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**: Deborah Paul (dlpaul@illinois.edu), Matthew Yoder (mjyoder@illinois.edu)
2. **Name of Collection Management System (CMS)**: TaxonWorks
3. **Website**: https://taxonworks.org/
4. **Company or group responsible for maintaining CMS**: The Species File Group (SFG) https://speciesfilegroup.org/
5. **Long-term funding structure for maintaining CMS (e.g., grants, membership, private)**: See https://speciesfilegroup.org/about.html. The SFG is funded by an endowment to the University of Illinois Champaign Urbana, this includes 9 FTE positions and significant discretionary funds. The endowment provides a relatively rare environment within the field of biodiversity informatics, i.e. *fixed funding* for research and development of biodiversity informatics tools, methods, and communities.
6. **Brief summary highlighting the market niche for this CMS**: TaxonWorks (TW) is a web-based serving taxonomists, biodiversity scientists, natural history collections and more. It allows you to capture, organize, and enrich your taxonomic, nomenclature, collections, and bibliographic source data; share it with collaborators; and package it for analysis and publication. The software has been in production for over 5 years and is very actively updated by a core team of 5 with numerous other community and internal contributions. Anyone in the world can and is encouraged to contribute to the open source code base, https://github.com/SpeciesFileGroup/taxonworks. Talk with us and the TW community on gitter (https://gitter.im/SpeciesFileGroup/taxonworks).

USABILITY QUESTIONS

7. **Restrictions on types of collection objects and/or disciplines (e.g., cannot handle anthropology)**: TW currently best addresses data needs for taxonomists and natural science collections (i.e. curation of collection objects of biological origin). It would not, at the moment, be suited for anthropology or cultural collections.
8. **Capacity for handling complex information related to taxonomic names (e.g. taxon concept mapping, recording annotations)**: TW offers perhaps the most sophisticated suite of tools for managing nomenclature of any CMS. It properly abstracts nomenclature from the classification of biological taxa (e.g. species) via "OTUs" (operational taxonomic units), i.e. taxon concepts. A rich citation model allows for all data to be properly attributed.

9. **Capacity for handling complex information related to geographic places and for facilitating tasks such as georeferencing**: TaxonWorks includes both Gazetteer based and in-application Georeferencing. It has hooks to services like Geolocate, it allows users to draw most geometric shapes and use them as georeferences, and it seeks to provide a truly spatial/computational internal model for querying, filtering, and asserting locality metadata.
10. **Capacity for handling complex information related to people (e.g. collectors, identifiers, loan agents)**: TW offers unparalleled features for storing people data, along with all identifiers for a given person that you might want (local identifiers, globally unique identifiers). In addition, people data are shared across instances of TW so that these data don't need to be managed / curated repetitively / duplicated across databases. Specific tasks facilitate the de-duplication of People records.
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**: TaxonWorks has a full suite of "matrix" management features allowing curators to add qualitative, quantitative, statistical, binary, media, and free-text classes of observations to either specimens or OTUs. It also allows users to create custom data-attributes (e.g. fields, or columns) for most classes of data.
12. **Capacity for facilitating linkages between collection objects and extended data stored elsewhere, such as a genetic data repository**: The TW "data attribute" mechanism allows for cross-references to any data that might be represented via a URI or other form. An API (https://api.taxonworks.org/#other-endpoints) facilitates computation, scripts, etc. that might link data curated in TaxonWorks to data accessible in other knowledge-bases via scripts or other applications.
13. **Capacity for facilitating collection management transactions, such as loans, accessions, and transfers**: TW includes a loan-management model that allows for the loans of collection objects, lots, or OTUs based records. This includes basic functionality like producing simple loan forms, and updating in batch returned specimens with new determinations.
14. **Capacity for facilitating physical collections care including tracking storage locations and condition reporting**: TaxonWorks includes an extensible "Container" classification that models locations, and data have been assigned to this model, however a fully fledged user-interface that takes advantage of this model is not yet ready.
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**: TaxonWorks has sophisticated mechanisms for both storing and relating images to CollectionObjects. It has multiple tasks for breaking down images and referencing parts of them to both metadata and the collection objects themselves. It includes multiple mechanisms for drag-dropping images, alleviating the need for costly curation steps like renaming images. 3D images are not supported, though interest is out there.
16. **Capacity for mobilizing collection object data (e.g., publish directly to an IPT, or export custom text files)**: A wealth of exposure exists via the powerful filtering capabilities of the TaxonWorks API (see above), this includes returning a row of data as a DwC object. The DwC export is being greatly extended, beta tests on datasets of 40k+ specimens for around 50+ DwC attributes have been delivered to curators for inspection. Plans are to publish straight to GBIF in the future.
17. **Capacity for mobilizing collection object media (e.g., serve publicly online via a stable URI)**: The TaxonWorks API has independent endpoints for identifiers and collection objects, facilitating true digital-specimen approaches. These are already available, though "batch" curation features such as applying classes of identifiers need to be built to increase efficiency in this regard.
18. **Ability for users to customize the CMS**: In addition to the customizations possible via the current user-interface for a whole suite of annotations (e.g. data attributes, notes, confidence/quality levels, topics, and tags), anyone can contribute to the open-source codebase.

