# CALIFORNIA PHENOLOGY TCN – QUARTERLY REPORT – NOVEMBER 2020

*Assembled by Katie Pearson, 2 November 2020*

## PROGRESS IN DIGITIZATION EFFORTS:

Figure 1 shows our progress in imaging, transcribing, georeferencing, and phenologically scoring the target specimens for our project, explained more in detail in the following sections.
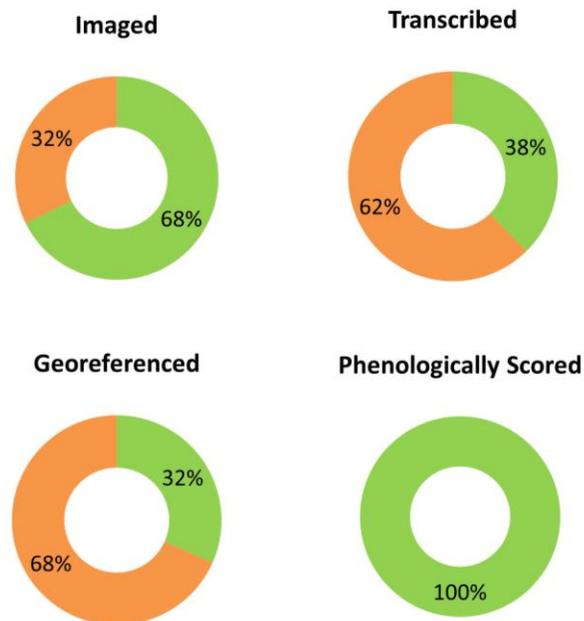
**Figure 1**

## TRANSCRIPTION

An estimated 113,000 specimen records have been transcribed across the CAP Network since the beginning of the project. This is approximately 38% of the goal number of transcriptions.

## GEOREFERENCING

We have georeferenced approximately 95,000 specimen records, which is 32% of our goal. Progress toward this goal was greatly accelerated by a new workflow. We recognized that many of the specimens in our aggregated database are duplicates, and that many of these duplicates had already been georeferenced. To leverage this pre-existing data, we developed an R script that retrieves the existing georeference data and returns the data in a separate spreadsheet. We cleaned and vetted these data and uploaded them to the appropriate collections. As a result, we applied georeferences to over 10,000 specimens in the span of about 2 weeks.

## IMAGING

Nine of our 22 herbaria (41%) have completed their imaging goals (Figure 2). Of the remaining herbaria, five 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. As a result, our numbers of imaged, transcribed, and georeferenced specimens have continued to increase. Figure 2 shows the current state of CAP imaging as of October 29, 2020.
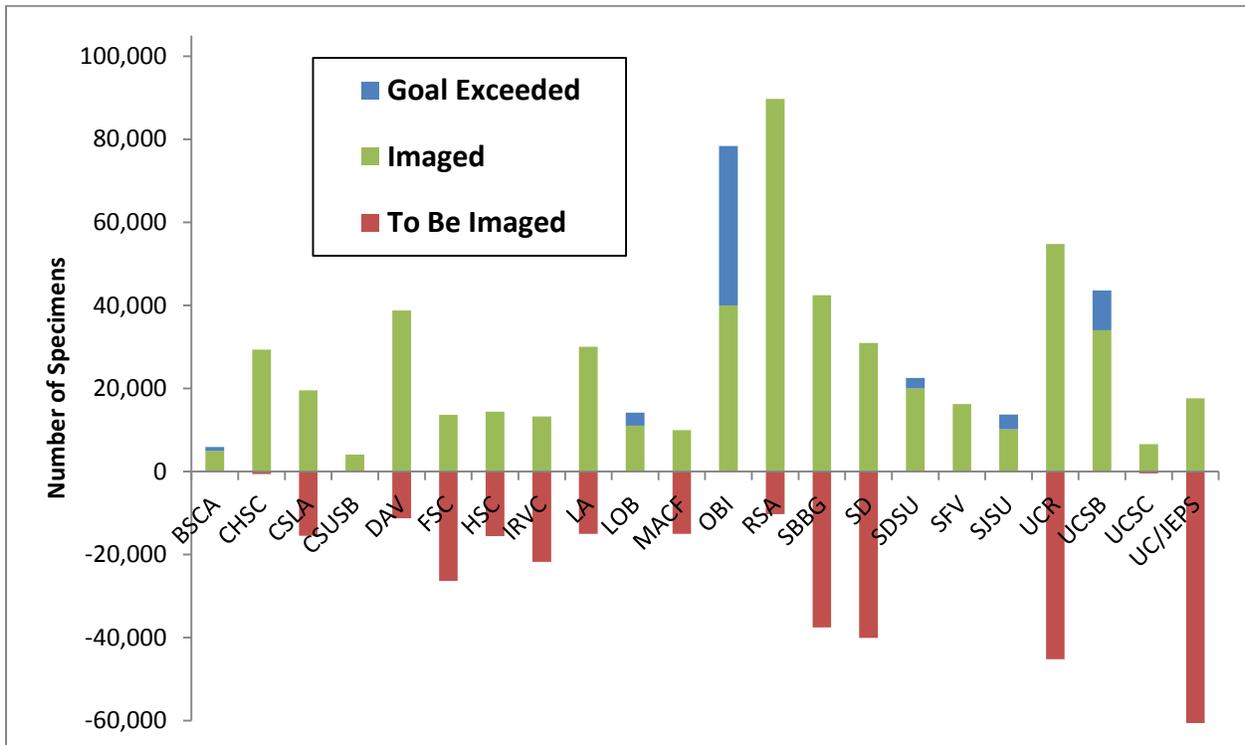


