

Response to the 2016 Report of the iDigBio External Advisory Board

Prepared by:

Larry Page, iDigBio Project Director

David Jennings, iDigBio Project Manager

Prepared on:

March 10, 2017

Summary

iDigBio's Project Director, Project Manager, Principal Investigators, and Project Evaluator met in-person with the External Advisory Board (EAB) on November 1, 2016,¹ at the ADBC Summit VII held in Chattanooga, Tennessee. The meeting was observed by Roland Roberts, the NSF Program Director for iDigBio. As a result of the meeting (and as required by iDigBio's Cooperative Agreement² with NSF), the EAB provided a report documenting concerns and recommendations to iDigBio regarding sustainability, with consideration given to operating after the NSF-ADBC program, and, more specifically, increasing data use by research as well as education-outreach activities.

To facilitate discussions and actions internally, iDigBio first provided the EAB Report to the Steering Committee for consideration. Then, to facilitate follow-up, iDigBio created issue tickets in its Redmine project management system for each concern raised in the report.³

This document describes iDigBio's responses to the concerns and recommendations raised in the 2016 Report of the iDigBio External Advisory Board.

¹ https://www.idigbio.org/wiki/index.php/IDigBio_External_Advisory_Board#November_1.2C_2016

² General Programmatic Terms and Conditions for the National Resource for Digitization of Biodiversity Collections: Integrated Digitized Biocollections Cooperative Agreement (NSF 10-603)

³ https://www.idigbio.org/redmine/search?utf8=%E2%9C%93&q=%222016+Report+of+the+iDigBio+External+Advisory+Board%22&scope=all&all_words=&all_words=1&titles_only=&issues=1&commit=Submit

Data Use

1. Accurately quantifying data use is critical to the assessment of the progress of iDigBio.

a. Concern

The development of a more targeted suite of data use metrics would better facilitate an understanding of factors influencing short-term and long-term trends in access to iDigBio data. While an increasing trend of data access through the iDigBio portal appears to be consistent since project initiation, the trend is difficult to discern as presented. A rigorous statistical analysis should allow for examination of general trends of data use over time while partitioning out the immediate impact of workshops and other outreach activities. This dichotomy will help to illustrate different facets of success for iDigBio.

Response

iDigBio requires input from the EAB regarding what metrics might be useful in assessing data use beyond those that iDigBio is already collecting (see discussion below). In addition, iDigBio requests recommendations and examples of the type of statistical analysis envisioned. Lastly, iDigBio welcomes recommendations on where it would be best to display on the portal the aggregated data usage statistics across all recordsets.

iDigBio has been collecting and displaying monthly data use statistics for every recordset⁴ in the search portal since January 15, 2015. These statistics are displayed in the Data Use tab on each recordset page. The metrics used for collection of data use statistics are:

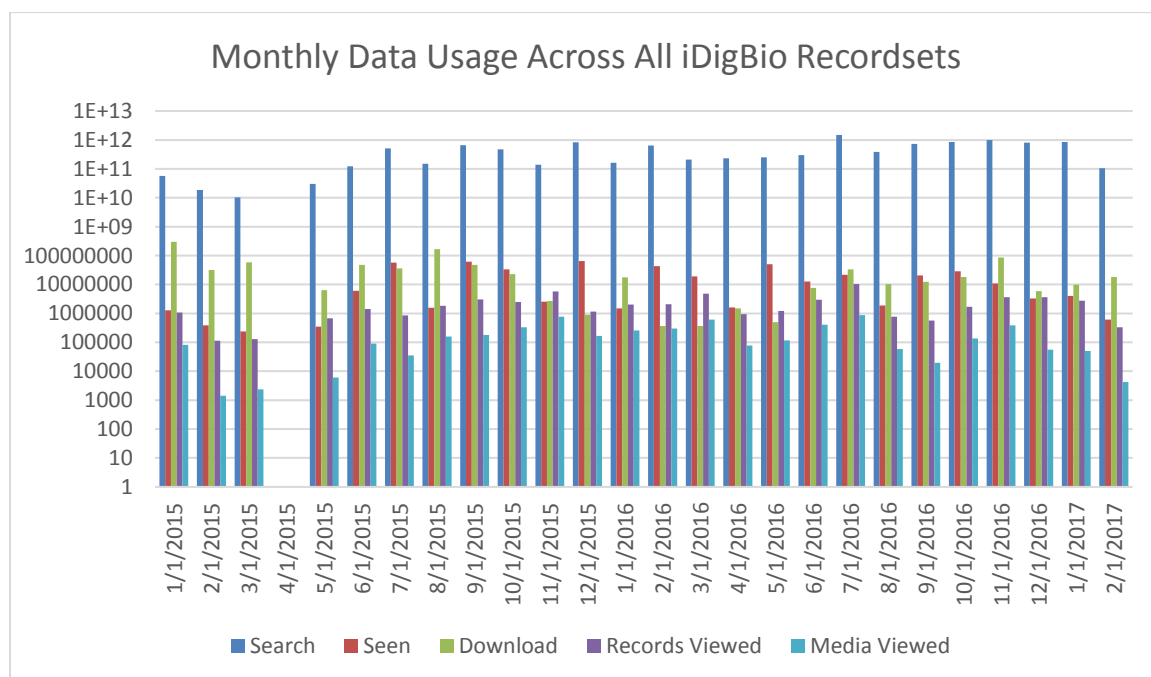
Metric	General Meaning	Specific Definition
Search	What is responsive to search queries?	How many records from the recordset matched a search query?
Seen	What is displayed to users?	How many records from the recordset visually appeared in the search results in a browser window?
Download	What is downloaded by users?	How many records from the recordset were downloaded?
Records Viewed	What specimen records are directly viewed by users?	How many specimen records from the recordset were opened and viewed in full detail?
Media Viewed	What media records are directly viewed by users?	How many media records from the recordset were opened and viewed in full detail?

⁴ <https://www.idigbio.org/portal/publishers>

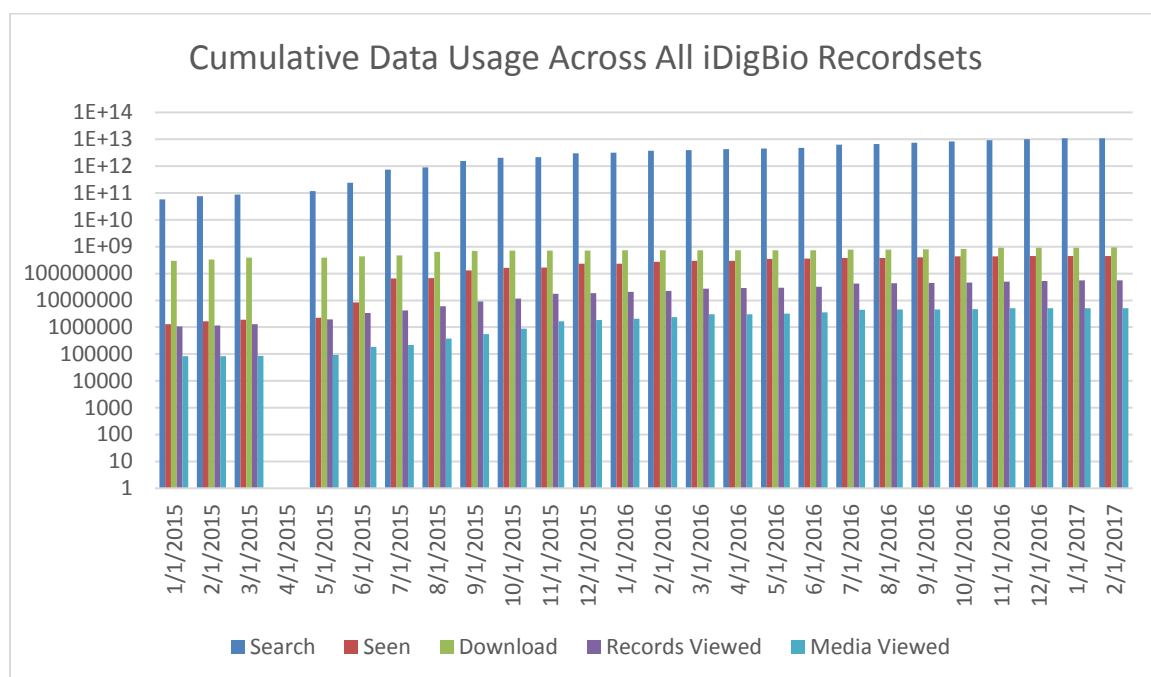
Additionally, for all of the above metrics, iDigBio collects (but does not currently report due to complexity of filtering coupled with size of the dataset) semi-anonymized query statistics (location only, not IP address).

For its semiannual and annual reports⁵ to NSF, iDigBio aggregates the data use statistics across all recordsets to provide a comprehensive view of data use across the iDigBio search portal. iDigBio currently does not display this aggregate information in the portal. The following are the aggregate data usage across all recordsets since the beginning of statistics collection (1/15/2015 thru 3/3/2017):

Search	10,977,317,245,741	Seen	449,527,218
Download	936,519,133	Records viewed	56,206,536
		Media viewed	5,168,987



⁵ https://www.idigbio.org/redmine/projects/administration/wiki/Historical_Report_Archive



As was noted during the EAB meeting, spikes in data usage may be coincident with events, workshops, or webinars that iDigBio hosts or is otherwise involved in. iDigBio is assuming the EAB is requesting information on any correlation between data use spikes and iDigBio events. However, iDigBio would like more information regarding what exactly is of interest to the EAB.

As an additional source of information on data use, iDigBio has been considering a mini survey associated with the data download feature to collect information on the types of data use.⁶ However, there has been lively debate within iDigBio about how useful this information will be (i.e., self-reported vs. system generated), what categories to collect (i.e., general vs. specific vs. length of survey), whether this creates an impediment to open data access, whether there are any IRB considerations, and whether our Privacy Policy as stated in our Terms of Use⁷ will need to be modified.

b. Concern

Calculation of metrics on data use from the iDigBio portal that are directly comparable to the metrics information from GBIF would help with understanding trends in data use as well as provide a comparative framework for project assessment. Specifically, iDigBio should consider adopting practices for tracking data use and communicating results in collaboration with GBIF

⁶ <https://www.idigbio.org/redmine/issues/2030>

⁷ <https://www.idigbio.org/content/idigbio-terms-use-policy>

and other relevant aggregators to ensure that metrics are comparable. Use of similar data collection standards and methods will enable data use statistics to be aggregated across portals. In this way, a more complete picture of data use can be assembled. We envision this involving collaborative meetings between iDigBio and GBIF to develop a consensus reporting structure that benefits both parties. This would also benefit the data providers by documenting the use of their data in a comparable way.