IMPLEMENTATION QUESTIONS

19. **Computer infrastructure (hardware, software) required**: Full details on installation are at https://github.com/SpeciesFileGroup/install_taxonwork. The development team is dedicated to supporting those who wish to install and manage their own copies of TW. TaxonWorks is accessed through a web-browser, Firefox and Chrome are supported. One instance of TW supports multiple projects. There is no one instance of TW, anyone can run an instance. The Species File Group manages several instances of TW for its collaborators and is open to new collaborators joining, see https://github.com/SpeciesFileGroup/services/wiki/TaxonWorks-in-production-at-the-Species-File-Group . Personal copies can technically be run off-line. Containerized (Docker) versions are available. Development and production environments have been described for Mac OS, Linux, and Windows, though the latter (Windows version) is not as tested as the others. Physical requirements are minimal, 8GB for runtime, databases size starts at 250mb, images are stored externally and corresponding space is required.
20. **In-house IT expertise required**: For curators none, because TW is cloud-based and accessed through commonly used browsers. For developers you need basic server management skills, advanced skills can take advantage of Kubernetes and other deployment tools. The TW community is encouraging and supporting the vision of "hybrid" users, those who both use and contribute to development, for example users with Javascript, Ruby (or other scripting languages), technical writing skills, etc. can make technical contributions to the project that benefit not only their projects but the whole TW community.
21. **Estimated costs for initial set up**: For production managers, a virtual machine with 8GB of memory for runtime (as low as $5-10/month), or a linux machine with minimum 8GB of memory, both minimal costs. For curators none, except time to learn the software.
22. **Estimated costs for ongoing expenses such as membership or upgrades**: None right now. For SFG instances see https://github.com/SpeciesFileGroup/services/wiki/TaxonWorks-in-production-at-the-Species-File-Group . Other providers may charge for support as they see fit. The long-term vision is to have an open software development community online that manages this product collaboratively. (e.g. see how Blender works, a free and open source 3D creation suite https://www.blender.org/), and have organizations determine whether to charge for their instances of TW.
23. **Migration or other new user services offered:** TaxonWorks has multiple in-application mechanisms for batch-loading data, including a top-notch DwC importer. Scripted approaches, commonly used by developers for bespoke data are also used. Otherwise, users start from empty projects that reference data shared in that instance of TW, for example, people, sources, geographic gazetteers, repositories are shared across individual instances of TW. The SFG offers all users, new or existing,weekly open office hours for help with nomenclature in TW or digitization in TW, see https://speciesfilegroup.org/events.html
24. **Example institutions/collections using your CMS**: Illinois Natural History Collection, Penn State's Frost Museum, North Carolina State Insect Collection, New Hampshire Insect Collection, AntWeb.
25. **Representative for potential users to contact**: Deborah Paul, Matt Yoder, Tommy McElrath, Dmitry Dmitriev.
26. **Best resources to point potential users to (e.g., presentations, brochures, recorded webinars)**: The SFG has a full time Community Liaison supporting TaxonWorks, contact her (Deborah Paul, dlpaul@illinois.edu). Chat with the community at https://gitter.im/SpeciesFileGroup/taxonworks. See also. Help documentation is being aggregated at https://github.com/SpeciesFileGroup/taxonworks_doc, these, videos, and other resources will be federated at https://docs.taxonworks.org.