On the Importance of Scientific Collections
A Series by
Natural Science Collections Alliance

Biological Specimens Go Online: The New Digital Frontier of Collections

By Eve S. McCulloch

Natural history collections contain a wealth of information about the organisms inhabiting our planet. The knowledge produced from collections fundamentally shapes our understanding of the variety of life on Earth, the effects of climate change on species distributions and extinction risk, and the threat posed by invasive species and pathogens, among many other issues. With this information, we can advance and support basic science, economy and trade, public health and safety, agriculture, and national security.

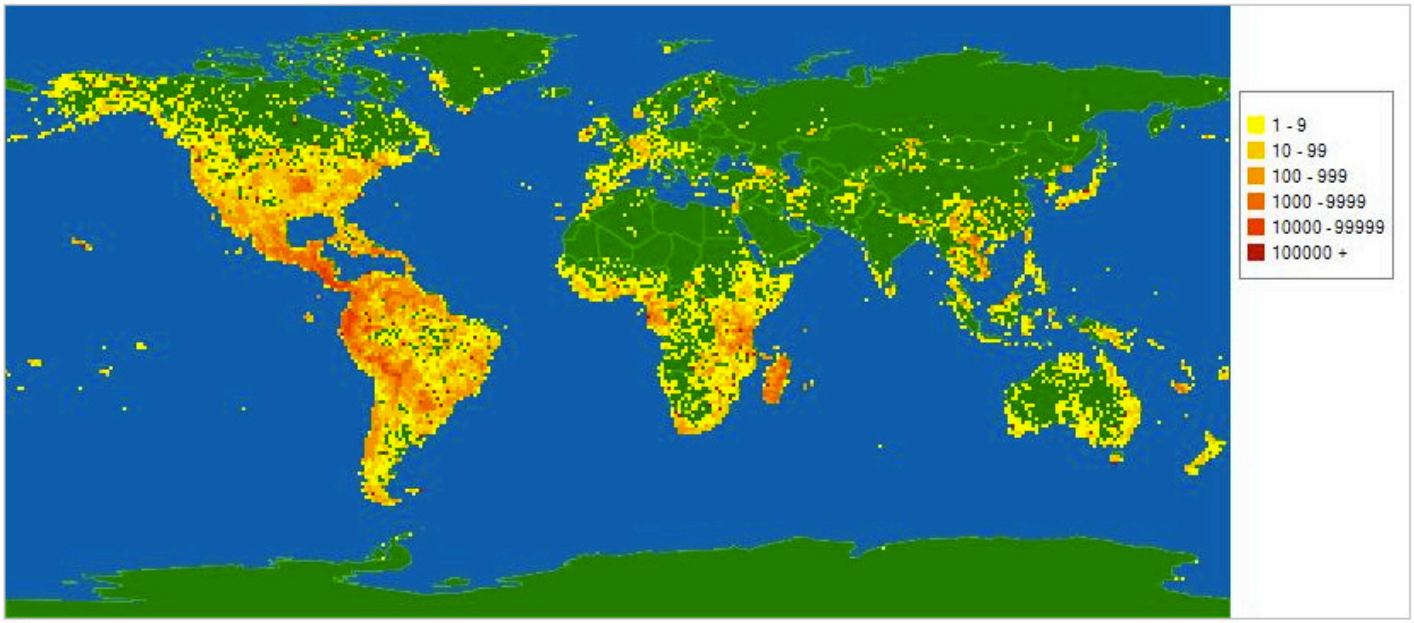
This information, however, is often decentralized, shared among a multitude of institutions located around the world, and retrieving the information can be challenging. This is beginning to change as technological advances revolutionize storage, access, and use of biological collections data.

Uses and Users of Digitized Plant Collections

The Missouri Botanical Garden is one example of a collection going online, with its massive plant database Tropicos. The Garden's herbarium is the second largest in the U.S., and sixth largest in the world—and Tropicos contains data on most of its specimens. Tropicos is the largest single botanical database freely available online, with 4.1 million records of plant species globally, including more than 360,000 type specimens, from which new species are described. The site averages over 40,000 access requests daily. Its scientific collections are worldwide in scope. The value of such a large collection is incalculable, and the specimens are just the beginning. With them, researchers can address issues of immediate societal and scientific importance, including understanding and preserving biodiversity.

Digitized scan of a plant specimen with associated data (location, collector, date), and a photo of the live plant, both from Tropicos. With Tropicos, users can link millions of specimen records to the place and date each was collected, and in many cases access associated data. Credit: Missouri Botanical Garden.

