CALIFORNIA PHENOLOGY TCN – QUARTERLY REPORT – APRIL 2021

Assembled by Katie Pearson, 29 April 2021

PROGRESS IN DIGITIZATION EFFORTS

Figure 1 shows our progress in imaging, transcribing, georeferencing, and phenologically scoring the target specimens for the original 22 CAP institutions, explained more in detail in the following sections.
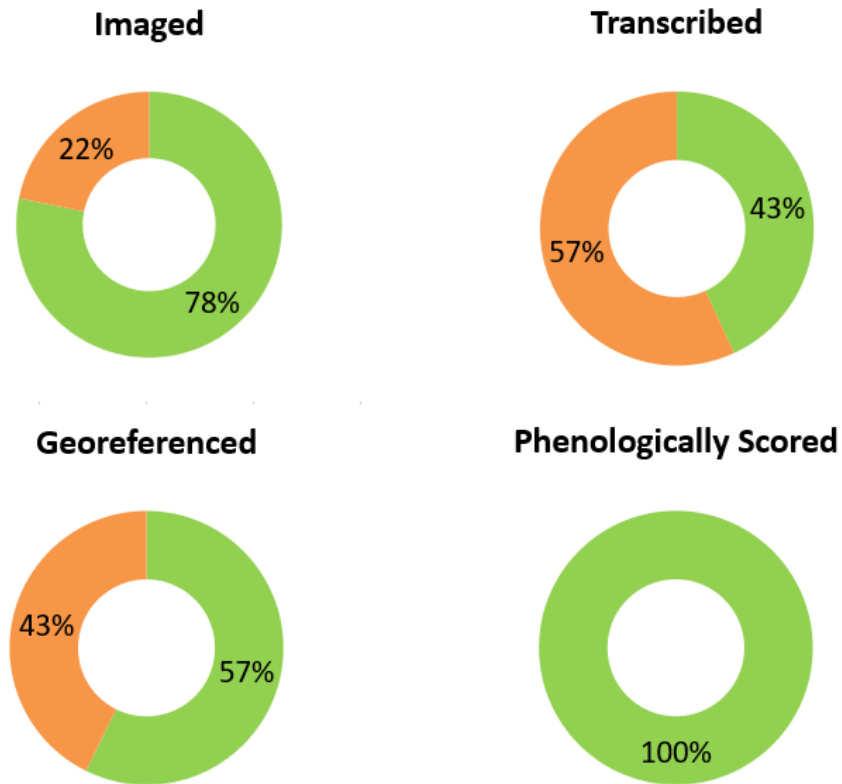
Imaged

22%
78%

Transcribed

43%
57%

Georeferenced

43%
57%

Phenologically Scored

100%

Figure 1. Progress in each of our four major digitization goals. Totals represent the original goals of the CAP grant: 902,400 specimens imaged and phenologically scored, and 300,000 transcribed and georeferenced. Additional specimens to be digitized by the PEN grant are tabulated in the PEN section below.

TRANSCRIPTION

An estimated 129,000 specimen records have been transcribed across the CAP Network since the beginning of the project. This is approximately 43% of our goal.

Transcription has largely been accomplished by institutional volunteers in CCH2 and online volunteers in Notes from Nature, including those who attended WeDigBio events (see Education & Outreach).

GEOREFERENCING

We have georeferenced 171,890 specimen records, which is 57% of our goal. Georeferencing is conducted by trained staff and students at HSC, OBI, and SD, by naturalist volunteers are part of the "100 Club," and by undergraduate students in the cross-institution herbarium digitization course led by Cal Poly. The CAP 100 Club currently has 28 active members who have collectively georeferenced over 8,000 specimens since September 2020. The digitization course consists of 18 undergraduate students who have georeferenced over 260 records since March 30, 2021.

We recognized that thousands of specimens had township, range, section (TRS) data in the verbatim coordinates field of CCH2, yet these same records did not have latitude and longitude values. We created an R script to convert these TRS data into coordinates with uncertainty radii of 969 m (1 mi^2), using a conversion table provided by Nelson Rios at Yale. With this code, we were able to add georeferences to 35,465 specimens in CCH2. The code is publicly available on Zenodo (Pearson 2021, http://doi.org/10.5281/zenodo.4507032).

