# 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: 11: 10/18/2021-10/29/2021**

### 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’ve been working through and soliciting more feedback on the visualizer, which we are aggregating. We keep discussing which features to implement and which not to. This keeps changing based on discussions with different stakeholders. Our plan is to have a final list of features that we will implement in several weeks all at once. We’ve decided to do this because ideas such as not including metadata would really change the design. In the meantime, I’ve been working on paper to play around with the design (like we do with any DH project which includes creating wireframes) rather than changing the working prototype.

The addition of Tori to the team has been great because she brings a lot of insight about PPOC, which has been very helpful. I thought our 1:1 was particularly generative on 10/19. We continued to discuss the metadata issue. I appreciated Tori’s reservation about not using any metadata. The discussion led to a distinction between *interpretative* and *descriptive* metadata. Much of the data for this collection is descriptive rather than interpretative, which is a very helpful way to think about the data and why to include it. The computer vision work is designed to offer more descriptive metadata rather than interpretative. It’s not that it isn’t interpretative but not in the way that this distinction is meant to signal.

During this time frame, we also heard back from PPOC about the data. This email from Christa was incredibly helpful!

Phil has it right: since the source metadata for both presentations originates with the MARC record for the item (<https://lccn.loc.gov/97518653>), the data presented in each should be pretty much in sync.  The differences between how and what PPOC and loc.gov displays (field names, capitalization) arises from application-specific processing determined by their own business rules and presentation needs.

The loc.gov presentation of content that is in the PPOC application pulls the data from PPOC (rather than straight from the catalog – so it inherits data processing that has already happened for that display); there are additional elements generated for the loc.gov display (articles\_and\_essays references, citations, rights statements) that aren’t in the PPOC display, or the MARC record, for that matter – in the three examples above, the data is derived from the broader SOLR index,  assembled from existing MARC fields, or inherited from collection framework (if available).

Likewise, the PPOC display includes generated references like “Item is from [this group](https://www.loc.gov/pictures/search/?q=LOT%2011147&fi=number&op=PHRASE&va=exact&co=coll&sg=true)” and  “[Browse neighboring items by call number.](https://www.loc.gov/pictures/related/?&pk=97518653&st=gallery&sb=call_number#focus)” Dave Woodward would know more of the details of the processing that happens with records going into and coming out of PPOC.

I hesitate to call the metadata in one presentation more authoritative than the other – probably the most ‘authoritative’ version of the data is what’s in the MARC record itself, available from the LCCN permalink above (<https://lccn.loc.gov/97518653>).

Christa

I add the email here because this is the kind of information that will be helpful for the future as LOC thinks about documentation for facilitating the use of collections as data. I could imagine a page on LOC called something like “Collections As Data & Computational Methods”. A section about the details about the configuration of the data for each area of the library would be helpful, or at least PPOC.

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

Continuing to work on the details for the visualizer. We also will be working on the presentation for PPOC, which we will be meeting with on Nov 1. Based on previous conversations, we will be working to make sure this presentation is clear about the scope of the project (i.e. CV isn’t going to replace people, rather we need humans-in-the-loop!) and offers an introduction to the methods and why. We want to make sure everyone understands what we are doing, particularly if they aren’t as familiar with computational methods such as ML and CV.

### 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.*

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.*

Getting clarity about the data has been very helpful!

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

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

I bought a house and am moving on November 20th. Given these changes, I anticipate we might need a few more weeks to complete the visualizer.

### 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.*

Continuing to percolate on the relationship between LOC metadata and the results of computer vision in the visualizer. I’ve been rereading *Data Feminism* and been thinking through some of our work through that lens. I was reading the chapter “On Rational, Scientific, Objective Viewpoints from Mythical, Imaginary, Impossible Standpoints”. D’Iganzio and Klein quote Jonathan Stray, who states, “The constraints of truth leave a wide space for interpretation”. They go on to say, “When visualizing data, the only certifiable fact is that it’s impossible to avoid interpretation” (online, OA version, Ch 3). Yet, different interpretations and whose interpretations can hold different kinds of power. I’ve been thinking about this in the context of the visualizer in a few ways.

1. One is back to this question of showing the LOC metadata, which carries significant authority. LOC is a revered institution of knowledge. I’m wondering if we need a way to indicate this. I worried that putting this metadata visually on an equal playing field as the CV results lends the CV results too much authority. So, I’m thinking of adding a box around this metadata and a short description. This also seems to align with the idea of *provenance rhetoric*. It’s an idea developed by Hullmand and Diakopoulos, which D’Ignazio and Klein draw on and define as a rhetoric “designed to signal the transparency and trustworthiness of the presentation sources to end users” (online, OA, Ch 3).
2. Another piece is the authority of computational methods. While this is changing, computational methods are given a lot of authority since they are seen as neutral and unbiased. We’ve included the confidence % to begin to indicate that this is probabilistic. I think there is more we can do and thinking through that.

As D’Ignazio and Klein also note, citing many others, visualizations also have a rhetoric. It involves choices. I really like this idea of “framing effects” that they offer. As they write, “they have an impact on how people interpret the graphics and what they take away from them” (online, OA, Ch 3). I’ve been thinking about this particularly for the bar chart representing the results from the image segmentation algorithm.

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

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