8-4-2016

Creating Newspaper-Style Captions and Metadata to Describe Images That Cannot be Digitized due to Copyright Concerns

Eric Willey
Illinois State University, emwille@ilstu.edu

Recommended Citation

Willey, Eric, "Creating Newspaper-Style Captions and Metadata to Describe Images That Cannot be Digitized due to Copyright Concerns" (2016). *Faculty and Staff Publications – Milner Library*. 73.
https://ir.library.illinoisstate.edu/fpml/73

[SLIDE 1]

Good morning everybody, thank you all for coming, my name is Eric Willey and my talk is entitled "*Creating Newspaper-Style Captions and Metadata to Describe Images that Cannot be Digitized due to Copyright Concerns,*" or alternatively "*Buddy, Seriously, Read Your Session Title Out-Loud Before Submitting Because That is Way too Wordy*

[SLIDE 2]

 and now You're Going to Have to Get Past That With a Joke or Something

[SLIDE 3]

 and You Know You're Not Very Good at That

[SLIDE 4]

 so Save Yourself some Trouble, Huh?" Thank you, wasn't sure anyone would laugh at that one, which would have been bad.

So I came to archives through the history side of academia, and I was taught to tell 'em what you're gonna tell 'em, tell 'em, and then tell 'em what you told 'em.

[SLIDE 5]

So this is a talk about creating a guide, a word document using student workers that would provide captions and metadata for images and scrapbook materials which could not be digitized. It is not a collection show and tell, nor is it a how we done it good story. While I consider this project a success there were numerous problems and stumbling blocks. So what I am going to tell you is: every single thing I screwed up on this project, and even a very few things I got right probably due to the law of averages. Some of it might be blindingly obvious to people with more project management experience than me, but I hope everyone finds something useful so that your problems will be new and exciting and different than mine.

[SLIDE 6]

Very briefly, to explain why I did this, Lois Lenski was a mid-20th century children's and YA before anyone called it YA author who illustrated her own books. Won the Newberry Medal in 1946 for *Strawberry Girl*, one of her regional books, about children who lived in different parts of the United States. And what Lenski would do is she would go and talk and hang out with these children and ask them about their lives as part of her research. And she saved all this stuff, in addition to her original illustrations, manuscripts, photos, etc. Which is fantastic, except: Lenski couldn't decide where she wanted to donate this material. She knew she knew she wanted it to go to a former normal school, a teaching college, but couldn't pick one so she just sporadically donated small random batches to around fourteen former normal schools across the country.

Literally, Florida to California. So, with materials even from the same book widely dispersed, this becomes a nightmare for researchers when archivists do what I did and describe illustrations as, "Illustrations from whichever book" and a date. And I also wanted to make our patrons aware of the research scrapbooks Lenski kept, because these went way beyond the newsclipping kind of scrapbooks. And a biography of Lenski was just published in 2016, by Dr. Bobbi Malone, so Lenski is still on some people's radar but my institution prefers to have very clear copyright before digitizing.

So I got to thinking, and we provide metadata for scanned images, what if I just skipped the digitization portion? I mean, if we assume our metadata is good enough for someone to find an image in CONTENTdm, why wouldn't the same kind of metadata let them find a folder number with an image in it? And providing metadata doesn't violate copyright… yet.

[SLIDE 7]

I got a $2400 grant from the university through the library to hire two student workers at $10 an hour for 120 hours each to create newspaper photo style captions and metadata for these images and scrapbook materials. I consulted with Metadata Librarian Angela Yon, on which controlled vocabulary to use, and we decided Thesaurus of Graphic Materials was an appropriate controlled vocabulary for illustrations in a children's book. So, this is one thing I think I did right: Don't make your own controlled vocabulary. There are a lot out there, and probably one that will fit your needs and save you the headache of making your own. So we have our metadata scheme, have two students, and we get rolling, and we immediately rolled off in a different direction than what I had planned, because to paraphrase Helmuth von Moltke, "No processing plan survives contact with the collection."

[SLIDE 8]

I knew students would only be a working a few hours a week so I was worried about them forgetting training, and I express myself better in writing anyway, so I walked them through the process once, and gave them a written manual with links to several websites talking about things like of versus about, websites where they could see controlled vocabularies in action, a site on writing captions, and a workflow with links to the Thesaurus of Graphic Materials (TGM) and a saying I saw somewhere and was quite fond of "If you are using a controlled vocabulary, the terms are not suggestions."

Turns out, if they are not explained correctly to your workers those terms are totally suggestions. What I got back at the end of the project had very little resemblance to TGM. And I want to stress here, this way my failure, not the students. I failed to impart the importance of following the terms in the TGM exactly. So I think what happened was, my student workers saw a picture with one boy in it, saw boys in TGM, and said well it's not a picture of boys it's a picture of a boy, I'm going to put in boy. And this leads to item one in what I lovingly call, "Eric's Big Catalog of Inadequacies."

[SLIDE 9]

One: Especially if you don't work with the public much, remember that your students or volunteers do not understand this the way you do. More training, more examples (both good and bad), more going over stuff with them, and especially why it's important to do something a certain way. Do it until they can explain it back to you, even if they think you're just being persnickety.

Bad news is off to a bad start, and I have no idea. But Eric, why didn't you check this stuff? Where's your thinking?