Using Tropicos data, the Garden is working with a consortium of institutions to produce the "World Flora Online" database by 2020, which will be the first comprehensive overview of all the world's known plants. Tropicos also has projects dealing with ethnobotany—the study of the relationships between people and plants—and climate change. Jan Salick, an ethnobotanist at the Garden, runs one such project. Since 2000, she has been working in the far-reaches of the Himalayas, along a 1500-kilometer transect that runs from western Nepal to the eastern Himalayas in China. She describes the first time she went into the Himalayas, "it was drop-dead beautiful and culturally incredible. I was looking at [plant] diversity that was equivalent if not greater than some of the tropical regions in which I had worked, and I was immediately captured."



The Missouri Botanical Garden houses thousands of plants collected around the world. The number of specimen records found in Tropicos are shown for each region of the world. Credit: Missouri Botanical Garden.

But the Himalayas are changing. They show some of the fastest rates of climate change in the world—second only to the Earth's polar regions—yet unlike the poles the Himalayas have the greatest diversity of plants of any region in the world outside of the tropics.

"People say [the Himalayas] are the world's thermometer, a forerunner of what we can expect in the rest of the world," Salick explains. "If we have any hope of keeping ahead of climate change, we need to look in places where it is going faster than, say, here in Missouri." Moreover, what happens there has wide-reaching influence. "The Himalayas are a huge high wall with tremendous amounts of ice and snow. Sometimes called the 'third pole,' the Himalayas shape climate, ocean currents, and weather patterns. Asia would not have monsoons except for the Himalayas. We wouldn't have the circulation patterns in the Pacific Ocean except for the Himalayas. Knowing what is happening there affects [how we understand] what happens in other parts of the world as well."

As Salick and her team of Tibetan and Chinese botanists work to unravel the mysteries of the Himalayan plantworld, and discover how climate change is altering vegetation in these extreme alpine environments, their data are also available online via Tropicos. This means researchers around the world can build upon the work conducted by the Missouri Botanical Garden, without travelling to the Himalayas.

Mt. Khawa Karpo, a Tibetan sacred mountain in the eastern Himalayas. Credit: Robert Moseley.

Mapping Life on Earth

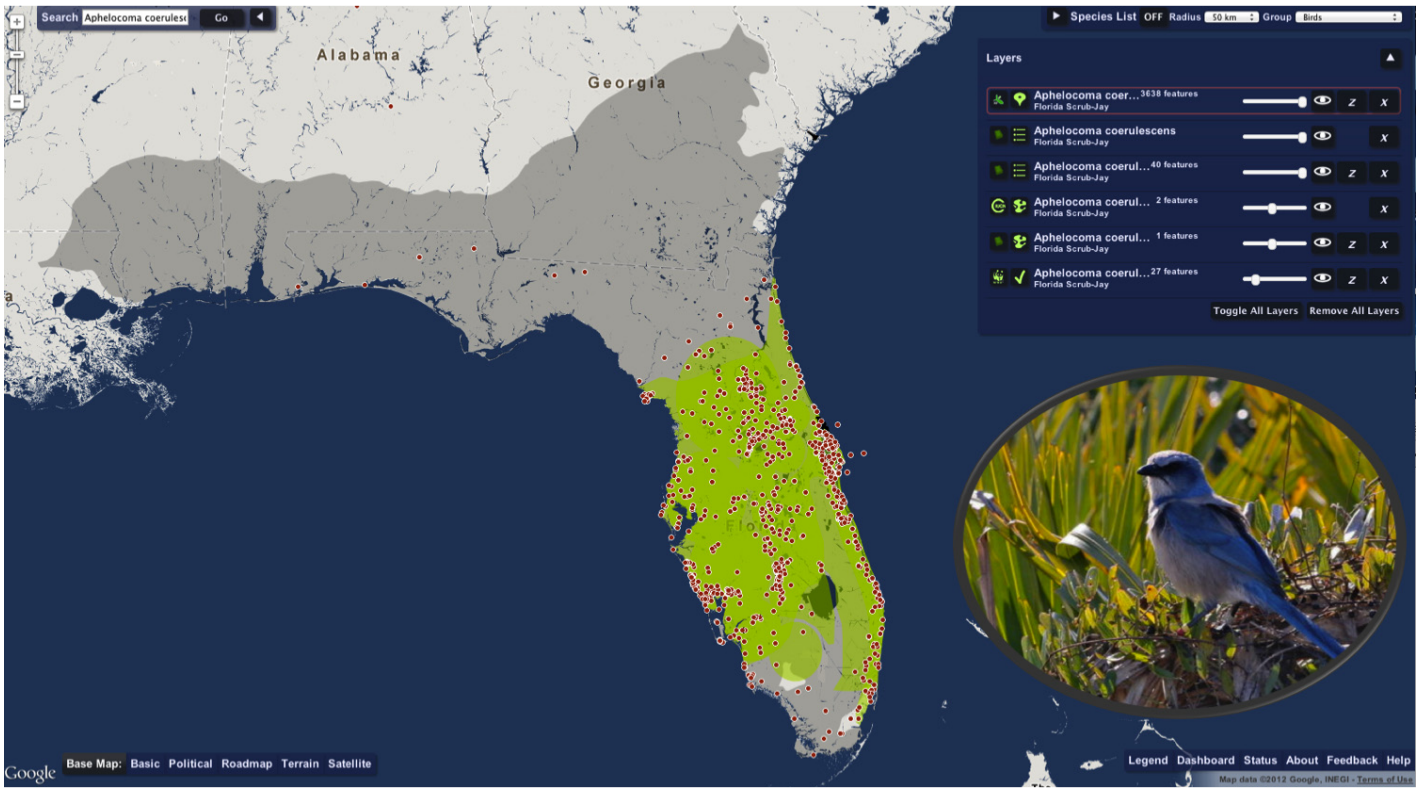
Many of the online collections databases coming of age will be broadly usable by the public, as well as by scientists. The laboratories of Robert Guralnick, curator and global change biologist at the University of Colorado at Boulder, and Walter Jetz, an ecologist at the project's lead institution, Yale University, have created such a multi-faceted tool. Map of Life is an online resource that aspires to help provide the best-possible maps of where species live.

