metadata creation including research, authority work, encoding, and supervisor review/editing.

* Avery Research Institute for African American History and Culture at the College of Charleston, SC
North Carolina State University Libraries

- Two partner institutions, both of whom employ minimal processing (Avery Research Institute for African American History and Culture at the College of Charleston, SC and North Carolina State University Libraries)

- 9 processors, 5 of whom had backgrounds in archives and 4 of whom had backgrounds in cataloging and had only recently been trained in archival processing as part of a new shared workflow between the two departments at NCSU. So there were three groups of processors that were under observation here: archivists at Avery, archivists at NCSU, and catalogers at NCSU.

- Data for 14 archival collections

* Time data was tracked to the minute for all aspects of metadata creation including research, authority work, encoding

Disclaimer

- Patterns in time data are not necessarily generalizable to other institutions.
- 14 data points is not many, especially split 3-ways! More data needs to be tracked before we can be sure that patterns we see are accurate.

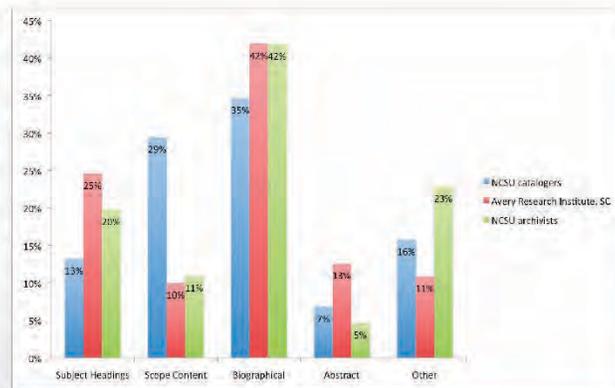
Patterns discussed in this time data are not meant to be generalizable to the entire archival community: your institution might have totally different metadata creation patterns than mine. 14 data points is not many especially split 3-ways. More data needs to be tracked before we can be sure that patterns we see are accurate. Each variable adds a df “degree of freedom” (the minimal number of values which should be specified to determine all the data points).

The usability data, on the other hand, is re-usable, so all you would have to do is a little time-study at your own institution to see how you stack up. However the usability study only tests one group of users: academic researchers (including post-docs, graduate students, faculty, and retired faculty), and advanced finding aid users.

How numbers were analyzed

- Numbers for time data were analyzed both as ratios of a whole and as real numbers.
- Time data for the Collection Inventory was analyzed both together and separately from other metadata because it throws off the ratios of time for other elements based on the size of the collection in linear feet.

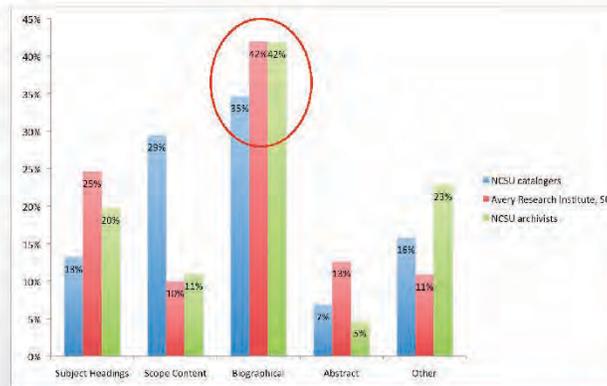
Timing analysis Average ratios without Collection Inventory



This chart shows the results of the timing analysis not including the Collection inventory.

