

Digitizing Virginia's Herbaria: A Call to Citizen-scientists

Andrea Weeks¹

George Mason University
Department of Biology and Ted R. Bradley Herbarium
4400 University Drive
Fairfax, Virginia 22030

ABSTRACT

A consortium of 11 Virginia herbaria was funded by the National Science Foundation in 2014 to create a publicly-accessible online database of its ca. 300,000 Virginian herbarium specimens, including high-resolution digital images of all specimens. Participation by citizen-scientists is integral to the project plan. Members of the public help transcribe label data from the high-resolution digital images of the herbarium specimens into database fields via the crowd-sourcing platform, Notes from Nature (www.notesfromnature.com). This paper reviews the project's progress to date, shares links to this emerging resource for research and education, and issues a call to citizen-scientists to assist in liberating these important historical data about Virginia's flora.

Key words: plants, citizen-science, crowd-sourcing, biodiversity informatics, southeastern United States.

INTRODUCTION

Our growing knowledge of the flora of Virginia is integrally tied to herbaria, research collections that house archivally prepared plant specimens gathered and identified by taxonomic experts. Every species treatment in the *Flora of Virginia* (Weakley et al., 2012) and most of the occurrence records in the *Digital Atlas of the Virginia Flora* (Virginia Botanical Associates, 2016) are based on these physical vouchers (Fig. 1). Twenty-five active Virginian herbaria hold over 500,000 plant specimens dating from the 19th century to the present day and the majority of these specimens were collected in the state (Fig. 2). These institutions are a lasting source of our knowledge about plants in the Commonwealth; they and the people who contribute specimens to these collections are the engines of discovery.

However, the contents of Virginian herbaria have not been electronically accessible to researchers or the public until recently. None served their databases online nor could they provide high-resolution digital images of their specimens for virtual inspection. This deficiency has precluded Virginian herbaria from fully

contributing to 21st century biodiversity informatics research, which aggregates records about the distribution, phenology, and ecology of plant species across large geographic regions through time, often for conservation purposes. Using these data to more accurately manage the preservation of native plant species in the face of habitat destruction, exotic species invasions, and rapidly shifting climate truly is a grand challenge for our times (Graham et al., 2004). Improving worldwide access to Virginian collections should also speed taxonomic discovery because over 80% of new species are recognized and uncovered for the first time in herbarium collections rather than in the field (Bebber et al., 2010).

As a first step toward liberating these data for the Commonwealth, the National Science Foundation has funded the digitization of 11 Virginian herbaria as part of a larger research collaboration among botanists in the southeastern US entitled, "The Key to the Cabinets: Building and Sustaining a Research Database for a Global Biodiversity Hotspot." The plan of digitization work includes collecting high-resolution digital images for all specimens and establishing publicly accessible databases of each herbarium's holdings according to best-practices developed by the natural history collections community (Nelson et al., 2015). The participation of citizen-scientists is integral to this plan.