Response

iDigBio welcomes collaborative data usage metrics, tracking, and communication. iDigBio agrees that the ultimate goal is a common set of data use metrics and unification of data collection processes across providers. iDigBio representatives are committed to leading the community discussion around harmonization of data use statistics with TDWG's Biodiversity Services & Clients⁸ (BSC) interest group.

Concerning metrics that are comparable with GBIF, iDigBio needs more input from the EAB regarding what metrics GBIF is collecting/providing that are particularly valuable and useful. iDigBio performed a comparison between GBIF and iDigBio with respect to data use metrics to better understand the landscape. iDigBio chose an example recordset, the FLMNH Ichthyology collection, that was known to exist in both GBIF⁹ and iDigBio¹⁰:

⁸ <http://www.tdwg.org/activities/biodiversity-services-clients/charter/>

⁹ <http://www.gbif.org/dataset/eccf4b09-f0c8-462d-a48c-41a7ce36815a>

¹⁰ <https://www.idigbio.org/portal/recordsets/c38b867b-05f3-4733-802e-d8d2d3324f84>

UF FLMNH Ichthyology

Occurrence dataset published by Florida Museum of Natural History

225,254
Occurrences
[View occurrences](#)

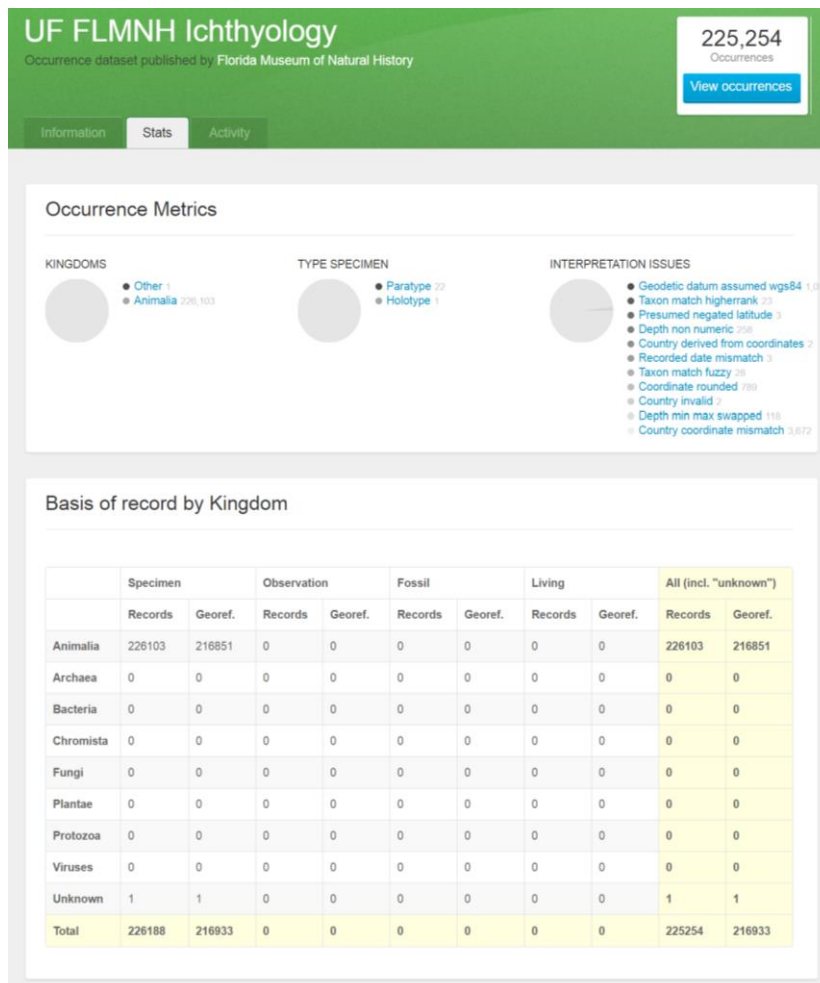
Information
Stats
Activity

14,930 download events

DOWNLOAD	doi:10.15468/dl.uvrjaz 9th March 2017
RECORDS	8 records from this dataset included at time of download
QUERY	TAXON <i>Carcharhinus altimus</i> (Springer, 1950) GEOREFERENCED true query latest data
DOWNLOAD	doi:10.15468/dl.d9hpfz 9th March 2017
RECORDS	225,194 records from this dataset included at time of download
QUERY	ALL DATA query latest data
DOWNLOAD	doi:10.15468/dl.vdip17 9th March 2017
RECORDS	622 records from this dataset included at time of download
QUERY	COUNTRY <i>French Guiana</i> query latest data
DOWNLOAD	doi:10.15468/dl.rkzcmi 9th March 2017
RECORDS	743 records from this dataset included at time of download
QUERY	COUNTRY <i>Suriname</i> query latest data
DOWNLOAD	doi:10.15468/dl.gtp7n1 9th March 2017
RECORDS	312 records from this dataset included at time of download
QUERY	COUNTRY <i>Guyana</i> query latest data
DOWNLOAD	doi:10.15468/dl.hftqe0 9th March 2017
RECORDS	3,780 records from this dataset included at time of download
QUERY	COUNTRY <i>Venezuela, Bolivarian Republic of</i> query latest data
DOWNLOAD	doi:10.15468/dl.iqiuqy 9th March 2017
RECORDS	2,413 records from this dataset included at time of download
QUERY	COUNTRY <i>Brazil</i> BASIS OF RECORD <i>Human Observation, Specimen</i> query latest data
DOWNLOAD	doi:10.15468/dl.xhcmho 9th March 2017
RECORDS	18 records from this dataset included at time of download
QUERY	COUNTRY <i>Uruguay</i> query latest data
DOWNLOAD	doi:10.15468/dl.ttcmlm 9th March 2017
RECORDS	336 records from this dataset included at time of download
QUERY	COUNTRY <i>Canada</i> query latest data
DOWNLOAD	doi:10.15468/dl.4wmnc 9th March 2017
RECORDS	1 records from this dataset included at time of download
QUERY	TAXON <i>Perca fluviatilis</i> Linnaeus, 1758 COUNTRY <i>France</i> query latest data

First
2
3
4
5
6
...
Last

GBIF Portal – FLMNH Ichthyology Collection – Activity Tab



GBIF Portal – FLMNH Ichthyology Collection – Stats Tab

The Activity tab on the GBIF dataset page includes the total number of download events as well as a list of every download event. Each entry in the download event list includes a DOI, a link to download the data in the event, and a link to query the latest data using the same search parameters as used in the download query. The Stats tab on the GBIF dataset page includes statistics about the dataset itself. One of the important items on this page is the Interpretation Issues, which indicate areas where GBIF has normalized the data provided.

Recordset

[Search Recordset](#)

UF FLMNH Ichthyology

Specimen Records: 225,130

Media Records: 0

iDigBio Last Ingested Date: 2017-02-20

The UF Fish Collection, dating to 1917, contains 214,205 lots and 2,300,803 specimens. Included are representatives of 8,250 species from 400 families. The collection includes 93 primary types and approximately 1,600 lots of secondary types representing 563 species. Also in the collection are 5,825 specimens of disarticulated and articulated skeletons representing 875 species. Especially notable are historic collections of large and important marine fishes as well as rapidly growing collections of freshwater fishes from Southeast Asia. In 2006, the museum expanded its program to archive frozen tissue samples with a newly established UF Genetic Resources Collection. Tissues of fishes are stored in -20°C freezers and number 4,150 samples of 900 species. All specimens and tissues are databased online and available for loan.

Contacts

Name Rob Robins
Role Ichthyology Collection Manager
Email rrobins@flmnh.ufl.edu

Name Rob Robins
Role Ichthyology Collection Manager
Email rrobins@flmnh.ufl.edu

Data Corrected Data Use Raw

The table below represents monthly iDigBio portal use statistics for this recordset. **Search** indicates in how many instances a record from this recordset matched a search query. **Download** indicates in how many instances a record from this recordset was downloaded. **Seen** indicates in how many instances a record from this recordset appeared (visually) in the search results in a browser window. **Records Viewed** and **Media Viewed** indicate how many specimen and media records were opened and viewed in full detail. Note: Monthly statistics aggregation began on Jan 15th 2015; therefore, the month of (01 / 2015) represents approximately half a month of statistics reporting.

Month of	Search	Download	Seen	Records Viewed	Media Viewed
01 / 2015	52,169,787	224,080	733	1,877	0
02 / 2015	160,296,710	223,895	2,372	1,035	0
03 / 2015	84,959,293	440,641	455	1,173	0
05 / 2015	257,737,259	213,932	2,486	7,539	0
06 / 2015	886,140,576	986,892	5,521	16,490	0
07 / 2015	2,446,086,020	221,546	21,805	10,133	0
08 / 2015	775,568,750	988,710	2,388	18,901	0
09 / 2015	3,259,157,586	225,407	1,580	32,974	0
10 / 2015	2,264,974,638	157,373	2,330	22,556	0
11 / 2015	658,529,023	1,132	3,958	55,848	0
12 / 2015	3,773,061,036	639	7,553	10,733	0
01 / 2016	710,692,040	182,569	21,690	21,787	0
02 / 2016	2,620,820,787	23,470	10,089	18,645	0
03 / 2016	814,797,222	7	575,613	50,747	0
04 / 2016	837,610,184	2,235	1,727	9,593	0
05 / 2016	2,076,558,364	162	4,088,241	14,646	0
06 / 2016	1,094,407,086	179,799	85,211	32,072	0
07 / 2016	5,540,219,855	0	572,945	105,109	0
08 / 2016	21,802,079,572	76,437	78,398	121,361	0
09 / 2016	43,830,266,977	421,936	1,461,523	132,822	0
10 / 2016	49,116,765,924	58,300	2,427,032	352,013	0
11 / 2016	59,774,600,576	36,843,185	773,682	718,454	0
12 / 2016	42,153,613,358	5,168,074	316,699	731,798	0
01 / 2017	48,410,109,197	998,276	485,411	637,908	0
02 / 2017	5,449,284,982	4,997,325	9,740	63,637	0

iDigBio Portal – FLMNH Ichthyology Collection – Data Use Tab

Recordset

Search Recordset

UF FLMNH Ichthyology

Specimen Records: 225,130

Media Records: 0

iDigBio Last Ingested Date: 2017-02-20

The UF Fish Collection, dating to 1917, contains 214,205 lots and 2,300,803 specimens. Included are representatives of 8,250 species from 400 families. The collection includes 93 primary types and approximately 1,600 lots of secondary types representing 563 species. Also in the collection are 5,825 specimens of disarticulated and articulated skeletons representing 875 species. Especially notable are historic collections of large and important marine fishes as well as rapidly growing collections of freshwater fishes from Southeast Asia. In 2006, the museum expanded its program to archive frozen tissue samples with a newly established UF Genetic Resources Collection. Tissues of fishes are stored in -20°C freezers and number 4,150 samples of 900 species. All specimens and tissues are databased online and available for loan.

