Symbiota
Mobilizing Biodiversity Data

February 12, 2018

Symbiota Overview

Symbiota is an open source *content management system* for curating specimen- and observation-based biodiversity data. It is built on the premise that a collaborative partnership of biodiversity informaticians, collection managers, and biodiversity research communities will be most effective in creating high quality biodiversity research resources with publicly useful portals. Symbiota is driven by a suite of tools and modules that are integrated with an underlying SQL database installed on a web server. The open source software promotes use-driven code innovation built on the backend of Symbiota, extending functionality through API collaborations. The modular structure and the ease of visualizing and sharing data has fostered large and diverse end-user communities to provide comprehensive biodiversity information.

Since 2012 the use of Symbiota has greatly expanded, with 73% of projects funded by the US National digitization program (NSF-ADBC) using Symbiota. A total of 38 Symbiota portals have mobilized 11 million records from 412 "live" collections, while 26 million additional records are served from 354 "snapshot" data providers *(see Appendix I)*. Vouchered records can be linked to images, tissues, DNA sequence data, as well as other taxonomic and ecological information. Portals are organized by end user communities and are typically structured by a taxonomic and geographic theme. Portals encompass vascular plants, lichens, bryophytes, algae, fungi, invertebrates, and to a lesser degree vertebrates. For some larger institutions portals are designed to include all the museum and herbaria collections at that single institution.

Symbiota has two fundamental and overlapping functions: **(1)** An online "database" (Figure 1**,** left panel) for data providers to enter and annotate biodiversity occurrence data and associated specimen data (e.g., genetic sequences, images, publications); and **(2)** A primary aggregator/publisher for any data provider, regardless of
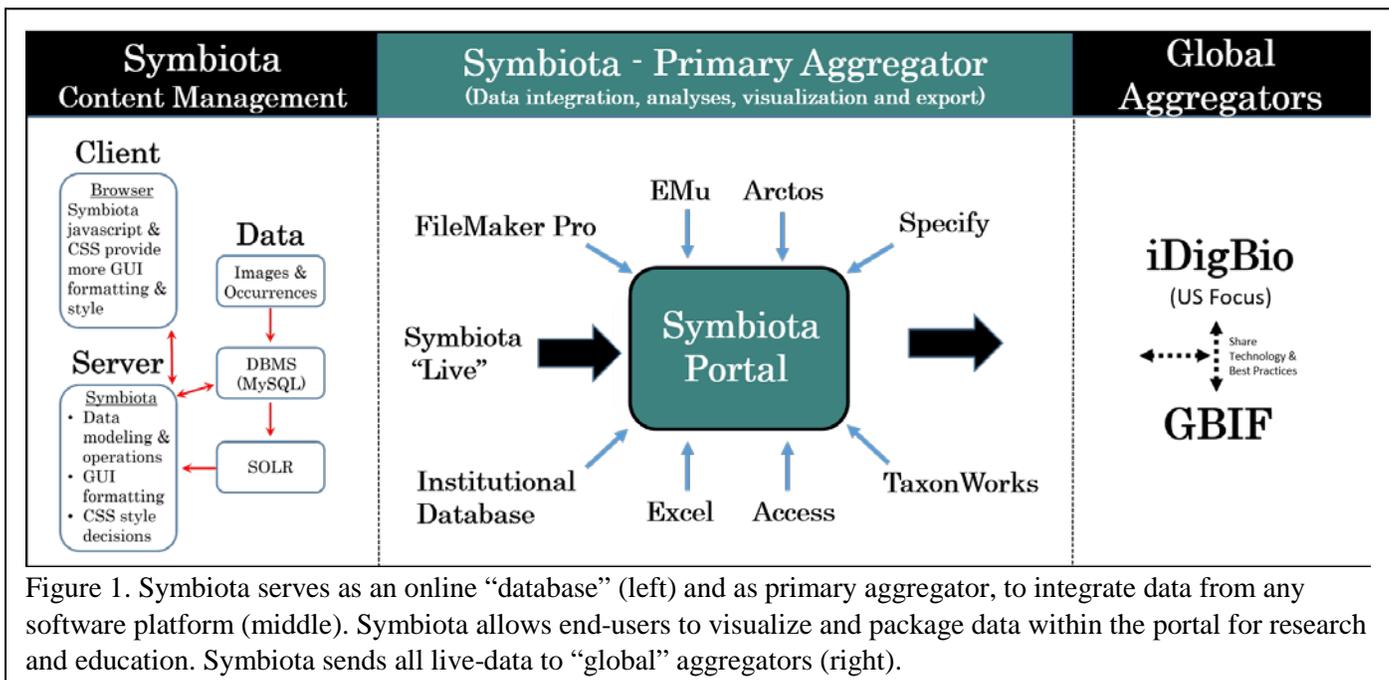

Figure 1. Symbiota serves as an online "database" (left) and as primary aggregator, to integrate data from any software platform (middle). Symbiota allows end-users to visualize and package data within the portal for research and education. Symbiota sends all live-data to "global" aggregators (right).

the software they use to enter and annotate data. Museums and herbaria that use Symbiota as an online

"database" to enter and edit data are referred to as "live collections", data providers that use another software platform as their primary database (e.g., Arctos) are referred to as "snapshot" collections. Both live and snapshot collections contribute to a portal by populating tables with information that is then used to help annotate incoming data and provide all users with more complete information on species. Finally, Symbiota automates the process of serving "live" data to global aggregators and the public through Darwin Core (DwC) archives (Figure 1, right panel).