Map of Life is still in the early demonstration phase, and is constantly growing and evolving, but already maps known distributions of approximately 46,000 species—including birds, mammals, and amphibians. The system integrates over 200 years of data from natural history collections, national park surveys, field guides, citizen scientists, and organizations such as the Global Biodiversity Information Facility, the International Union for the Conservation of Nature, and the World Wildlife Fund. It allows users to map the distribution of any given species, and separate data according to the source. Alternatively, users can search by location to get a species list for a focal area, complete with threat status, images, and links to further resources.

The potential utility for such digital data is vast. It is an education tool for children and adults. It helps expose holes in our knowledge of where species live, and provides focus for future biodiversity research. It is a resource for decision makers, stakeholders, conservationists, and land managers. For example, one could pinpoint a proposed region slated for development in order to assess which species exist in the area and might be impacted, or to analyze how species-rich that area is compared to others.

One issue at the heart of Map of Life is assessment of species' threat status. "Threat status is not a one-time designation, but changes as our planet changes. Such status information matters for governments, for those working in protected areas, and very much to the general public who ultimately want to know if their children or grandchildren will see iconic species such as polar bears, pandas, or pikas," say Guralnick and Jetz. Presently, threat status is determined using data at a much coarser resolution, but "Map of Life may ultimately help provide data, information, and knowledge for such assessments in important novel ways."

Guralnick and Jetz are already developing Map of Life so users will not only be able to access data, but will be able to contribute data. "Map of Life is growing by leaps and bounds… [and] we have more to do to continue integrating Map of Life deeply into the fabric of scientific and general public access points to biodiversity knowledge."

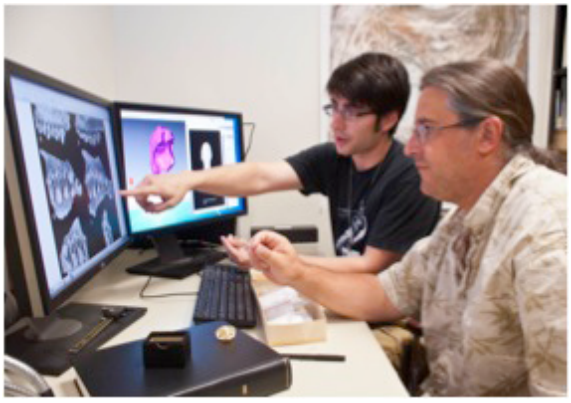


The distribution of the Florida Scrub Jay is shown in Map of Life. It is one of a handful of bird species endemic to the continental U.S. The species has recently declined in numbers, and is listed as "vulnerable" (IUCN). Pink points represent specimens from collections. Dark-gray and lime-green areas illustrate what is known of its distribution, based on data from different sources. Image insert: Mary Keim.

