METHODS AND TECHNIQUES

Back to the future: A refined single-user photostation for massively scaling herbarium digitization

Charles C. Davis, Jonathan A. Kennedy & Christopher J. Grassa

Department of Organismic and Evolutionary Biology, Harvard University Herbaria, Harvard University, 22 Divinity Avenue, Cambridge, Massachusetts 02138, U.S.A.

Addresses for correspondence: *Charles C. Davis, cdavis@oeb.harvard.edu; Jonathan A. Kennedy, jonathan_kennedy@harvard.edu; Christopher J. Grassa, c.j.grassa@gmail.com*

DOI https://doi.org/10.1002/tax.12459

Abstract The digitization of herbarium collections is greatly transforming plant biodiversity science, yet most herbarium data remain inaccessible. Here, we present a novel, single-user photostation and associated workflow for efficiently mobilizing herbarium specimens. Our apparatus represents a significant improvement to existing technology and is scalable to a variety of digitization tasks from small to massive collections.

Keywords biodiversity science; collections; digitization; herbaria; natural history museums; plant diversity

■ BACKGROUND

Nearly 390 million plant specimens reside in the world's herbaria (Thiers, 2017–) and represent our best source of geographical, temporal, and morphological variation of plants. Most of these data remain largely inaccessible. The digitization and online mobilization of herbarium specimens has greatly helped to bridge this impasse and ignite revolution in the biodiversity sciences (Nelson & al., 2015; Thiers & al., 2016; Drew & al., 2017; Soltis, 2017; Nelson & Ellis, 2019; Hedrick & al., 2020). These efforts have mobilized millions of specimens with significant economies of scale and accelerated advances in scientific investigations, including phenological studies of climate change, species range assessments, and biotic interactions (Willis & al., 2017; Meineke & al., 2018, 2019; Hedrick & al., 2020; Pearson & al., 2020). In addition, the use of natural history collections to answer scientific questions using only their digitized representation, rather than the physical specimen itself—i.e., Digitization 2.0 sensu Hedrick & al. (2020)—has sparked the integration and development of new scholarly disciplines and lines of inquiry not previously possible. Despite these exciting new directions, however, Digitization 1.0 sensu Hedrick & al. (2020)—i.e., the generation of digitized products from the physical specimen—remains an active area of innovation and development. This relates to both hardware and workflow innovations as well as their integration with advancements in software. Along these lines, innovations in these areas have greatly increased the cost-effectiveness of digitizing herbarium specimens and enabled the successful mobilization of entire collections and whole floristic regions (Pignal & Michiels, 2012; Van Oever & Gofferjé, 2012; Slijkhuis, 2014; Heerlien & al., 2015; Schorn & al., 2016; Sweeney & al., 2018). Here, we present a novel photostation and workstation design for imaging herbarium specimens that represents a dramatic improvement upon existing approaches and is scalable for large and small institutions alike.

■ CONVEYOR BELT TECHNOLOGIES AND HERBARIUM DIGITIZATION

The digitization of flattened herbarium specimens has historically been accomplished using copy stands, light boxes, or inverted flatbed scanner setups, which are operated from a single-user interface (see detailed summaries in Nelson & al., 2015). The first two methods are quite similar and rely on cameras for image capture. The differences in these approaches mainly relate to the use of a stand to mount the camera plus high-intensity fluorescent or strobe lighting. In contrast, a light box typically does not use a copy stand. Instead, image capture is contained within an enclosure that restricts environmental light sources and offers more control for diffuse lighting. Inverted scanners were applied widely as part of the Global Plants Initiative to digitize all type specimens (https://plants.jstor.org/). This approach produces high-quality images but is very slow and cannot easily accommodate bulkier material. A major breakthrough in the digitization of herbaria was the application of conveyor belt technologies (Pignal & Michiels, 2012; Van Oever & Gofferjé, 2012; Slijkhuis, 2014; Heerlien & al., 2015; Schorn & al., 2016; Sweeney & al., 2018). Such conveyor belt setups are industrial-scale, high-throughput systems dedicated especially to vascular plant herbarium specimens mounted to sheets. A central feature of these systems is an automated conveyor belt apparatus, which allows image capture and specimen handling to occur simultaneously, and which prioritizes imaging and the capture of at least minimal specimen

Article history: Received: 18 Sep 2020 | returned for (first) revision: 22 Oct 2020 | (last) revision received: 22 Nov 2020 | accepted: 5 Dec 2020
Associate Editor: Dirk C. Albach |

data. To date, these efforts have been successfully deployed at a massive scale at the Muséum national d'Histoire naturelle Paris (P) (Slijkhuis, 2014; Le Bras & al., 2017), the Naturalis Biodiversity Center (NL) (Slijkhuis, 2014; Rogers, 2016), the United States National Herbarium (US), and at our home institution, the Harvard University Herbaria (HUH) (Schorn & al., 2016; Sweeney & al., 2018). It is also worth mentioning that the vendor Picturae (https://picturae.com/en/what-we-offer/digitizing-collections), who offers solutions using a conveyor belt workflow (Slijkhuis, 2014), was contracted for the successful digitization of P (Le Bras & al., 2017), NL (Heerlien & al., 2015), and more recently US.

