

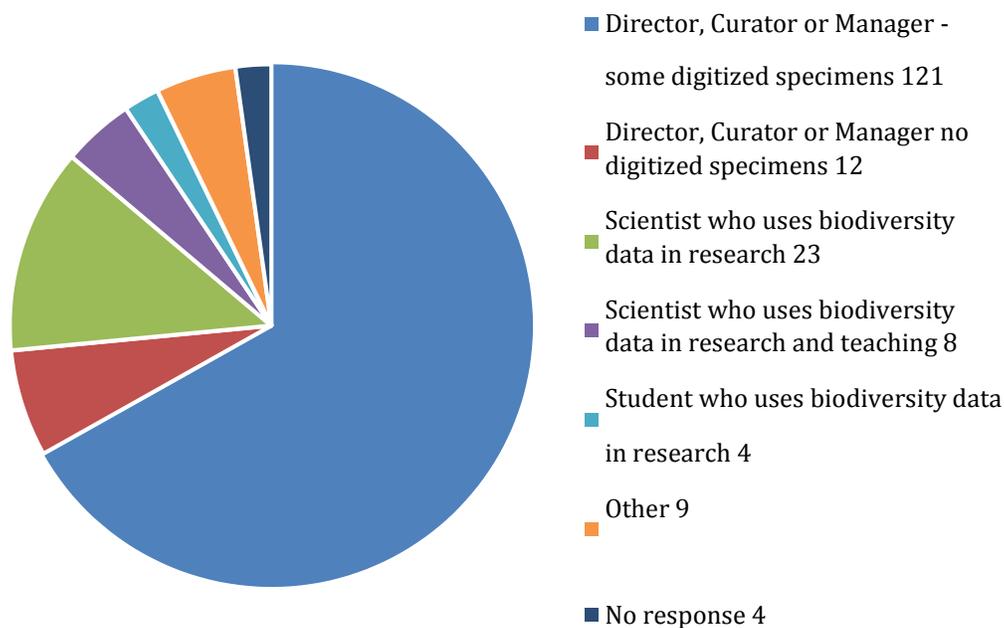
Appendix C: Stakeholder Outreach Summaries

1. SURVEY RESULTS: Harnessing Biodiversity Collections Data for Addressing National Challenges

This survey, initiated on 7 February 2018, was intended to elicit information for a stakeholder vision of how to maximize the value of biodiversity collections data for collections management, research and education in the future. BCoN is coordinating the development of this vision in response to requests from NSF and others to inform the development of a new funding program to succeed NSF's current ADBC program. The effort will culminate in a workshop in October 2018 where we will consolidate the feedback into a strategic plan for 2020-2030. The strategic plan, to be issued in early 2019, will be available for public comment and modified in response to comments made before it is finalized.