¹email: aweeks3@gmu.edu

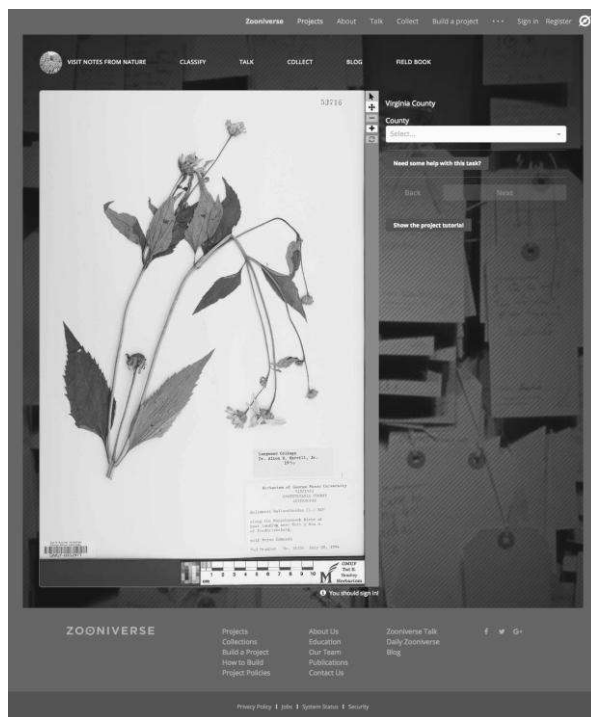


Fig. 1. Screenshot of an herbarium specimen image on the “Plants of Virginia” expedition hosted by the Notes from Nature crowd-sourcing platform (<http://www.notesfromnature.org>). This site allows citizen-scientists to transcribe label data from images of herbarium specimens to build a publicly-accessible database of the contents of Virginian herbaria (<http://www.serneportal.org>).

Members of the public transcribe label data from the high-resolution digital images into database fields via the online platform Notes from Nature (www.notesfromnature.com). This method of accessing labor across the internet, which is called crowd-sourcing, is expected to help digitize over 4.5 million southeastern US herbarium specimens from 106 US herbaria.

HERBARIUM DIGITIZATION IN VIRGINIA: INSTITUTIONAL PARTICIPANTS AND DATA MANAGEMENT

The 11 Virginian herbaria included for digitization as part of the “Key to the Cabinets” grant comprise both mid- and small-sized collections distributed across the state and are estimated to house ca. 300,000 specimens collected in the Commonwealth. These include herbaria at Virginia Polytechnic Institute and State University (VPI, 75,600; Index Herbariorum acronym from Thiers [2016], estimated number of Virginian specimens, respectively), Longwood University (FARM, 58,650),

George Mason University (GMUF, 45,500), Lynchburg College (LYN, 44,800), Virginia Military Institute (VMI, 21,000), University of Richmond (URV, 14,000), Lord Fairfax Community College (LFCC, 14,000), James Madison University (JMU, 12,600), Virginia Commonwealth University (VCU [housed at Lewis Ginter Botanical Garden], 11,900), the City of Alexandria Herbarium (AVCH, 4,500), and Bridgewater College (BDWR, 4,000). Two digital imaging stations rotate among these herbaria, which allow rapid capture of high-resolution digital images of herbarium specimens during the creation of skeletal database records. Digitized specimens are barcoded, and the unique barcode value links skeletal database records with the high-resolution image files. Curators of the collections and undergraduate student-workers primarily collect these data and manage their dissemination.

The high-throughput data collection and management workflow balances the need to share data publicly while protecting data about legally-protected or otherwise sensitive species (Fig. 3). In step 1, a digital image of each specimen is taken with a 25 megapixel digital SLR camera and a minimal database entry about the specimen is created online containing its barcode number, species name, and the state in which it was collected. Each herbarium has a publicly accessible database within the SouthEast Regional Network of Expertise and Collections (SERNEC) portal (www.serneportal.org), whose administrative controls are under the direction of the individual curator (step 2). For each specimen, the high-resolution digital image and the database record are displayed together. In step 3, data are filtered to exclude detailed information about legally-protected or otherwise sensitive taxa from being viewed by the public. This customizable filter, represented by the dotted lines in Figure 3, obscures the high-resolution digital image of the specimen and any detailed locality data. Curators evaluate direct requests for sensitive data from the public on a case-by-case basis and can adjust data access permissions for individual researchers within the SERNEC portal. In step 4, the high-resolution digital image of each specimen is sent to Notes from Nature for the full transcription of its remaining label data. In step 5, the Notes from Nature citizen-science platform guides participants through transcribing label data into database fields including: county, locality, ecological information, date collected, collector, and collector number (Fig. 1). Label data from each specimen are transcribed three times by different users and a computer algorithm creates the most correct transcription, which is returned to the SERNEC portal database in step 6. In step 7, all image and text data

