iDigBio Collection Management System (CMS) Information Gathering

Thank you for taking the time to consider and respond to the following questions. iDigBio will make your responses available to the natural history collections community, both as an online resource available to anyone, and as a reference specifically for participants in our "Introduction to Biodiversity Specimen Digitization" course. This resource will serve as an update to a similar survey we did in 2012: https://www.idigbio.org/content/biological-collections-databases.

Please return your completed survey to Erica Krimmel (ekrimmel@fsu.edu).

BASIC QUESTIONS

1. **Name and email of person responding to this survey**: Katie Pearson (katelin.d.pearson24@gmail.com), Ed Gilbert, Nico Franz
2. **Name of Collection Management System (CMS)**: Symbiota
3. **Website**: https://symbiota.org; open Slack space: symbiota-support.slack.com; https://github.com/BioKIC/Symbiota-light
4. **Company or group responsible for maintaining CMS**: Biodiversity Knowledge Integration Center (BioKIC)-Arizona State University (https://biokic.asu.edu)
5. **Long-term funding structure for maintaining CMS (e.g., grants, membership, private)**: Symbiota is an open source software product being developed by various projects. At ASU's BioKIC, Symbiota portal services (including new development) are being supported in part by funding from iDigBio 3, the 30-year National Ecological Observatory Network / NEON Biorepository, select Thematic Collections Networks projects, and other sources.
6. **Brief summary highlighting the market niche for this CMS**: Symbiota is an adaptable, customizable software that enables self-defined and self-governed *communities of practice* to form, collaborate, and produce globally integrated, high-quality biodiversity data resources. Symbiota is ideal for small- to medium-sized collections that lack IT support and resources. Symbiota has also been used to create location-based data portals (e.g., http://www.cal-ibis.org/) and institutional management systems across collections (e.g. https://bellatlas.umn.edu). Data in web-based Symbiota portals can be instantly searchable via public search features, and there are many tools to connect these data to other collections and to aggregators (e.g., iDigBio, GBIF).

USABILITY QUESTIONS

7. **Restrictions on types of collection objects and/or disciplines (e.g., cannot handle anthropology)**: Symbiota has generally been used for biodiversity data, and the current fields align with Darwin Core standards; however, other disciplines could be supported with further development.

8. **Capacity for handling complex information related to taxonomic names (e.g. taxon concept mapping, recording annotations)**: Symbiota portals each include an internal Taxonomic Thesaurus that is used for quality control and resolving taxonomic discrepancies. The thesaurus is functional and capable of managing synonymy (i.e., taxa can be assigned as synonyms to other taxa); however, complex mapping of taxonomic concepts is not currently supported.

9. **Capacity for handling complex information related to geographic places and for facilitating tasks such as georeferencing**: Symbiota portals each contain a Geographic Thesaurus that can be used for quality control of place names and coordinates. Many tools have been developed for management and creation of geographical data, including a GEOLocate plugin that can be used to individually georeference specimens as they are transcribed or in batch via the Batch Georeferencing Tool, a Cross-Collection Batch Georeferencing Tool, and a Google Maps plugin that can be used to view and edit specimen georeferences individually or via the Batch Georeferencing Tool.

10. **Capacity for handling complex information related to people (e.g. collectors, identifiers, loan agents)**: The ability to manage and links Agents across the system is in development which will include collectors, identifiers, etc.

11. **Capacity for handling complex information related to extended data facets such as traits of (e.g. morphometrics) and interactions between (e.g. parent-child) collection objects**: An occurrence trait module has recently been developed and allows scoring of individual occurrences for specimen traits. These traits can be defined by the portal administrator or manager. Currently, this module is being used to capture phenological traits in a standard data format, but it has the potential to model additional occurrence traits (e.g., measurements). A module that supports Associated Occurrences is also being developed, which allows the user to link specimens to other specimens in the database, external occurrences, or observational taxon references. These linkages can be created as one of many types of relationships (e.g., part of, subsample of)

12. **Capacity for facilitating linkages between collection objects and extended data stored elsewhere, such as a genetic data repository**: The Genetic Resources module can be used to link occurrences to genetic resources (e.g. BOLD, GenBank). The new Associated Occurrences module (described above) allows for linking to associated occurrence residing within the portal or external system. Symbiota portals have well-established Duplicate Clustering tools that can be used to identify and link specimen duplicates, including exsiccatae duplicates. These tools can be used to import and share data, even among collections, during data entry and georeferencing to facilitate efficient digitization.