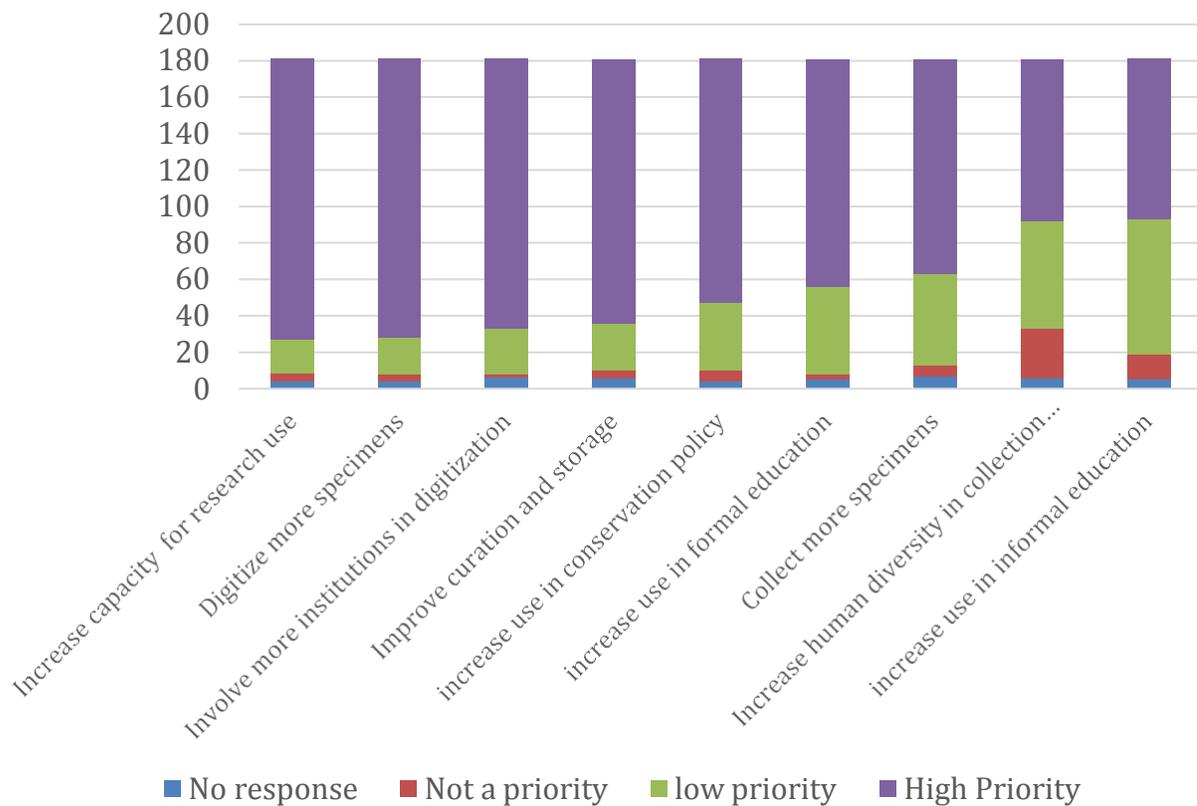
Participants in the Survey

There were 181 respondents to the survey, who self-identified in the following categories:



Priority Future Goals

Respondents were asked to rank a set of future goals for biodiversity collections according to whether they considered the goal was high priority, low priority or not a priority.



Results indicate that increasing the capacity for research, digitizing more specimens and involving more institutions in the digitization process were all viewed approximately equally as high priority, and increasing human diversity in collections and increasing the use of collections in informal education were ranked lowest priority, in approximately equal numbers.

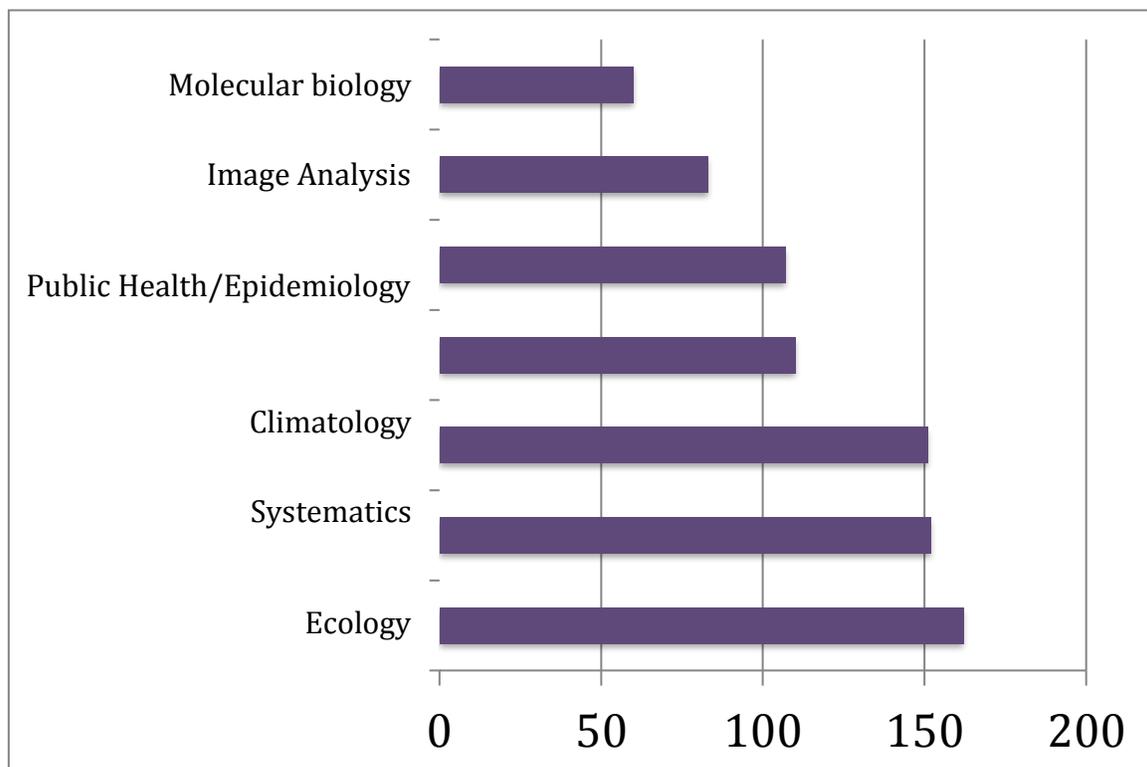
Other Goals

Participants were given the opportunity to suggest other goals that should be a high priority for biodiversity collections. The responses fell into five over-arching categories:

- Train more systematists, collections professionals
- Identify other (e.g., non- NSF sources of funding for collections digitization)
- Improve data quality, especially specimen identifications
- Develop new digitization technologies
- Broaden the user community for digitized specimen data
- Focus future digitization on: historical collections; international collections; small collections; acoustic collections

Which Research Areas Would Benefit from Greater Integration and Collaboration with Biodiversity Collection Data?

