Harnessing Natural History Collections Data for Ad

This spreadsheet compiles the responses to the survey named above.
Responses are grouped by category for each of reading the document
Numbers assigned are for convenience of reference only -- they do no

Category	Number	1. What would you like to accomplish with biodiversity collections or derived data that you cannot do now (or cannot do easily)?
Collections Management	1	know what we DON'T have well-represented in our collections
Collections Management	2	Reduce the backlog of unprocessed specimens.
Collections Management	3	As part of the next initiatives I would like to see increased focus on conservation and curation of the actual, physical specimens themselves. In other words, repacketing; updating their nomenclature; and improving accessibility within the physical space of the herbarium.
Collections Management	4	Better tracking of accession history such as annotations, was the specimen absorbed by another institution, was it an exchange and from where, and are there any legacy identification number used to refer to a particular specimen?
Collections Management	5	I would like to know more about collectors.
Collections Management	6	Accurately summarize, keep track of and properly attribute all specimens collected by a given person
Collections Management	7	Find out more about the people behind the collections 9the collectors, taxonomists, geologists, etc.) and their lives
Collections Management	8	Summarize the scientific value of an individual collection to help administrators, donors and others understand the inherent value of the collection
Collections Management	9	Be able to determine derivatives and/or rleated specimens at our own and other collections, especially paleo-related coal ball specits where some have been traded with other institutions

Collections Management	10	Connect specimens in published literature to museum voucher specimens.
Data Improvement	11	Download ready to use data, without obvious errors, field mis-matches, etc. Downloads can require extensive cleaning which makes it difficult to use datasets with undergraduates
Data Improvement	12	Standardization of taxonomy of digitized specimens with flexible concept-mapping
Data Improvement	13	scan a specimen barcode and correct specimen data and/or identification
Data Improvement	14	Help to visualize biocollections data quality needs by vizualizing the distinct terms for each value where DwC suggests a controlled vocab (some 23 terms) - this way data quality metrics are easy to track and the taxonomic group specialists could easily see how they can contribute to better data standardization for more robust research.
Data Improvement	15	Reach agreement across collections, aggregators, publications on how to cite and attribute resources so credit goes where it needs to, where it must.
Data Improvement	16	An annotation system for digital specimeen data records that maintains original data as well as all possible corrections and interpretations of the data in the record
Data Improvement	17	Easily identify super high quality datasets that have already been vetted
Data Improvement	18	I want a way to 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's possible on iNaturalist)

Data Improvement	19	Increase taxonomic determination of specimens and have a taxon concept resolution service
Data Improvement	20	I would like to have more prominant messaging for idigbio users that citation of the individual institutions from which data are obtained is expected. We cite genbank numbers and R packages- we should be citing individual museums so that these museums have a more tangible record to bring to their institutional administrators. It's great to see that idigbio citations are shooting up, but that doesn't help individual museums
Data Improvement	21	I am a taxonomic and nomenclatural authority, and I am presently unable to rapidly fix errors in the names of organisms in data aggregators.
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.
Digitization	23	An easy way to crowdsource georeferencing
Digitization	24	Automatically digitize an entire drawer of pinned insects AND their labels.
Digitization	25	have it all digitised! ... so that researchers can use it to solve science questions that can make informed decisions about the evironment
Digitization	26	I would like to upload all my data and images to a national data base.

Digitization	27	Seamless import collections data from Symbiota portals to in-house database platforms (Specify, KE, etc.)
Digitization	28	Way to archive specimen derived data, and associated publications, with digitized specimen metadata record
Digitization	29	Associate specimen-derived data (e.g., leaf traits) with the original specimen records in a way that is searchable
Digitization	30	Allow or inform users who automatically go to aggregator sites that unique data are also available at "feeder" sites
Digitization	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
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
Education	33	Use the resources to generate knowledge (research), training and science communication
Education	34	Use aggregated collections images in education and outreach
Education	35	Not with the data but with the people, 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.
Education	36	Use the resources in courses to help train educators, users, workforce, and next generation of researchers