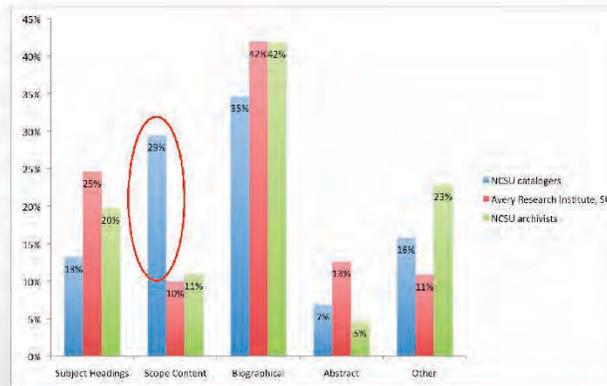
Timing analysis

Average ratios without Collection Inventory



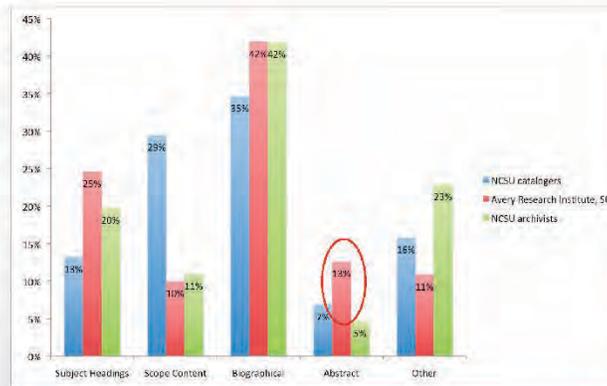
Note the large percentage of time all three groups are spending on Biographical notes. This is more than any other non-Inventory metadata element. Remember we saw that the value rating for Bios was the lowest – though of course, we are only testing one facet of value from many.

Timing analysis Average ratios without Collection Inventory



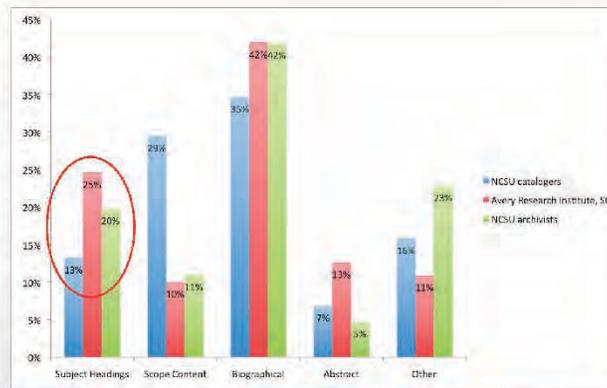
Catalogers are spending a much higher percentage of time of the Scope Note than the archivist groups. Catalogers are spending a roughly similar % of time on the Bio and Scope, whereas the archivists are spending 4 time more time on Bio than the Scope.

Timing analysis Average ratios without Collection Inventory



Another interesting observation is that Avery archivists are spending a higher percentage of their time on the Abstract than either group at NCSU.

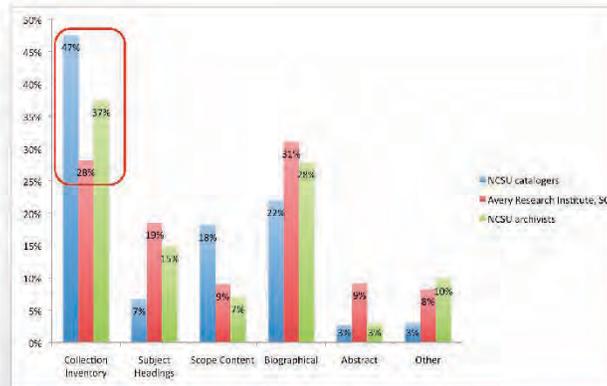
Timing analysis Average ratios without Collection Inventory



One last item of note is that the catalogers are spending a smaller percentage of their time on subject analysis than either group of archivists.

Timing analysis

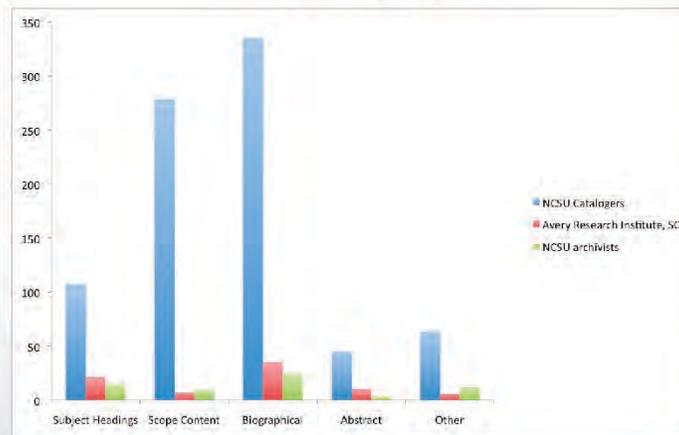
Average ratios including Collection Inventory



Let's look at the data including the averages for the Collection Inventory.

The averages for the Inventory are over here on the left. Now these will change based on the variety of the sizes of collections in your sample, but still interesting. It's interesting to see that Avery is actually spending a greater % of their time on Biographical Notes than Inventory for this sample. The Catalogers are spending a greater percentage of time in the inventory than anyone else, though that might not hold across a larger sample. Basically what we are seeing though is that on average, the Collection Inventory is the metadata element that people are spending the most time on. However: are we spending enough time to equal the value rating?