Participants indicated which of the Research Areas listed above would benefit from greater integration with digitized data. Digitized specimen data was considered most relevant to Systematics, Ecology and Biogeography.



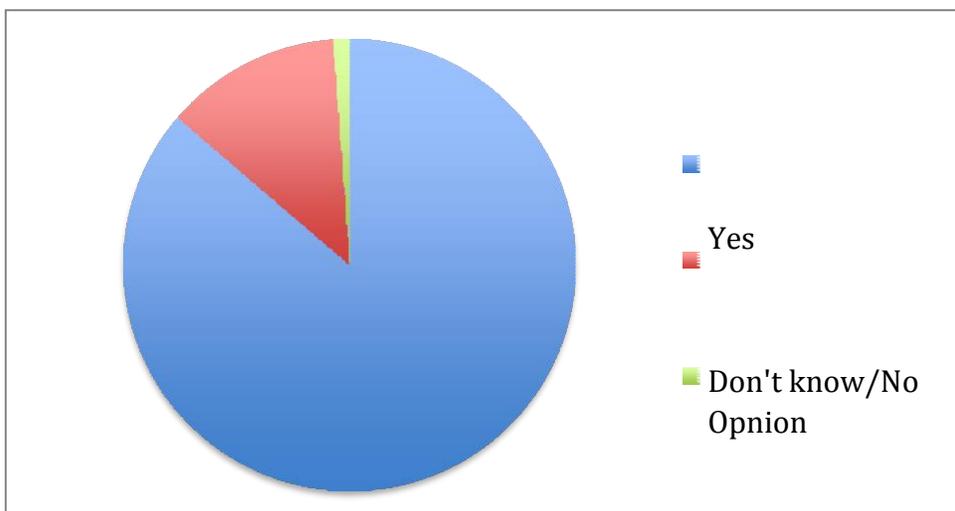
Other Fields that could Benefit from Greater Integration and Collaboration with Digitized Collections?

Participants were asked to list other fields that could make use of digitized specimen data. These are summarized in the table below:

Agriculture	Evolutionary Biology
Anatomical and Health Sciences	Fisheries Science
Architecture	Geology
Art	History
Conservation Biology	Human Evolution
Ecology	Invasive species control
Economics	Paleoecology
Education	Paleontology
Ethology	Stratigraphy