Education	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.
Policy	38	More unrestricted person*hours for small institutions that are chronically understaffed
Policy	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.
Policy	40	Engage the Dept of Homeland Security for issues of port of entry and biosecurity.
Policy	41	Achieve sustainable financial support for our collections, databases and at leasat the digital library for our private, non-profit mseum
Policy	42	Engage with communities beyond biology to address grant challengers. For example, USDA, NASA, EPA may offer joint funding opportunities to leverage TCN data
Research	43	Easily identify specific specimens with associated geocoded locality data, genetic data, and media files for integrative biological research.
Research	44	Construction of species distribution models from digitized collections data
Research	45	Generate local checklists (town or county or state) with the most up-to-date taxonomy
Research	46	Access to comprehensive regional KEYS to families with photos to support the descriptions would be fabulous!
Research	47	Have DNA barcodes of all type collections.

Research	48	Compare unknown DNA sequence data to comprehensive regional flora/fauna reference library for species determination with high probability
Research	49	Use biodiversity collections to estimate the historic and current range of aquatic organisms.
Research	50	find all records for a species in one place
Research	51	Share CT data
Research	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).
Research	53	Implement phylogeny changes to records already in the database (force changes in taxonomy back onto records already logged)
Research	54	Search a single, global data portal that combines the resources of GBIF, iDigBio, ALA, DiSSCO, and other aggregators.
Research	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
Research	56	Know how "complete" a given dataset is (are all records of a given taxon digitized, a subset, none, etc.?)

Research	57	Create affordable (perhaps free) DNA barcoding services available to unfunded researchers doing taxonomic investigations.
Research	58	I want to access DNA barcoding data from indeternined specimens
Research	59	Create distribution data to track future range shifts and extinctions.
Research	60	Examine plant distributions visually across a region to investigate the movement of non native species anf the potential decline of rare species.
Research	61	Build meta-datasets of phenotypic characters of specimens making up collections.
Research	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.
Research	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.
Research	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
Research	65	Have specimen images orientated in standard ways so that morphometric data can be easily retrieved from them

Research	66	Trace the evolution of a clade from from deep time to modern, including recent past from lake cores and archeological sites
Research	67	I would like to be able to harvest phenological data from a wide variety of taxa from herbarium specimens.
Research	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.
Research	69	More effectively examine the intersection of geology with modern and past occurrence data to test to what extent distributions are driven by geology
Research	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).
Research	71	Map distributions and track traits of cenozoic fossil vertebrates from North America
Research	72	Have a character database to pair with species distribution databases
Research	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 coordiinates
Research	74	Extract trait data from images in an automated way
Research	75	More advanced mapping? To identify botanical black holes, both location and time-ranges

The survey opened on 24 September and closed on 5 October, 2018.
: -- in many cases, responses could belong to multiple categories.
ot indicate priority or order of receipt.

2. List ONE currently unavailable action or tool that you would need to in order to do this
a tool that overlays range maps with observational data with specimen data (we do have iterations of this but they all need work)

Staff to process material.

An expert in the taxonomy/nomenclature of a group to put modern names on specimens (not re-identification; just updating names). Also, money to fund repacketing and remounting of herbarium specimens with new archival materials.

We 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.

Centralized collector database with name versions, dates & places of collection

A unique identifier for every distinct collector - i.e., Orcid ID for collectors

Lind collectors to biographical information, where they worked, gender, accomplishments.

Need to know the number of unique specimens nationwide and the distribution of collections geographically and taxonomically across the country

Ways to link acros museum/institution collections

Requirement of specimen or repository UUIDs in publications

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

More spophisticated tools for "digital annotation"

specimen codes that are shared across different institutional databases; software that allows modifications

Need a visualization tool connected to the world's biocollections data. See https://github.com/tdwg/dwc-qa/tree/master/data

Policy-level meetings with IT staff included, across projects/programs to work toward an agreed format, agreed requirements and expectations..

Annotation tools for web portals

Easy dataset publication tool or some other workflow that incentives rich datasets

the technical communication infrastructure

An application that manages taxon concepts

this could be something 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 (similar to what you see in the citations sections, for example, of the Flora of North America project). I appreciate that the citation style is not an area of current consensus (e.g. between vertebrates, plants ....) but idigbio is uniquely situated to take a leading role in getting credit to data providers.

