

# 2018-11-20 Meaty email thread

Hi Megan,

Now that I'm back from Denver I just wanted to follow up on scheduling. Is next Tuesday still good for a visit? What's the best time for me and Josh to be there to see the process (scanning, spreadsheet, quality control) in action—or at least “emulate” it—and be able to brainstorm with you all?

Also, to follow up on some questions I was remiss about before:

- I stand corrected. We don't actually know the number of files—a major part of the project is for us to better understand the true extent of the files we have. We've estimated 52K based on our previous pilot. (Sorry I didn't fully grasp this before as I'm still relatively new to the museum.)
- Let's discuss timeline and deliverables. We really want to make this work with you.

Thanks,

Sherri

**From:** Berger, Sherri

**Sent:** Tuesday, November 13, 2018 6:14 PM

**To:** Meghan O'Brien <[Meaghan@TheCrowleyCompany.com](mailto:Meaghan@TheCrowleyCompany.com)>

**Cc:** Brady Wilks <[BradyW@TheCrowleyCompany.com](mailto:BradyW@TheCrowleyCompany.com)>; Gorman, Joshua M. <[GormanJ@si.edu](mailto:GormanJ@si.edu)>; Cutler, Alicia M. <[cutleram@si.edu](mailto:cutleram@si.edu)>; Kurasz, Stephanie <[kuraszS@si.edu](mailto:kuraszS@si.edu)>; Blanchard, Melanie <[BlanchardM@si.edu](mailto:BlanchardM@si.edu)>; Pat Crowley <[Pat@TheCrowleyCompany.com](mailto:Pat@TheCrowleyCompany.com)>

**Subject:** Re: clarifications, next deadline, return visit

Thanks so much Meghan. See below for comments on comments :-)) and pls excuse weird formatting via phone.

Sent from my iPhone

On Nov 13, 2018, at 3:52 PM, Meghan O'Brien <[Meaghan@TheCrowleyCompany.com](mailto:Meaghan@TheCrowleyCompany.com)> wrote:

Hi Sherri,

Sorry for the delay in getting back to you—

To answer your questions and a little more information:

1. The spreadsheet you sent back was great, with one exception: please make sure that accession numbers retain their full 6-digit extension—meaning that when a number ends with a zero, that zero stays put. It looks like right now they are getting cut off due to the formatting of the cells. Thank you! **This is a weird one, we are seeing the following zeros on our end, so we are not sure why this is occurring, but we are aware of the issue.**

Hm I wonder if it's how the file was saved? Perhaps if it is saved as csv it won't do that -- or just make sure the formatting of that column is "text"? We can try to investigate too on the receiving end. Thank you!

1.

1. Currently you are sending back text files as separate text files for each *image*. Would it be possible to receive a single text file per *accession number/file/PDF*? I'm not sure at what point in the process you are doing the OCR, so I don't know whether that would be difficult or not. Let me know and we can discuss. **We are looking into this, but we should be able to provide you with a single OCR/TXT file per digital file, not per image. Stay tuned on this.**

Super, thanks!

1.

Meanwhile: our expectation, based on the rate you estimated on the recent call, is that the images/metadata for Shipment 3 (20 boxes) will be delivered to us on December 3. Can you confirm that is a realistic due date? **We do not think this is going to be possible, but we will know more by the end of the week, see below for more thoughts on the timeline.**

OK, could you let us know what a realistic date would be given your current rate of throughput? We thought we calculated this Correctly based on what Brady said in the call he hoped to work up to, but perhaps our math was wrong. Throughput questions aside, it would be helpful to know when this next batch will be ready.

And finally: would you be amenable to a return visit, by Josh and me, to see the process of scanning these materials? We are both available Tuesday, November 27, if there is a good time on that date. As you know, we are keen to refine this mass digitization process and make it as quick and efficient as possible. If we were able to see you all in action, I think we could get a better idea of ways we can better prep the files, and perhaps develop other creative solutions together. What do you think? **Yes, definitely, that is a good plan. What time would you like to come in?**

Right now both Josh and I have held the whole day. What's the best time for us to be there to see in action and discuss with the folks who have been working on this content and could help us think through this?