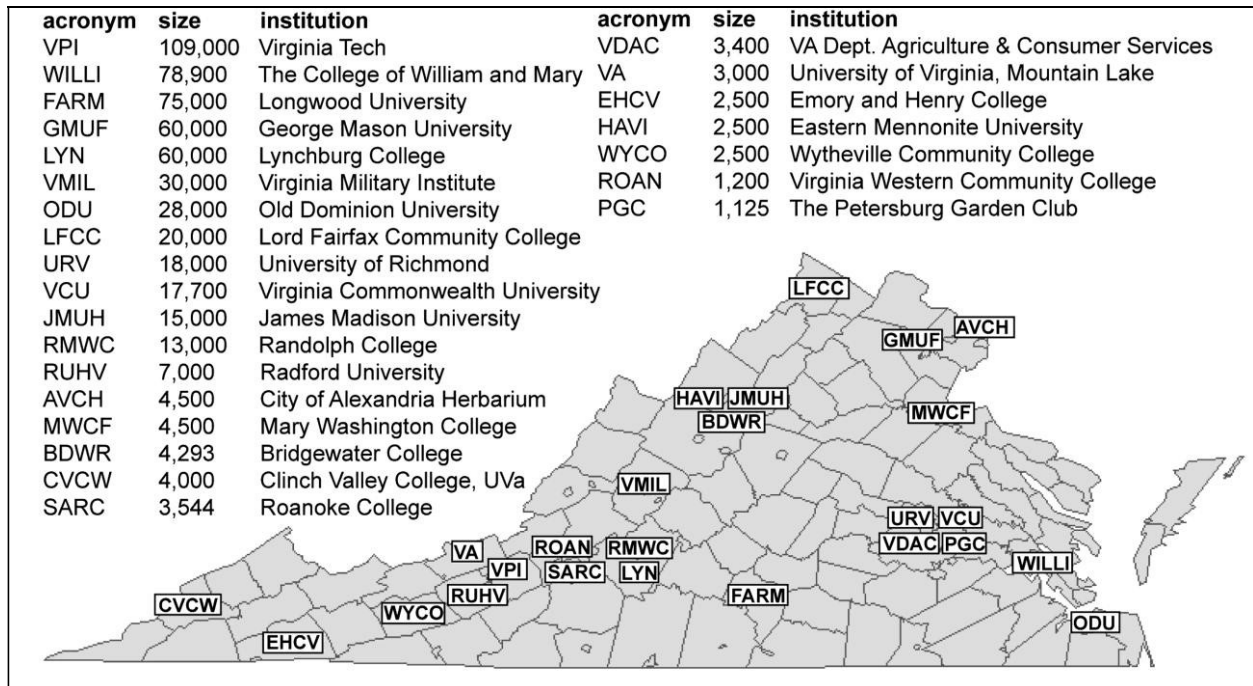


Fig. 2. Distribution and size of the 25 active herbaria in Virginia. Data compiled from *Index Herbariorum* (Thiers, 2016).

within from the SERNEC portal database are shared with the Global Biodiversity Information Facility (www.gbif.org), an international clearinghouse for biodiversity specimen data that is freely accessible to the public (step 8).

PROGRESS TO DATE: DIGITIZATION IN VIRGINIA

In the two years since the inception of the “Key to the Cabinets” herbarium digitization project all steps in the data collection and management workflow have become operational and there is evidence that the new database resource is leading to new inquiries and insights about the specimens. Within Virginia, seven of the original 11 herbaria have completed imaging (GMUF) or are in the process of imaging their specimens (VPI, FARM, LYN, URV, JMUH, AVCH). Two additional herbaria housed at the College of William and Mary (WILLI) and the University of Mary Washington (MWCF), now share their data through the SERNEC portal, as well. As of September 2016, 145,376 database records have been created and 50,170 high-resolution digital images of Virginia-collected herbarium specimens have been uploaded to the site. For example, high-throughput digitization of GMUF took approximately 393 hours at an average rate of 92 specimens/hour and yielded 35,351 unique database entries and images on the GMUF SERNEC portal

database. Since completing the digitization process, GMUF has received requests for loans of specimens from researchers who have been able to preview images of targeted specimens. Surprisingly, the number of Virginian specimens held by GMUF was revealed to be 22% smaller than originally estimated. Anecdotal reports from other herbaria in the process of digitizing suggest this trend may be common and result in a smaller final dataset for Virginia than originally anticipated.

