iDigBio Collection Management System (CMS) Information Gathering
Please return your completed survey to Erica Krimmel (ekrimmel@fsu.edu).

BASIC QUESTIONS

1. **Name and email of person responding to this survey**:

   James Beach, beach@specifysoftware.org, for information on Specify's capabilities, contact: Theresa Miller (support@specifysoftware.org), for Specify Consortium membership inquiries contact: Norine Spears (membership@specifysoftware.org).

2. **Name of Collection Management System (CMS)**: Specify 6, Specify 7.

3. **Website**: www.specifysoftware.org

4. **Company or group responsible for maintaining CMS**:

   The Specify Collections Consortium (SCC) is an international, non-profit organization of biological collections institutions and research repositories. A Board of Directors and advisory committees, whose members are largely from SCC institutions, provide governance and strategic oversight. Outside Board members include representatives from GBIF, DiSSCo, and iDigBio. Staff at the University of Kansas build, maintain, and support the Consortium's software and services.

5. **Long-term funding structure for maintaining CMS (e.g., grants, membership, private)**:

   The Consortium's primary source of revenue is institutional membership fees. That sustains the Consortium's core activities. The SCC obtains additional funding from grants and contracts.

6. **Brief summary highlighting the market niche for this CMS**:

   Specify Software handles the curatorial processing and publishing of data associated with specimens in biological collections, including museums, herbaria, and biorepositories of tissue and DNA samples. Specify manages data for collections management tasks associated with museum transactions and mobilizes collections data through the Specify Web Portal, and through IPT to caches such as GBIF and iDigBio. Specify 7 also integrates through APIs with aggregators and online services for retrieving records of information related to specimen catalog entries.

7. **Restrictions on types of collection objects and/or disciplines (e.g., cannot handle anthropology)**:

   Specify processes and mobilizes data from all biological collections disciplines including strong support for paleontology. Some members have successfully adapted Specify to manage collections of geological samples. Specify does not currently support anthropological, archaeological, or living collections.

8. **Capacity for handling complex information related to taxonomic names (e.g. taxon concept mapping, recording annotations)**:

   Specify's approach to taxonomic names is to provide a dynamic, editable, taxon tree of ranks and names from which determinations can be created and linked to specimen records. Additional names can be applied to specimens from nomenclatural or taxonomic synonyms, hybrids, or shelving names. Names from alternative classifications can be included (queried, edited, printed, etc.) as non-current determinations for any specimen record. Although multiple taxon names can be applied to specimens, Specify is not a classification management system; its design is centered on the application of taxon names as determinations of specimens.

9. **Capacity for handling complex information related to geographic places and for facilitating tasks such as georeferencing**:

   Specify comprehensively manages geographic information about collecting localities. Political units are organized hierarchically in a tree-structured visualization with standard and custom ranks, ranging from continents and oceans, to countries, states and provinces, and to any lower or intermediate geographical entities required. The Geography Tree allows for ISO codes, synonymies, alternate names, moving, merging and editing nodes. Specify comes loaded with authoritative geographic names from GeoNames. For georeferencing, Specify contains an embedded GeoLocate plugin and Specify 7 displays maps of collection localities with several optional mapping layers. Locality information is highly structured for performing precise queries. Specify 6 exports georeferenced localities in KML for visualization in Google Earth. A new API-based integration in Specify 7 with the Lifemapper Project shows predicted distribution models for the species in your collection.

10. **Capacity for handling complex information related to people (e.g. collectors, identifiers, loan agents)**:

Specify software is thoughtfully designed to process information on people and to record and provide attribution for their contributions as individuals or in groups. It comprehensively tracks all roles and functions that people play in a collection, including: registrars, collectors, determiners, loaners, borrowers, preparers, "gifters", data authors and editors, and others.

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**:

    Specify has a few different ways to store and manage trait data. For Collection Objects and Preparations our data model supports the customization of numerous 'stub' fields available for naming the trait entities you would like to score and record. Specify also supports a more extensible and flexible approach to storing and managing trait data associated with specimens and preparations using "key/value pairs". Using this method in Specify one can create an unlimited number of trait entities and record unlimited values for each trait observation.

    For hierarchical biotic interactions, in Specify 6 we support parent-child links between two Collection Objects as well as links between specimens in different institutional collections in the same database in order to associate data records for example relationships like tissues:vouchers; plants:pollinators; hosts:parasites, and others.