Okay, my plan was the two students would generate the captions and TGM terms for around twenty-images. They would then hand each other their lists, check each other, making sure terms appeared in TGM, see if captions made sense, and help each other. Except, students have busy lives. Unexpected things happen. One student worker worked for two hours, and then quit partway through the first semester. So I couldn't tell the other student worker to stop, but they weren't getting checked, because if I had time to do this I would have done it instead of writing a grant. But I got a third student worker, but by now student worker 1 was way ahead. I decided the new worker would generate captions and terms until the end of fall semester to get good and familiar with the process, then I'd ask them both stop in the spring and they'd both go over all the captions.

Except at the end of the fall semester, that student decided they wouldn't be coming back in the spring. And I want to stress, both of these students made the right call. My 5 hour a week ten dollar an hour job should not have taken precedence over their education or whatever else was going on in their life in any way shape or form. But, I still had a project, and the only thing I could imagine salvaging was being able to stand up here and say, "Then I got another student worker, and that one burned down, fell over, and then sank into the swamp." But luckily that fourth student worker was great, really saved my neck in terms of working extra hours to get things done and use up the hours I had requested in my budget.

[SLIDE 10]

Which brings us to item 2 in Eric's Catalog of Inadequacy: Plan for Attrition. Not only will the people you end this with maybe not be the people you start it with, but if your project needs two people, have a plan for what you're going to do when there's only one. And all the usual things apply, keep track of where people are putting data, get it from them periodically, but also look at that plan, and ask yourself, what if a person drops out? What do I do then? How do I keep this project going to avoid losing the time and effort that has been invested.

Doing this might save you having to proof all your metadata at the end, by yourself. Which by this point, I had pretty much decided what I was going to have to do. But I still thought I had solid stuff that wouldn't require much work. So, I asked the students to just copy and paste their

excel sheets to Google Docs for formatting. At which point I found out a couple of things. One, Google Docs is hard to re-format in. Versions makes it difficult to make changes and just close out and re-open the document if you don't like the results. You can do it, you just have to be a lot more careful that you're looking at the version you want to be looking at. Also, I realized that both of my students had laptops, and they had wanted to do their work on them, sure, why not, there's limited terminals in Special Collections anyway. Except they had different versions of Excel. And that led to slightly different formatting in Google Docs.

So, item 3 in my catalog, make sure your people are working on the same software and same version, or test for compatibility issues.

[SLIDE 11]

There are some proprietary formats out there, and you might have a lot of trouble merging data between some of them. My problems with this were fairly minor I just dumped both spreadsheets into my version of Excel, saved as, and got on with it. But thinking back, if one of them had been on a Mac or Unix system, depending on how they saved that file there could have been major problems. You know that feeling of dread you get when someone tries to load a Keynote presentation on one of these laptops and there's a real long pause? Yep, kind of like that.

Speaking of Excel, it has some very bad habits. For instance, if someone types "Automoblies" instead of Automobiles, the next time someone enters Auto, Excel will help them by suggesting Automoblies. It is auto-correct with that passive-aggressive "Hey, I did exactly you told me to do, and if that wasn't what you wanted me to do, that's not my fault."

[SLIDE 12]

So, point 4, think about your software carefully and train your students or volunteers on it. I gave them some basic stuff on Excel, and they were familiar with it generally, about like most people where if you want a big table you do that in Excel even if you don't need a spreadsheet. But I recommend thinking very carefully about how your software will impact your data and your project. There are a lot of cool tools out there, some of them even free, and often we just grab MS Office Suite because it's ubiquitous. But consider every annoying trick it has ever played on you, and how that might break something. And then warn people, train them. Help that not happen.

At this point, we're finally coming around the bend. And I saw this problem coming, the student workers had sometimes asked me what the final product was going to look like, and they knew it was going to be a book that contained their captions and metadata, but when I said, "It's your project too, let's figure that out, what do you think it will look like." The silence was deafening. I think they were okay with doing the data, because I broke that down into small chunks for them. When I just said, "We're gonna make a book!" that was too big of a project. I should have broken

that down as well. So, final thing, and this kind of ties back into the first, remember to make all phases of the project manageable for your volunteers.

[SLIDE 13]

Break every portion of the project down into small, manageable chunks for your students and volunteers. That way they can see step by step what they're going to do and don't have to step outside of their comfort zones too far.

So I ended up doing the proofing, formatting, and making the book because of the initial plan being thrown so far off that by the time we were done with captions and metadata, spring classes were over and the students were gone. Which wasn't a big deal, but I'd rather they had those experiences too.

So what did we do right? These aren't as funny so I can get through them pretty quickly. We created a 43 page guide which listed captions for 618 illustrations or scrapbook items, and about 2913 metadata terms. This is freely available for download from our IR, and I will include a link in the finding aid for the collection as soon as we have a working Content Management System for our finding aids again.

[SLIDE 14]

We also made the excel file, in a comma separated value format available, I have no idea what anyone will do with this, but why not? I made a quick word cloud, it was kind of interesting. Lenski drew a lot of boys, girls, and trees, not as many fish.

[SLIDE 15]

It's not under any copyright, people know they can do anything they want with it. It's intended to help people who want to find things in the collection, but I always hope people will do something cool.

[SLIDE 16]

You know, it's a little Pollyanna but work with students and volunteers, their enthusiasm really is contagious and its very rewarding.

[SLIDE 17]

And finally, read your session title out loud because, oh well, you know, maybe next time. Thank you very much, my name is Eric Willey, it has been a pleasure talking to you.