# 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: Lauren Tilton**

**Reporting Period: 1: 5/31/2021-6/11/2021**

### 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.*

For our work through June 11, 2021, we worked on getting the data. This included metadata and the images.

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

The plan is to (1) incorporate each file into a database (i.e. each image is a different file and needs to be part of larger database), (2) check the data, and (3) deal with any data cleaning issues.

### 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 began by turning to the API. However, this was a challenge. There were two websites with instructions with different levels of information that aren’t linked to each other (at least we couldn’t find it). They included:

* <https://www.loc.gov/apis/json-and-yaml/> : It provides information about making a request. It doesn’t have the tutorial or additional info that the GitHub has.
* <https://libraryofcongress.github.io/data-exploration/index.html> : This info is quite helpful and more detailed. However, it doesn’t include any information about the rate limit.
* <https://www.loc.gov/apis/micro-services/image-services/>: We went to this one and it was like <https://www.loc.gov/apis/json-and-yaml/> but just for images. We did not use the IIIF version. The level of detail is helpful.

The main challenge was that we tried to set it up and then ran into strict rate limits. Given the scale of our requests, this was a huge issue. It was going to take us days for each collection. So, we decided to bypass the API and scrape the data. It took a bit of work to figure out the URL scheme and then we were able to download all the data without being stopped. It took several days still because we had to fetch each record one by one, but this was still substantially shorter and less complicated than dealing with the API.

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.*

It would have been easier to go to a site or make a request to LOC where we could have received all the data at once. An example includes the Rijsmuseum API (<https://data.rijksmuseum.nl/object-metadata/api/>). They have a “download” section with zip files with multiple formats (<https://data.rijksmuseum.nl/object-metadata/download/>) for the entire collection.

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 currently expect changes to the timeline 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.*

It was great to speak with the LOC about this challenge a few weeks ago. It was also interesting to discuss with colleagues open data practices. Along with discussing the Rijks, a colleague brought up Clinical Trials.gov that has all of their data available for quick and easy download.

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