Why Symbiota Works

Symbiota offers an efficient and cost-effective mechanism for documenting biodiversity occurrences from specimens and observations by providing tools to share, visualize and synthesize biodiversity data (**Table 1**). The basic functions are reliable and easy to learn and use, whereas more seasoned users can integrate the advanced functions making powerful data products leading to the generation of data-driven publications. Symbiota provides specific data packaging functions (e.g., checklist management), visualization tools (occurrence mapping and specimen images), analyses (computer-vision image processing through Fieldguide), education materials (identification keys and quizzes), and options to quickly configure and download datasets in .csv format (**Figure 2**). Thus, Symbiota portals allow researchers to compile and screen data sets for further analysis, or educators to integrate biodiversity data for use in the classroom.

Table 1. Examples of Symbiota Features

"Live" Data Provider Features	End User Features
Open source code for relational database (SQL)	Download entire biodiversity datasets in 2-3 clicks
Virtual portal set-up & backend support – no dedicated IT support required at each institution	High-resolution images and species profile pages
Shallow learning curve for basic functions	GIS mapping functions
Allows multiple users from remote locations	Rapid generation of data papers
Built-in OCR/NLP and georeferencing tools	Interactive, region-sensitive identification keys
Powerful data/products management tools	Create collaborative regional taxon checklists
Facilitates data mobilization to iDigBio, GBIF	"Name that species" & Flash Card Quiz games

Using Symbiota minimizes the impediments to sharing and managing collection records. Front-end management of collections in a portal is intuitive, as is managing individual collection for each institution. Each institutional collection in a portal is recognized separately and controls its own data management but data from all the collections are hosted on the same server and used in responding to queries. By offering both live data management and periodical data snapshot support, Symbiota portals allow diverse collections to share and integrate their holdings in spite of different preferred in-house management systems. By design, Symbiota is very de-centralized and each portal has unique characteristics while still capitalizing on the code updates that are available to everyone.

Symbiota Community Activities

In 2015, the **Symbiota Working Group (SWG)** was formed to coordinate support and increased functionality to existing portal communities and facilitate communication with data providers wishing to join existing portals or create new portals. Symbiota activities can be broadly divided into (**1**) new *software development;* (**2**) *support services* for creating Symbiota portal instances, adding collections to existing portals as well as maintenance support as needed; and (**3**) *end-user collaborations* to promote research and education use of information. A core group of developers provide the necessary code to make Symbiota work, while a large end-user community generates research and education products, provides the stimulus for new ideas and implementation of increased functionality, and develops webinars and online support documentation, including

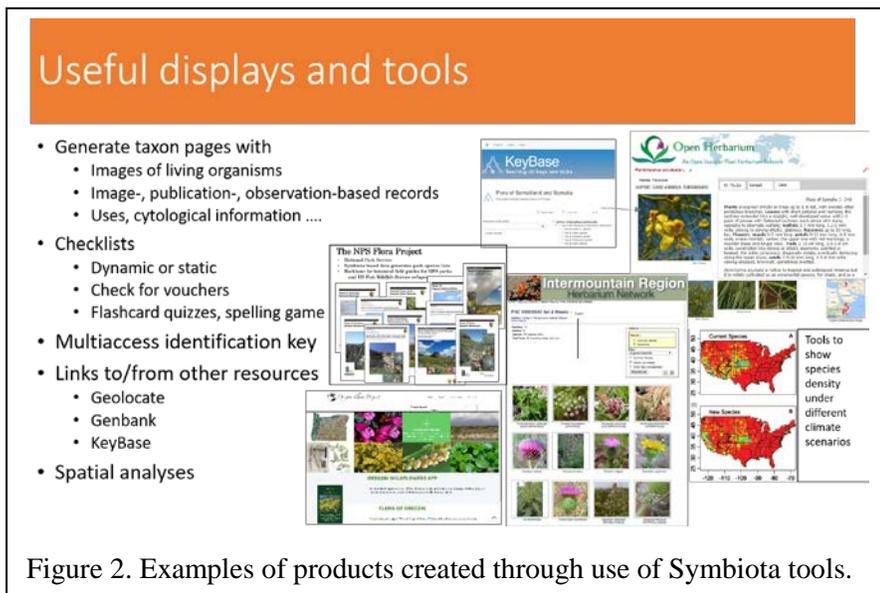data management plans. The SWG also promotes collaborations with outside developers interested in developing Symbiota code or APIs that allow their applications to interact with the Symbiota code and/or specific portals. This effort ensures the growth and sustainability of the diversifying Symbiota community. We will continue sharing both expertise and training across portals; identifying community priorities for new development, including added-value functions and modules; and leveraging new collaborations to seek funding for new software development and research-education use of data.