Contacts

Name	Rob Robins	Name	Rob Robins
Role	Ichthyology Collection Manager	Role	Ichthyology Collection Manager
Email	rrobins@flmnh.ufl.edu	Email	rrobins@flmnh.ufl.edu

Data Corrected

Data Use

Raw

This table shows any data corrections that were performed on this recordset to improve the capabilities of iDigBio Search. The first column represents the correction performed. The last two columns represent the number and percentage of records that were corrected. A complete list of the data quality flags and their descriptions can be found here. Clicking on a data flag name will take you to a search for all records with this flag in this recordset.

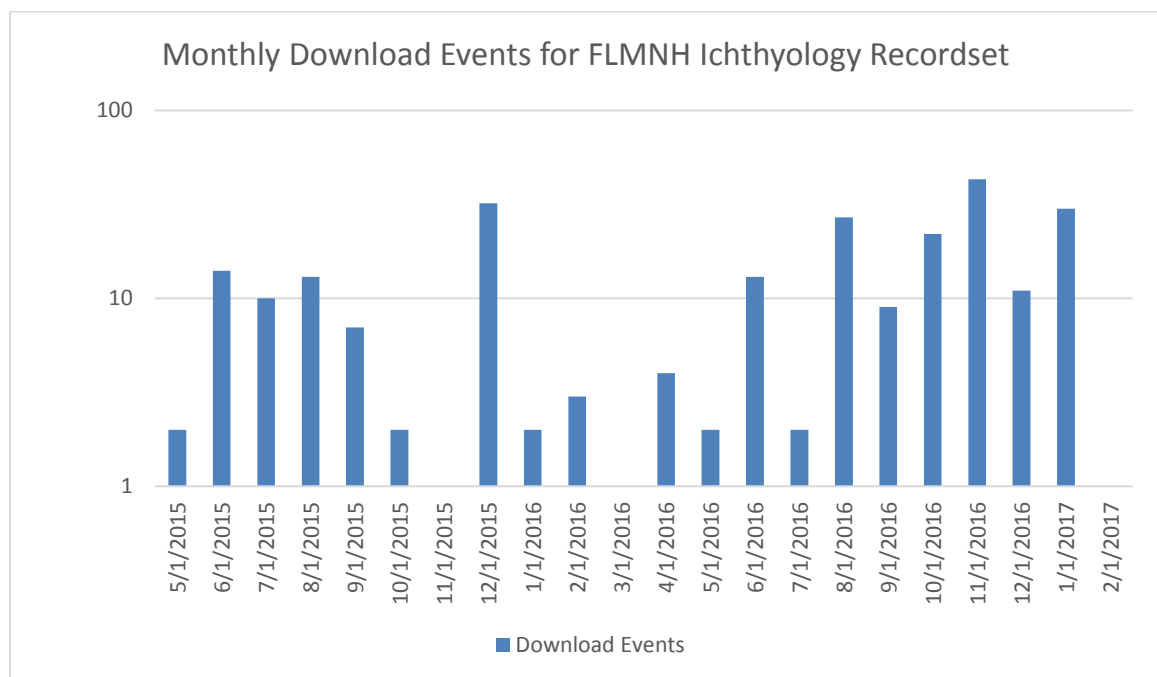
Flag	Records With This Flag	(%) Percent With This Flag
dwc_kingdom_added ⓘ	225008	99.946
dwc_phylum_added ⓘ	225008	99.946
dwc_datasetid_added ⓘ	222892	99.006
dwc_parentnameusageid_added ⓘ	222892	99.006
dwc_taxonid_added ⓘ	222892	99.006
dwc_taxonomicstatus_added ⓘ	222892	99.006
dwc_taxonrank_added ⓘ	222892	99.006
gbif_canonicalname_added ⓘ	222892	99.006
gbif_genericname_added ⓘ	222892	99.006
gbif_taxon_corrected ⓘ	222892	99.006
dwc_scientificnameauthorship_added ⓘ	222243	98.718
gbif_reference_added ⓘ	222158	98.668
dwc_taxonrank_replaced ⓘ	222110	98.659
gbif_vernacularname_added ⓘ	219687	97.582
geopoint_datum_missing ⓘ	219553	97.523
idigbio_isocountrycode_added ⓘ	203768	90.511
dwc_country_replaced ⓘ	159275	70.748
dwc_multimedia_added ⓘ	133150	59.144
dwc_originalnameusageid_added ⓘ	131981	58.824
dwc_continent_replaced ⓘ	97097	43.129
rev_geocode_eez ⓘ	96517	42.872
dwc_family_replaced ⓘ	17417	7.736
dwc_specificepithet_replaced ⓘ	17141	7.614
dwc_genus_replaced ⓘ	7251	3.221
dwc_order_replaced ⓘ	6358	2.824
rev_geocode_mismatch ⓘ	2418	1.074
taxon_match_failed ⓘ	2226	0.989
rev_geocode_failure ⓘ	707	0.314
dwc_stateprovince_replaced ⓘ	439	0.195
scientificname_added ⓘ	350	0.155
rev_geocode_corrected ⓘ	228	0.101
dwc_taxonremarks_added ⓘ	183	0.081
rev_geocode_lat_sign ⓘ	165	0.073
dwc_class_replaced ⓘ	155	0.069
rev_geocode_eez_corrected ⓘ	125	0.056
dwc_infraspecificepithet_added ⓘ	110	0.049
rev_geocode_lon_sign ⓘ	45	0.02
geopoint_0_coord ⓘ	24	0.011
geopoint_similar_coord ⓘ	15	0.007
rev_geocode_flip_both_sign ⓘ	7	0.003
rev_geocode_flip_lon_sign ⓘ	5	0.002
rev_geocode_flip ⓘ	4	0.002
datecollected_bounds ⓘ	2	0.001
dwc_infraspecificepithet_replaced ⓘ	2	0.001
rev_geocode_both_sign ⓘ	2	0.001

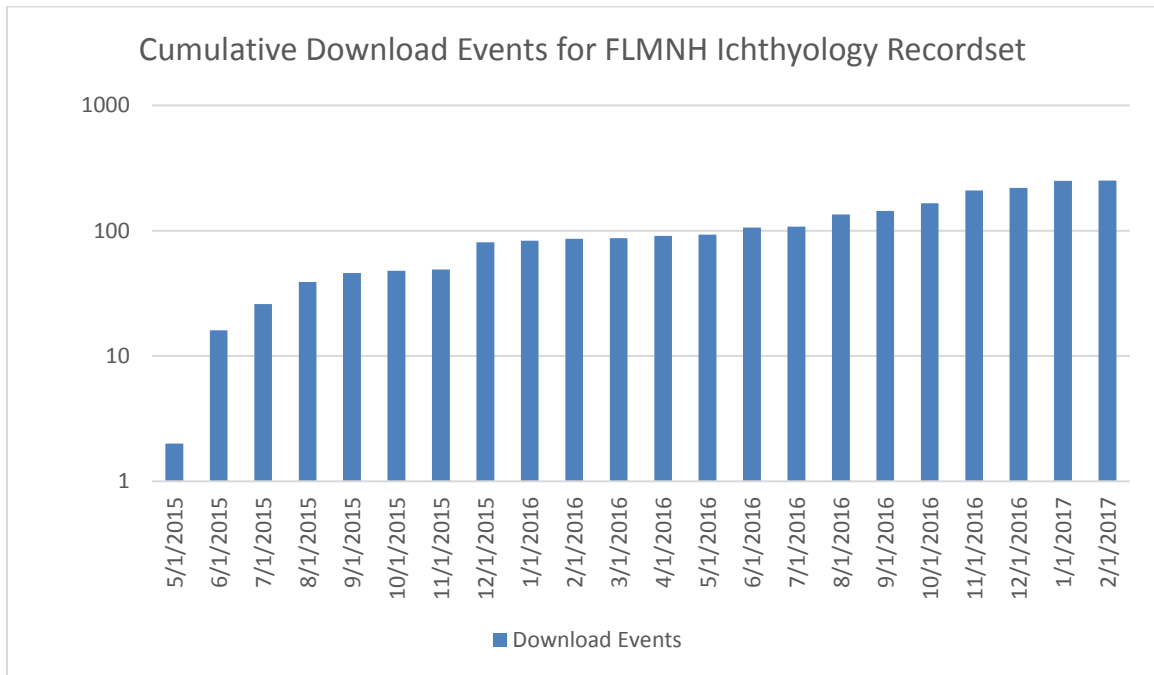
iDigBio Portal – FLMNH Ichthyology Collection – Data Corrected Tab

iDigBio's recordset page includes a Data Use tab that lists the monthly data use counts for Searches, Downloads, Records Viewed, and Media Viewed (see definitions in Data Use 1(a) above). The Data Corrected tab on the iDigBio recordset page lists the Data Quality Flags, which are analogous to GBIF's Interpretation Issues.