The HUH have played a role in innovating rapid, efficient imaging and transcription workflows (Schorn & al., 2016; Sweeney & al., 2018). In particular, as part of the "Mobilizing New England Vascular Plant Specimen Data to Track Environmental Changes Project" (NEVP), a Thematic Collections Network (TCN) funded through the U.S. National Science Foundation's Advancing the Digitization of Biodiversity Collections (ADBC) program (award 1208835), the HUH (in collaboration with colleagues at Yale University [Patrick Sweeney] and the University of South Carolina [Bili Starly]) pioneered a novel and cost-effective conveyor belt system to digitize (image, transcribe, and georeference) the entirety of the HUH's ~350,000 New England vascular plant holdings (Schorn & al., 2016; Sweeney & al., 2018). This was a dramatic improvement in imaging speed and efficiency from previous efforts in the herbarium community. At the same time, these conveyor belts were expensive and required a large space footprint. In deploying the conveyor belts at the HUH, the physical space requirements meant the digitization area was not immediately co-located with the collection. The distance between the collection and the digitization area was a barrier to reducing transit time and also increased the risk of specimen damage during transit. In the case of the HUH, the digitization area was located in the same building and in a stringent pest-management zone. However, had the available space necessitated locating the digitization area further from the collection, this would have introduced the need for additional pest-control procedures, adding significant time and labor to the workflow. Moreover, each conveyor belt relied on the coordinated effort of minimally two staff members, and absences caused a measurable decline in efficiency. With further analysis, our team discovered that many of the positive efficiencies of the conveyor belt system were the result of ergonomic and workflow-specific customizations that were not specific to the use of a conveyor belt system. Taking this entire workflow into consideration, we noted that the presence of a conveyor belt did not reduce the necessary physical motions for an operator to stage and de-stage a specimen for imaging compared to other single-user workflows, nor did it increase the speed of that activity. Our team hypothesized that a single-user photostation, better optimized for herbarium imaging workflows, would produce similar or greater efficiencies while resolving the space and coordination challenges of the conveyor belt system. At the same time, the team conceived of leveraging new software tools in the digitization process, so that the imaging workflow could be designed and optimized independent of additional data capture-related activities.

■ RETURNING TO A SINGLE-OPERATOR PHOTOSTATION FOR RAPID HERBARIUM SPECIMEN IMAGING

With continued investment in process and workflow analysis, we developed a single-operator photostation specifically designed and fabricated for herbarium specimen digitization (Fig. 1). This photostation design has now been in operation at the HUH since mid-2018 (first prototype deployed fall of 2017) and has greatly exceeded expectations for increased imaging efficiencies while also realizing the benefits of a small, flexible, and scalable system. In a nutshell, we have taken herbarium digitization efforts 'Back to the future', returning to a more traditional single user approach, but with much greater efficiency that is scalable to small and large collections alike. Here, we present our workflow and novel photostation design in the hopes that similar institutions, large and small, can benefit from our efforts.

Our photostation design presented in Fig. 1 can be built using available consumer hardware. The framing is constructed using Grainger 6000 series extruded aluminum components (https://www.grainger.com/), which have a tensile strength (205–250 MPa) and elastic modulus (70 GPa) that yields a robust and stable design. Specifically, the majority of the upper frame is 1″ T-slotted extruded 6105-T5 aluminum; the camera support is constructed from 3″ T-slotted extruded 6560-T6 aluminum. The lower cubby is made from laminated wood. Automated height-adjustable Fully Jarvis standing desk legs are attached to the base of the photostation. While the photostation frame is a custom construction, the attached electronics and photographic equipment are typical of herbarium imaging workflows (e.g., DSLR camera, attached computer with image capture software, LED lighting). Multiple camera technologies can be attached to the photostation and easily upgraded over time. The HUH presently utilize the Canon 5DSR camera body, which is capable of 50 MP images, equivalent to a 520 ppi image of a standard herbarium sheet, and ZEISS Milvus 50 mm f/2 macro lens. Genaray 5600K LED lighting with a >90 CRI rating and a controlling computer (Mac Mini) with small tablet-sized monitor (mounted at eye level) are included in the hardware configuration. Capture One software is used for image capture on the photostation as well as post-processing on a separate computer. Construction of the photostation requires tools for cutting wood and extruded aluminum framing. Drilling is necessary to attach the frame and legs to the wooden cubby. Assembly of the upper frame and attachment of the electronic components are accomplished with only a screwdriver and Allen wrench. Institutions without local expertise can work with local builders or hackerspaces for the construction. The HUH worked with a local builder to construct the final design presented here. The total

