

Digitizing Hidden Special Collections and Archives: Amplifying Unheard Voices

Session 2: Assessing Collections for Digitization
March 13, 2024



Links mentioned:

- [Program Homepage](#)
- [Core Values](#)
- [Apply for an Award](#)
- [Native land](#)
- [Application Guidelines \(view only\)](#)
- [Application Guidelines \(make a copy, Google account required\)](#)
- [DLF's Digitizing Special Formats wiki](#)
- [Projects funded through Digitizing Hidden Collections, 2015-2020 and DHC: Amplifying Unheard Voices, 2021](#)
- [Registration for the Applicant Support Series Session 3, March 20](#)
- [Webinar Survey](#)

Questions and Answers

All questions asked in chat and via the Q&A box during the live webinar are copied below. Some questions were answered live during the March 13, 2024 webinar and are marked. Any questions answered live may include additional references or clarification. If you have any other questions, email the CLIR Grants team at hiddencollections@clir.org.

Q: The collection we have proposed has been surveyed and inventoried at the box level. We are doing a deeper survey for this grant application but we also hope to receive funding to hire someone to further help with processing and selecting for digitization at the item-level. Is this a good approach for these types of collections?

A: (Mitch Fraas):The part I can't answer that I want to refer to the CLIR team is the cost and budget financial implications of the cost of metadata and how that fits into your overall plan and budget. Some archivists and researchers interacting with materials digitally, often expect item level interaction. I cannot say that it's always a good thing, and there are many reasons why this is not always tenable. I think we cannot say that it is not common, and that is not really what people expect, even if we sort of wish people followed archival practice and we're okay with box level description. There are a lot of great archival reasons why item level description is not tenable, but nonetheless it is often very helpful. For example, if you're trying to make an argument about engaging specific audiences, and you can say, we think that no one is going to look through a 1000 page PDF of financial images at the item level. We don't think we can do

that. The justification is it's described as financial minutes, it's got 1,000 images, it's a thousand page PDF, and we just think that is just not going to engage audiences. We think that is not going to do it for the purpose of this grant. In other cases, you can make an argument about the box labeled with a particular technical report on athletic achievements or whatever. So it's just like a thousand page report about a very specific subject. Someone's gonna go to that and flip through it and skim and read, and they sort of know what it's about. It's a concrete thing. Then, doing some sort of item level description volume one is about this, volume 2 is about that. The budget implications and the grant application things are significant, so I'll let the CLIR folks answer.

A: (Program Officer): The program allows for funds to be used to prepare collections for the digitization process. Staffing is an allowable expense for projects and is a good approach to handling this expense. Competitive applications ensure they are offering an equitable wage for their area and include job duties, wage, hours, and fringe benefit amount. The reviewers are primarily interested in seeing that your budget detail and narrative align with the project timeline and digitization plan.

A: (Mitch Fraas): The one thing I'd add quickly, the Muslim world grant we did, we were able to make the case for using some of that money for metadata work, because we felt that Arabic script, metadata work was essential and we could not do the project without describing these things. The way we sort of justified that was saying, the institutions would pick up and greatly reduce their internal digitization costs to help offset this. To say, we know that this is a significant expense, so instead of charging \$2 and whatever 47 cents per page. The universities are gonna say, we're charging a dollar and 43 cents, whatever it is, so that we had to make some big trade offs to make that work possible within the limits.

Q: What advice do you have for those of us who are digitizing materials on a specific subject from across multiple collections? What should we consider when selecting materials for digitization?

A: (Mitch Fraas): That is a good question. The medical hub grant, actually, was sort of like that. Often people were picking material relating to Philadelphia medical education, and it was one folder in a collection of someone's papers who was in Philadelphia for a month and wrote about medical education. The big question is how are you doing the selection? In some cases, there are very obvious ways of topic selection. On this topic, we will choose this topic. If you are making a choice on this topic, then we can't choose everything. What is the criteria? Is it justifiable? Does it make sense? Are we going to focus on the first generation? Narrowing down criteria and I would prefer, making clear to people what this topic is from, and what area it is, but related to a collection. For example, we are going to do everything created before 1990. Is it justifiable? Does it make sense? We are going to focus on the first generation only and apply for another grant for future generations. Within the digitization project make clear the provenance or source of the materials.

Q: Our audio collection was donated by the producer of the collection based on his ownership rights and there were no performance releases collected from the musicians. Our digitization workflow will include metadata work as well as gathering releases from the musicians or their estates. This affects access outcomes. There are 200+ reels with that number of performances. Will you have a webinar on this scenario?