Taking Digitization Nation-wide

These individual efforts by innovative people and forward-thinking organizations have experienced substantial and accelerating use, yet those interested in following suite often lack resources or know-how to do so. Estimates indicate there are approximately one billion biological specimens in U.S. collections, at more than a thousand institutions. The next great challenge is coordinating nation-wide digitization.

In recognition of the huge payoffs of digitizing natural history collections, the National Science Foundation (NSF)—the federal agency charged with keeping the U.S. at the leading edge of scientific discovery—is playing a major role in supporting innovation to address the cyberinfrastructure challenge, particularly to handle large quantities of complex data, the so-called "big data" problem. Integrated Digitized Biocollections (iDigBio) is funded by NSF's Advancing Digitization of Biological Collections program, and stands at the heart of the federally supported digitization initiative.



Credit: Florida Museum of Natural History.

iDigBio is an administrative hub and houses a common portal for accessing digitized collections data. The program works to stimulate interest within the scientific community in digitization, and assist Thematic Collections Networks and other institutions that curate natural history collections to increase accessibility of their specimens and associated data. Despite being a relatively young organization, iDigBio has taken the lead in design and support of software, protocols, and community standards necessary to streamline the process of capturing collections data, and has published tested workflows (see Additional Reading). In the last year, iDigBio has organized workshops covering themes such as "train-the-trainers," "public participation in digitization of biodiversity," "IT standards," "Developing Robust Object to Image to Data (DROID workflows," among others.

Many other groups are improving accessibility for specimen data, post-digitization. VertNet is one notable such organization; it has helped mobilize millions of digitized vertebrate records using a fast, cost-effective, and scalable cloud-based data platform. The Global Biodiversity Information Facility is another major player, providing a platform for approximately 400 million species occurrence records, and still growing.

Looking Ahead

Biological collections represent roughly 250 years of scientific dedication and work. They hold immense value for research and education, and for addressing agricultural, public health, environmental, and economic concerns. Tapping the vast potential for scientific and societal benefits depends on researchers, educators, and other stakeholders having the ability to efficiently access biocollections and their associated data. There is an urgent need for an aggressive, sustained, and coordinated large-scale effort to digitize biocollections and make them widely available online. A host of institutions and individuals are progressing towards that end, in part with federal support from the National Science Foundation. The framework for digitization is in place, the payoffs of digitization are innumerable, and continued support is critical. Now is the time for stakeholders to work together to implement new technologies, and maintain the momentum of the digitization movement.

Additional Resources

Integrated Digitized Biocollections	www.idigbio.org
Map of Life	www.mappinglife.org
Tropicos®	www.tropicos.org

VertNet: Across four portals (MaNIS [mammals], HerpNET [reptiles and amphibians], ORNIS [birds], and FishNet2 [Fish]), a global consortium of institutions provides open access to approximately 85 million vertebrate records. www.vertnet.org

Global Biodiversity Information Facility: Through a global network of countries and organizations, GBIF promotes and facilitates the mobilization, access, discovery, and use of information about the occurrence of organisms over time and across the planet. Presently provides access to approximately 400 million indexed records, from 10,000 datasets. www.gbif.org

Further reading

Zookeys 209 (2012) Special Issue: "No Specimen Left Behind: Mass Digitization of Natural History Collections."

Jetz, W., J.M. McPherson, and R.P. Guralnick (2012): Integrating biodiversity distribution knowledge: toward a global map of life. *Trends in Ecology & Evolution*.

"The Grand Challenge: Unifying the Nation's Biodiversity Collections through Digitization." www.idigbio.org/content/grand-challenge-uniting-nation's-biodiversity-collections-through-digitization

Beach J., S. Blum, M. Donoghue, L. Ford, R. Guralnick, M. Mares, B. Thiers, M. Westneat, Q. Wheeler, B. Wiegmann, and the Network Integrated Biocollections Alliance (2010): A Strategic Plan for Establishing a Network Integrated Biocollections Alliance. http://digbiocol.wordpress.com/brochure/

Credit: Florida Museum of Natural History.

About NSCA

The NSC Alliance is a nonprofit association that supports natural science collections, their human resources, the institutions that house them, and their research activities for the benefit of science and society. Our 100 institutional members are part of an international community of institutions that house natural science collections and utilize them in research, exhibitions, academic and informal science education, and outreach activities.