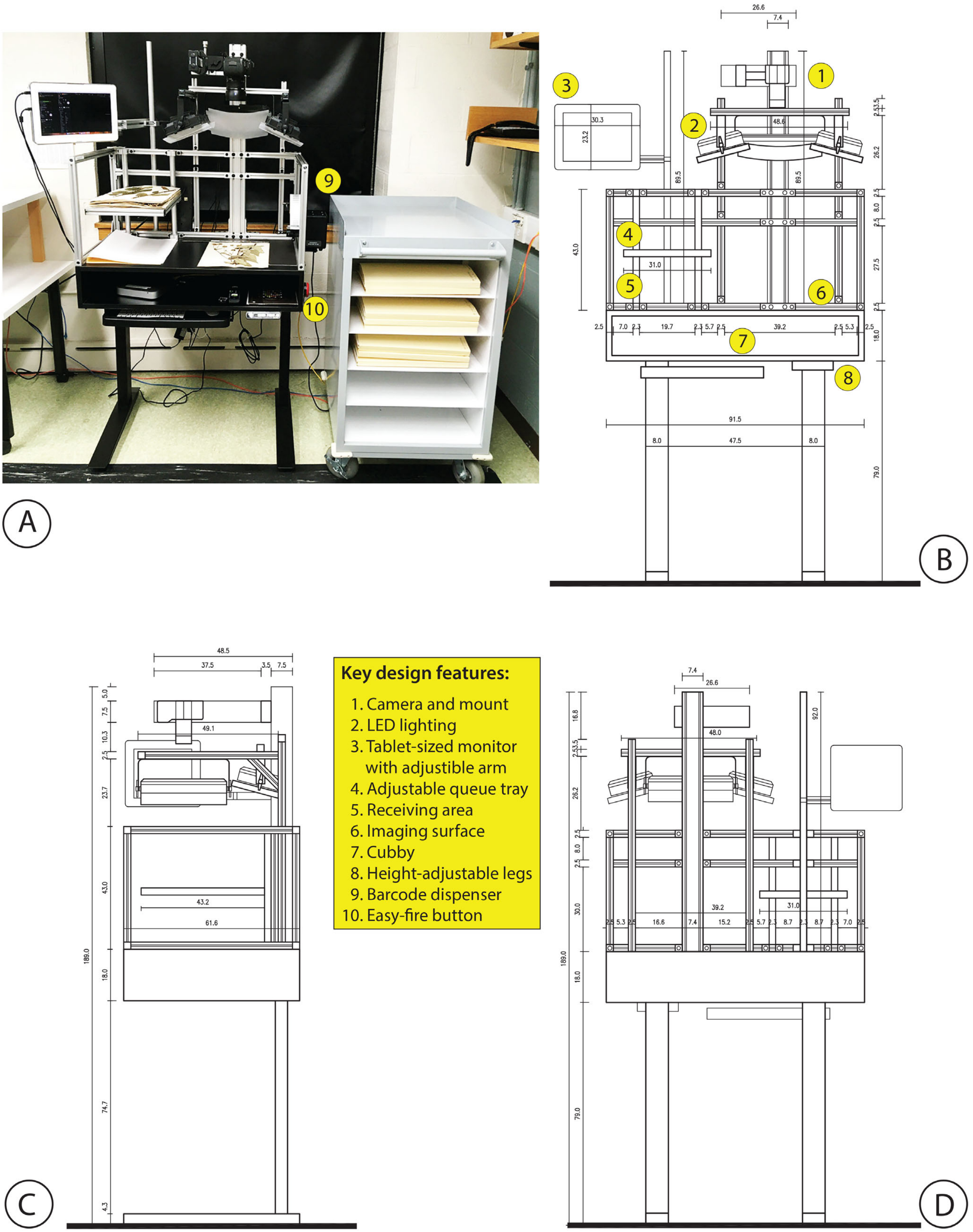

Fig. 1. Single-user photostation design. **A**, Photograph of photostation. **B–D**, Technical drawings of front (**B**), side (**C**), and (**D**) rear view of station. Measurements in metric units (cm). Key details highlighted in yellow with numbered legend. Built-in jig for mounting specimens with ruler and color checker, automated barcode dispenser, and easy-fire button shown only in photograph on imaging table.

cost of the materials, electronic components, and software is under $10,000, while the cost of labor will vary by region.

■ PHOTOSTATION WORKFLOW

Alongside our novel photostation hardware design, we also greatly optimized our workflow (Fig. 2). This imaging workflow can be accomplished by a single person and takes place as follows: (1) A batch of specimens are brought to the digitization area in a manner appropriate for the specific collection and distance. In the case of the HUH, specimen folders are removed from cabinets and loaded into an herbarium cart that is (2) transported a short distance to the digitization area, generally to a room on the same floor. (3) Before imaging, specimens are checked for damage, minor repairs are performed as needed, and specimens with severe damage are set aside. Any specimen missing a barcode has one affixed. As the specimens are inspected, they are stacked so that their order is reversed prior to imaging. (4) An imaging session begins by logging into the computer attached to the photostation, launching the photo capture software and creating a new session named for the operator and date. Once the attached camera and lights are turned on, a level is placed underneath the camera lens to check for any tilt along the horizontal and z-axis, the camera lens is focused using a focus chart, the camera settings are checked (shutter speed, ISO, aperture), and a test image is captured using a large color checker (X-Rite ColorChecker Digital SG). To begin imaging, a specimen folder is placed on a "queue" shelf directly next to the imaging stage. (5) The specimens are removed from the folder and kept on the queue shelf while the folder is placed on the imaging stage. An image is captured using an easy-fire button adjacent to the stage, attached to the outside of the photostation frame. This placement helps ensure the hand of the operator is out of the camera's view. (6) After hearing the audible shutter-sound (screen flash can also be used for the hearing impaired), the operator moves the folder to the area immediately below the queue shelf (the receiving area). The waiting stack of specimens are similarly imaged one by one, in rapid succession, as are any species covers. A built-in jig, sized for standard 16.5″ × 11.5″ herbarium sheets, helps to quickly guide the placement of the specimen sheet. Images of the folder, species covers, and specimens will be preserved in sequence for later transcription. The HUH logo and mini color checkers are already attached to the jig to eliminate additional placements by the operator. If a specimen did not receive a barcode ahead of imaging, a barcode is retrieved from an automatic barcode dispenser attached directly to the right of the photo stage. After an image is captured, the operator can inspect the image using the monitor attached to the photostation. Because image capture can occur many times faster than the images can be rendered to the monitor, operators are only asked to inspect images at the start of the session and only occasionally thereafter. (7) As specimens are imaged, they are stacked in the receiving area to the left of the photostage and under the queue tray. This stacking also corrects the ordering that was reversed during the curation step so the specimens can be easy refiled and (8, 9) returned to the collection.