IMAGING

Eleven of our 22 herbaria (50%) have accomplished their imaging goals (Figure 2). Of the remaining herbaria, eight have been able to resume imaging since the COVID-19 shutdowns. The other herbaria have used this time offsite to process images, georeference specimens, and transcribe specimens. Figure 2 shows the current state of CAP imaging as of April 29, 2021.
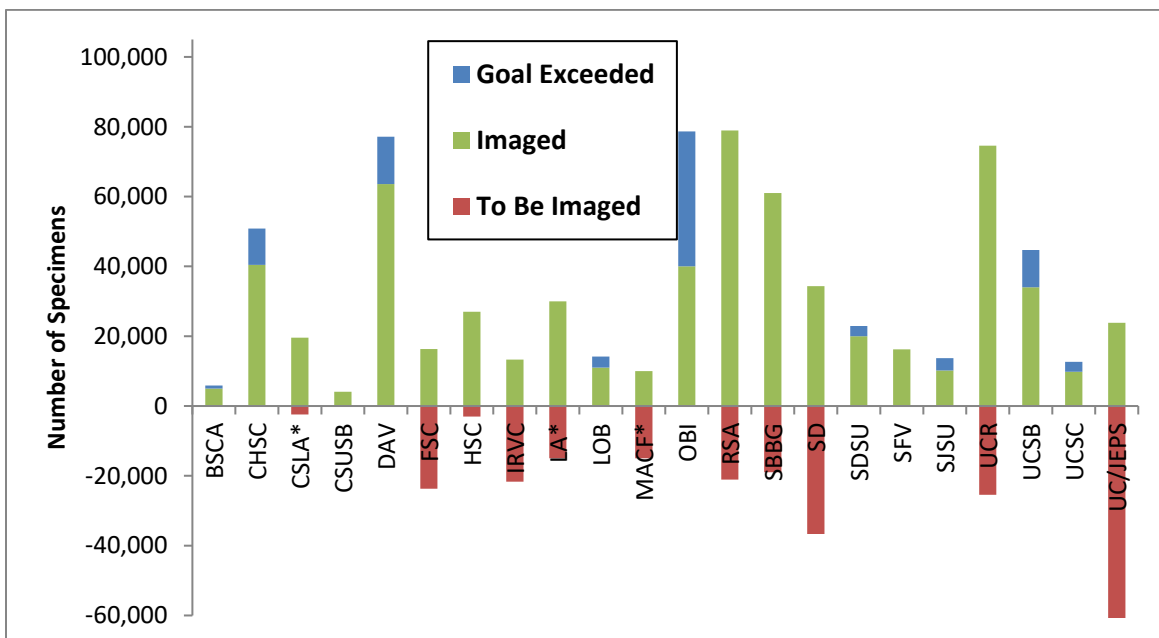

Figure 2. Herbarium specimen imaging progress. Green portions represent the number of specimens that have been imaged, while blue portions represent the number of specimens that have been imaged beyond the expected target specimens. Red bars below the zero line indicate the number of target specimens that have not yet been imaged. Asterisks indicate herbaria that are currently unable to image due to COVID-19-related prohibitions.

PEN PROGRESS

Specimen data from PUA were successfully extracted from an RBase database, coalesced by the PM, and are now managed live in CCH2. Of the 71,333 specimens, 70,163 (98%) are already georeferenced and none are imaged. We used the trait coding from text tool to phenologically score 55,633 (78%) of the specimens, which brings our total number of phenologically scored specimens to 1,430,615 as of April 29, 2021, which is 136% of our goal even with the additional PEN institutions.