Figure 2. Examples of products created through use of Symbiota tools.

Improving Capacity

Symbiota was created in 1999 and has received several NSF grants to develop the basic framework. Symbiota is currently supported by digitization and innovation grants coupled with contract work to create additional functionality in existing portals and develop new portals. Thus, the core code continues to be actively developed with dedicated resources in place through 2021 to ensure Symbiota functionality is improved. This will extend support to new biology domains (e.g., paleobiology, sociobiology), as well as data fields specific to ecological monitoring programs. For existing users there is a high priority for both basic functions like increasing the speed of data entry and batch cleaning tools as well as developing higher-end functions like image analyses, phenology calculators and biotic association linking. Much of the expected increase in functionality will come from API development to take advantage of specific features like taxonomic servicing that are being developed outside of Symbiota.

Expanding Symbiota beyond North America

Symbiota portals have supported hundreds of North American biodiversity collections to share their holdings on-line. We are confident that Symbiota is a strong option for data providers outside North America, especially for institutions that do not want to support dedicated IT personnel for a biodiversity content management system. We can create new collections on an existing portal or help an organization establish their own portal. Our primary motivation for reaching out to institutions and collections is that using a common Symbiota portal is a great way to promote regional collaborations in research and education among collections within a portal, and ultimately global-scale collaboration among portals.

If you are interested in knowing more about Symbiota, please contact SWG lead coordinators Neil Cobb (Neil.Cobb@nau.edu) or Andrew Miller (amiller7@illinois.edu).

Key Symbiota Websites

Homepage: http://symbiota.org/   Code @ GitHub: https://github.com/Symbiota
Citable publication: http://bdj.pensoft.net/articles.php?id=1114
Google Group (support): http://symbiota.org/docs/google-group/
Symbiota Working Group: https://www.idigbio.org/wiki/index.php/Symbiota_Working_Group

APPENDIX I   List of Known Symbiota Portals (Updated September, 2017)

Portal Name	Portal Web Address (URL)
Aquatic Invasives	http://greatlakesinvasives.org/portal/index.php
Arctic Lichen Flora	http://lichenportal.org/arctic/index.php
Consortium of Midwest Herbaria	http://midwestherbaria.org/portal/index.php
Consortium of North American Bryophyte Herbaria	http://bryophyteportal.org/portal/
Consortium of North American Lichen Herbaria	http://lichenportal.org/portal/index.php
Consortium of Northeastern Herbaria (CNH)	http://portal.neherbaria.org/portal/
Consortium of Pacific Herbaria	http://www.pacificherbaria.org/
Consortium of Small Vertebrate Collections (CSVColl)	http://csvcoll.org/portal/
CoTRAM – Taxonomic Resource for US Myrtaceae	http://cotram.org/
Documenting Ethnobiology in Mexico and Central America	http://demca.sites.gettysburg.edu/
Frullania Collaborative Research Network	http://bryophyteportal.org/frullania/
Herbario Virtual Austral Americano	http://herbariovaa.org/
Intermountain Region Herbaria Network (IRHN)	http://intermountainbiota.org/portal/index.php
InvertEBase Data Portal	http://invertebase.org/portal/index.php
Lepidoptera of North America Network (LepNet)	http://symbiota4.acis.ufl.edu/scan/portal/index.php
Macroalgal Consortium Herbarium Portal	http://macroalgae.org/portal/index.php
Madrean Archipelago Biodiversity Assessment (MABA) - Fauna	http://madrean.org/symbfauna/projects/index.php
Madrean Archipelago Biodiversity Assessment (MABA) - Flora	http://madrean.org/symbflora/projects/index.php?proj=74
Mid-Atlantic Herbaria Consortium	http://midatlanticherbaria.org/portal/
Minnesota Biodiversity Atlas	http://bellatlas.umn.edu/
Monarch (California Academy of Sciences)	http://monarch.calacademy.org/
MyCoPortal	http://mycoportal.org/portal/index.php
Neotropical Entomology	http://symbiota.org/neotrop/entomology/index.php
Neotropical Flora	http://symbiota.org/neotrop/plantae/index.php
North American Network of Small Herbaria	http://nansh.org/portal/index.php
Northern Great Plains Herbaria	http://ngpherbaria.org/portal/index.php
OpenZooMuseum	http://openherbarium.org/ozm/
Oregon Flora	http://symbiota.oregonflora.org/portal/index.php
Red de Herbarios del Noroeste de México	http://herbanwmex.net/portal/index.php
SEINet	http://swbiodiversity.org/seinet/index.php
Smithsonian Tropical Research Institute Portal (STRI)	http://stricollections.org/portal/
SE Regional Network of Expertise & Collections (SERNEC)	http://sernecportal.org/portal/
Symbiota Collections of Arthropods Network (SCAN)	http://symbiota4.acis.ufl.edu/scan/portal/index.php
The Lundell Plant Diversity Portal	http://prc-symbiota.tacc.utexas.edu/index.php
The Open Herbarium	http://openherbarium.org/index.php
University of Colorado Herbarium	https://botanydb.colorado.edu/
Virtual Flora of Wisconsin	http://symbiota.botany.wisc.edu