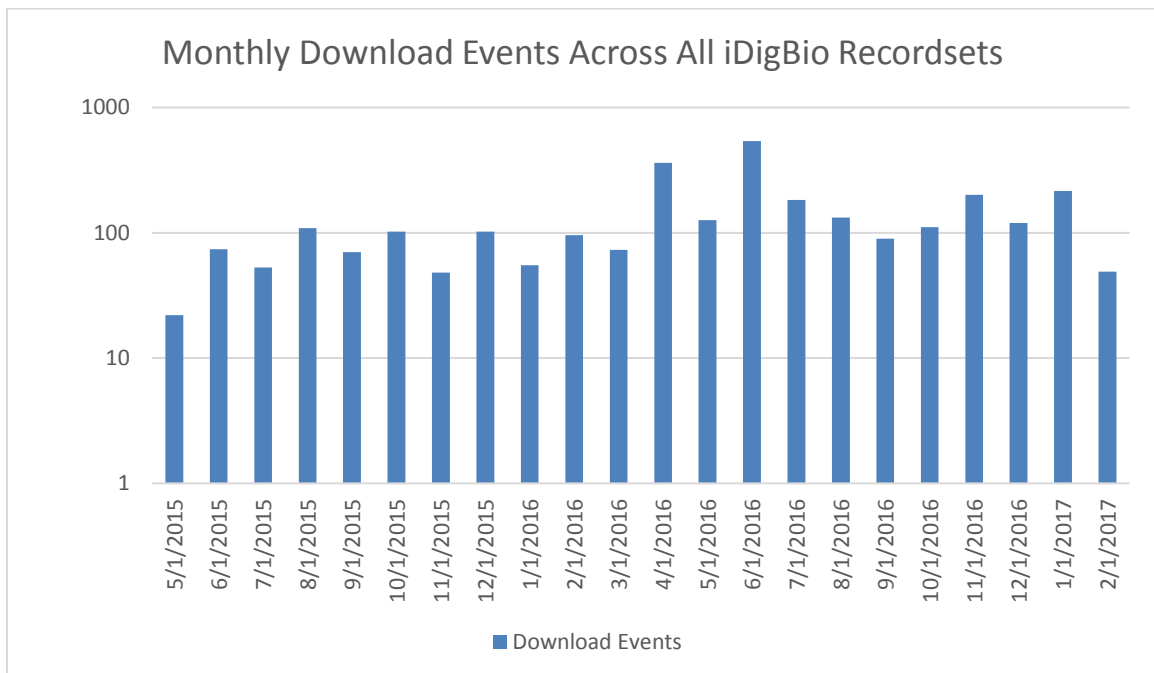
Based on the above comparison, iDigBio believes the information we currently provide, at least in terms of public facing metrics, is broader than GBIF's metrics reporting. Although, there may be additional metrics that GBIF makes available to data providers with a login.

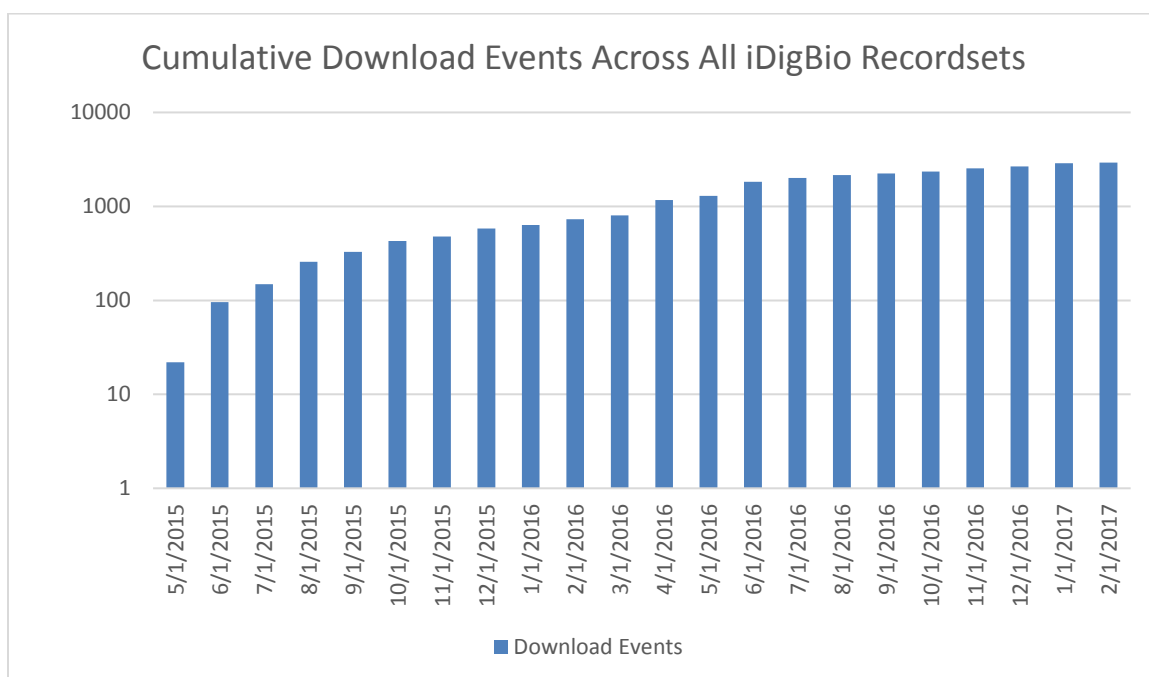
iDigBio has been collecting download event data that would be comparable with GBIF's activity data, but this information has not been exposed information publically. The following are graphs of the download events for the FLMNH Ichthyology Collection since the beginning of statistics collection (1/15/2015 thru 3/3/2017)





The following are graphs of the aggregate download events across all recordsets since the beginning of statistics collection (1/15/2015 thru 3/3/2017):





c. Concern

iDigBio would benefit from the development of approaches and tools that would allow for a better understanding of data use. In particular, the EAB feels that data tracking that provides the distinction between iDigBio portal-based data use and overall data use through other portals would improve understanding of data use and help justify sustainability plans.

Response

iDigBio promotes open-access to sharing of biodiversity data, and iDigBio's goal is to make data for scientific specimens widely available in electronic format. As such, iDigBio is a registered data publisher¹¹ with GBIF, facilitates IPT registration for new iDigBio data providers, and hosts its own IPT¹² for those providers that do not have adequate IT resources to host their own.

An unfortunate consequence of the open-sharing approach is the lack of an attribution chain. It is often not possible to track use across all portals due to the lack of appropriate fields in DwC to document all of the entities that should be attributed. For example, suppose iDigBio invests resources to mobilize and ingest a new dataset, including help with setting up an IPT at the host institution, and the data are then shared with GBIF. Suppose a user then searches and finds the data in GBIF, and uses them in a publication. It is highly likely that GBIF will be attributed as the

¹¹ <http://www.gbif.org/publisher/a86e9e36-12ec-49a4-a94c-c0c981fffb71>

¹² <http://www.gbif.org/installation/c6c9f784-8af9-4631-9912-dde88e38fafb>

source of the data (not iDigBio and potentially not the host institution). Proper attribution of data is a community problem where the ultimate responsibility lies with the researcher.

Although data citation and attribution is a community problem, iDigBio is attempting to help lead the charge in addressing it through partnerships with entities such as SPNHC¹³ and TDWG¹⁴. In addition, iDigBio has held discussion sessions on this important topic at its last two annual summits.¹⁵¹⁶

d. Concern

The EAB would appreciate additional data on use of the iDigBio portal that spans a longer time period, allowing assessment of whether portal use is increasing year over year.

Response

In its annual and semiannual reports to NSF, iDigBio has been providing data use graphs that cover the current fiscal year. However, iDigBio also generates graphs that cover data use since the beginning of data use collection. The graph provided in the response to Data Use 1(a) above represents the "all time" graph on a weekly basis.

iDigBio will start providing these data to the EAB when it submits annual and semiannual reports to NSF. A copy of the metrics supplied with iDigBio's most recent semiannual report to NSF is attached to this response.

2. The EAB feels strongly that expanding data use and establishing links to groups that have not traditionally used natural history collections data are critical to the justification of future funding and fundamental to any sustainability plan.

a. Concern

Along with sustainability, further developing the use of data generated by the TCNs and organized by iDigBio should be a priority. The EAB recognizes the challenges associated with balancing the development of the taxonomic strengths of iDigBio and the TCNs with the need to engage more diverse and nontraditional groups of researchers. Nevertheless, the engagement of a broad range of research groups beyond taxonomists and systematists should try to include ecologists, molecular biologists, earth systems scientists, conservation biologists, and climatologists.

¹³ <http://www.spnhc.org/>

¹⁴ <http://www.tdwg.org/>

¹⁵ https://www.idigbio.org/wiki/index.php/IDigBio_Summit_2016

¹⁶ https://www.idigbio.org/wiki/index.php/IDigBio_Summit_2015

Response

iDigBio is promoting the use of biodiversity data broadly and without regard to the source.

iDigBio believes that promoting the use of TCN data above other national data sources would be counter to its cooperative agreement with NSF: *"The primary goal of this project is to create and maintain a national biodiversity collections resource based on the existing biodiversity collections housed in non-federal U.S. institutions."* iDigBio believes the responsibility for specific promotion of TCN data falls on the TCNs in fulfilling their postulated research objectives. That said, iDigBio does, in effect, prioritize TCN data mobilization and ingestion (and, therefore, more quickly enables use) by virtue of the numerous resources that iDigBio makes available to TCNs, including the strong partnership iDigBio has developed with Symbiota¹⁷.

iDigBio is actively working to engage research communities other than systematists. For example, since 2014, iDigBio has organized symposia or workshops or given presentations on use of biodiversity data at annual/national/international meetings of the Botanical Society of America (N=3), Geological Society of America (3), SPNHC (3), Ecological Society of America (3), Entomological Society of America/ECN (2), Association of Southeastern Biologists (2), American Society of Mammalogists (2), Joint Meeting of Ichthyologists and Herpetologists (2), Organization of Biological Field Stations (2), Pacific Science Congress (1), and International Congress of Entomology (1). In addition, iDigBio convened a symposium "Collections for the 21st Century" specifically emphasizing use of specimen-based data for research in areas other than systematics. The following is a list of all iDigBio-hosted events focused on use of specimen-based data:

Date	Event	Meeting	Location
May 5-6, 2014	Symposium: Collections for the 21 st Century		Gainesville FL
July 26-30, 2014	Symposium: Digitized natural history collections records in traditional research, collaborative research and big data research	Botanical Society of America Annual Meeting	Boise ID
Nov 15-19, 2014	Symposium: Out of the field and into the lab: the state of the art of sorting biodiversity samples and processing to publication	Entomological Society of America/ECN Annual Meeting	Portland OR