In order to fix errors, I would 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.
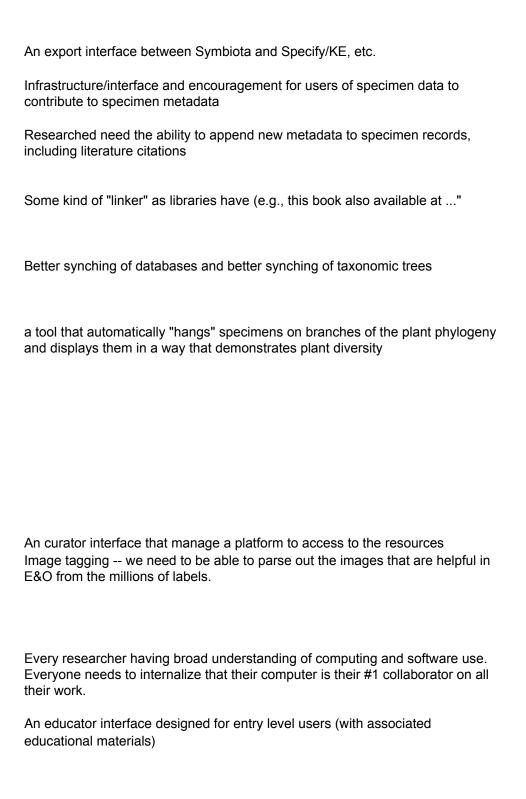
The worldwide community needs CoL+ with GNUB, GNA, etc functionality added - as starting point to create this resource.

A module or tool in established crowdsourcing platforms (Digivol, Notes From Nature) for crowdsourced georeferencing or integration of GeoLocate with either of these

Conveyor or robot driven imaging station for insect drawers.

understanding by researchers that digitisation can help them not take away money from curation/research activities

An easier way to upload; current method is extremely time-consuming and onerous.

An export interface between Symbiota and Specify/KE, etc.

Infrastructure/interface and encouragement for users of specimen data to contribute to specimen metadata

Researched need the ability to append new metadata to specimen records, including literature citations

Some kind of "linker" as libraries have (e.g., this book also available at ..."

Better synching of databases and better synching of taxonomic trees

a tool that automatically "hangs" specimens on branches of the plant phylogeny and displays them in a way that demonstrates plant diversity

An curator interface that manage a platform to access to the resources
Image tagging -- we need to be able to parse out the images that are helpful in E&O from the millions of labels.

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.

An educator interface designed for entry level users (with associated educational materials)

A conversation with K-12 community that does ot create more work for teachers but make them want to collalborate and create lesson plans and activities

Grants focused on increasing staff at struggling institutions that are not tied to specific research projects

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
I don't know how collections data may address issues that are important to homeland security

A data monetization plan, tied to a list of potential funders for our region and the nation and the world

Information needed to to address the issues that are important to these agencies

Enable researchers to associate derivative data (measurements, photos, etc.) to specimens without having to go back through the data publisher (i.e., by emailing the GenBank numbers of photos to the collection manager)
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

Synonymy list

Mentor sites online to answer taxonomic questions would be great.
Trained staff for tissue collection and sequencing. (and funding)

Build comprehensive regional DNA reference library

Cleaned and vetted location data for specimens.

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 datasets from multiple portals
A shopping cart on our database where users can download ct data and complete associated paperwork.

LInking data lists to options of graphic outputs.

Option or code in database to force new taxonomic tags over old ones

A common API and interface.

Perhaps even colour coding the points of occurrence (e.g., blue for historical (>50 years ago) versus red for recent)

Digitization status for each contributing collection

Funding, or free access for submitting samples.

DNA barcoding program focusing on indets

Consistent databasing completeness across institutions

Mapping of specied distributions by county across the entire country.

A broadly accessible phenotype database for herbarium collections.

Automated color extractor/picker tool from within Symbiota image viewer so I
don't have to pull all images out into another program

A high-level meeting to agree on a way forward.