Imaging has begun at SFSU, albeit with limited hours due to COVID-19 precautions. Imaging is ongoing at UNLV and OSC, which have achieved 50% and 22% of their imaging goals, respectively. CDA was approved as a subcontract as part of the NSF award as of February, but funds are still being handled by their respective offices and have not yet been liberated for purchasing equipment. CDA will begin imaging once they have acquired their equipment. The remaining two PEN institutions will begin imaging later in 2021 (SHTC) and 2022 (PUA), as previously scheduled. Figure 3 shows the current imaging progress at PEN institutions.
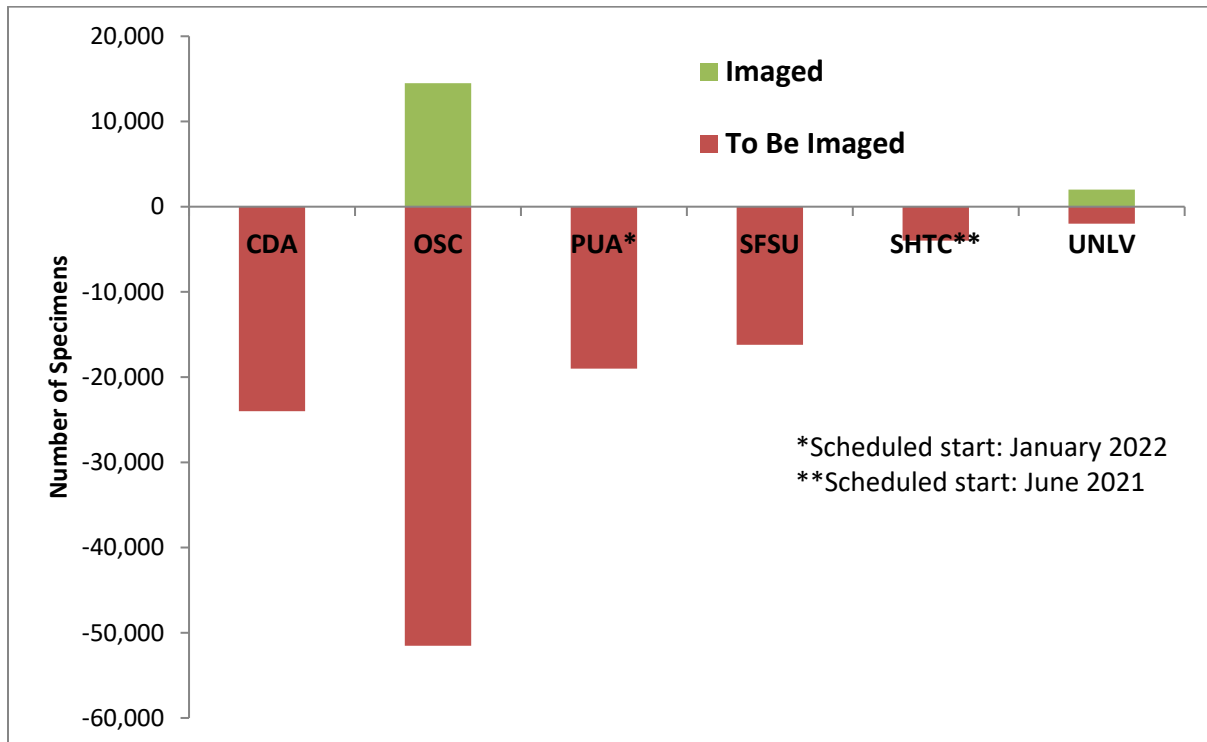

Figure 3. Herbarium specimen imaging progress for the seven PEN institutions. Green portions represent the number of specimens that have been imaged, while red bars below the zero line indicate the number of target specimens that have not yet been imaged. For SD, the total number of specimens to be imaged, including those added as part of the PEN grant, is indicated in Figure 2; therefore, SD is not included in this figure.

PORTAL DEVELOPMENT

With the help of Chris Tyrrell (Milwaukee Public Museum), we are continuing to build new functionality to the CCH2 portal that can be integrated into other Symbiota portals. Chris developed a way to search the portal by specimens that are georeferenced, which is now functional in our portal (see screenshot below).

Chris also developed a way to search specimens by trait criteria, specifically by phenological traits in our portal. This functionality is versatile; it can be used by any portal to search for any trait that the portal has enabled. This tool will be added to the CCH2 portal once it has been fully vetted by Ed Gilbert, Symbiota developer (see screenshot below).