¹⁷ <http://symbiota.org/docs/>

Date	Event	Meeting	Location
May 17-23, 2015	Symposium: Specimen full circle: collection to digitization to data use	SPNHC Annual Meeting	Gainesville FL
July 25-29, 2015	Workshop: Ecological niche modeling: a crash course at Botany 2015	Botanical Society of America Annual Meeting	Edmonton, Canada
Aug 10-14, 2015	Ignite Session at ESA: Enhancing ecological research with iDigBio specimen data	Ecological Society of America Annual Meeting	Baltimore MD
Nov 1-4, 2015	Symposium: Using digitized data in geological and paleontological research	Geological Society of America Annual Meeting	Baltimore MD
Nov 14-18, 2015	Talk Session: Using digitized insect data in research – included the following presentations (among others): "Introduction to using digitized data" "Using digitized fossil Coleoptera to study ecological and evolutionary response to global climate change"	Entomological Society of America/ECN Annual Meeting	Minneapolis MN
Dec 2-3, 2015	Workshop: Using biodiversity specimen data to study global change		St Louis MO
Dec 14, 2015	NSCA: Digitization – New tools for increasing the use of natural history collections for research, education, and informed decision-making (Page/Mast from iDigBio)		
March 12-13, 2016	Workshop: Coding phenological data from herbarium sheets		Berkeley, CA
June 20-25, 2016	Symposia: An international conversation on mobilizing natural history collections data and integrating data for research	SPNHC Annual Meeting	Berlin, Germany
June 24-28, 2016	Workshop: Geometric Morphometrics	American Society of Mammalogists Meeting	Minneapolis MN

Date	Event	Meeting	Location
July 18-22, 2016	Mini-workshop: Mobilizing natural history collections data for research use	Island Biology 2016	Terceira Island, Azores
July 30-Aug 3, 2016	Using digitized herbarium data in research: a crash course	Botanical Society of America Annual Meeting	Savannah, GA
Aug 7-12, 2016	Symposium: Leveraging the power of biodiversity specimen data for ecological research	Ecological Society of America Annual Meeting	Ft Lauderdale, FL
Sept 25-30, 2016	Symposium: Data without borders Symposium: Entomology in the digital age	ICE 2016 (International Congress of Entomology)	Orlando, FL
Oct 4-7, 2016	Workshop: Georeferencing for research use		Santa Barbara CA

This is an important area in which iDigBio will continue to focus in the coming years.

b. Concern

The production of Commentary and/or Review publications in specific disciplinary journals that highlight the research potential of data housed in iDigBio could provide an important stimulus for data use, as has been demonstrated in the past for other biodiversity informatics approaches (e.g., Kozak et al, 2008, "Integrating GIS-based environmental data into evolutionary biology" *Trends in Ecology & Evolution* 23:141–148). The 'Digital Data in Biodiversity Research Conference' in June 2017 at the University of Michigan presents an opportunity to publish a conference proceedings in a well-recognized journal.

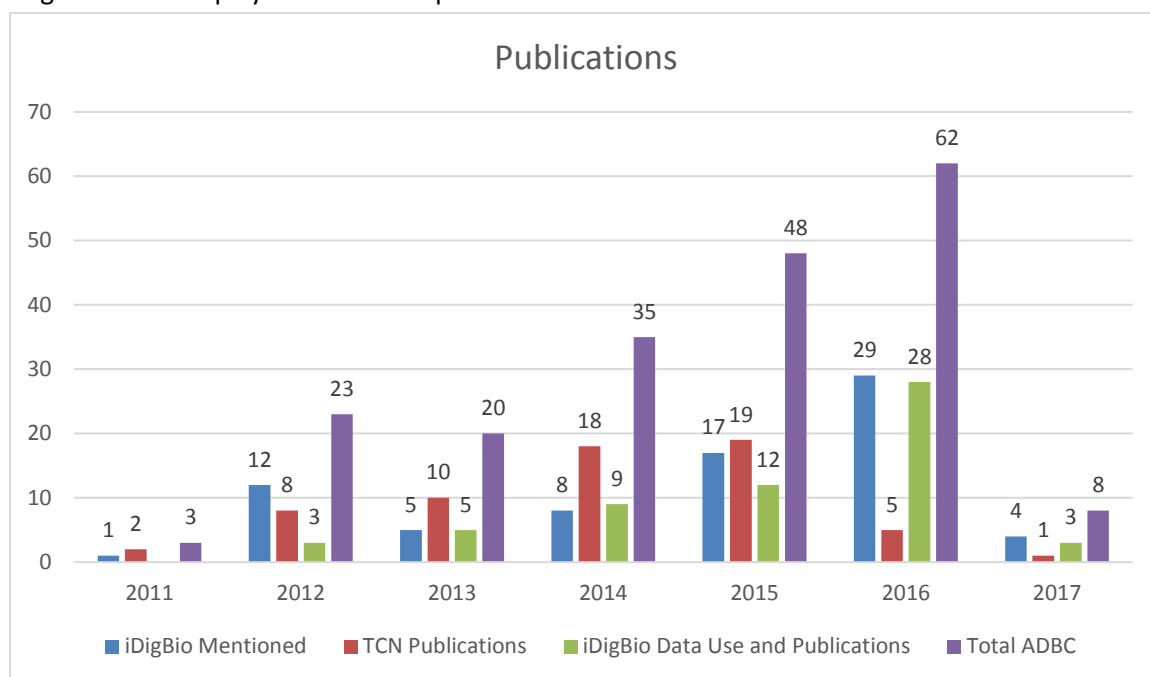
Response

iDigBio agrees with these ideas and is fully supportive. Publishing conference proceedings from the "Digital Data in Biodiversity Research Conference" is an excellent suggestion.

Articles authored by iDigBio PIs/staff on the research potential of data housed in iDigBio include "Digitization of biodiversity collections reveals biggest data on biodiversity" by Page, et al. in *Bioscience* in 2015; and "Old plants, new tricks: phenological research using herbarium specimens," by Willis, et al, in press in *Trends in Ecology and Evolution*. A manuscript highlighting results of the iDigBio workshop, "Using biodiversity specimen-based data to study global change" held in December 2015 is nearing completion.

A few notes about other activities that iDigBio facilitates to promote and highlight the research potential of the data housed in iDigBio:

- iDigBio regularly searches the literature to discover scientific papers that reference use of data from the iDigBio portal. These publications, along with other publications that mention iDigBio, are tracked and made available through iDigBio's public bibliography on Mendeley¹⁸. iDigBio welcomes suggestions from the EAB on where this information might best be displayed for consumption.



- iDigBio has worked with Pensoft to introduce a feature that allows authors to import occurrence records into a taxonomic manuscript by specifying an occurrence data identifier from GBIF, BOLD systems, or iDigBio, which serves to improve visibility of data use.¹⁹
- iDigBio has incorporated a Research Spotlight²⁰ section into its monthly newsletter, *The iDigBio Spotlight*²¹. The Research Spotlight highlights specific research uses of digitized collections data, including some of the research being produced by TCN's and iDigBio's investigators, graduate students, and post-doctoral associates.

¹⁸ <https://www.mendeley.com/groups/7777901/idigbio-public-library/papers/>

¹⁹ <http://blog.pensoft.net/2015/10/20/streamlining-the-import-of-specimen-or-occurrence-records-into-taxonomic-manuscripts/>

²⁰ <https://www.idigbio.org/tags/research-spotlight>

²¹ <https://www.idigbio.org/content/e-newsletter-archive>

c. Concern

Identify leaders within the non-traditional broader research disciplines to help develop the potential applications of iDigBio data, possibly via conferences or invitation to iDigBio summits.

Response

iDigBio agrees with the importance of this objective and has assumed that leaders of various disciplines, such as basic and applied ecology, conservation biology, and evolution other than systematics, were among the individuals attending our symposia and workshops. Examples of such efforts since 2014 are included in the response to Data Use 2(a) above. If the EAB can suggest another more targeted strategy for identifying leaders, we will consider it in the future.

d. Concern

Forward social media such as spotlight iDigBio articles to other societies and social media outlets. Do not rely solely on iDigBio sites for information dissemination.

Response

iDigBio agrees with these ideas and is fully supportive.

iDigBio currently attempts to amplify its announcements and news through many partner originations and projects, but we are always looking to connect with new societies and groups. iDigBio staff regularly post event announcements on multiple listservs outside iDigBio, including SPNHC (NHCOLL-L), Ecology (ECOLOG-L), HERBARIA@NACSE.ORG, TDWG, GCG (Geological Curators Group in the UK), and MUSEUM-L. We also ask for help in promoting our news and events on social media through our various partner organizations such as QUBES (Quantitative Undergraduate Biology Education and Synthesis), EOL (Encyclopedia of Life), BCoN, VertNet, NSF Biology, and ESA (Ecological Society of America).

iDigBio also reciprocates by promoting other projects and organizational information as long as it aligns with the general message of the iDigBio project. iDigBio uses the strategy of cultivating relationships and asking for help with promotion than it is to just simply forward material or post on other societies' social media pages. In addition, audience is an important consideration. For example, for iDigBio's upcoming undergraduate educator workshop, we reached out to QUBES and ESA for help in advertising because both entities have teaching faculty as part of their regular audience. One of the tasks for each iDigBio event is to decide where to advertise and whom to partner with to better amplify the message.

The following are a few examples of outside of community advertising we have done:

- Data Nuggets²² (we are now on their data resources page)
- Blog Post for Spatial Reserves²³
- Brain Scoop Webinar²⁴
- BetterMe Libraries of Life video²⁵

e. Concern

Consideration should be given to whether the iDigBio portal will evolve to include research tools that can engage a broader community of scientists.

Response

The iDigBio portal currently includes a link to our Research Collaborations²⁶ page, which highlights several websites, tools, and workflows that are currently taking advantage of the data being digitized at U.S. and global institutions and made available by iDigBio through its data services. The page also includes a call for collaboration in an effort to keep the list growing. Additional information is provided in the response to Data Use 1(g) below.

f. Concern

Members of the iDigBio team should offer to give seminars at universities with strong programs in disciplines needing engagement.

Response

iDigBio agrees with these ideas and is fully supportive.