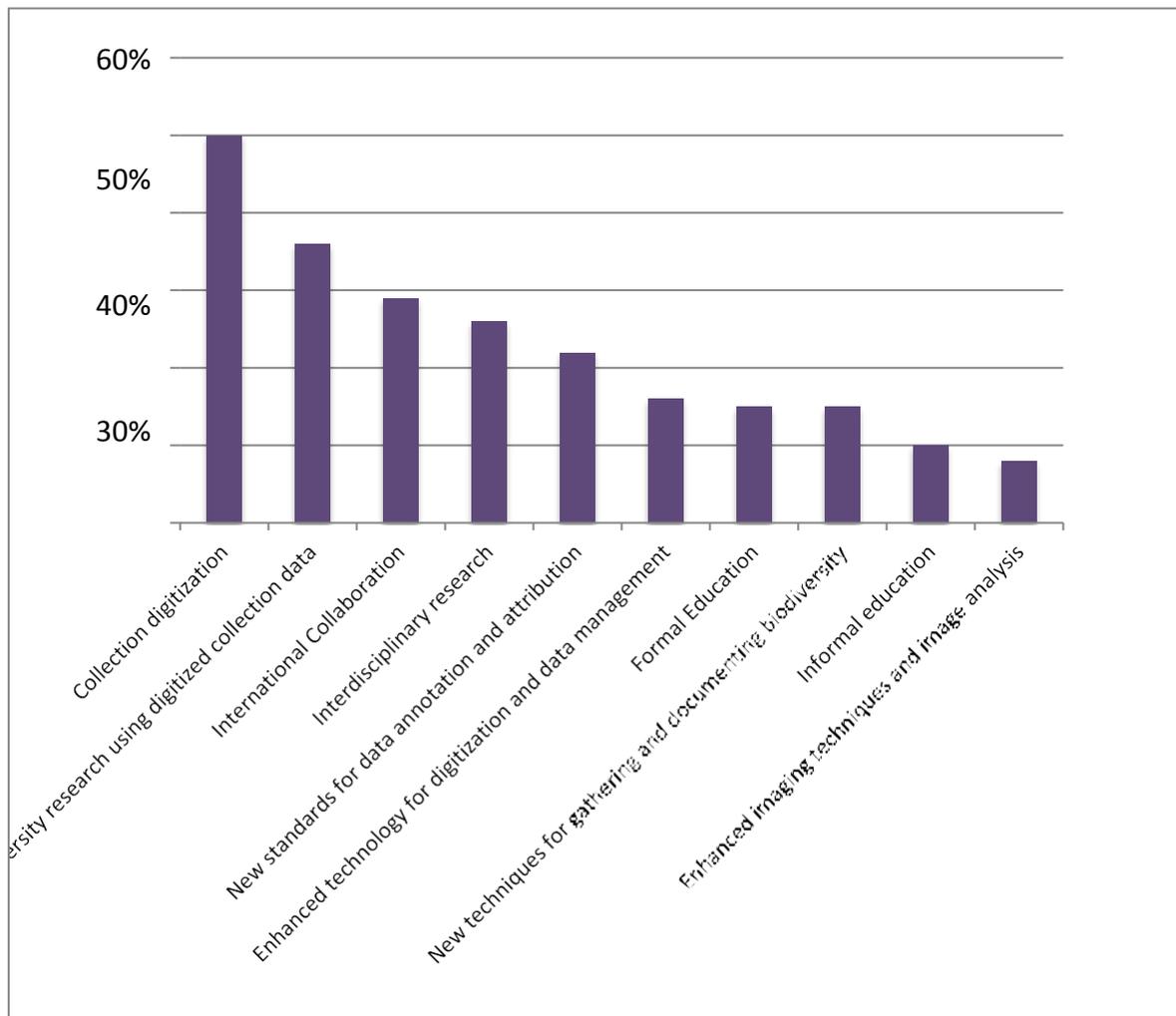
A Future Funding Program

Participants were asked, “Do you think another NSF funding program is necessary for the digitized biodiversity data to reach its full potential for research, education and conservation objectives?” The response was strongly in favor of a new NSF funding program.



Highest Priority Focus for a New Funding Program

Respondents were asked to rank a set of activities that could be part of a new funding program. The chart below focuses on only those items that were ranked as highest priority.



Half of all respondents ranked continued collection digitization as the highest priority activity, followed by biodiversity research using digitized collection data and international collaboration.

2. Harnessing Natural History Collections Data for Addressing National Challenges: The Biodiversity Collections and Data Wish List

This spreadsheet compiles the responses to the survey named above. The survey opened on 24 September and closed on 5 October, 2018. Responses are grouped by category for each of reading the document -- in many cases, responses could belong to multiple categories. Numbers assigned are for convenience of reference only -- they do not indicate priority or order of receipt.

Question: *What would you like to accomplish with biodiversity collections or derived data that you cannot do now (or cannot do easily)?, List ONE currently unavailable action or tool that you would need to in order to do this; Who could help you achieve the accomplishment listed in Question 1?*

Organization of Categories:

Collections Management, responses 1 – 10
 Data Improvement, responses 11 – 21
 Digitization, responses, 22-31
 Education, responses 32-37
 Policy, responses 38-42
 Research, responses 42-75

Collections Management

1. Know what we DON'T have well-represented in our collections. Needed: a tool that overlays range maps with observational data with specimen data (we do have iterations of this but they all need work). Partners: taxonomist, information scientist, data scientist
2. Reduce the backlog of unprocessed specimens. Needed: Staff to process material. Partners: 1. Institutions providing funds. 2. Museum administrators using money as directed rather for other projects.
3. Increased focus on conservation and curation of the actual, physical specimens themselves. Needed: Expert in the taxonomy/nomenclature of a group to put modern names on specimens (not re-identification; just updating names). Also, money to fund repackaging and remounting of herbarium specimens with new archival materials. Partners: 1) taxonomist (2) collections manager (3) library/book/museum conservator. -- Scott LaGreca, Duke University,
4. Better Tracking of accession and annotation history of specimens. Needed: need fields for legacy identifiers, and exchange history. I think annotations are well covered in Symbiota, but data from other programs may be hard to import. Partners: IT to add fields in a way that does not complicate more routine data entry and imports.
5. I would like to know more about collectors. Needed: Centralized collector database with name versions, dates & places of collection. Partners: biographers, historians
6. Find out more about the people behind the collections -the collectors, taxonomists, geologists, etc.) and their lives. Needed: Link collectors to biographical information, where they worked, gender, accomplishments. Partners: 1) Historians of science, geography 2) Librarians, gender study researchers 3) scientists focused on history of discovery –
7. Accurately summarize, keep track of and properly attribute all specimens collected by a given person. Needed: A unique identifier for every distinct collector - i.e., Orcid ID for collectors. Partners: Database specialist.