**Figure 2.** Herbarium specimen imaging progress. Green portions represent the number of specimens that have been imaged, while blue portions represent the number of specimens that have been imaged beyond the expected target specimens. Red bars below the zero line indicate the number of target specimens that have not yet been imaged.

## NEW PEN INSTITUTIONS

In September 2020, a PEN grant was awarded to six new CAP institutions (CDA, OSC, PUA, SFSU, SHTC, and UNLV) and one existing CAP institution (SD). We held a kick-off meeting with these new collaborators on September 15th, and the PM met with each institution during the following week to ensure participants were prepared to begin the digitization process.

Data from SFSU and SHTC had previously been imported into the CCH2 portal. Data from UNLV are now regularly imported into CCH2 from SEINet, where the dataset will continue to be managed live. CDA data were newly mapped and imported into CCH2, where they will now be managed live. We plan to

establish a link between the OSC Symbiota-based data portal soon, as well as upload data from PUA, which will be managed live in CCH2 as well.

CDA and UNLV will begin imaging once they have received their equipment and hired their technicians. We predict that imaging will commence no later than January 2021, barring additional coronavirus-related shutdowns. SFSU and OSC, which already had imaging equipment, will begin to image once they have hired their technicians and they are allowed to work in the herbarium. SHTC and PUA will conducting imaging, one after another, once UNLV has completed their goals and transfers the imaging station to California (Figure 3).
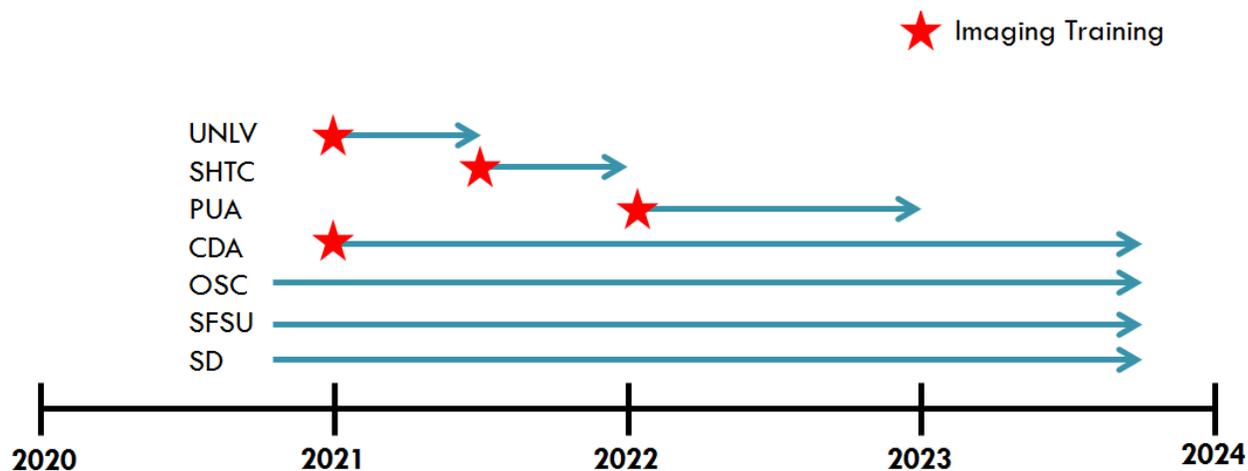


**Figure 3.** Expected timeline for imaging herbarium specimens at each of the seven PEN institutions. Note that imaging at SFSU and OSC has been delayed due to COVID-19 shutdowns and will likely re-commence in early 2021.