Although iDigBio staff regularly give talks at a diverse array of locations, iDigBio does need to shift focus towards those audiences that are not already fully engaged. At the most recent meeting of the Internal Advisory Committee (IAC), it was proposed that the TCN members help iDigBio staff to present our data and our projects to non-traditional departments – whenever possible – in conjunction with travel already planned (to workshops, conferences, etc.). In addition, iDigBio has asked the TCN IAC members to plan to do the same type of outreach to nontraditional groups (engineering departments, information science departments, applications

²² <http://datanuggets.org/resources/data/>

²³ <https://spatialreserves.wordpress.com/2016/09/11/dusting-off-the-spatial-data-hidden-in-museum-collections/>

²⁴ <https://www.idigbio.org/content/education-and-outreach-webinar-series-brainscoop-insight-using-videos-engaging-outreach>

²⁵ <https://www.facebook.com/bettermeproductions/videos/1106685936124443/>

²⁶ <https://www.idigbio.org/content/idigbio-collaborations-enabling-research>

developer groups, etc.) whenever traveling to another conference, workshop, symposium related to this endeavor.

When iDigBio staff give a talk, these occurrences are documented externally in iDigBio's annual and semiannual reports to NSF. Many of these occurrences are also documented internally via External Collaboration Reports²⁷, which are disseminated to the Steering Committee for project-wide communication and have a ticket created in iDigBio's Redmine²⁸ project management system to facilitate follow-up actions.

g. Concern

Explore ways to stimulate and fund the creation of tools for researchers (especially for those not familiar with museum collections) to easily integrate their data into the iDigBio portal. For example, SiBBr in Brazil has been investing in the development of simple tools to assist those for whom the IPT is too difficult: <https://ferramentas.sibbr.gov.br/mycena/>. Additionally, continued development of tools to package iDigBio data into research products would be helpful.

Response

As noted in the response to Data Use 2(e) above, iDigBio maintains a Research Collaborations²⁹ page, which highlights websites, tools, and workflows that are currently taking advantage of the data being digitized at U.S. and global institutions and made available by iDigBio through its data services. iDigBio is working to keep this list growing.

In addition, iDigBio provides an API³⁰ (Application Programming Interface) as an abstraction layer to facilitates reuse and mashup of aggregated data without needing to understand the complex underlying details of the back-end data storage. iDigBio has also created an R³¹ and Python³² packages³³ that interfaces with the iDigBio search API allowing download of specimen records, review of metadata, etc.

Lastly, iDigBio is actively partnering with other entities to work towards development of other tools for researchers. For example, GUODA³⁴ (Global Unified Open Data Access) is a collaboration among developers and technical staff at EOL, iDigBio, and Jorrit Poelen. The initial

²⁷ <https://www.idigbio.org/content/idigbio-external-collaboration-reporting>

²⁸ <https://www.idigbio.org/redmine/projects/idigbio-collaboration-reporting/issues>

²⁹ <https://www.idigbio.org/content/idigbio-collaborations-enabling-research>

³⁰ https://www.idigbio.org/wiki/index.php/IDigBio_API

³¹ <https://cran.r-project.org/web/packages/ridigbio/index.html>

³² <https://pypi.python.org/pypi/idigbio>

³³ <https://cran.r-project.org/web/packages/ridigbio/index.html>

³⁴ <http://guoda.bio/>

group met during the June 2015 iDigBio API hackathon and since have been meeting regularly to discuss projects using Apache Spark³⁵ and biodiversity datasets to build tools to explore broad biodiversity questions.

Sustainability

1. Promoting broad use of iDigBio data across a variety of research disciplines and developing a good understanding of who uses the data and how they use it are essential prerequisites to development of a strong sustainability plan.

a. Concern

The recommendations above on diversifying the user community are also relevant to sustainability because a diverse user base creates a broader range of possibilities for seeking support (funding, partnerships, and collaborative tool development) that is the foundation of a sustainable business model.

Response

iDigBio agrees with these ideas and is fully supportive.

iDigBio is actively working to engage research communities other than systematists. For example, since 2014, iDigBio has had a distinct presence, organized symposia, or given presentations on use of biodiversity data at annual/national/international meetings of the Botanical Society of America (N=3), Geological Society of America (3), SPNHC (3), Ecological Society of America (3), Entomological Society of America/ECN (2), Association of Southeastern Biologists (2), American Society of Mammalogists (2), Joint Meeting of Ichthyologists and Herpetologists (2), Organization of Biological Field Stations (2), Pacific Science Congress (1), and International Congress of Entomology (1). In addition, iDigBio convened a symposium “Collections for the 21st Century” specifically emphasizing use of specimen-based data for research in areas other than systematics and phenology.

As noted in the response to Data Use 2(a) above, iDigBio is actively working to engage research communities other than systematists. For example, since 2014, iDigBio has organized symposia or workshops or given presentations on use of biodiversity data at annual/national/international meetings of the Botanical Society of America (N=3), Geological Society of America (3), SPNHC (3), Ecological Society of America (3), Entomological Society of America/ECN (2), Association of Southeastern Biologists (2), American Society of Mammalogists (2), Joint Meeting of Ichthyologists and Herpetologists (2), Organization of Biological Field Stations (2), Pacific Science

³⁵ <http://spark.apache.org/>

Congress (1), and International Congress of Entomology (1). In addition, iDigBio convened a symposium “Collections for the 21st Century” specifically emphasizing use of specimen-based data for research in areas other than systematics.

Also, in an effort to increase data use beyond research, iDigBio is planning a conference for spring 2017 directed at staff of governmental agencies and NGOs concerned with conservation of natural resources. The goal is to increase awareness of the value of the huge amount of occurrence data and other forms of data associated with specimens in collections.

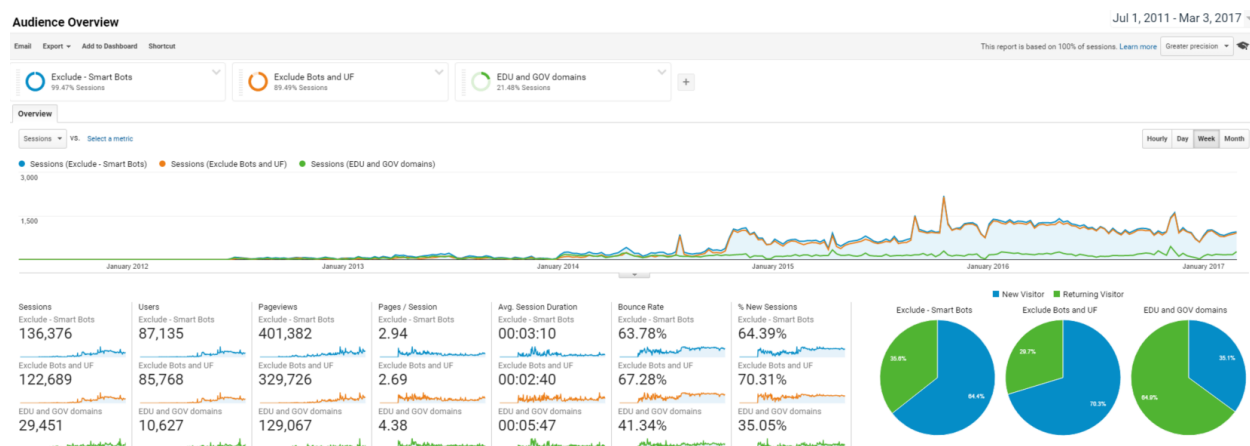
This is an important area in which iDigBio will continue to focus in the coming years.

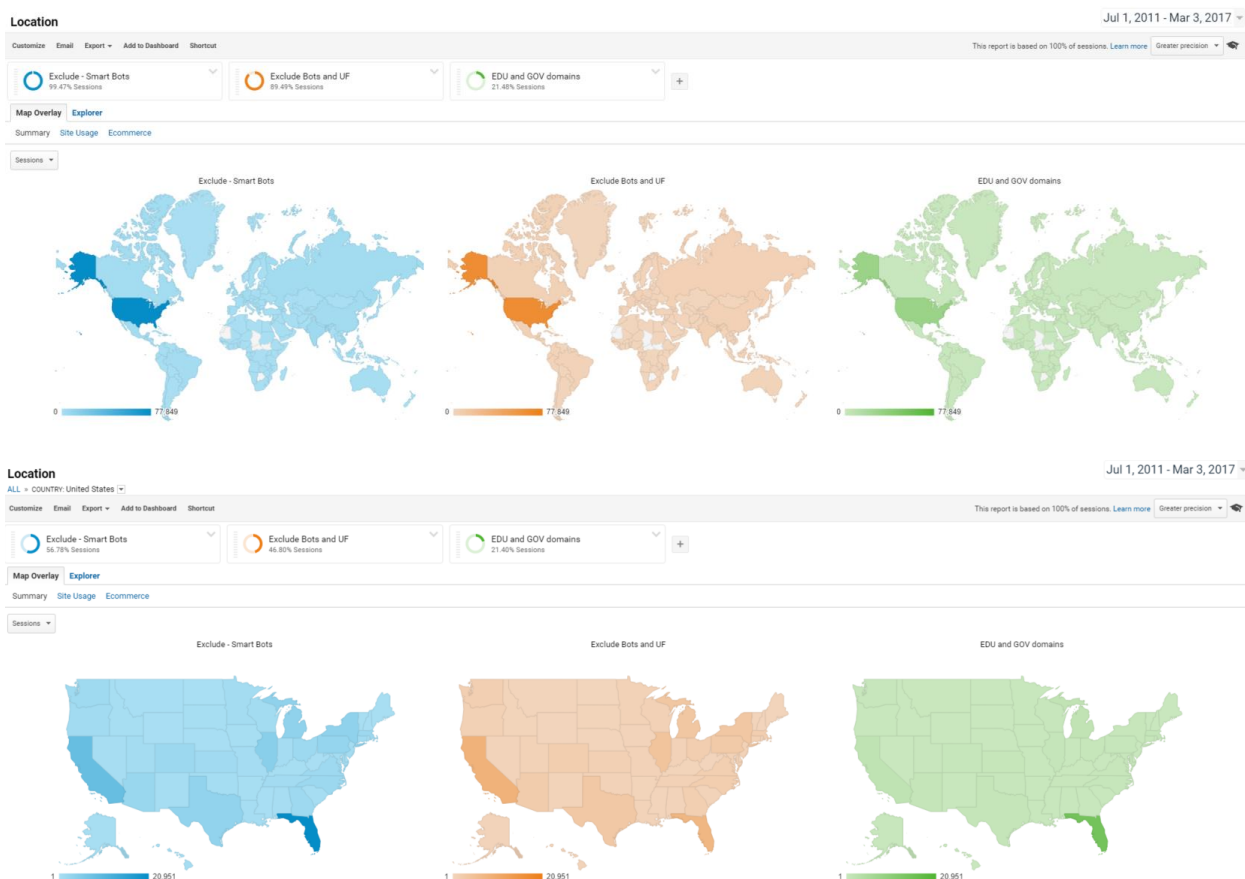
b. Concern

Usage tracking should include the number of unique visitors per month and the geographical location and institutional affiliation of users. This information can be gathered through Google Analytics.