8. Summarize the scientific value of an individual collection to help administrators, donors and others understand the inherent value of the collection. Needed: The number of unique specimens nationwide and the distribution of collections geographically and taxonomically across the country. Partners: 1) data miners, graphics experts 2) science communicators 3) experts on special taxonomic groups that know the history of an area or taxon well, and/or natural history of science

9. Be able to determine derivatives and/or related specimens at our own and other collections, especially paleo-related coal ball species where some have been traded with other institution, Needed: Ways to link across museum/institution collections. Partners: 1) subject experts 2) collections staff 3) database managers

10. Connect specimens in published literature to museum voucher specimens. Needed: Requirement of specimen or repository UUIDs in publications. Partners: 1) Publishers to implement 2) researchers to adopt a standard of UUIDs in publications 3) biodiversity informaticians to manage data and APIs.

Data Improvement

11. Download ready to use data, without obvious errors, field mismatches, etc. Downloads can require extensive cleaning which makes it difficult to use datasets with undergraduates. Needed: Better mechanism to get data providers to improve their data uploads, e.g. including the minus in west longitudes, mapping data to the correct Darwin core field. Partners: Not sure- possibly a software designer. -- Jean Woods, Delaware Museum of Natural History.

12. Standardization of taxonomy of digitized specimens with flexible concept-mapping. Needed: More sophisticated tools for "digital annotation". Partners: Need software design carefully informed by systematics.

13. Scan a specimen barcode and correct specimen data and/or identification. Needed: specimen codes that are shared across different institutional databases; software that allows modifications. Partners: Collection administrators, software designers.

14. Help to visualize biocollections data quality needs by visualizing the distinct terms for each value where DwC suggests a controlled vocab (some 23 terms). Needed: Need a visualization tool connected to the world's biocollections data. See <https://github.com/tdwg/dwc-qa/tree/master/data>. Partners: 1) software developers 2) researchers representing different taxonomic groups 3) BIS TDWG and SPNHC.

15. Reach agreement across collections, aggregators, publications on how to cite and attribute resources so credit goes where it needs to, where it must. Needed: Policy-level meetings with IT staff included, across projects/programs to work toward an agreed format, agreed requirements and expectations. Partners: 1) Aggregators 2) Collections 3) Publishers.

16. An annotation system for digital specimen data records that maintains original data as well as all possible corrections and interpretations of the data in the record. Needed: Annotation tools for web portals. Partners: 1) collections institutions to accommodate annotation layers in their databases 2) web portal developers to implement annotation layers.

17. Easily identify super high quality datasets that have already been vetted. Needed: Easy dataset publication tool or some other workflow that incentives rich datasets. Partners: Funding focus on high quality datasets as an end result, taxonomic expertise included in digitization process.

18. Communicate between people via aggregators, e.g. if I could flag a subset of my dataset for "needs to be reviewed by a taxonomic expert," or if a taxonomic expert could "follow" a species or region (similar to what is possible on iNaturalist). Needed: the technical communication infrastructure; Partners: taxonomists, programmers, collection managers.

19. Increase taxonomic determination of specimens and have a taxon concept resolution service. Needed: An application that manages taxon concepts. Partners: Software designers and biologists as taxonomists.

20. I would like to have more prominent messaging for iDigBio users that citation of the individual institutions from which data are obtained is expected. Needed: could be as easy as an auto email for anybody who downloads data that has the (automatically generated) list of museums, a nice note that citation of both idigbio and the individual institutions is expected, and clear directions on expected format for citations. Partners: Directors of big museums who could indicate which kind of metrics would be most useful- in text citations? something more like genbank- with tables of individual ID#s in an index, etc .

21. Ability to easily fix errors of names of organisms in aggregators. Needed: either need editorial privileges for a large number of independent online resources, OR a message board where errors in those resources can be posted and flagged so as to notify the managers of those resources that they have problems - and the postings would remain until the problems were fixed. Partners: This would need a networked organization of those who CAN modify the content of individual data sources.