Real Time



Real time data: while ratios are useful in one way, they tell us nothing about the actual amount of time that is being spent on metadata creation. The Y-axis shows real time in minutes.

The catalogers are spending more time on metadata creation for all elements than the archivists at either institution (though they might not be spending a lot of time compared to archivists at other institutions!). Particular areas of concern are the Biographical Note and Scope Note. Keep in mind that there is a particularly small dataset for the NCSU catalogers in this study, so it could well be that the time difference isn't as dramatically high an average over a larger data set. Another issue is that the catalogers are currently in training, so this data represents the first several collections that they have processed.

Really interesting is that the archivists at different institutions are spending fairly similar average amounts of time on various metadata element creation. There was no reason to have expected that at all.

My first question when I saw the disparity in NCSU cataloger and archivist time was, "are the catalogers spending a similarly disparate amount of time on the entire act of archival processing as archivists? The answer is no, they are not. Last year, archivists at NCSU processed apprx. 430 linear feet per FTE and catalogers processed apprx. 404 linear feet per FTE. So the catalogers are doing an excellent job of processing at the same speed as their counterpart archivists (and trainers) overall. Keep in mind that the cataloger statistic also includes their training time, because it was the first year that they were involved with processing.

So this disparity is really only in metadata creation and implies to us that the catalogers may need more training from NCSU archivists in how to do minimal processing as applied to metadata, or simply a more constant feedback loop from archival supervisors in this part of training.

Cost to Value? Issues and Questions

1. Compared to use, a disproportionately high % of time is spent creating Biographical Notes.
 - Data has different meaning within the context of different institutions and their goals
 - We might want to regulate time spent on metadata, but only for metadata creation that exceeds a certain institution-specific **real time baseline**.

–Data acquires meaning in context: do we care if we spend a high % of metadata creation time on the Biographical Note if that translates to real numbers that are not “significant” by our institutional standards? We might want to regulate time spent on metadata, but only for metadata creation that exceeds a certain **real time baseline**.

Data acquires meaning in context

- Examples in real numbers from NCSU:
 - **Collection A**
Biographical Note = 51% of total metadata time
= 24 minutes
 - **Collection B**
Biographical Note = 43% of total metadata time
= 9.8 hours

Cost to Value? Issues and Questions

2. Abstract is high value, users often go there first, and some use only the Abstract and never Scope and Content Note.
3. Users emphasize importance of Collection Inventory. Are we spending enough time there to equal the value rating?

2. I wouldn't necessarily say that we need to spend a higher percentage of time on the Abstract to equal its value rating (remember that in fact the time allocated to Scope and Content Note should really be double counted for the Abstract because of the tendency to cut and paste). However, research results lend a new weight to the critical nature of the contents of the Abstract. We need to really think about what goes in there as well as how issues with metadata duplication with the Scope might be dealt with.

3. No comments.

What to do? User opinions

Follow-up email question to all participants:

Q: If you had the option of archivists reallocating a percentage of the time spent writing Biographical Notes to fleshing out the Collection Inventory instead, would you leave as is or reallocate time?

The two participants whose behavioral scores ranked the Biographical Notes the highest both answered “yes.”

I wrote a follow-up question to all participants. Not many answered, but the two participants who used Biographical Notes the most were among the three who answered. Both said that they would sacrifice time that went into the Biographical Note if it were to be redirected into the Collection Inventory.

What to do? User opinions

Participant 1: “While the biographical notes are useful, a more detailed account of the collection is more important to me. As long as there is a brief biographical sketch (not as detailed).”

Participant 2: “A more detailed description in the Container List is preferable to me. The Bio is a quick entry point for the type of research I do, but I end up having to do it myself anyway, in part because I need other information not generally covered, and in part because of trust issues.”

Important issues: trust, the archivist acting as the researcher when in fact the researcher must replicate this work themselves.

Next steps: study other facets of value

- This study examined only one facet of value. In order to form a complete picture of value for metadata, further studies must be conducted.

Thank you!

- Joyce_chapman@ncsu.edu
- Final report of the ALA Task Force on Cost/ Value Assessment of Bibliographic Control (6/2010) <http://connect.ala.org/node/106017>