The imaging workflow described above is intended to be performed by a single person and occurs very rapidly. After the initial setup procedure and inspection of the first image, capture of subsequent images can occur in as quickly as 4 seconds per image. Achievable *average* imaging rates will depend on factors related to the collection, such as the proportion of fragile specimens or number of species covers. We have also observed that taxonomic identity can be a factor in the rate of image production, with thin, densely stacked specimens (e.g., grasses and orchids) more quickly imaged than thicker, woody specimens (e.g., many rosid tree species). The photostation design can also accommodate bulky specimen types (e.g., fruits and cacti), by adjusting the camera settings to

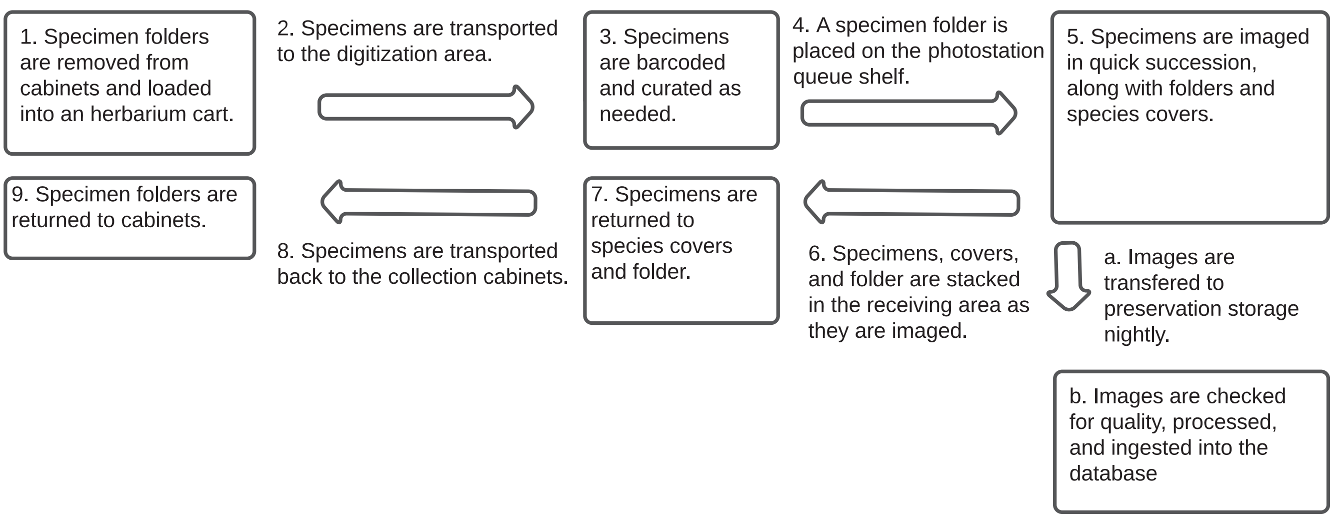

Fig. 2. Harvard University Herbaria digitization workflow.

create a large depth of field. Quality control steps can be performed by the operator, other digitization or curatorial staff, or by local informatics staff. Training new digitization staff to use the photostation is accomplished by existing staff and typically requires less than one day. The photostation needs relatively minimal technical support; after the initial configuration of the attached computer and image capture software, support needs are generally limited to replacing faulty electronics and occasionally correcting the alignment of the camera arm and camera. Since deploying this design in 2017, the HUH have experienced no camera, lens, or computer failures. However, we have replaced approximately five LED light panels due to failure over time.

■ ERGONOMIC ENHANCEMENTS

Ergonomics were considered extensively during the design of the photostation because the workflow calls for significant time performing highly repetitive motions, especially specimen placement and removal. To increase comfort, the photostation is built with adjustable-height legs controlled with an electronic panel preset for up to seven users. Operators can stand on an ergonomic mat or use a chair or stool of their choice.