PROGRESS TO DATE: PROTECTING DATA ABOUT SENSITIVE VIRGINIAN PLANT TAXA

A key step in completing the data management workflow included creating the filters that redact from public view detailed information about sensitive Virginian plant taxa. The SERNEC portal includes two tiers of filters: a higher-level filter that can redact specimen data held by any participating herbaria and a lower-level filter that is specific to individual herbaria. In all cases, curators have full authority to redact or release records of any specimen in their collection. The rationale for creating higher-level filters is to provide a minimal level of protection at the outset of the digitization process for sensitive plant taxa in all participating SERNEC herbaria. For example, detailed locality information and images of all federally-protected plant species are automatically redacted by

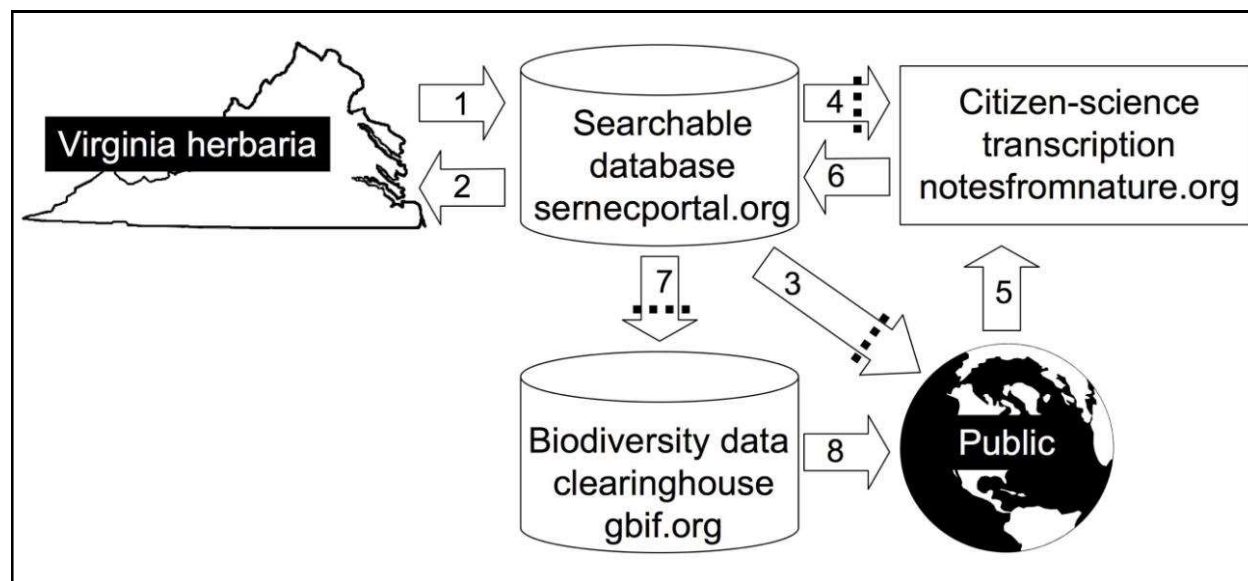


Fig. 3. The data collection and management workflow for herbarium digitization, charting the liberation of herbarium specimen data held by individual herbaria to the public at large. Dotted lines represent customized filters that protect the image and detailed locality data about sensitive plant taxa from unregulated exposure to the public.

the SERNEC portal. Although redacted data are not viewable by the public, the data remain in the database and can be accessed by individuals who are given additional permissions within the herbarium collection that holds the specimens of interest. Herbarium curators receive requests from potential researchers and control permissions.

The higher-level sensitive taxon filter for Virginia-collected plants was expanded to include the 12 taxa that have threatened or endangered status only within the state as well as 35 taxa that are commonly harvested from the wild for medicinal or horticultural purposes (Table 1). As a consequence of both Federal and state-level filters, the images and detailed locality data for 60 sensitive plant taxa native to Virginia are automatically and immediately redacted from public view across the SERNEC portal. Filters also capture taxonomic synonyms since sensitive specimens are often filed under different names depending on the herbarium.

