# Researcher Bi-Weekly Reporting Template

The purpose of this document is to help prompt reflection and capture the progress, barriers, and needs you have as you undertake your project. We’ve included some preferred and suggested prompts in each section below. This document is to be completed every two weeks and submitted according to the reporting schedule.

Please append the filename with the date of submission and your initials (ex: CCHC-Biweekly-Reporting-Template-**2021-06-01**-**od**.docx). Once completed, please submit to [**LC-Labs@loc.gov**](mailto:LC-Labs@loc.gov)and CC Jaime Mears ([jame@loc.gov](mailto:jame@loc.gov)).

**Name:**

**Reporting Period: 10: 10/4/2021-10/15/2021**

### No Work Completed

We anticipate that there may be periods in your project during which you make little progress. In this situation, we ask you to check the “Nothing to report” box. After the “Reason” prompt below, please briefly indicate why there was nothing to report. Where possible, we’d still like to learn more about the reasons behind any unexpected events or barriers you encounter. You can share additional details in the “Obstacles/Needs” section before submitting this report.

Nothing to report

Reason:

### Activity and Progress

*We’d like to hear about your progress over the last couple of weeks. Please feel free to expand on the questions below.*

1. **What did you and your team work on over the last two weeks? Why?**

*We’d love to hear about your decision-making, the technologies (ie. cloud services, hardware, and software) you’ve been using, why you’ve adopted them, modifications or hacks you’ve had to create, prototyping and testing, and anything you’d like to share about your current in-progress deliverables.*

We focused on preparing for the CCHC Advisory Meting and then processing the conversations. For the meeting, I prepared the slides about the project as well as a challenge. We focused on the issues with the API. They included:

* Accessing the amount of images and getting blocked.
  + This is particularly the case for FSA. There seems to be a limit to how deep the API can go in terms of pages.
* Unclear relation to other sources of LOC data, and therefore it is unclear which is the primary data.
  + This is the issue that we discussed the week before and are going to get clarity from PPOC.
  + It is also an issue because we are trying to think through the primary source of metadata for the visualizer.
* Unclear when data has been updated since we can’t find clear documentation of change history.
  + It can be challenging to know when and which data has been changed. For example, we know having worked with FSA that the data looked differently 10 years ago. Now, PPOC is using the M # to put the strips of photos back together to then suggest more metadata such as captions. Are there other changes like this?

While we offered these challenges, the actual conversation in the breakout shifted. We should have just focused on the visualizer in our presentation, which was the focus of the discussion. We shared and it led to a very interesting discussion about metadata. At first, we were discussing which version of the data and then if we surface all the data. This led to the question about whether to reveal the metadata at all. For example, John Hessler loved not having the metadata. He argued that it was more experimental and made people ask questions such as what am I looking at and what are these connections? He thought it created more serendipity. This also brought up questions about context. Do the images just become decontextualized? The metadata does important work to situate them in time, space, and other features. The discussion was short, so we only began to scratch the surface of this. Taylor and I have been thinking about it and discussing all week!

We were also asked to debrief about the meeting over Slack. Including what I wrote here:

* I found them all really interesting. I thought Adromeda's presentation about having to think through the results of her models to be fascinating. I found Lincoln's provocation about the "monocrop" and the recap of the conversation to be really interesting.  I also found it really helpful to get the big picture of where the project is and the other pieces. It helped me better understand how we continue to connect into the goals/ aims of CCHC.
* We shifted gears very quickly to the visualizer, which I shared with the group. I've been really thinking about John Hessler's comments in our breakout group about ADDI.  He really like the idea of there being minimal metadata from the LOC on the visualizer. He thought the way it made the user think differently about the pictures was really interesting.   It was as if the metadata got in the way of seeing the photos in the new ways that the computer vision algorithms were asking of us.  I'm still thinking through this relationship of metadata to CV results/as the primary way to see the photos. The line of thought made me think later about how powerful text is and how authoritative it is in search and discovery. Still very much thinking through this and welcome others thoughts.
* Yes! The discussions brought in a lot of perspectives from different positions that were helpful to think about and hear. I think because they were all part of a dynamic conversation about the projects and then data and ethics that it's hard to tease out a specific point. Rather, it was about the ways of thinking/responses to the projects beached on peoples’ areas of expertise that was generative.
* I reached out to Liz Lorang in the chat during the session to get some more feedback from her since we didn't quite have enough time in one of our break outs to finish our conversation.  Curious to talk more about this mono crop/ least common denominator issue, which I found to be a very powerful articulation of a major disciplinary debate that I'm also interested in.  One of the major questions I left thinking about is how these projects connect to / speak to the last set of prompts  about data, ethics, institutions, and power. Those are huge questions that we just scratched the surface of.

1. **What are your plans for the next two weeks? What excites you about those plans?**

* Continuing to think through the role of metadata in the visualizer.
* Hoping to get clarity about the metadata from PPOC.
* Looking into the image storage issue for the visualizer.

### Obstacles/Needs

1. **What barriers, if any, have you encountered over the past two weeks? Were these barriers expected?**

*Examples: unexpected costs; API limitations; organizational barriers; data issues; technical gaps or challenges; collections knowledge.*

I don’t think there have been many barriers. Lots of interesting discussions to keep processing!

1. **What would resolve these barriers (or would have resolved them, in retrospect)?**

*We’d love to know what you need help with! This is an opportunity for LC staff to learn from you about how to best support this work going forward.*

I keep returning to the issue of documentation. I’m learning so much in the 1:1 conversations and so much institutional knowledge is held by people. It would be great to get some of this written down and documented! I also recognize that there are only so many hours in the day.

1. **What changes, if any, do you anticipate to your proposed timeline? Why?**

*Do you expect any significant blockers to your progress?*

I don’t anticipate changes at this moment.

### Reflection

1. **What happened over the last two weeks that was particularly thought-provoking?**

*Examples: surprises; places where help from Library staff was critical; incorrect assumptions; things you learned; things that, in retrospect, you wish you’d known.*

The Advisory Board meeting was interesting! Between the question about metadata in the visualizer and the monocrop idea, it’s been a thought-provoking week. I don’t love the term monocrop, but I take the general point. I’m not as critical of the monocrop approach as Lincoln. I think that is shaped by the (inter)disciplinary lens that we come to this issue with. It was a reminder about the importance of audience. It’s also interesting that Andromeda is coming at this question as well by trying to find everything about reconstruction across collections. I’ve also been thinking about Andromeda’s work a lot because there are a lot of parallels with ADDI. She’s kind of coming from the issue of ML results from another angle. I’m using ML methods out-of-the-box and seeing if we can use them and/or reframe them, with attention to if the results are problematic, irrelevant, or nonsensical. What I understood is that she is training the algorithms to interpret in a way that helps one analyze across a large, multivalent collection that shares a general scope (i.e. reconstruction). Because of the materials that she’s working with, the algorithm is surfacing the racist ideas in the archive (if I understood her correctly). This is, of course, no surprise. Given the way that machine learning works, is this an algorithm that surfaces racism in the archive as well as learns to be racist? Is this just a central tension of this kind of computational approach given that ML is designed to learn from the data? I left kind of worrying about what the larger implications of an algorithm trained on these materials, which are infused with all kinds of problematic ideologies that the algorithm might be learning. But maybe I’m misunderstanding, and I’m very much off here. Regardless, this has made me reflect on the algorithms that I’m using and larger questions about the role of ML in our work.

1. **Anything else you’d like to add?**

As always, thank you to everyone for all their work!