Significant consideration was also given to lighting. Many herbaria use light boxes or copy stands for their imaging workflows. Lightboxes allow for controlled lighting of the photostage without interference from environmental light sources, such as windows and overhead lighting. For the same reason, copy stand setups, which are usually not enclosed, are often used in a dark room. Because the quickness of staging and de-staging the specimen is a central factor for imaging efficiency, our photostation also uses an open design and is subject to room lighting conditions. In our initial testing, we used a dark room, and a fabric hood was placed over the photostation lights to reduce the brightness experienced by the operator. Operators were also supplied with optional OSHA-compliant partially tinted glasses to further reduce eyestrain. This configuration was functional but less than ideal for operator comfort. We ultimately abandoned the hood and dark room and installed overhead lighting (5000K) that reasonably matched the color temperature of the lighting attached to the photostation (5600K). Mixed lighting (i.e., multiple light sources with differing chromatic properties) can impede the accurate rendering of color. This happens when the light sources illuminate the image unevenly, causing areas of the image to have differing white balance values, thus making it impossible to render color accurately across all areas of the image. However, if all light sources evenly illuminate the image target, the average white balance will remain consistent across the image. In our photostation deployment, we found that typical overhead lighting evenly illuminates the stage, in conjunction with the attached lighting equipment, to maintain a consistent white balance for accurate color reproduction. Overhead lighting also illuminates the stage with significantly weaker intensity than the photostation lighting, thus having a smaller impact on the final white balance value (~5400K). To test the color accuracy of this configuration, we captured six images using the X-Rite ColorChecker Digital SG, changing the orientation of the color checker in each image so that each color swatch would appear in a different location. We then compared the HSL (hue, saturation, lightness) values for 10 of the grayscale swatches. The HSL values were computed from the RGB values taken from the images, and lightness was ignored as a confounding factor in favor of comparing only hue and saturation. We found that hue and saturation remained consistent across the swatches regardless of position on the stage, confirming that the stage was being illuminated with a consistent white balance. In light of these results, we conclude that environmental lighting can be designed so to not interfere with accurate color capture and that testing the lighting conditions of a new location or configuration can be done with relatively minimal effort. By changing our configuration to allow for overhead lighting, we significantly improved operator comfort and overall job satisfaction while avoiding the time cost of repeatedly opening, closing, and navigating around a lightbox.

We conducted numerous tests with the help of HUH curatorial staff throughout the development of our first two prototypes and final design. We concluded that 2–3 hours was the upper limit for comfortably imaging in a single session. Imaging at or above two hours, more than once per day, was considered to be both physically and mentally taxing. Fortunately, owing to the overall efficiency of the system, HUH staff regularly perform only one or two imaging sessions per week, averaging between 800–1500 specimen images per 1–3 hours of imaging work. This produces a sufficient number of images so that the remaining work-week can be dedicated to transcribing those images.

■ PHOTOSTATION PRODUCTION RATES

Images produced by the photostation are timestamped, creating a lasting record of the system's imaging rates. To determine our real-world efficiencies, we analyzed the time data from batches spanning multiple operators and taxonomic groups to draw strong conclusions about the average efficiency of the imaging workflow. Using data from eight full-time and five part-time digitizers, spanning approximately two years (corresponding to the deployment of the final iteration of the photostation design), we calculated an average imaging rate of 8 seconds per image. This rate does not include photostation setup, which is performed by the operator in about five minutes. The above rate also does not include time spent transporting, curating, and refiling specimens, but does include images taken of folders and species covers. Many of these factors are highly variable between institutions, and we believe our timing analysis is comparable to those published for other imaging workflows.

The speed of imaging is governed by the physical speed of the operator, therefore decisions regarding personnel, work organization, and staff management can make significant impacts on achieving higher efficiencies. We found that there was notable variation in the imaging speed among staff, with some staff handling specimens almost twice as quickly as the slowest operators. However, most staff clustered slightly below our reported average. Also, due to variations across the collection (e.g., specimen thickness or condition), we were unable to fully separate these factors to determine operator speed. Our slowest imaging rates (ca. 14 seconds per image) are still a marked improvement over past efforts and, to ensure the careful handling of specimens, we chose not to ask staff to accelerate their imaging rate beyond what they feel was comfortable. Newly trained staff should not be expected to achieve maximum speeds immediately, but it is our experience that staff begin achieving personal maximal efficiency after only 2–3 weeks of daily imaging, sometimes sooner. Image quality does not noticeably vary between operators due to standard camera settings, setup procedures, photo stage guides, and limited activity required for image capture (staging and destaging the specimen). While staff management can make a significant impact on overall productivity, the HUH did not make significant changes to its staff management practices to accommodate this project, and similar efficiencies should be attainable at other institutions.

Additionally, the described workflow does not include data entry, nor does it rely on the presence of previously captured data. Images captured by the photostation (camera RAW files) are transferred to HUH preservation storage nightly and ingested into the HUH specimen database to be used for data entry using a locally developed transcription web application. Quality control of images is performed daily by selecting a random sample of images from each batch. Images are inspected for deviations in focus, exposure, color, orientation, the presence of hands or fingers, and any other noticeable issues. After inspection, images are processed in batch with minor adjustments for white balance, exposure, levels, and lens correction. Derivative images (JPEG, DNG) are also created during this post-processing stage. Performing these steps takes the operator approximately five minutes per batch, while the creation of derivative images may take 30 minutes to two hours (unattended), depending on the available computing power and size of the batch.

After image post-processing is complete, images are passed into an informatics pipeline for additional transformations, data extractions, and ingest into the HUH database and transcription application. Barcodes are automatically read from the image by software during ingest, and the image is linked if there is an existing database record. In cases where multiple specimens are affixed to the same sheet, the system will identify each barcode and link the image to all pre-existing records. Other components of the pipeline under active development include automatic image border cropping, label identification, and deep learning handwriting recognition models. The HUH database can be accessed at https://data.huh.harvard.edu/databases/specimen_index.html; for example specimen images produced by the workflow described here see the following: ex1, ex2, ex3, and ex4.