We pushed forward the production of a traits tab that can be viewed on the public page of any specimen. The format of this tab will undergo future development, but it currently displays all traits associated with an occurrence (see screenshot below).

SHARE AND IDENTIFY BEST PRACTICES AND STANDARDS (INCLUDING LESSONS LEARNED)

As previously described, we created an R script to convert these TRS data into coordinates to the precision of 969 m (1 mi^2) uncertainty, using a conversion table provided by Nelson Rios at Yale. With this code, we were able to add georeferences to 35,465 specimens in CCH2. The code is publicly available on Zenodo (Pearson 2021, http://doi.org/10.5281/zenodo.4507032).

We have continued to hone the use of the code to extract georeference data from botanical duplicates and have created a new version with documentation, again publicly available on Zenodo (Pearson 2021, https://zenodo.org/record/4637000).

OBI beta tested a process for quickly acquiring georeferences for specimens from prolific collectors. We downloaded all specimens collected by Dave Keil (>20,000 specimens) from the OBI collection and sorted the spreadsheet according to collector number. We were then able to apply coordinates from georeferenced specimens with similar locality strings to specimens that previously lacked coordinates. We also applied the metadata associated with these coordinates to the newly georeferenced specimens. OBI georeferenced 2,273 records using this technique. We plan to apply a similar technique to prolific collectors of other herbaria.

We are working with the California botanical community at large to develop a CCH policy on protecting specimen data for sensitive taxa (e.g., threatened species). As part of these efforts, we have solicited statements from the botanical community about their opinions on this matter, i.e., whether specimen data should be redacted for some or any taxa, and collated their responses here (https://www.capturingcaliforniasflowers.org/sensitive-taxa.html). In May, we will hold a community informational meeting in which we hear statements from key stakeholders and present currently available best practices documentation (e.g., the GBIF Best Practices for Generalizing Sensitive Species Occurrence Data).

IDENTIFY GAPS IN DIGITIZATION AREAS AND TECHNOLOGY

Feedback from members of our georeferencing "100 Club" has indicated that the batch georeferencing tools could use more development to be more efficient.

We are developing a way to search for specimens based on taxon-level traits, such as CNPS rarity ranking.

SHARE AND IDENTIFY OPPORTUNITIES TO ENHANCE TRAINING EFFORTS

In March 2021, the ASU Symbiota team developed a new label-making tool for Symbiota portals. We created a video tutorial and written guide for using this new tool, available here (https://youtu.be/ult9QAWax9I) and here (https://www.capturingcaliforniasflowers.org/uploads/1/6/3/7/16372936/newlabelprintingguide.docx), respectively. We also updated the Guide to Using a Symbiota-based Portal with these and other new portal adjustments (https://www.capturingcaliforniasflowers.org/symbiota.html).

We have continued to support the work of our "100 Club" of naturalist georeferencers through email and by sending them promotional "I Digitize California Plants" stickers. On April 8, we held a 100 Club virtual co-working session in which we georeferenced and socialized. We will hold similar events monthly to promote member engagement.

During the winter 2021 quarter, we led an online digitization course in which 30 students from 9 institutions learned how to transcribe specimens and work in CCH2. We are leading a similar course for the spring 2021 quarter in which many of the same students have been trained to georeference specimens in the GEOLocate Collaborative Georeferencing (CoGe) platform. This course meets synchronously via Zoom, once per week for 2 hours. As part of the course, the students also give a short presentation on an herbarium-related topic. The 18 students in this course have georeferenced 265 specimens as of 29 April 2021.

In preparation for the spring 2021 version of this course, we developed an alternative georeferencing training course that includes a CoGe module (https://www.capturingcaliforniasflowers.org/georeferencingcourse-coge.html) featuring a new "Georeferencing in GEOLocate CoGe" video (https://youtu.be/h1JfJuSC-eg).

Oscar Vargas, herbarium director at digitizing institution HSC, is leading a class of students in georeferencing in the CCH2 portal. The students were trained using the online georeferencing course followed by a hands-on session with the PM.