Digitization

22. Would be great to finally see all taxonomic names (old, new, temporary, phylocode, etc.) in one place and the ability to pick a name from that list to associate with a given specimen in a collection database. Needed: The worldwide community needs CoL+ with GNUB, GNA, etc. functionality added - as starting point to create this resource. Partners: 1) funders (governments - since we all need it) 2) researchers to contribute 3) collections software that links to this resource

23. An easy way to crowdsource georeferencing. Needed: A module or tool in established crowdsourcing platforms (Digivol, Notes From Nature) for crowdsourced georeferencing or integration of GeoLocate with either of these. Partners: 1) Experts on Geolocate 2) Experts on Digivol or Notes From Nature 3) Science educators with expertise in crowdsourcing natural history data.

24. Automatically digitize an entire drawer of pinned insects AND their labels. Needed: Conveyor or robot driven imaging station for insect drawers. Partners: 1) Photogrammetry experts 2) Robotics experts 3) Industrial company to put it together.

25. Have it all digitized! ... so that researchers can use it to solve science questions that can make informed decisions about the environment. Needed: understanding by researchers that digitization can help them not take away money from curation/research activities. Partners: Curators who digitise? examples showing that digitization helps further curation & access to the collection. Someone who could redesign the way to upload images in particular.

26. I would like to upload all my data and images to a national database. Needed: An easier way to upload; current method is extremely time-consuming and onerous.

27. Seamless import collections data from Symbiota portals to in-house database platforms (Specify, KE, etc.).
Needed: An export interface between Symbiota and Specify/KE, etc. Partners: Software designer.

28. Way to archive specimen derived data, and associated publications, with digitized specimen metadata record.
Needed: Infrastructure/interface and encouragement for users of specimen data to contribute to specimen metadata. Partners: Database manager and data portal designer

29. Associate specimen-derived data (e.g., leaf traits) with the original specimen records in a way that is searchable. Needed: Researched need the ability to append new metadata to specimen records, including literature citations. Partners: 1) data scientists 2) functional trait database people (e.g., TRY, BIEN).

30. Allow or inform users who automatically go to aggregator sites that unique data are also available at "feeder" sites. Needed: Some kind of "linker" as libraries have (e.g., this book also available at ...) Partners: 1) Programmers 2) IT specialists.

31. Easy search across databases -- example -- I can find different specimens using the Consortium of California Herbaria and GBIF and see many duplicates in GBIF. Needed: Better synching of databases and better synching of taxonomic trees. Partners: 1)taxonomists 2)bioinformaticians.

Education

32. Use the data as a tool for teaching college level classes on plant evolution and diversity. Specifically, displaying a subset of the data (local plants) displayed on a phylogenetic tree. Needed: a tool that automatically "hangs" specimens on branches of the plant phylogeny and displays them in a way that demonstrates plant diversity. Partners: software designers

33. Use the resources to generate knowledge (research), training and science communication. Needed: An curator interface that manage a platform to access to the resources. Partners: 1) Collection manager designed to such interface; 2) technician to digitalize the resources; 3) bioinformatician.

34. Use aggregated collections images in education and outreach. Needed: Image tagging -- we need to be able to parse out the images that are helpful in E&O from the millions of labels. Partners: IT, Collections people, Educators.

35. I would like to be able to interact with people with biodiversity questions as peers through the code that would answer their question instead of having problems vaguely described with the expectation that the "computer people" will just figure it out. Needed: Every researcher having broad understanding of computing and software use. Everyone needs to internalize that their computer is their #1 collaborator on all their work.. Partners: K-12 and undergraduate curricula expecting that computer use, specifically the ability to have a computer implement your idea, is a modern life skill that pervades all disciplines

36. Use the resources in courses to help train educators, users, workforce, and next generation of researchers. Needed: An educator interface designed for entry level users (with associated educational materials). Partners: Implementation scientist to determine needs of educators 2) software designer 3)science educators.

37. Outreach to teachers, especially K-6, to inspire younger generations to pursue biodiversity preservation and conservation. Additionally, make a case for the use of biodiversity databases to non- biology majors, like math, statistics, English, etc. Needed: A conversation with K-12 community that does not create more work for teachers but make them want to collaborate and create lesson plans and activities. Partners: 1) Education standards 2)local teachers 3)people in the biodiversity community with experience in formal education.

Policy

38. More unrestricted person*hours for small institutions that are chronically understaffed. Needed: Grants focused on increasing staff at struggling institutions that are not tied to specific research projects. Partners: 1) Grant agencies 2) professional societies 3) people good at talking to and convincing administrators.

39. Enforcing the voucher system (depositing original and tracking with a unique identifier) established in the biodiversity collections community also in commercial/corporate research & development practices. Needed: Legal and policy reforms that mandate for-profit researchers to disclose the origin of biodiversity research material they used and deposit samples and data with public collection institutions. Partners: Legal and policy reforms that mandate for-profit researchers to disclose the origin of biodiversity research material they used and deposit samples and data with public collection institutions.