■ KEY DESIGN AND ENHANCEMENT FEATURES TO EARLIER IMAGING APPROACHES

Approaching our design as a custom fabrication, as opposed to assembling general-purpose components (i.e., light box, copy stand), allowed us to make numerous improvements that were critical to our overall efficiency gains and demonstrates the value of investing in customized hardware and process fine-tuning, particularly for large-scale digitization efforts. Efficiencies gained using this new photostation are achieved via numerous ergonomic customizations specific to the handling and dimensions of standard herbarium sheets, the inclusion of components that have a high return on investment at scale (e.g., automatic barcode dispenser), and the use of automated image processing steps to replace manual activity (e.g., image recognition for barcode data entry). Key customizations include: platform sizing for vascular sheets to eliminate manual turning of specimens, adjustable shelving to easily separate in-process from completed specimens (queue tray versus receiving area), remote camera shuttering, height adjustment for operator comfort and efficiency, eye-height tablet display for quickly assessing image quality, and built-in automatic barcode dispenser. Additionally, the workflow streamlines image processing steps by moving these activities to a later stage where they can be performed in batch and with the aid of automation. Many herbarium workflows partially decouple imaging from data capture (Heerlien & al., 2015; Thiers & al., 2016). However, the need for minimal data capture, even skeletal record creation, can add seconds to per-specimen digitization rates, significantly reducing workflow efficiency and sometimes necessitating poorly supported software additions. By fully decoupling the imaging and data capture stages of the digitization workflow, relying instead on feeding images to the backend informatics pipeline, we were able to focus on fine-tuning the imaging process while reducing workflow complexity and technical support challenges.

In addition to these enhancements, the shape and size of the photostation is a significant benefit of our design. Measuring only 189 cm tall, 92 cm wide, and 62 cm deep and requiring only access to a power outlet, institutions adopting this design are likely to find a suitable space near their collection to deploy this hardware. At the HUH, a single digitization area requires only enough space to accommodate the photostation, operator, herbarium cart, and worktable (Fig. 1). Multiple digitization areas can be organized throughout a building or moved as digitization progresses through the physical collection. Using this design, specimens are imaged and immediately refiled in small batches in close proximity to the collection. This reduces the time specimens are kept out of the collection and simplifies workflow coordination. This photostation design also requires no complex installations or long-term space commitments, nor does it require undertaking a large-scale digitization effort to be useful. The photostation can be economically used for projects of all sizes, including mass digitization, incremental imaging of smaller or prioritized subcollections, or in conjunction with ongoing accessioning activity. It can be moved or placed into storage in response to changing institutional priorities or available funding. Additional units can also be constructed to increase the overall rate of digitization or respond to new projects as they arise. We also note that while these features are useful for institutions undertaking in-house digitization, our design is equally relevant to vendors that would seek to offer a digitization solution that is adaptable to a variety of collection facilities and institutional budgets. Importantly, with this hardware and workflow we have dramatically increased the cost-effectiveness of imaging while eliminating the space and labor obstacles associated with the conveyor belt design, thus facilitating the deployment of lightweight, mobile photostations throughout the collection range.

■ COMPARISON OF IMAGING EFFICIENCY TO OTHER APPROACHES

The imaging workflow summarized above has been in operation at the HUH since mid-2018 (first prototype deployed fall of 2017) with six deployed photostations used by 13 staff. HUH digitization efforts draw from five full-time curatorial assistants that contribute part-time to digitization activities and eight term-based curatorial assistants working full time on digitization. HUH staff have accumulated many thousands of hours of digitization time from which we have analyzed our imaging performance to make clear comparisons in efficiencies gained relative to previously implemented imaging workflows, including the conveyor belt system described above and traditional single-user interfaces. It should be mentioned that herbarium digitization workflows are diverse and challenging to summarize (Nelson & al., 2015). However, Nelson & al. (2012) and Thiers & al. (2016), who adopted commonly applied single-station approaches, reported average imaging rates of around 100 images per hour (or ca. 36 seconds per specimen) per imaging station. The conveyor belt workflow described by Sweeney & al. (2018) was tested for “image-only” throughput where only a barcode or QR code was scanned in addition to the image capture. These tests achieved an impressive 20 seconds per image (or ca. 180 specimens per hour), a substantial improvement over previous efforts using single user interfaces. In contrast, our refined workflow and single-user photostation captures images with comparable minimal data capture (a barcode) at an estimated average of 8 seconds per image. Requiring half the staff to operate, our average rate of image capture is a four-fold increase in efficiency over the best reported times of the previous best-of-breed conveyor belt system.

■ BROADER IMPLICATIONS OF HIGH-THROUGHPUT IMAGING

Herbaria undertaking digitization projects must address difficult questions about project scope and focus, and importantly how to retrieve relevant specimens from the collection in a way that is efficient and cost-effective. Current funding sources for herbarium digitization, such as NSF ADBC, often require a thematic focus but many important and novel research themes are not congruent with typical models of herbarium organization, which are usually organized by family or geography. Answering even simple questions (e.g., specimen counts) about dispersed subsets of the collection is labor intensive and accuracy is often questionable. These challenges continue as projects are undertaken because staff must sort through numerous specimens to identify only those of relevance to the project, at a significant loss to overall efficiency.