PROGRESS TO DATE: OUTREACH AND EDUCATION IN VIRGINIA

Participation by citizen-scientists on the crowd-sourcing platform Notes from Nature is integral to the success of the “Key to the Cabinets” herbarium digitization project as is the full dissemination of the digitized data to researchers around the world. Digital images of specimens from Virginian herbaria needing

label transcription are presented on the Notes from Nature platform within projects entitled, “Plants of Virginia.” The projects bundle ca. 5,000 specimen images following a theme, such as taxonomy, geography or ecology, and are termed “expeditions” by the Notes from Nature platform. Results from completed Plants of Virginia expeditions indicate that thousands of label transcriptions can be accomplished in a matter of days via this method of crowd-sourcing. In the last two years, presentations and workshops about the project and the Notes from Nature crowd-sourcing platform have been organized for non-academic interest groups including the Virginia Native Plant Society and the Virginia Master Naturalists, for whom Notes from Nature participation now fulfills continuing education credit. Presentations and workshops have also been delivered to academic and scientific research audiences at the annual meetings of the Association of Southeastern Biologists, Virginia Academy of Science, Virginia Association for Environmental Education, and the Virginia Botanical Associates, which includes staff from the Virginia Natural Heritage Program. Two school lesson plans incorporating the Notes from Nature crowd-sourcing platform appropriate for grades 6-8 and 9-12 also have been developed and disseminated online (<http://sernec.appstate.edu/education-outreach>). Each lesson plan incorporates several of Virginia’s Standards of Learning for science and biology.

Table 1. List of native Virginian plant taxa filtered by the SERNEC database portal (www.sernecportal.org). If herbarium specimens of these taxa were collected from the state of Virginia and are present within the database, the filter will automatically obscure from public view the herbarium specimen image and detailed locality information. Synonyms for the listed taxa (not included here) are also captured by the filter. LE = legally endangered; LT = legally threatened.