A world network of collections updated daily

Scale bars are standard, but a set of orientation and lighting instructions might
be formalized

The data in the gap between paleo collections

Phenological standards (e.g., DarwinCore field and semantic foundation)

Computing resources and an interface located in the same place as the data (either move the data somewhere or add computation to the data)

Effective workflows (for dummies!) using R to

Maps of different described above, in layers to overlay searches for one or more taxa.

We cannot get NSF funds for cenozoic vertebrates because so many of the fossils are still owned by BLM, NPS, USFWS, Forest Service, etc.

Character database

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

Machine learning/conv. neural networks and algorithms

Add layers to map?

ions and Data Wish List

3. Who could help you achieve the accomplishment listed in Question 1?	4 OPTIONAL. Your name, institution, and email address
taxonomist, information scientist, data scientist	
1. Institutions providing funds. 2. Museum administrators using money as directed rather for other projects.	
1) taxonomist (2) collections manager (3) library/book/museum conservator	Scott LaGreca, Duke University, scott.lagreca @duke.edu
IT to add fields in a way that does not complicate more routine data entry and imports.	
biographers, historians	
database person	
1) Historians of science, geography 2) Librarians, gender study researchers 3) scientsits focused on history of discovery	Lena Struwe
1) data miners, graphics expertse 2)science communicaors 3) experts on special taxonomic groups that know the history of an area or taxon well, and/or natural history of science	Lena Struwe
1) subject experts 2) collections staff 3) database managers	Margaret Landis

1) Publishers to implement 2) researchers to adopt a standard of UUIDs in publications 3) biodiversity informaticians to manage data and APIs.	Kathy Hollis Smithsonian NMNH hollisk@si.edu
Not sure- possibly a software designer	Jean Woods, Delaware Museum of Natural History, jwoods@delmnh.org
Need software design carefully informed by systematics	ALan Weakley (NCU) weakley@unc.edu
Collection administrators, software designers	Jack Longino, University Utah, jacklongino@gmail.com
1) software developers 2) researchers representing different taxonomic groups 3) BIS TDWG and SPNHC	Deborah Paul, Florida State University, dpaul@fsu.edu
1) Aggregators 2) Collections 3) Publishers	Deborah Paul, Florida State University, dpaul@fsu.edu
1) collections institutions to accommodate annotation layers in their databases 2) web portal developers to implement annotation layers	
Funding focus on high quality datasets as a end result, taxonomic expertise included in digitization process.	Katja Seltmann (UCSB)
taxonomists, programmers, collection managers	

Software designers and biologists as taxonomists

Directors of big museums (Berkeley, UF, Smithsonian etc) 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	Ben Carter, San Jose State University, benjamin.carter@sjsu
This would need a networked organization of those who CAN modify the content of individual data sources.	Doug Yanega, UC Riverside, dyanega@ucr.edu
1) funders (governments - since we all need it) 2) researchers to contribute 3) collections software that links to this resource	Deborah Paul, Florida State University, dpaul@fsu.edu
1) Experts on Geolocate 2) Experts on Digivol or Notes From Nature 3) Science educators with expertise in crowdsourcing natural history data	Crystal Maier, Field Museum, cmaier@fieldmuseum.org
1) Photogrammetry experts 2) Robotics experts 3) Industrial company to put it together.	
curators who digitise? examples showing that digitisation helps further curation & acess to the collection	
Someone who could redesign the way to upload images in particular.	Katharine Gregg, George B. Rossbach Herbarium, West Virginia Wesleyan College, gregg@wvwc.edu

Software designer	Rich Rabeler, MICH, rabeler@umich.edu
Database manager and data portal designer	
1) data scientists 2) functional trait database people (e.g., TRY, BIEN)	Mason Heberling
1)Programmers 2)IT specialists	
1)taxonomists 2)bioinformaticians	Amanda Fisher
software designers	
1) Collection manager designed to such interface; 2) thecnician to digitalize the resources; 3) bioinformatician	Moisés Escalona, Pontifícia Universidade Católica do Rio Grande do Sul (PUCRS), moises.escalona@gmail.com
IT, Collections people, Educators	
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. Implementation scientist to determine needs of educators 2) software designer 3)science educators	