At the HUH, the development of our photostation greatly propelled our effort to undertake a key mission outlined in the 2018 HUH Strategic Plan: digitizing the entirety of our North American vascular plant holdings. This represents ca. 1.5 million specimens in our collections including among the oldest and most extensive collections of North American plants in the world. To date, HUH staff have imaged and databased over 500,000 specimens using the described photostation and workflow. This ambitious goal has further benefited by funding from several completed and ongoing NSF-funded TCNs, including: the above mentioned NEVP, “Using Herbarium Data to Document Plant Niches in the High Peaks and High Plains of the Southern Rockies: Past, Present, and Future” (SoRo), “Digitizing ‘endless forms’: Facilitating Research on Imperiled Plants with Extreme Morphologies” (PoE), and “American Crossroads: Digitizing the Vascular Flora of the South-Central United States” (TORCH). More immediately, our success with implementing this workflow organization was essential in remaining operational during recent COVID shutdowns, in which HUH staff were unable to return to campus for three months. In advance of the shutdown, HUH staff prioritized imaging activities and created a backlog of images that allowed approximately 13 staff to continue working full-time on transcription and georeferencing activities while remote. At the time of this writing (9/2020), the HUH is operating under a low-density occupancy plan. However, because of our imaging speeds, and as noted above, staff only need to return to the office once per week to generate enough specimen images for full-time data capture activities.

Our photostation has been a boon to productivity and to our current NSF-funded efforts and to the larger goal to digitize the entirety of our North American holdings. Our capacity to affordably image broad sections of the collection with minimal data capture has benefitted the HUH in numerous ways. In cases where we have already imaged a significant portion of a collection that is relevant to a project, we can generate precise counts by taxon or geography (county-level and above) to accurately estimate the labor required and funding necessary for further digitization (usually full-label transcription and georeferencing). Using our backend software, it is trivial to present staff with project-specific subsets of specimen images for full transcription and georeferencing. By targeting entire families or continents for “initial” digitization, the HUH has been able to participate in many overlapping digitization efforts without managing multiple workflows.

We have also identified that our efforts have stimulated tremendous interest in these data, signaling that there is no shortage in the demand for herbarium specimen data. Between 2017, when the North America digitization project was launched, and 2019, the HUH saw a 260% increase in citations through GBIF. This is despite the fact that over 50% of these newly created ca. 500,000 records include only a taxon and state (or county). Our experience suggests that by advancing the development of extremely cost-effective solutions to herbarium specimen digitization, even with minimal data capture, the community can increase their utility and service to scientific advancement, far beyond what we do today.

■ CONCLUSIONS

The inability to quickly and cost-effectively image herbarium specimens has been an impediment to the broader digitization of the world's herbaria. Herbaria have made tremendous strides in mobilizing these data in the past decade, which has been greatly facilitated by community involvement and broad participation. We feel this photostation marks a tremendous improvement to current efforts. It can be easily scaled to larger operations by producing more individual stations or reduced accordingly for smaller projects. In addition, the return to a single-user interface obviates the need for conveyor belt apparatuses, which are often expensive (especially when a vended solution is weighed), space consuming, and especially reliant on a constant work staff. All of these factors are liable to be substantial barriers to entry for most institutions. We feel that this advancement is particularly relevant and timely because there exist more than 390 million herbarium specimens in the world and only a small fraction of these specimens have been imaged to date (Sweeney & al., 2018). Here, we freely distribute the design and plans of this apparatus to expand efforts to more easily and efficiently liberate collections that have been hidden from view for decades to centuries.

■ AUTHOR CONTRIBUTIONS

CCD initiated, supervised, and worked to fund the effort to expand specimen digitization during his tenure as HUH Director (2013–2019) and as Curator of Vascular Plants (ongoing). JK and CJG conceived the original idea and high-level design for the photostation hardware and imaging workflow with significant intellectual input from CCD and HUH staff. CJG designed and constructed the first photostation prototype with input and improvements by JK. JK supervised the day-to-day management of the project (ongoing), researched imaging technologies and configurations, developed the software components, and performed the data analysis. CCD and JK wrote the manuscript with final acceptance by all authors. — CCD, https://orcid.org/0000-0001-8747-1101; JK, https://orcid.org/0000-0002-4902-2011; CJG, https://orcid.org/0000-0002-2705-4872

■ ACKNOWLEDGEMENTS

We acknowledge Michaela Schmull, Anne Marie Countie, and the HUH curatorial staff who greatly helped to facilitate our design and implementation of the photostation. Special thanks to Anthony Brach, Ellie Taylor, and Emma Tanner, who provided important details on the photostation workflow. This work was largely conducted on the traditional territory of the Wampanoag and Massachusetts people. Funding was provided by the HUH, and by the ADBC program of the U.S. National Science Foundation (awards 1208835 [NEVP], 1702322 [SoRo], 1802209 [PoE], and 1902078 [TORCH]). We are indebted to L. Caria for his assistance with Figure 1.

■ LITERATURE CITED

Drew, J.A., Moreau, C.S. & Stiassny, M.L.J. 2017. Digitization of museum collections holds the potential to enhance researcher diversity. *Nature Ecol. Evol.* 1: 1789–1790. https://doi.org/10.1038/s41559-017-0401-6

Hedrick, B.P., Heberling, J.M., Meineke, E.K., Turner, K.G., Grassa, C.J., Park, D.S., Kennedy, J., Clarke, J.A., Cook, J.A., Blackburn, D.C., Edwards, S.V. & Davis, C.C. 2020. Digitization and the future of natural history collections. *BioScience* 70: 243–251. https://doi.org/10.1093/biosci/biz163

Heerlien, M., Leusen, J.V., Schnörr, S., Jong-Kole, S.D., Raes, N. & Hulsen, K.V. 2015. The natural history production line: An industrial approach to the digitization of scientific collections. *J. Computing Cult. Herit.* 8: 3. https://doi.org/10.1145/2644822