Scientific name	Common name	Conservation Status
<u>Federally threatened or endangered</u>		
<i>Aeschynomene virginica</i> Britton & Stern et Poggenb.	Sensitive Joint-vetch	LT
<i>Amaranthus pumilus</i> Raf.	Sea-beach Amaranth	LT
<i>Arabis serotina</i> Steele	Shale Barren Rock Cress	LE
<i>Betula uber</i> (Ashe) Fernald	Virginia Roundleaf Birch	LT
<i>Cardamine micranthera</i> Rollins	Small-anthered Bittercress	LE
<i>Echinacea laevigata</i> (C.L. Boynt. & Beadle) S.F. Blake	Smooth Coneflower	LE
<i>Helenium virginicum</i> Blake	Virginia Sneezeweed	LT
<i>Helonias bullata</i> L.	Swamp-pink	LT
<i>Houstonia montana</i> Small	Roan Mountain Bluet	LE
<i>Iliamna corei</i> Sherff	Peters Mountain Mallow	LE
<i>Isotria medeoloides</i> (Pursh) Raf.	Small Whorled Pogonia	LT
<i>Platanthera leucophaea</i> (Nutt.) Lindl.	Prairie Fringed Orchid	LT
<i>Ptilimnium nodosum</i> (Rose) Mathias	Harperella	LE
<i>Rhus michauxii</i> Sarg.	Michaux's Sumac	LE
<i>Schwalbea americana</i> L.	Chaffseed	LE
<i>Scirpus ancistrochaetus</i> Schuyler	Northeastern Bulrush	LE
<i>Spiraea virginiana</i> Britton	Virginia Spiraea	LT
<u>Virginia-only threatened or endangered</u>		
<i>Boltonia montana</i> J.F. Townsend & V. Karaman-Castro	Valley Doll's-daisy	LE
<i>Carex juniperorum</i> Catling, Reznicek, & Crins	Juniper Sedge	LE
<i>Clematis viticaulis</i> Steele	Millboro Leatherflower	LT
<i>Corallorhiza bentleyi</i> Freudenstein	Bentley's Coralroot	LE
<i>Fimbristylis perpusilla</i> Harper ex Small & Britt.	Harper's Fimbristylis	LE
<i>Ilex collina</i> Alexander	Long-stalked Holly	LE
<i>Isoetes virginica</i> N.E. Pfeiffer	Virginia Quillwort	LE
<i>Juncus caesariensis</i> Coville	New Jersey Rush	LT
<i>Nuphar sagittifolia</i> (Walt.) Pursh	Narrow-leaved Spatterdock	LT
<i>Panax quinquefolium</i> L.	Ginseng	LT
<i>Scirpus flaccidifolius</i> (Fern.) Schuyler	Reclining Bulrush	LT
<i>Trifolium calcaricum</i> Collins & Wieboldt	Running Glade Clover	LE
<u>Other Virginian species commonly wild-harvested</u>		
<i>Actaea racemosa</i> L.	Black Cohosh	none
<i>Aletris farinosa</i> L.	Colicroot	none
<i>Allium burdickii</i> (Hanes) A.G. Jones	White Ramps	none
<i>Allium tricoccum</i> Ait.	Red Ramps	none
<i>Caulophyllum thalictroides</i> (L.) Michx.	Blue Cohosh	none
<i>Chamaelirium luteum</i> (L.) Gray	Devil's-bit	none
<i>Cypripedium acaule</i> Aiton	Pink Lady's-slipper	none
<i>Cypripedium candidum</i> Muhl. ex Willd.	Small White Lady's-slipper	State Rare
<i>Cypripedium kentuckiense</i> C.F. Reed	Kentucky Lady's-slipper	State Rare
<i>Cypripedium parviflorum</i> Salisb. var. <i>pubescens</i> (Willd.) Knight	Large Yellow Lady's-slipper	none
<i>Cypripedium reginae</i> Walter	Showy Lady's-slipper	State Rare
<i>Echinacea pallida</i> (Nutt.) Nutt.	Pale Purple Coneflower	none
<i>Echinacea purpurea</i> (L.) Moench	Purple Coneflower	none
<i>Endodeca serpentaria</i> (L.) Raf.	Virginia snakeroot	none
<i>Galax urceolata</i> (Poir) Brummit	Galax	none
<i>Hydrastis canadensis</i> L.	Goldenseal	State Watch-list
<i>Liatris helleri</i> T.C. Porter	Heller's blazing-star	none
<i>Sanguinaria canadensis</i> L.	Bloodroot	none
<i>Sarracenia flava</i> L.	Yellow Pitcher Plant	State Rare
<i>Sarracenia purpurea</i> L.	Northern Pitcher Plant	State Rare
<i>Shortia galacifolia</i> Torrey & A. Gray var. <i>galacifolia</i>	Shortia	none
<i>Trillium cernuum</i> L.	Nodding Trillium	State Rare
<i>Trillium erectum</i> L.	Red Trillium	none
<i>Trillium flexipes</i> Raf.	Drooping Trillium	State Rare
<i>Trillium grandiflorum</i> (Michx.) Salisb.	Great White Trillium	none
<i>Trillium luteum</i> (Muhl.) Harbison	Yellow Toad-shade	none
<i>Trillium nivale</i> Riddell	Snow Trillium	State Rare
<i>Trillium pusillum</i> Michx. var. <i>virginianum</i> Fern.	Virginia Least Trillium	State Rare
<i>Trillium sessile</i> L.	Sessile Trillium	none
<i>Trillium sulcatum</i> Patrick	Southern Red Trillium	none
<i>Trillium undulatum</i> Willd.	Painted Trillium	none