12. **Capacity for facilitating linkages between collection objects and extended data stored elsewhere, such as a genetic data repository**:

    The SCC is advancing the development of a digital specimen architecture, a global network of linked data and analytical resources for collections, integrated through application program interfaces or APIs. This vision for the biological collections data community is championed in the U.S. by BCON with its Extended Specimen concept and in Europe by the DiSSCo initiative and GBIF. The Specify Project is engaged in the development of technologies for this global data network, and Specify platforms will be early adopters of the integration technologies that underlie this transformative vision for global biological collections computing. Today, Specify software has numerous fields for http linkouts to remote data resources related to particular specimens, preparations, tissues and DNA extractions. Http links embedded in Specify data bring up the web pages at remote resources and show the data related to objects in your collection.

13. **Capacity for facilitating collection management transactions, such as loans, accessions, and transfers**:

Specify excels at managing the details associated with collection management transactions including loans, gifts, borrows, exchanges, accessions, repository agreements, deaccessions, and disposals. Collection management transactions have been a well supported and regularly extended, core capability of Specify for over 20 years.

14. **Capacity for facilitating physical collections care including tracking storage locations and condition reporting**:

Specify handles collection storage locations with a hierarchical tree display so the location of collection objects and their individual parts and preparations can be linked to any custom ranked location (e.g. buildings, rooms, galleries, cabinets, drawers, etc.) in the tree. Browsing the Storage Tree will show the location of specimens for each named area. The software allows for two storage locations for a collection object, one for its current location and a second for a previous or permanent location. Notes and remarks complete storage location history as needed. Specify keeps detailed, item-level preparation counts, and statuses, for tracking the involvement of specimens or sub preparations in transactions.

Specify has purpose-designed data tables for treatment events (on specimens), specimen conservation events, and for conservator descriptions. These tables and their fields enable documentation of the history of conditions to which the object has been exposed, or how it should be treated, for storage or display.

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**:

The SCC offers a media asset server that stores, indexes, and links images and other types of media files with related specimen information in a Specify database. Specify's major data types and forms enable file attachment links to images, videos, sounds, spreadsheets, PDFs, and other types of media. Specify uses helper applications for complex media formats--popup web helper apps for viewing and manipulating media attachments. The Specify CMS has an embedded image viewer and uses thumbnails of uploaded images to enable quick browsing through media files within the application. And Specify 6 supports mass uploading of images and automatically generating thumbnails, which can then be situated anywhere on data forms for clicking through to original size versions.

The SCC does not distribute or support a Digital Asset Manager (DAM) but the Specify 7 platform architecture uses API-based integration for all communication between its web interface and "back-end" database server processes. That makes Specify 7 potentially integrable with API-based DAMs for media management. As institutions and campuses

require DAM integration for museum media, we will work with members to pursue this type of integration. The Specify Web Portal, a separate software application, displays a customizable set of search data fields for queries that then present collection object text records along with their occurrence maps and associated media attachments. Finally, the Specify CMS and Web Portal have the capability to link out to remote resources related to collection objects through URLs in Specify web link fields.

16. **Capacity for mobilizing collection object data (e.g., publish directly to an IPT, or export custom text files)**:

    Specify 6 exports collection object data files for incorporation into a DwC Archive for publishing through IPT. It has a built-in graphical user interface for custom mapping of a collection's data concepts to elements in the DwC schema. Specify also supports exporting the results of queries to external csv files for mobilizing data to other databases, and includes a built-in connector to Symbiota for transferring collection object data to the platform. Specify 7 has integrated most IPT server functions and is capable of mapping collections data to the DwC schema, packaging it into a DwC Archive, and advertising it to aggregators for assimilation. Specify 7 supports any DarwinCore extension through the addition of additional data field mappings to a TDWG standard extension with an XML configuration file within the software.

17. **Capacity for mobilizing collection object media (e.g., serve publicly online via a stable URI)**:

    The Specify Web Portal serves all or a selected set of collection object records based on export filters that sites choose. The Web Portal integrates with the Specify media asset server for display of images associated with collection objects. Static URLs are assigned to media objects in the Web Portal, so it is possible to reference and access them through an http link from a remote application. Images from the Web Portal can be copied and saved by remote access users for subsequent local use. In Specify 7, which is completely based on APIs for communications, it will be possible in the future to access media in Specify 7, through an API call. In addition, for access to media in the reverse direction from a remote resource for inclusion within Specify, the CMS supports weblink fields for multiple major data types for URLs that link and retrieve remote media files.

18. **Ability for users to customize the CMS**:

    Specify is fully customizable to reflect the data types and workflows of importance to a collection. Data forms have outstanding flexibility for data field placement and layout. Data fields can be represented in various ways on Specify data forms, including as picklists, grids, and as text, real, integer, logical, or date field data entry boxes. Specify's

user interface has outstanding flexibility to adapt to requirements of a discipline, a collection, a particular project, or even to a certain preparation type. Fields can be re-labeled and re-purposed to meet a collection's specific requirements and Specify's interface can be localized into any Unicode supported language.