1)Education standards 2)local teachrs 3)people in the biodiversity community with experience in formal education	Kimberly Cook
1) Grant agencies 2) professional societies 3) people good at talking to and convincing administrators	Jess Miller-Camp, University of California, Riverside, jessmc@ucr.edu
1) lawmakers/policymakers; 2) international legal scholars; 3) biodiversity data scientists	aspeciosus@yahoo.com
Experts in biosecurity, taxonomists who can develop identification tools	griccardi@fsu.edu
1)Marketing experts 2)Financial planners 3)Administrators who promote collections support	Larry Stevens
none given	Jason Knouft
NCBI staff; Data Dryad staff; bioinformaticians.	David Blackburn, University of Florida, dblackburn@flmnh.ufl.edu
1) Scientist who is an expert in species distribution modeling 2)software developer	cmaier@fieldmuseum.org
Taxonomists and programers	
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	
Molecular biologists	

1. Taxonomist to identify exemplar specimens to build reference library. 2. Technical officer to generate DNA sequences to high standard. 3. datasystem manager to utilise VOUCHERED DNA sequences for comparison in analyses. 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.	Prof Michelle Waycott, State Herbarium of South Australia, michelle.waycott@sa.gov.au

Duncan Elkins, University of Georgia |
software designer	Jean Woods, Delaware Museum of Natural History, jwoods@delmnh.org
???	
an science communicator, data vision expert, and designer collaborating with the national dataportal iDigBio or similar	Lena Struwe, Rutgers University, lena.struwe@rutgers.edu
1) taxonomist to identify differences between database and current accepted phylogenies, 2) database software engineer, 3) someone familiar with database for implementation	
Combined efforts of aggregators.	Gil Nelson, iDigBio, gnelson@bio.fsu.edu
software designer	Jana Vamosi, UCalgary, jvamosi@ucalgary.ca
database manager include as standard information	

Any agency or university could sponsor unfunded researchers doing taxonomic investigations	Eric. H. Metzler, unpaid, unfunded volunteer researcher, National Park Service, US
(1) the program which unites efforts from multiple herbaria in relation to their indets; (2) outsource agencies to perform DNA barcoding; (2) outsouce agencies which perform actual DNA barcoding (and probably store DNA)	Alexey Shipunov, MISU
1) Funds for undergraduates to database 2) expert scientists to correctly identify specimens	
1) website database designers 2) scientific users to test the functionality 3) place to engage the public to help in these efforts	
1) web/database developer, 2) hands on the ground (student researchers), 3) long term hosting	Amanda Grusz, University of Minnesota Duluth, algrusz@d.umn.edu
Symbiota developer, herbarium curator/collections manager	Brooke Best, BRIT, bbest@brit.org
1) The Carpentries 2) current worldwide staff mobilizing collections data 3) researcher feedback on skills they need to use biocollections data	Deborah Paul, Florida State University, dpaul@fsu.edu
1) Collections community to identify and link collections 2)social media company/companies 3)software engineers	
Agreed standard illustration format for particular taxa formalized. Some such standards exist - such as lighting strongest from upper left.	Nigel Hughes, Uni. Cal. Riverside, nigel.hughes@ucr.edu

Engage with the scientists who have biological data in archeological collections	Pat Holroyd, Univ. Of California
developer of biodiversity standards	Katie Pearson, kdpearso@calpoly.edu
NSF XSEDE infrastructure	Matthew Collins, UF, mcollins@acis.ufl.edu
1. Some programmer effort 2. Data Carpentry type workshops to teach it	
Software designers, IT experts. Maps are very important for answering scientific questions and I think for engaging the community.	Dr. Michael G. Simpson, San Diego State University, msimpson@sdsu.edu
1) iDigBio PIs 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	Edward Davis edavis@uoregon.edu
Widespread use of Symbiota features built for lichens of New Mexico	Tom Nash
1) GIS professionals 2) other georeference/databae herbarium specimens 3) historic place name specialists	Jason Alexander Univ. California, Berkeley
1)AI/ML experts 2)ecologists who could identify and extract traits and tie to their ecological function 3) other IT experts	Pam Soltis
none listed	