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

As discussed in "Georeferencing" above, we developed a script that identifies and imports georeference data from duplicate specimens. Applying this script to our database increased the number of databased specimens by over 10,000.

To further accelerate our georeferencing progress, we instituted an undergraduate internship course led by Cal Poly (OBI) involving students at multiple CAP institutions, facilitated by the remote learning environment (Zoom). This course began in mid-September and currently involves 65 students from 6 institutions (IRVC, LOB, OBI, OSC, UCSB, and UCSC). There are two sections of the course, one meeting on Tuesdays for 2 hours and one meeting on Thursdays for 2 hours. The course is facilitated by the PM and two experienced OBI undergraduate assistants. The 65 students were trained in transcription (through Notes from Nature) and georeferencing (in our data portal, CCH2). Unfortunately, because of the complexity of the georeferencing task, it became apparent that georeferencing by these students was neither efficient nor very effective, so we pivoted to having the students primarily transcribe through Notes from Nature. A small percentage of students who produce high-quality georeferences are continuing to work on georeferencing.

Another strategy we employed to advance georeferencing is engaging persons who are already knowledgeable in California geography. To this end, we launched the CAP Network "100 club" (https://www.capturingcaliforniasflowers.org/100-club.html), a team of botanists and naturalists across the state whom we train to use the georeferencing tools in CCH2. We have identified members of the California botanical community through our network of faculty and staff, and we recruit individuals through targeted emails and announcements in relevant botanical societies. Many of these naturalists are botanists and herbarium specimen collectors, and they are now empowered to augment the data of their own specimens "live" in CCH2. This initiative has so far proven quite successful. We have recruited and trained 13 individuals who have georeferenced over 800 records since late September.

## IDENTIFY GAPS IN DIGITIZATION AREAS AND TECHNOLOGY

We are continuing to develop the functionality of our CCH2 data portal. Currently, we lack a way to search by, download, and visualize specimens' phenological scorings, though these data currently exist. We are also developing a way to track who downloads our data and for what purpose.

## SHARE AND IDENTIFY OPPORTUNITIES TO ENHANCE TRAINING EFFORTS

All digitization protocols were re-assessed because of the addition of 6 new digitizing institutions as part of the PEN grant. Protocols and equipment recommendations were updated on the website.

We created a new training video for using the taxonomic cleaning tool and an updated video for how to image specimens for the PEN participants, both available on our YouTube channel (https://www.youtube.com/channel/UCFjYwX5cjS_NaUkx3F-4lpg).

As described in the "best practices and standards" section above, we launched an herbarium digitization internship course across six CAP institutions with 65 students currently enrolled. These students have all been trained in transcription through Notes from Nature and georeferencing in CCH2. Because of the relatively low quantity and quality of georeferences from these students, we pivoted to transcription in Notes from Nature. We maintain high student engagement by playing transcription games such as "Habitat bingo" (https://biospex.org/bingos) and encouraging them to share their screens to ask questions. Once during the course, each student also conducts a 5-minute presentation on a primary research paper of their choice that involves herbarium specimens. Each week, we also have a number of students give a 2-minute summary of their activities from the previous week. At the end of the course, students will complete a formative assessment of their learning and provide feedback about the course.

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

CAP leadership has continued to lead the new Consortium of California Herbaria Administrative Committee. This committee drafted a data sharing and downloading agreement with Calflora, met in early September to ratify it, and sent it to Calflora for consideration. We are now working with Calflora to ingest CCH2 data into their system.

We are partnering with the UC California Naturalists program to recruit volunteer transcribers and georeferencers. We met with California Naturalist leadership on August 31, 2020, and are now building a plan to disseminate promotional materials, present at the next instructors' meeting, and lead virtual training workshops for instructors and volunteers.