LOOKING TO THE FUTURE: A CALL TO CITIZEN-SCIENTISTS

Bringing to light more than 200 years of detailed information about the botanical diversity of Virginia is both a logistical challenge and a remarkable opportunity to advance research and education about the state's flora. The "Key to the Cabinets" herbarium digitization project is employing many technological advances that will partially address this grand challenge in a reasonable timeframe, including efficient digital image processing, flexible information architecture, and distributed computing. Yet, arguably none has greater power to speed this process than crowd-sourced label transcription (Ellwood et al., 2015), such as that facilitated by the Notes from Nature platform. During crowd-sourcing, many people contribute a small amount of labor each to accomplish a large task collectively and often quite rapidly. Viewed from this perspective, crowd-sourcing is a resource-efficient strategy for collecting data. However, crowd-sourcing is also an innovative method for outreach and education. Engaging volunteers in crowd-sourcing can build communities of citizen-scientists, expand awareness of and support for the scientific merits of the research project outside of academia as well as provide a launching pad for training the next-generation of scientists. The Notes from Nature platform facilitates communication among all stakeholders in addition to guiding rapid transcription of label data by volunteers. Readers of *Banisteria* are encouraged to participate in Notes from Nature to hasten the process of liberating important historical data about Virginia's diverse flora. A tutorial is available within the "Plants of Virginia" expeditions and an instructional webinar on the transcription process is available through the Virginia Master Naturalist's website: <http://www.virginiamasternaturalist.org/continuing-education.html>.

ACKNOWLEDGEMENTS

This work was supported in part by National Science Foundation award no. 1410086. Any opinions, findings, and conclusions or recommendations expressed in this material are mine and do not necessarily reflect the views of the National Science Foundation. I also wish to thank the Virginia Native Plant Society for a research award that supported launching the herbarium digitization initiative in the Commonwealth in 2015.

LITERATURE CITED

- Bebber, D. P., M. A. Carine, J. R. I. Wood, A. H. Wortley, D. J. Harris, G. T. Prance, G. Davide, J. Paige, T. D. Pennington, N. K. B. Robson, & R. W. Scotland. 2010. Herbaria are a major frontier for species discovery. *Proceedings of the National Academy of Sciences of the United States of America* 107: 22169-22171.
- Ellwood, E. R., B. Dunckel, P. Flemons, R. Guralnick, G. Nelson, G. Newman, S. Newman, D. Paul, G. Riccardi, N. Rios, K. C. Seltmann, & A. R. Mast. 2015. Accelerating digitization of biodiversity research specimens through online public participation. *BioScience* 65: 386-396.
- Graham, C. H., S. Ferrier, F. Huettman, C. Moritz, & A. T. Peterson. 2004. New developments in museum-based informatics and applications in biodiversity analysis. *Trends in Ecology and Evolution* 19: 497-503.
- Nelson, G., P. Sweeney, R. Rabeler, L. Wallace, D. Allard, H. Brown, J. R. Carter, M. Denslow, E. Ellwood, C. Germain-Aubrey, E. Gilbert, E. Gillespie, L. Goertzen, B. Legler, B. Marchant, T. Marsico, C.A. McCormick, A. Morris, Z. Murrell, M. Nazaire, C. Neefus, D. Paul, B. Ruhfel, T. Sasek, P. S. Soltis, J. Shaw, K. Watson, A. Weeks, & A. R. Mast. 2015. Digitization workflows for flat sheets and packets of plants, algae, and fungi. *Applications in Plant Science* 3(9): 150065. 9 pp.
- Thiers, B. 2016. Index Herbariorum: A global directory of public herbaria and associated staff. New York Botanical Garden's Virtual Herbarium. <http://sycamore.nybg.org/science/ih/>. (Accessed April 2016).
- Virginia Botanical Associates. 2016. Digital Atlas of the Virginia Flora. Virginia Botanical Associates, Blacksburg, VA. <http://www.vaplantatlas.org>. (Accessed 10 September 2016).
- Weakley, A. S., J. C. Ludwig, & J. F. Townsend. 2012. *Flora of Virginia*. B. Crowder (ed.). Foundation of the Flora of Virginia Project Inc., Richmond, VA. Botanical Research Institute of Texas Press, Fort Worth, TX. 1,554 pp.