Sub-comment: We are applying for the grant to pay for this hunt for releases.

A: (Program Officer) This will be best addressed in our Rights, Ethics, & Reuse Applicant Support Webinar Session 4. This session will include information on how you will address nominated materials from an access perspective and what rights issues you may have and their explanation of what rights you have. This activity is something that can happen during the process, however it can be rather lengthy. You may want to begin this process before you are awarded funding so you have a clearer understanding of the rights, ethics and reuse prior to the project start date.

Q: Did you leave out materials that were part of the same collections to avoid copyright violations? Or did you digitize and set up controlled access?

A: (Mitch Fraas) In none of these CLIR Grants did we leave out for logical material for copyright reasons. In the Marian Anderson collection, we chose to just do her home recordings. Because we knew the rights were gonna be okay on that. We definitely chose not to do some other audio because of copyright. Within those logical series, like archival series, we did them all and just avoided an entire archival series. Controlled access is something we are currently struggling with. We have somewhat controlled access in the reading room. Our long term solution is for controlled access. I think in the future that'll be a big part of determining things and allow more flexibility. Instead of just saying we'd love to digitize the series of these recordings in their entirety, but half of them are problematic. We might be able to say, yes, we'll do it because we have an access solution, as opposed in 2017, we're just gonna avoid that whole series for that reason. This will be covered in the rights, ethics and reuse session, but there are different risk tolerances. I would say a lot of academic university libraries vary wildly on how risk tolerant they are and than smaller and community organizations. I think there are some smaller organizations who have a pretty high tolerance for risk, knowing that they aren't a billion dollar institution that's likely to attract the ire of a recording label. It doesn't mean there's no risk and it's not problematic. But I think within legality there's a lot of risk. It's not a question always of strict, "are we allowed to"? It's sometimes a tough judgment call. So, I'm sure that will come up, and the kind of institution you're at, and who you work with will affect that.

A: (Program Officer) The rights, ethics and re-use session will clarify many of your questions. The new iteration of the program may create some instances where the ethical disclosure requires this strict determination. In these cases reviewers will want to see you understand the rights, ethics and reuse issues and have a plan in place, or if your organization requires some fees associated with it. There is a place for that information on the collections form, and the explanation/justification for these decisions.

Q: Following up on your answer to one of the last questions about level of metadata, if we are digitizing 63,000 images we don't necessarily need to create 63,000 metadata records, one for each negative? We can possibly do one metadata record for each envelope of negatives since the envelopes contain several negatives on the same subject. Is that acceptable?

A: (Mitch Fraas) The reviewers may have different opinions. I think largely, given our internal debates here, I know we have curators, people with different opinions. We have made decisions that the envelope, folder, contact sheet level make a lot of sense. When I was thinking of an example I saw the other day, which is from a particular event. A musician was visiting a high school, and someone had taken 130 photos.. So, it's a lot of photos. We're not gonna put 130 photos with individual metadata. What people want to know is, here are the 130 photos of that event on August 3rd, 1985, with this person, and you can scroll through the 130 photos that are slightly different head positions of people. And I think that makes a ton of sense.

Q: How did you avoid feeling overwhelmed coming up with quantity estimations? lol did you have someone help you come up with those numbers?

A: (Mitch Fraas) I think I had our digitization head, he was helpful given his experience and he could eyeball time estimates or a gut check on what it would entail. In retrospect this could have been great for a work study student to do more sampling (and not every institution has this), and counting pages on a couple of these. Having more people involved, there was a lot more off the cuff I would have done differently, knowing more now. This was during the pandemic and to CLIR's credit, we were in contact with them all the time, when we hit our image limit, where other institutions were closed and no digitization going on, we could employ our folks to do digitization. Something we realized pretty quickly and communicated about it. There are solutions. There was communication with the PIs and with CLIR all along so the estimation error wasn't something discovered 3 years into the project. It was something we realized pretty quickly and were able to come up with a budget plan with CLIR on whether it needed a change order, or whatever the approval was. I shouldn't be so cavalier about it. It's an example of where estimation could matter, and also know there are solutions. It's not a disaster. There are ways to work around it.

(Program Officer) If awarded funding, we have a grant modification process you can use. If you find estimation issues, you can submit a grant modification to change the deliverables of your project and increase or decrease your collections list. Sometimes, the digitization costs are less than expected; a project may want to add additional materials that fall within the original project scope. This is another instance that a grant modification for change in deliverables may occur.