13. **Capacity for facilitating collection management transactions, such as loans, accessions, and transfers**: Symbiota portals include a Loan Management tool that allows tracking and documentation of loans, gifts, and exchanges. These tools can be

used to track the loan status of individual specimens, and they can be used to generate loan paperwork as well.

14. **Capacity for facilitating physical collections care including tracking storage locations and condition reporting**: These functions are not very extensively supported by Symbiota.

15. **Capacity to manage media (e.g., 2D images, 3D images, audio, video), and/or to work in sync with a dedicated Digital Asset Management System**: Symbiota portals have the basic ability to link out to media stored in external management systems, and many tools have been developed to batch harvest image data from servers and external services (e.g., CyVerse).

16. **Capacity for mobilizing collection object data (e.g., publish directly to an IPT, or export custom text files)**: Symbiota portals contain well-developed import and export modules that support ingestion of Darwin Core compliant data (including through an IPT and export as DwC-A, CSV, or other formats. Portals also include a Darwin Core Archive publishing module that functions similar to an internal IPT, and portals can publish data directly to certain aggregators (e.g., GBIF) through the administration control panel.

17. **Capacity for mobilizing collection object media (e.g., serve publicly online via a stable URI)**: Features in development through the NEON Biorepository project.

18. **Ability for users to customize the CMS**: Symbiota is an open-source software project that can be significantly customized via modifications to configuration files and/or the backend code. Substantial re-configuring requires knowledge of PHP, JavaScript, and HTML.

IMPLEMENTATION QUESTIONS

19. **Computer infrastructure (hardware, software) required**: Maria DB or MySQL, Web server (e.g. Apache), PHP.

20. **In-house IT expertise required**: Significant IT expertise is needed to establish and maintain a portal. Once the portal is established, the clients and data managers only need an up-to-date browser to manage datasets. Implementation of the software is best done through a collaboration where one entity establishes the portal that provides services for a specific collection community.

21. **Estimated costs for initial set up**: Symbiota software is free and open source with no initial cost. IT support and personnel are needed to establish and maintain the portal. Once the portal is established, it can be provided as a service to a specific set of clientele.

22. **Estimated costs for ongoing expenses such as membership or upgrades**: Not currently implemented or planned for the great majority of academic and public contributors and users.

23. **Migration or other new user services offered:** We invite potential new users to consider joining an existing consortium to receive maximum support. Contact a representative for suggestions.

24. **Example institutions/collections using your CMS**: SEINet (seinet.org), Consortium of California Herbaria (cch2.org), VertNet (vertnet.org), Symbiota Collections of Arthropods Network (SCAN; scan-bugs.org), Consortium of North American Bryophytes (bryophyteportal.org), Consortium of North American Lichens (lichenportal.org), MycoPortal (mycoportal.org), PteridoPortal (pteridoportal.org), SouthEast Regional Network of Expertise and Collections (SERNEC; sernecportal.org), and many others (

25. **Representative for potential users to contact**: Katie Pearson (katelin.d.pearson24@gmail.com)

26. **Best resources to point potential users to (e.g., presentations, brochures, recorded webinars)**: https://symbiota.org/docs/, https://www.capturingcaliforniasflowers.org/symbiota.html, https://www.youtube.com/channel/UC7gIMVLRnTA6ES3VTsci7iQ, https://www.youtube.com/channel/UCFjYwX5cjS_NaUkx3F-4lpg