Le Bras, G., Pignal, M., Jeanson, M.L., Muller, S., Aupic, C., Carré, B., Flament, G., Gaudeul, M., Gonçalves, C., Invernón, V.R., Jabbour, F., Lerat, E., Lowry, P.P., Offroy, B., Pimparé, E.P., Poncy, O., Rouhan, G. & Haevermans, T. 2017. The French Muséum national d'histoire naturelle vascular plant herbarium collection dataset. *Sci. Data* 4: 170016. https://doi.org/10.1038/sdata.2017.16

Meineke, E.K., Davis, C.C. & Davies, T.J. 2018. The unrealized potential of herbaria for global change biology. *Ecol. Monogr.* 88: 505–525. https://doi.org/10.1002/ecm.1307

Meineke, E.K., Davies, T.J., Daru, B.H. & Davis, C.C. 2019. Biological collections for understanding biodiversity in the Anthropocene. *Philos. Trans., Ser. B* 374: 20170386. https://doi.org/10.1098/rstb.2017.0386

Nelson, G. & Ellis, S. 2019. The history and impact of digitization and digital data mobilization on biodiversity research. *Philos. Trans., Ser. B* 374: 20170391. https://doi.org/10.1098/rstb.2017.0391

Nelson, G., Paul, D., Riccardi, G. & Mast, A.R. 2012. Five task clusters that enable efficient and effective digitization of biological collections. *ZooKeys* 209: 19–45. https://doi.org/10.3897/zookeys.209.3135

Nelson, G., Sweeney, P., Wallace, L.E., Rabeler, R.K., Allard, D., Brown, H., Carter, J.R., Denslow, M.W., Ellwood, E.R., Germain-Aubrey, C.C., Gilbert, E., Gillespie, E., Goertzen, L.R., Legler, B., Marchant, D.B., Marsico, T.D., Morris, A.B., Murrell, Z., Nazaire, M., Neefus, C., Oberreiter, S., Paul, D., Ruhfel, B.R., Sasek, T., Shaw, J., Soltis, P.S., Watson, K., Weeks, A. & Mast, A.R. 2015. Digitization workflows for flat sheets and packets of plants, algae, and fungi. *Applic. Pl. Sci.* 3(9): 1500065. https://doi.org/10.3732/apps.1500065

Pearson, K.D., Nelson, G., Aronson, M.F.J., Bonnet, P., Brenskelle, L., Davis, C.C., Denny, E.G., Ellwood, E.R., Goëau, H., Heberling, J.M., Joly, A., Lorieul, T., Mazer, S.J., Meineke, E.K., Stucky, B.J., Sweeney, P., White, A.E. & Soltis, P.S. 2020. Machine learning using digitized herbarium specimens to advance phenological research. *BioScience* 70(6): biaa044. https://doi.org/10.1093/biosci/biaa044

Pignal, M. & Michiels, H. 2012. Switching to the fast track: Rapid digitization of the world's largest herbarium. In: Botany 2011 – Colombus, Ohio. Botanical Society of America. http://collections.mnhn.fr/wiki/attach/Visit_October2012/Paris-Herbarium-Digitization_2012-07-12.pdf

Rogers, N. 2016. Museum drawers go digital. *Science* 352: 762–765. https://doi.org/10.1126/science.352.6287.762

Schorn, C., Weber, E., Bernardos, R., Hopkins, C. & Davis, C. 2016. The New England Vascular Plants Project: 295,000 specimens and counting. *Rhodora* 118: 324–325. https://doi.org/10.3119/15-34

Slijkhuis, O. 2014. Mass digitization of a scientific biodiversity collection. Pp. 283–288 in: Pavlov, R. & Stanchev, P. (eds.), *Digital presentation and preservation of cultural and scientific heritage*, vol. 4. Sofia: Institut po matematika i informatika – Bŭlgarska akademiya na naukite.

Soltis, P.S. 2017. Digitization of herbaria enables novel research. *Amer. J. Bot.* 104: 1281–1284. https://doi.org/10.3732/ajb.1700281

Sweeney, P.W., Starly, B., Morris, P.J., Xu, Y., Jones, A., Radhakrishnan, S., Grassa, C.J. & Davis, C.C. 2018. Large-scale digitization of herbarium specimens: Development and usage of an automated, high-throughput conveyor system. *Taxon* 67: 165–178. https://doi.org/10.12705/671.9

Thiers, B. 2017–. Index Herbariorum: A global directory of public herbaria and associated staff. New York Botanical Garden's Virtual Herbarium. http://sweetgum.nybg.org/ih

Thiers, B.M., Tulig, M.C. & Watson, K.A. 2016. Digitization of The New York Botanical Garden Herbarium. *Brittonia* 68: 324–333. https://doi.org/10.1007/s12228-016-9423-7

Van Oever, J.P. & Gofferjé, M. 2012. 'From Pilot to production': Large scale digitisation project at Naturalis Biodiversity Center. *ZooKeys* 209: 87–92. https://doi.org/10.3897/zookeys.209.3609

Willis, C.G., Ellwood, E.R., Primack, R.B., Davis, C.C., Pearson, K.D., Gallinat, A.S., Yost, J.M., Nelson, G., Mazer, S.J., Rossington, N.L., Sparks, T.H. & Soltis, P.S. 2017. Old plants, new tricks: Phenological research using herbarium specimens. *Trends Ecol. Evol.* 32: 531–546. https://doi.org/10.1016/j.tree.2017.03.015