SHARE AND IDENTIFY COLLABORATIONS WITH OTHER TCNS, INSTITUTIONS, AND ORGANIZATIONS

PM Pearson has been able to apply the georeference duplicate searching code to institutions of the GLOBAL TCN as part of her duties in that role. This has led to some of the improvements in the georeferencing code described previously in this report.

After finding our digitization documentation on our website, digitization managers at North Dakota State University and Polly Hill Arboretum independently contacted us about each acquiring one of the custom camera mounts that we developed to fit the Ortery Photosimile 50 lightbox. We manufactured these mounts and provided them to NDSU and Polly Hill at cost.

We have continued to increase the quality of our data, especially georeferences, by examining records that had previously been flagged by Calflora. California CCH2 data will soon be ingested by Calflora and therefore be searchable on their site.

SHARE AND IDENTIFY OPPORTUNITIES AND STRATEGIES FOR SUSTAINABILITY:

On February 16th, CAP held a meeting of PIs to discuss how to engage undergraduates more consistently in herbarium-related activities. The goal is not only to enhance education and student involvement in herbaria, but also increase the visibility and value of the herbaria to administrators and colleagues.

As we near the last year of CAP funding, CAP leadership has begun to plan for the sustainability of tasks currently completed by the PM beyond 2022. We are drafting necessary documentation to enable the continuation of certain tasks (e.g., Notes from Nature expeditions) and planning training opportunities for PIs and collaborators. One of our primary strategies is to build up the Consortium of California Herbaria as an organized, democratic society with regular meetings and task groups.

SHARE AND IDENTIFY EDUCATION AND OUTREACH (E&O) ACTIVITIES:

The PM shares updates on the project and phenology-related news via the network Twitter account (@CalPhenologyTCN).

Three blog posts were written and published to the CAP website: https://www.capturingcaliforniasflowers.org/blog-recap. Blog posts are publicized via Twitter and the "Herbarium Junkies" Facebook page.

On February 16th, CAP held a meeting of PIs to discuss how to incorporate herbarium-related activities into undergraduate courses or create herbarium-focused independent study and research courses. PIs shared their current approaches, and potential improvements and other approaches were discussed as a group. Participants agreed to share their developed educational materials via a Google Drive folder.

As previously described, we led an online digitization course in which 30 students from 9 institutions learned how to transcribe specimens and work in CCH2. In the spring quarter, we are leading a similar course which many of the same students (and some new students) have learned to georeference specimens in the GEOLocate Collaborative Georeferencing (CoGe) platform. Currently, 16 students and 2 volunteers from seven institutions are participating in this course.

For winter quarter 2021, PI Yost led a phenological research course at Cal Poly similar to our spring 2020 course, assisted by the new CAP-funded postdoctoral student (Natalie Love), a graduate student at UCSB (Tadeo Ramirez Parada), and the PM. Ten students completed this course, including analyzing changes in flowering time with climate, writing a research paper, and presenting a research poster virtually. PI Mazer (UCSB) is now running this course with 15 students at UCSB for the spring 2021 quarter.

The UC Riverside Botanical Garden Newsletter published an article about an indispensable herbarium volunteer at the UC Riverside herbarium, thereby publicizing UCR's digitization work (https://drive.google.com/file/d/1BzkomSBifYfsmtkIf-tPsXrKUyhg9j5j/view).

The CAP Network participated in the spring WeDigBio event by leading two virtual events on April 8th ("Herbarium Happy Hour") and April 10th ("Botany Brunch"). These events each included a presentation on the use of herbaria for research, education, and conservation and a virtual tour of the Robert F. Hoover Herbarium at California Polytechnic State University. Participants then learned to transcribe specimen records in Notes from Nature and played transcription games. The Thursday event engaged 55 participants, and the Saturday event engaged 42 participants. Over the WeDigBio weekend (April 8-11th), volunteers transcribed 1,149 specimens.

Four Notes from Nature expeditions were completed during this quarter involving 8,642 specimens from UC Los Angeles, Humboldt State, Cal State LA, and Fresno State. The data were cleaned and imported into CCH2. We launched three new expeditions in February and March 2021 consisting of 5,943 specimens from Cal Poly, Cal State LA, and UC Los Angeles.