Response

In its annual and semiannual reports to NSF, iDigBio has been providing Google Analytics data for both the website and portal. The data provided include all sessions (excluding smart bots), all sessions excluding UF (and excluding smart bots), and visits by country and by state. In addition, iDigBio has created other segments for potential future use, such as sessions by EDU and GOV domains. The following are the Google Analytics for the portal by week covering "all time" (7/1/2011 thru 3/3/2017):





iDigBio recommends that it start providing these data to the EAB on the same intervals as its annual and semiannual reports to NSF. A copy of the metrics supplied with iDigBio's most recent semiannual report to NSF is attached to this response.

c. Concern

Periodic online surveys can identify the research domains and research questions of interest to the community of data users. By identifying the communities that most value iDigBio/TCN data, tools and expertise, we can gain a better understanding of how those user groups might be able to help sustain key elements of iDigBio and the broader digitization effort into the future.

Response

iDigBio has been gathering this type data via its annual Community Survey and/or the post-Summit surveys. iDigBio will include this as one of the questions for this year's Community Survey.

iDigBio has internally discussed the possibility of increasing the frequency of the Community Survey to semiannual. However, there are significant concerns over survey fatigue, which will diminish participation. As a result, iDigBio will keep the Community Survey on an annual schedule.

As mentioned in the response to Data Use 1(a) above, iDigBio has been considering the addition of a mini survey associated with the data download feature in the portal. We could consider a similar type of survey on the website to address research domains and topics of interest. However, there are many factors to consider, including: how to ensure we are gathering useful information, what categories should we be collecting (i.e., general vs. specific vs. length of survey), are there any IRB considerations, and does our Privacy Policy as stated in our Terms of Use need to be modified.

2. Careful definition of iDigBio's distinct, future role within the global biodiversity data ecosystem will be critical to ensure that: (a) no essential function is lost after the current funding ends, and that (b) resources for sustaining digitization and data access infrastructure are used efficiently to avoid duplication of effort.

a. Concern

Coordination with GBIF and other data providers will be essential in defining the unique role of the iDigBio cyberinfrastructure that will need to be preserved after the current funding period. A deep discussion with these groups on current areas of strength and future plans will help clarify where iDigBio's future focus should be. This should be an ongoing process involving regular meetings that include GBIF, iDigBio and other groups, and collaboration on areas of mutual interest.

Response

iDigBio agrees with this assessment and, as outlined in the Sustainability Plan³⁶ submitted with the proposal to NSF for renewal of iDigBio, national and international coordination and collaborations are key to the long-term success of iDigBio and other elements of NIBA/ADBC. Efforts in this area must go beyond defining the unique role of the iDigBio cyberinfrastructure.

In a major effort toward enhanced coordination and increasing collaborations, a team consisting of Volker Mosbrugger, John LaSalle, Rosemary Gillespie, Donald Hobern, James Hanken, and Larry Page have organized a workshop, "Exploring Synergies and Sustainability for Biodiversity Information Systems (BISs)," to be held March 29-31, 2017, at the Senckenberg Gesellschaft für Naturforschung in Frankfurt, Germany. The primary goal of the workshop is to develop a

³⁶ https://www.idigbio.org/redmine/projects/administration/wiki/IDigBio_Sustainability_Plan

coordinated strategy for future development of Biodiversity Information Systems, of which 25 will be represented at the workshop. Principal topics include: identifying unique roles, removing redundancies, identifying gaps in data and services, sharing technology and other resources, improving communication, and identifying strategies for sustainability.

This workshop, which is an outgrowth of the iDigBio workshop, “Using Biodiversity Specimen-Based Data to Study Global Change” held at the Missouri Botanic Gardens in December 2015, has the potential to clarify the future role for iDigBio/ADBC and thus to be critically important in development of a long-term sustainability strategy. The workshop has the attention of NSF Program Directors who see improvements, including identifying key roles in the various funded efforts to improve access to information on biodiversity, as a top priority.

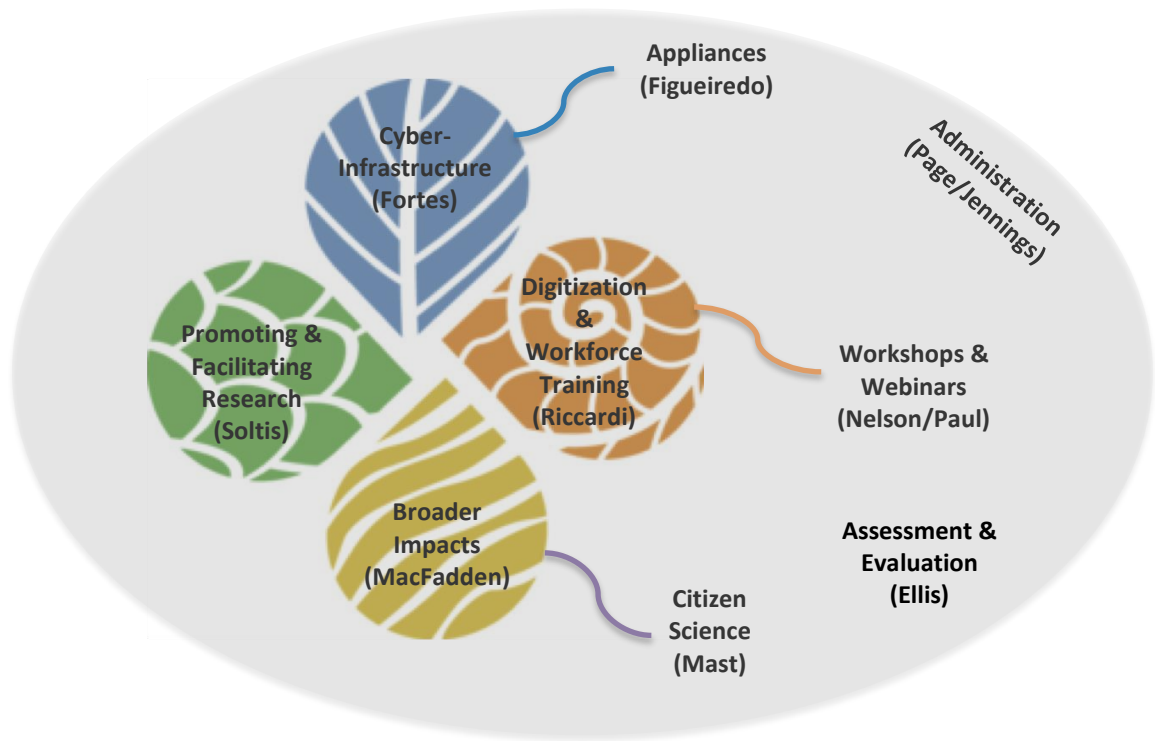
3. Sustainability planning will need to separately consider iDigBio’s four core areas of activity (engaging the collections community, digitization, database/informatics, and research/education) to determine what must be sustained in each area and develop plans that target each area.

a. Concern

iDigBio should identify specific metrics to assess progress within each of the four activity areas and set achievable and measurable goals for the remainder of the funding period. This will have a dual benefit of focusing efforts for the remainder of the funding period and forming the basis for an estimate of the work within each area that is not likely to be completed under the existing funding.

Response

iDigBio is organized into five domains to achieve its goals: (1) Project Administration & Management, (2) Education & Outreach, (3) Serving the Research Community, (4) Digitization, and (5) Cyberinfrastructure. The Principal Investigators (PIs) provide leadership within each domain, and the Project Director and Project Manager provide overall leadership for the project.



The project incorporates several oversight committees to provide governance, advice, and leadership concerning goals, strategies, implementation, activities, and progress. The Project Director, in consultation with the Project Manager, Executive Committee, and Steering Committee, makes project decisions that affect scope, budget, and/or risk. The Project Manager effects day-to-day planning, execution, and decision-making.



iDigBio's Strategic Plan³⁷, which was submitted with the proposal to NSF for renewal of iDigBio, includes strategies and objectives for each domain of iDigBio. In addition, the Strategic Plan describes iDigBio's assessment and evaluation plan, including the definition of key metrics used to evaluate the impact of iDigBio.

A copy of iDigBio's Strategic Plan is attached to this response.

b. Concern

Sustainability planning should focus on each of the four activity areas of iDigBio. A draft plan for each area should include 1) assessment of the need for continued activity in each of the four areas after the end of the current funding period; 2) how the activity might change assuming the proposal goals are met; and 3) a set of ideas including both federal grant funding and other revenue streams that could contribute to sustaining the anticipated future activity (based on perceived value to users, as derived from data use metrics). Assessment of need in each area should include discussions with a range of stakeholders including data providers and aggregators and a variety of data users.

³⁷ https://www.idigbio.org/redmine/projects/administration/wiki/IDigBio_Strategic_Plan

Response

iDigBio's Sustainability Plan, which was submitted with the proposal to NSF for renewal of iDigBio, provided our view of the sustainability landscape. Expansion of the EAB, with a particular focus on increasing expertise on sustainability, was, in large part, done in recognition of the need to think more broadly about developing sustainability strategies. iDigBio agrees that needs assessments, including discussions with a range of stakeholders including data providers, data aggregators, and data users, need to be conducted as an initial step to adequately plan for and implement sustainability. Two upcoming events appear likely to provide particularly useful information toward developing sustainability strategies: (1) the workshop involving 25 biodiversity data-focused organizations to take place later this month (see the response to Sustainability 2(a) above), which will help identify the unique roles played by iDigBio, and (2) the conference schedule for Spring 2017 directed at staff of governmental agencies and NGOs (see the response to Sustainability 1(a) above), which will focus on data use outside the normal uses of institutional collections data. We look forward to further discussions with the EAB on other mechanisms for moving forward.

