Critic's Choice Essay

VIRTUAL HERBARIA COME OF AGE

Wayne Elisens
Department of Microbiology and Plant Biology
Oklahoma Biological Survey
University of Oklahoma

These are exciting times for natural history collections. An international effort is underway to make images and data of biological specimens available in electronic format via digitization. These initiatives are an effort to bring natural history collections *out of the dark* of museum and herbarium cabinets and *into the light* of public access for use by stakeholders in government, academia, biodiversity organizations, business, and K-12 education. The democratization of information contained in natural history collections through images and online databases is an important new development to better investigate our natural world and solve important social and environmental problems (Scoble 2010).

For herbarium collections, digitized images and data from specimens are generally referred to as a virtual herbarium. What exactly do we mean by digitization of natural history collections? For plants, digitizing collections transforms herbarium specimens into digital images and label data sorted (parsed) into its component units such as names, locations, collectors, dates, habitats, and reproductive state. All data and images are fully searchable and distributed in electronic format, such as virtual herbaria. There are several outstanding examples of virtual herbaria already online, such as Australia's Virtual Herbarium (http://avh.ala.org.au/) and the New York Botanical Garden's Virtual Herbarium (http://sciweb.nybg.org/science2/VirtualHerbarium.asp.html). According to the US Interagency Working Group on Scientific Collections (IWGSC 2007), plant specimen data distributed via virtual herbaria would have a profound impact on science education and investigations of environmental change and quality, invasive species, public health, national security, bioscience research, and many other issues (NIBA 2010).

Why digitize natural history collections? Think for a moment about the incomparable treasure trove of biodiversity information contained in the world's natural history collections. If we focus just on plants, herbarium specimens document most of what is known about the world's plant species diversity and represent a 200+ year record of what species were present at a given *location* and at a given *time*. Herbaria collections not only document the different kinds of plants constituting a flora, but they record valuable information about where they occurred and when they were flowering or fruiting. Plant specimens provide a spatial and temporal window into the dynamic processes of plant diversity, introduction and spread of exotics, expansion and contraction of species ranges, and changes in time of flowering and fruiting. Is a digitized herbarium specimen as valuable as the specimen itself? Certainly not, and this is one of the main reasons for maintaining natural history collections in museums and herbaria. The primary rationale for digitizing specimens is **access**. Scientists throughout the world will have greatly enhanced access to digitized specimens, which greatly adds to their value for research and education.

In the US, a key component is in place to assist efforts to digitize biological research collections (e.g., herbarium specimens) – the Integrated Digitized BioCollections resource

https://doi.org/10.22488/okstate.17.100093

(iDigBio; www.idigbio.org). Tools and training provided by iDigBio are funded by the National Science Foundation, who also established a 10-year funding program entitled Advancing Digitization of Biological Collections (ADBC) to aid conversion of biodiversity collections into electronic formats. These advances in funding and infrastructure were established using recommendations of the National Science and Technology Council, who recognized the importance of biocollections for national science infrastructure.

With an estimated 90 million herbarium specimens in U.S. herbaria (Tulig et al. 2012), is it feasible to construct a US Virtual Herbarium comparable to Australia's Virtual Herbarium based on "only" 6 million specimens? At a minimum, digitization of biocollections involves specimen imaging, image processing, electronic data capture, and georeferencing of locality descriptions (Nelson et al. 2012). Mass digitization methods continue to be refined and automated (Beaman and Cellinese 2012), but it is unlikely that all US herbarium specimens can be digitized in a 10-year timeframe. However, a recent survey conducted by the US Virtual Herbarium project (Barkworth and Murrell 2012) indicated that ca. 30% of herbarium specimen labels were already databased. While it appears that much digitization has occurred at the individual herbarium or regional level, there has been no coordinated national effort to expedite digitization of biocollections. However, the iDigBio mission aims to fill that void and has a major objective to facilitate access to US biocollection data. This goal is certainly feasible, especially since a **global** portal for digitized images and data from natural history collections already exists – GBIF, the Global Biodiversity Information Facility (www.gbif.org). GBIF currently serves up more than 300 million specimen records.

Herbaria throughout the country are actively engaged in efforts to image and database information and to present them in searchable online formats. International standards and best practices for data capture have been established and are being implemented by the collections community nationwide. Luckily, botanists in Oklahoma demonstrated great foresight by establishing a data portal for digitized herbarium label data for specimens collected in Oklahoma – the Oklahoma Vascular Plants Database (OVPD; Hoagland et al. 2004). With the OVPD as a firm foundation and a collaborative network in place among curators in the state and region, Oklahoma herbaria are poised to expand their digitization efforts. This endeavor will help develop the concept of a virtual herbarium to maturity and will undoubtedly enhance the value and access of real herbaria.

LITERATURE CITED

Barkworth, M. E. and Z. E. Murrell. 2012. The US Virtual Herbarium: Working with individual herbaria to build a national resource. *ZooKeys* 209:55-73.

Beaman, R. S. and N. Cellinese. 2012. Mass digitization of scientific collections: New opportunities to transform the use of biological specimens and underwrite biodiversity science. *ZooKeys* 209:7-17.

Hoagland B. W., A. K. Buthod, I. H Butler, P. H. C. Crawford, A. H. Udasi, W. J. Elisens, and R. J. Tyrl. 2004**.** Oklahoma Vascular Plants Database. Oklahoma Biological Survey, University of Oklahoma, Norman. Available from: http://www.oklahomaplantdatabase.org).

IWGSC, Interagency Working Group on Scientific Collections. 2007. Scientific collections: Mission-critical infrastructure for Federal Science Agencies. Available from: http://www.whitehouse.gov/files/documents/ostp/NSTC%20Reports/Revision_1-22_09_CL.pdf).

Nelson, G., D. Paul, G. Riccardi, and A. R. Mast. 2012. Five task clusters that enable

efficient and effective digitization of biological collections. *ZooKeys* 209:19-45.

NIBA, Network Integrated BioCollections Alliance. 2010. A strategic plan for establishing a network integrated biocollections alliance. Available from: http://digbiocol.wordpress.com/brochure/

Scoble, M. J. 2010. Natural history collections digitization: Rationale and value. *Biodiversity Informatics* 7:77-80.

Tulig, M., N. Tarnowsky, M. Bevans, A. Kirchgessner, and B. M. Thiers. 2012. Increasing the efficiency of digitization workflows for herbarium specimens. *ZooKeys* 209:103-113.