40. Engage the Dept. of Homeland Security for issues of port of entry and biosecurity. Needed: I don't know how collections data may address issues that are important to homeland security. Partners: Experts in biosecurity, taxonomists who can develop identification tools.

41. Achieve sustainable financial support for our collections, databases and at least the digital library for our private, non-profit museum. Needed: A data monetization plan, tied to a list of potential funders for our region and the nation and the world. Partners: 1)Marketing experts 2)Financial planners 3)Administrators who promote collections support.

42. Engage with communities beyond biology to address grant challengers. For example, USDA, NASA, EPA may offer joint funding opportunities to leverage TCN data. Needed: Information needed to address the issues that are important to these agencies. Partners: none given.

Research

43. Easily identify specific specimens with associated geocoded locality data, genetic data, and media files for integrative biological research. Needed: Ability to associate derivative data (measurements, photos, etc.) to specimens without having to go back through the data publisher. Partners: to associate derivative data (measurements, photos, etc.) to specimens without having to go back through the data publisher.

44. Construction of species distribution models from digitized collections data. Needed: A simple tool or interface to allow an educated user to construct SDMs from collections data, whether from an aggregator (iDigBio, GBIF) or from personally acquired data. Partners: 1) Scientist who is an expert in species distribution modeling 2) software developer.

45. Generate local checklists (town or county or state) with the most up-to-date taxonomy. Needed: Synonymy list. Partners: Taxonomists and programmers.

46. Access to comprehensive regional KEYS to families with photos to support the descriptions would be fabulous!. Needed: Mentor sites online to answer taxonomic questions would be great. Partners: Regional master taxonomists who are willing to share knowledge with those not as experienced; trainers in technology that is used to populate the comprehensive databases.

47. Have DNA barcodes of all type collections. Needed: Trained staff for tissue collection and sequencing. (and funding). Partners: Molecular biologists.

48. Compare unknown DNA sequence data to comprehensive regional flora/fauna reference library for species determination with high probability, Needed: Build comprehensive regional DNA reference library. Partners: 1. Taxonomist to identify exemplar specimens to build reference library. 2. Technical officer to generate DNA sequences to high standard. 3. Data system manager to utilize VOUCHERED DNA sequences for comparison in analyses.

49. Use biodiversity collections to estimate the historic and current range of aquatic organisms. Needed: Cleaned and vetted location data for specimens. Partners: Taxonomists to check IDs and update species names. GIS professionals to identify and correct location data or interpret text locations where no coordinates are present

50. Find all records for a species in one place. Needed: consolidation of data portals- each portal has a different set of contributors and the data downloads are all formatted differently/have different data fields, create a meta search capability that combines. Partners: Software designers.

51. Share CT data. Needed: A shopping cart on our database where users can download ct data and complete associated paperwork. Partners: ???.

52. Provide simple summaries as graphic visuals, tables, etc., of the data for a set of specimens, and this should include some possibilities for quality control (so not just a map or table of GBIF records or similar). Needed: Linking data lists to options of graphic outputs. Partners: A science communicator, data vision expert, and designer collaborating with the national data portal iDigBio or similar.

53. Implement phylogeny changes to records already in the database (force changes in taxonomy back onto records already logged). Needed: Linking data lists to options of graphic outputs. Partners: 1) taxonomist to identify differences between database and current accepted phylogenies, 2) database software engineer, 3) someone familiar with database for implementation.