**We think the issues that are most impacting the throughput are: often very few pages per folder, and envelopes within folders containing pages which need to be removed/reinserted.**

We can talk about this and ways we might be able to prep in advance or perhaps not scan envelopes within envelopes.

**This doesn't seem very challenging but is much more time consuming than opening a standard file folder with loose pages within...let's think about whether anything can be done here.**

**Also: we'd like to make the suggestion that going forward, we only use/have two sizes of "backing pages" that we use for the onion skin paper: 8.5x11" OR 11x17" and operators will use the smallest page logically for backing. We are finding some of the operators are taking the time to cut backing pages/or use multiple sizes of backing pages to create the best look. I think if we can agree to two sizes of backing paper across the entire collection, this will help to streamline.**

That sounds just fine to me given that this is a mass dig project for non-collections items. Perhaps you could do a sample of this for us and we could review the images just to make sure this is ok for our purposes? I really expect it will be.

**Finally: regarding the schedule...**

**Most importantly: is the 12 month time frame a hard/written in stone contract date? (I believe we received the award in early August) Or, might there be a possibility for an extension? If it is a hard deadline, then starting Dec. 1, through (end?) of August, we'd need to turn around 40 boxes per month. Do you think you'd be able to have 40 boxes per month/be able to review 40 boxes per month on the backend? Let us know your thoughts on this important detail?**

Let me check with Josh and Melanie about both the contract requirements and our highest possible throughput.

**Finally: do you have any way of knowing the total number of files/accession numbers in this collection? I have the number: 52,500 files in my records, but I wonder if you know the actual number of files?**

To be sure I know what you're asking: how many physical files there will be? I believe for the most part it's a one-to-one accession file per physical file /envelope -- but some accessions do have multiple files. Melanie will have to weigh in on this. Sorry I don't know better off the top of my head!

**Please let us know your thoughts on all of the above.**

Thanks so much,

Meghan

Meghan O'Brien  
240-215-0224 x 221  
717-515-9091 (m)

**From:** Berger, Sherri <BergerSh@si.edu>  
**Sent:** Monday, November 12, 2018 6:04 PM  
**To:** Meghan O'Brien <Meghan@TheCrowleyCompany.com>; Brady Wilks <BradyW@TheCrowleyCompany.com>  
**Cc:** Gorman, Joshua M. <GormanJ@si.edu>; Cutler, Alicia M. <cutleram@si.edu>; Kurasz, Stephanie <kuraszS@si.edu>; Blanchard, Melanie <BlanchardM@si.edu>  
**Subject:** clarifications, next deadline, return visit

Hi Brady and Meghan,

Thanks very much for sending over the most recent deliverables last week. We have two requests coming out of our Q/A:

1. The spreadsheet you sent back was great, with one exception: please make sure that accession numbers retain their full 6-digit extension—meaning that when a number ends with a zero, that zero stays put. It looks like right now they are getting cut off due to the formatting of the cells. Thank you!
1. Currently you are sending back text files as separate text files for each *image*. Would it be possible to receive a single text file per *accession number/file/PDF*? I'm not sure at what point in the process you are doing the OCR, so I don't know whether that would be difficult or not. Let me know and we can discuss.

Meanwhile: our expectation, based on the rate you estimated on the recent call, is that the images/metadata for Shipment 3 (20 boxes) will be delivered to us on December 3. Can you confirm that is a realistic due date?

And finally: would you be amenable to a return visit, by Josh and me, to see the process of scanning these materials? We are both available Tuesday, November 27, if there is a good time on that date. As you know, we are keen to refine this mass digitization process and make it as quick and efficient as possible. If we were able to see you all in action, I think we could get a better idea of ways we can better prep the files, and perhaps develop other creative solutions together. What do you think?

Thanks!

Sherri