# 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: 6: 8/9/2021-8/20/2021**

### 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 began thinking about the prototype and what kind of data should be in it. We had worked out the computer vision metadata and then turned to the extant archival metadata from LOC. We had the MARC records, but the problem is they don’t entirely align between collections. So, after the discussions of the APIs in the group and with the API team, we decided to go back to API and try again.

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

We are planning to get a working version of the visualization software running on one collection. It will be exciting to see the annotations over the images and have a more streamlined way of viewing the output of the computer vision algorithms.

### 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.*

We tried to use the API and got it working better with the documentation on the LOC site (not the GitHub). It worked for a bit but then we hit another restriction where we couldn’t get the data we needed. The issue was “depth”. We were paginating through the FSA and it will only let you go through a certain depth. So we were only able to get to about page 700 and ~90,000 images (which is just half). It’s not a buffering or waiting issue. The URL will always fail. Here is the URL:   <https://www.loc.gov/collections/fsa-owi-black-and-white-negatives?c=150&fo=json&sp=900>. There is no marc record.

There are also two pages for each image and two marc records. What’s challenging is there are so many different pages. For example, there is <https://www.loc.gov/item/2017798400/> and then the same image from the PPOC collection <https://www.loc.gov/pictures/collection/fsa/item/2017798400/>. Then, there are the MARC records. For example:

Same Photo - Marc 1: <https://www.loc.gov/pictures/item/2017798400/marc/> (XML link from PPOC database: https://www.loc.gov/item/2017798400/)

Same Photo – Marc 2: <https://lccn.loc.gov/2017798400/marcxml> (XML link from LOC aggregate: <https://www.loc.gov/item/2017798400/>)

So there are the two versions on the website, the MARC records, and the API. The API has lowercase names and the locations are different formats from the others, for example. The two different website pages are different too. One has each location separated and the other one does not. One has an actual "city" field and the other does not. There may be more inconsistencies, but we haven’t methodically checked it. We don’t have a full version of the LOC aggregate because the lccn site is fickle to scrapping. The XML server has lots of restrictions that don’t seem to be documented.

I think this made more sense after talking with the API team since they were talking about how different database and pulling into different parts of the site, and how PP has its own world and then another database is pulled into aggregate LOC search (if I understood correctly).

All to say, it would be helpful to have documentation about the different versions and what info is where, and if one version is the “main” and/or latest version of the data.

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.*

Removing the depth restriction to let us go further in with the API. Otherwise there is no way to get the entire dataset we need for the project through the API.

While less necessary, for the tasks of the grant, we would love to be able to access the XML Marc data from the lccn.loc.gov server. The data is super structured, but the rate limiting and need to fetch each record individually makes this unfeasible for these collections.

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

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

None at this time.

### 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.*

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