54. Search a single, global data portal that combines the resources of GBIF, iDigBio, ALA, DiSSCO, and other aggregators. Needed: A common API and interface.. Partners: Combined efforts of aggregators.
55. I would like to be able to easily see how collection effort and collection frequency is changing over time in order to see which species are declining. Needed: Visual representation using colour coding the for the points of occurrence (e.g., blue for historical (>50 years ago) versus red for recent). Partners: Software developers.
56. Know how "complete" a given dataset is (are all records of a given taxon digitized, a subset, none, etc.?). Needed: Digitization status for each contributing collection. Partners: database manager include as standard information.
57. Create affordable (perhaps free) DNA barcoding services available to unfunded researchers doing taxonomic investigations. Needed: Funding, or free access for submitting samples. Partners: Any agency or university could sponsor unfunded researchers doing taxonomic investigations.
58. I want to access DNA barcoding data from undetermined specimens. Needed: DNA barcoding program focusing on indets. Partners: (1) the program which unites efforts from multiple herbaria in relation to their indets; (2) outsource agencies to perform DNA barcoding; (2) outsource agencies which perform actual DNA barcoding (and probably store DNA).
59. Create distribution data to track future range shifts and extinctions. Needed: Consistent databasing completeness across institutions. Partners: 1) Funds for undergraduates to database 2) expert scientists to correctly identify specimens.
60. Examine plant distributions visually across a region to investigate the movement of non -native species and the potential decline of rare species. Needed: Mapping of species distributions by county across the entire country. Partners: 1) website database designers 2) scientific users to test the functionality 3) place to engage the public to help in these efforts.
61. Build meta-datasets of phenotypic characters of specimens making up collections. Needed: broadly accessible phenotype database for herbarium collections. Partners: 1) web/database developer, 2) hands on the ground (student researchers), 3) long term hosting.
62. I want automated flower color analysis from images compared to colors mentioned on labels to build DB of spp, gen, fams that exhibit color change upon drying and also maybe compare over time and techniques/conditions. Needed: Automated color extractor/picker tool from within Symbiota image viewer so I don't have to pull all images out into another program. Partners: Symbiota developer, herbarium curator/collections manager.
63. Set up a worldwide museum staff network of trained Carpentries instructors - so that collections can address their own biodiversity informatics data skills and literacy needs in a systematic and sustainable way. Needed: A high-level meeting to agree on a way forward. Partners: 1) The Carpentries 2) current worldwide staff mobilizing collections data 3) researcher feedback on skills they need to use biocollections data.

64. To know on any given day what new biodiversity collections had been added to world's collections and how these compare with existing ones. Needed: A world network of collections updated daily. Partners: Collections community to identify and link collections 2) social media applications customized for the purpose

65. Have specimen images orientated in standard ways so that morphometric data can be easily retrieved from them. Needed: Scale bars are standard, but a set of orientation and lighting instructions might be formalized. Partners: Agreed standard illustration format for particular taxa formalized. Some such standards exist - such as lighting strongest from upper left.

66. Trace the evolution of a clade from deep time to modern, including recent past from lake cores and archeological sites. Needed: The data in the gap between paleo collections, Partners: Engage with the scientists who have biological data in archeological collections.

67. I would like to be able to harvest phenological data from a wide variety of taxa from herbarium specimens. Needed: Phenological standards (e.g., DarwinCore field and semantic foundation) Partners: developer of biodiversity standards.

68. I would like to be able to run analysis on data in place: upload and evaluate a NN model, train a model, apply ML algorithm, etc. Needed: Computing resources and an interface located in the same place as the data (either move the data somewhere or add computation to the data). Partners: NSF XSEDE infrastructure.

69. More effectively examine the intersection of geology with modern and past occurrence data to test to what extent distributions are driven by geology. Needed: Effective workflows (for dummies!) using R. Partners: 1. Some programmer effort 2. Data Carpentry type workshops to teach it.

70. Correlated georeference data for a given taxon with ecological data, mapping in a layered, GIS-type format and including soil type, rock type, topography, vegetation type, average precipitation by season, and bioregions (by different systems). Needed: Maps of different described above, in layers to overlay searches for one or more taxa. Partners: Software designers, IT experts. Maps are very important for answering scientific questions and I think for engaging the community.

71. Map distributions and track traits of cenozoic fossil vertebrates from North America. Needed: We cannot get NSF funds for cenozoic vertebrates because so many of the fossils are still owned by BLM, NPS, USFWS, Forest Service, etc. Partners: 1) iDigBio Pls 2) Congressional representatives and senators from stakeholder states: OR, CA, WY, NE, NV, WA, CO, UT, MT, KS, TX, FL etc. 3) Domain scientists from the cenozoic vertebrate paleo community -- folks who would be in such a TCN.

72. Have a character database to pair with species distribution databases. Needed: Character databases. Partners: Widespread use of Symbiota features built for lichens of New Mexico.

73. A geographical hierarchical georeference quality control scoring program/app that uses machine learning to assign specimen label data to a hierarchical geolocation database via location data coordinates. Needed: A GIS tool, a machine learning program that examines label data to extract and assign hierarchical values from label data then also a GIS layer that is created from many polygon and point files to create hierarchy. Partners:1) GIS professionals 2) other georeference/database herbarium specimens 3) historic place name specialists.

74. Extract trait data from images in an automated way. Needed: Machine learning/conv. neural networks and algorithms. Partners: 1) AI/ML experts 2)ecologists who could identify and extract traits and tie to their ecological function 3) other IT experts.

75. More advanced mapping? To identify botanical black holes, both location and time-ranges. Needed: Add layers to map? Partners: none listed.