IMPLEMENTATION QUESTIONS

19. **Computer infrastructure (hardware, software) required**:

To operate Specify 6, a minimal hardware setup would include a recent Windows or MacOS office computer, desktop or laptop. For shared, multi-user access to a collection database with Specify 6, a local (campus), server computer (real or virtual) running MySQL or MariaDB would be required. For Specify 7 a Linux server is needed (real or virtual). With Specify Cloud, no hardware or software is needed beyond a desktop computer with a web browser. Disk storage requirements largely depend on the number and size of attached media files.

20. **In-house IT expertise required**:

For Specify 6, our (Java) platform that runs identically on Windows, MacOS and Linux platforms with a back-end MySQL or MariaDB data manager, typical tasks requiring IT expertise would be the ability to administrate desktop computers, install new software, configure software parameters and adjust network settings. For Specify 7, our web-based platform, installing a Specify 7 server locally will require a Linux systems administrator with proficiency in installing and configuring software on Linux server machines. For a "Zero IT" approach, the Consortium offers "Specify Cloud", a Specify 7 hosted, cloud-native service that requires no local IT expertise and no software beyond a desktop computer or Chromebook, and a network connection with a web browser.

21. **Estimated costs for initial set up**:

Specify is open source licensed and is free to download and use. The Consortium's Help Desk, technical support, training, and prioritized software development activities are funded through the membership fees and are available only to member institutions. Four levels of membership allow collection institutions to choose the desired level of technical support and level of engagement in the Consortium's governance activities. The most economical membership level costs $1,000 USD/collection/year. Collections interested in running their own building or campus Specify server have access to technical support for customizing Specify's user interface, printed specimen labels and reports. The cost for these one-time setup services are priced *à la carte* and typically sum to an additional

$1,000 based on the level of setup and support desired. The cost of converting legacy data sets to Specify can be modest or reach a few thousand dollars depending on the complexity and data quality of the legacy system. Members can also choose to perform data migration themselves.

For institutions with under-resourced IT or which prefer to avoid managing local computing hardware, the Consortium offers "Specify Cloud" which is Specify 7 web platform offered as "Software-as-a-Service". Specify Cloud hosting is available on secure servers around the globe, it includes automatic software updates, nightly backups, and systems management. Operating your collection database in the cloud with only a web browser costs about $250 to $500 per year, in addition to the SCC annual membership fees (see below).

22. **Estimated costs for ongoing expenses such as membership or upgrades**:

Specify Collections Consortium membership is renewed yearly, and it entitles members access to technical support services to participation in Consortium governance oversight. Institutions become members at one of four levels:

Associate ($1,000) and Solutions ($1,250 - $3,000) memberships are charged per collection, and offer comprehensive support and collaboration opportunities. Benefits include incidental help desk assistance and access to all events and additional support services such as Specify Cloud Hosting, legacy conversions, label and report creation.

Full Members ($5,000 per year) are leading collections that wish to proactively engage in shaping the priorities and technology directions of the Consortium with priority access to seats on the Board and Advisory Committees. These members also enjoy additional support benefits such as access to our technical staff, supplemental support such as label and report creation, and comprehensive assistance during data conversion projects.

Founding Partners pay $40,000/year for an unlimited number of collections. These are typically large institutions with robust biodiversity research collection activities. They commit to the open source software paradigm, showing deep collections community engagement. Their substantial support of the Consortium's core activities helps us support under-resourced collections with the same software platforms and capabilities. Founding Partners take an active, leadership role in Consortium governance and strategy. They are entitled to two permanent seats on the Board of Members and seats on both the Science and Technology Advisory Committees. Their status ensures them immediate access to Consortium technical staff, and a suite of technical support services benefiting their institutional IT and collections researchers.

23. **Migration or other new user services offered:**

The Consortium offers legacy data set conversions to Specify. It also offers setup, configuration and customization services, including the design of custom data entry forms, and formats for printed labels and reports. The SCC can also provide custom live tutorials and recorded training webinars for collections new to Specify.

24. **Example institutions/collections using your CMS**:

Specify databases are used in every biological collection discipline at large national museums, university collections, at small colleges and other specialized collections. For a list of SCC member institutions, see: https://www.specifysoftware.org/members/

25. **Representative for potential users to contact**:

For information on capabilities and use, contact: Theresa Miller at support@specifysoftware.org, for SCC membership info, including benefits and costs, contact: Norine Spears at membership@specifysoftware.org.

26. **Best resources to point potential users to (e.g., presentations, brochures, recorded webinars)**:

Go to www.specifysoftware.org or e-mail membership@specifysoftware for more specific information or to start a conversation. We would be happy to refer you to colleagues and Specify users at SCC member institutions.