Specimen data and images from the Sequoia and Kings Canyon National Parks Herbarium (THRI) were added to CCH2. THRI will now manage their collection live in CCH2.

The PM provided training materials and an orientation to CCH2 to California Native Plant Society rare plant botanists, who will work to improve data and georeferences regarding rare plants.

The lead PI and PM continue to participate in the BIOME Institute to further develop the course-based undergraduate research experience (CURE) we implemented in spring 2020. As part of this institute, we are developing additional educational materials including a video that explains the importance of attribution and citation when using biodiversity specimen data. This resource will be posted on QUBES and YouTube when complete.

## SHARE AND IDENTIFY OPPORTUNITIES AND STRATEGIES FOR SUSTAINABILITY:

The California Botanic Garden (formerly Rancho Santa Ana Botanic Garden) and the California Department of Food and Agriculture herbaria have switched to "live" management in CCH2. This will allow more streamlined digitization workflows in the future and likely improve the sustainability of the portal in general.

## 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.

To engage more educators and students with the specimen data in CCH2 and how they can get involved through Notes from Nature, we developed an educational exercise that walks students through both platforms. The exercise is now published on QUBES here: https://qubeshub.org/publications/2068/1 . We also sent this resource to botany and ecology faculty throughout the state via email.

As described in the "enhance training efforts" section above, we launched an herbarium digitization internship course that remotely engages 65 undergraduate students from 6 universities.

The PM created an informational video about the Robert F. Hoover Herbarium at Cal Poly and plants of the Central Coast of California as part of Biodiversity Week, hosted by the Central Coast State Parks Association. The recording of the presentation can be viewed here: https://www.youtube.com/watch?v=V8NCB413TII

The CAP TCN hosted two virtual WeDigBio events, engaging 95 participants across the western U.S. and resulting in 2,072 Notes from Nature transcriptions from October 15-18 (Figure 4).
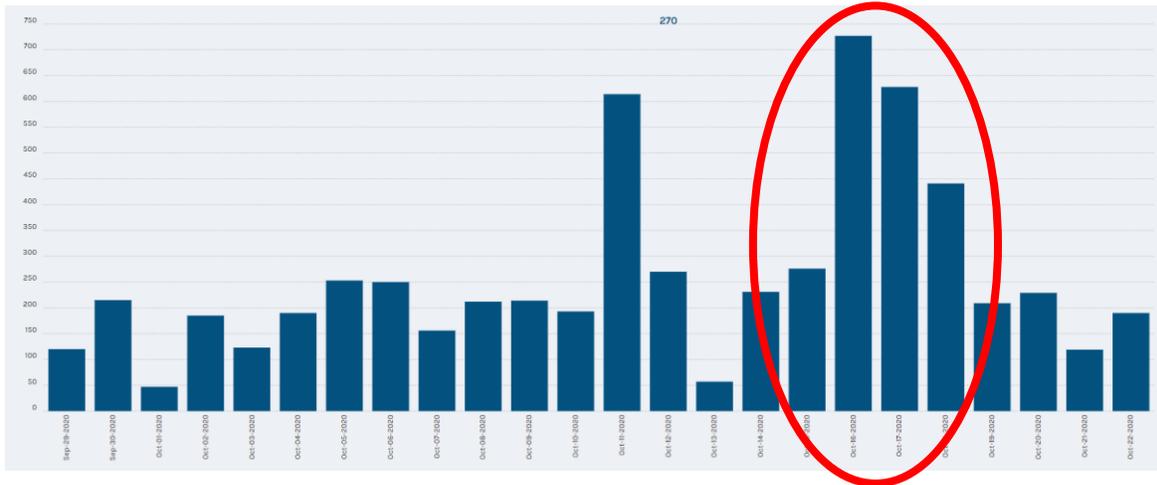


**Figure 4.** Number of transcriptions per day completed in Notes from Nature. The four days of WeDigBio are circled in red, demonstrating a significant uptick in transcriptions on those days total 2,072 transcriptions.

One Notes from Nature expedition was completed in late August 2020, resulting in 1999 specimens from CSLA being fully transcribed and their data imported into CCH2. We launched an additional expedition for IRVC in late August and a new expedition for CSLA in late October.