A copy of iDigBio's Sustainability Plan is attached to this response.

c. Concern

In order to develop better metrics on digitization progress, iDigBio should obtain more accurate estimates of numbers of specimens and data providers. There appear to be three classes of data that iDigBio ingests 1) TCN and PEN data from US institutions, 2) data from individual collections and non-ADBC funded programs in the United States, and 3) data from larger museums and programs outside the United States. Estimating the digitization trajectories for #1 and #2 (within reasonable margins of error) are critical in assessing the success of the ADBC program and for planning a strategy to implement beyond 2020. Understanding how many collections and how much data has been produced by #1 versus #2 is very important in knowing the direct impact of ADBC funding and who has been included in TCN funding and who is not likely to receive ADBC funding. Additionally, knowing the progress for the different major taxa is important in establishing sustainability plans with different priorities. For example, if 75% of vertebrate specimens will be digitized in US collections by 2021 but only 10% of arthropod specimens, then devising two separate plans might be appropriate. For arthropods, planning would focus on ways to greatly increase the rate of digitization of existing collections, whereas the priority for vertebrates would be to focus on ensuring that new specimens are routinely digitized and promoting use of the digitized data in a range of research areas. There should be a separate assessment for digitization by museums focused on paleo versus modern specimens as well as TCN efforts digitizing non-specimen records (e.g., vocal recordings). Finally, the vast majority of specimens exist within a small subset of museums, and so another key metric would be to

assess the relative progress of efforts within each of the larger institutions and develop strategies to ensure digitization efforts are maximized at each institution.

Response

iDigBio agrees that targeting certain collections/institutions outside those funded by NSF, and employing funding strategies not currently considered by NSF, have the potential to significantly increase the rate of digitization. Discussions have taken place between iDigBio and NSF about alternative strategies, including funding individual institutions with large non-digitized collections (mostly insects and other invertebrates), but NSF has resisted this approach, preferring to continue its multiple-institution strategy and focus on research questions for funding TCNs. Changes in funding strategies at NSF can have significant social as well as scientific impacts and need to come from the collections and research communities.

An iDigBio workshop to be held at SPNHC 2017 in Denver, “Directors Summit on Digital Data,” is designed to target the precise suggestion from EAB to address “strategies to ensure digitization efforts are maximized at each institution.” Twenty-five administrators of collection-holding institutions will discuss digital asset management goals and related topics.

With regard to metrics and tracking digitization progress, iDigBio actively maintains and updates³⁸ a list of natural history collections³⁹ in the United States and includes identifiers to link information housed in GRBio⁴⁰ and Index Herbariorum⁴¹. iDigBio uses this collections list as the foundation for making assessments decisions concerning data mobilization and ingestion. iDigBio employs the following efforts to gather more information about collections:

- a) iDigBio periodically reviews GRBio and Index Herbariorum for potential new collections. In addition, iDigBio reviews the IPT instances in GBIF for new data sources.
- b) When iDigBio ingests a dataset the first time, the provider is contacted and encouraged them to update the information we have about them (contacts, taxonomic range, geographic range, collection size). Unfortunately, only a small percentage of providers do that.
- c) In early 2015, iDigBio wrote to every collection on iDigBio's list that has not already mobilized its data. Of the 500 letters written, only 9 engendered a response, and only 4 have resulted in any data ingestion; the remainder are in the mobilization queues that we continue to manage.

³⁸ <https://github.com/iDigBio/idb-us-collections>

³⁹ <https://www.idigbio.org/portal/collections>

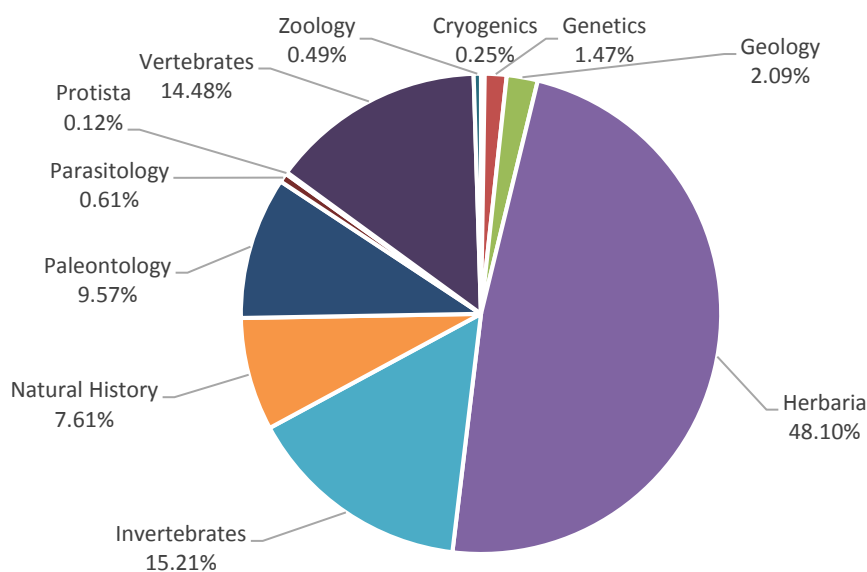
⁴⁰ <http://grbio.org/>

⁴¹ <http://sciweb.nybg.org/Science2/IndexHerbariorum.asp>

- d) We have discovered that the size of collections is not consistently available nor reliable and that community standards describing this property of collections are in early stages of development. In addition, domain expertise is often required to estimate the size and scope of collections that are missing from iDigBio's collections list as these are often the smallest and least responsive collections.
- e) Finally, we have discovered that when reconciling entries in the collections list against data published by contributing collections, the data are heterogeneous: Sometimes there is one entry for an institution (e.g., Natural History); other times the data are very granular and represent individual datasets (e.g., the lichen dataset in a herbarium).

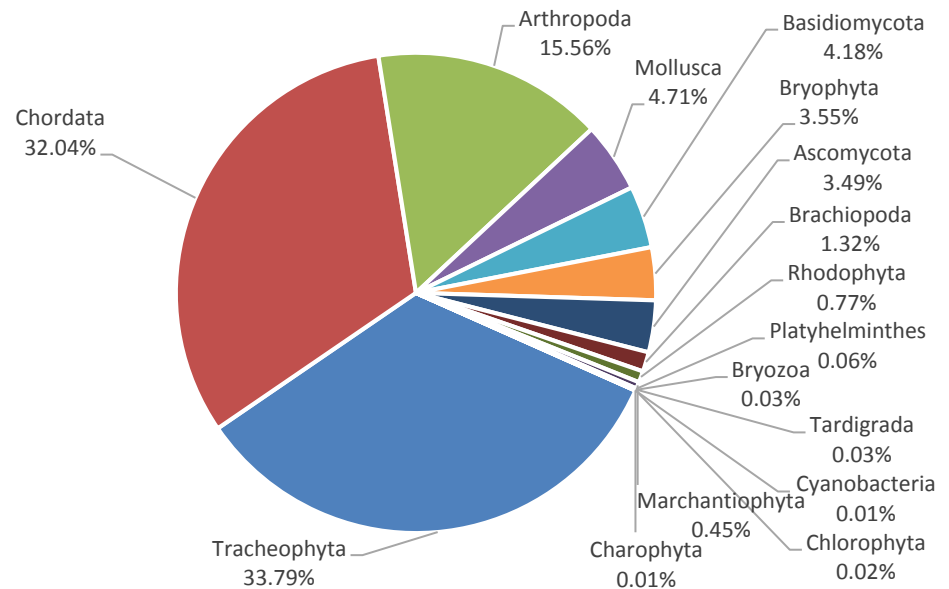
Based on iDigBio's understanding of natural history collections of the United States, we have:

- **Total number of collections: 1303**
- **815 collections that HAVE NOT yet provided data**
 - Some of those 815 are in TCNs/PENs but have not delivered data yet (especially herbaria and institutions with collections of invertebrates)
 - The 815 collections that HAVE NOT yet contributed data can be categorized as (based on their names or other, sometimes limited, information provided):



- **488 collections that HAVE provided data**
 - These collections are from 283 institutions that are contributing 629 recordsets

- Breakdown of specimen record counts by phylum for U.S. collections ingested by iDigBio:



Conclusion

iDigBio understands and respects the issues raised by the External Advisory Board in their 2016 report. In addition, iDigBio appreciates the time, effort, and concern of the EAB in ensuring the success of iDigBio and the national digitization effort. iDigBio will continue to work closely with the EAB to address improvements in the areas identified.

Attachments

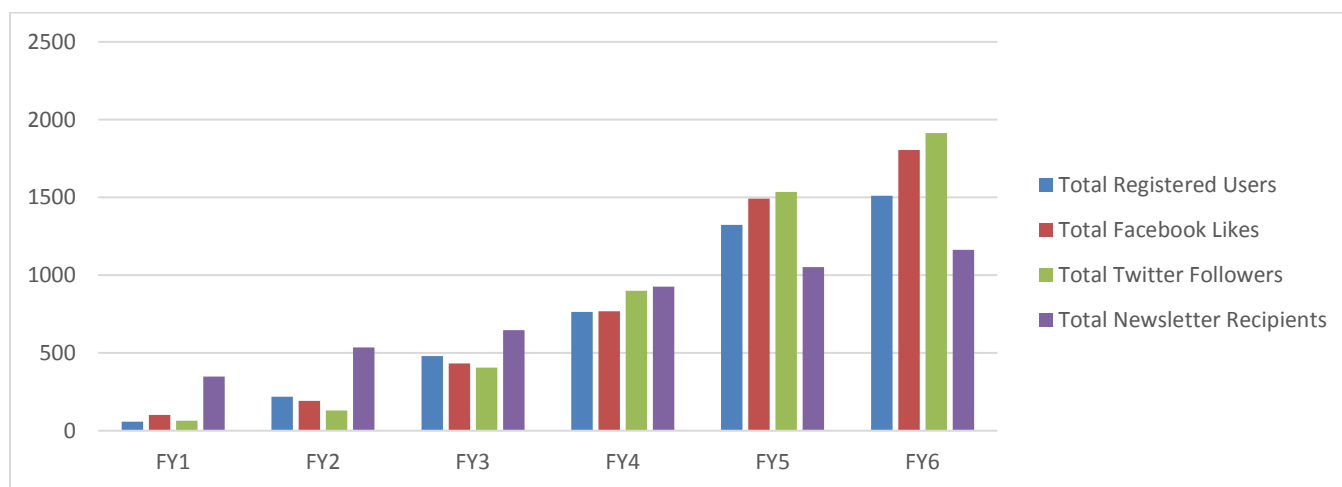
- Web Presence Statistics from iDigBio's FY6 Semiannual Report
- iDigBio's Strategic Plan
- iDigBio's Sustainability Plan

TABLE OF CONTENTS

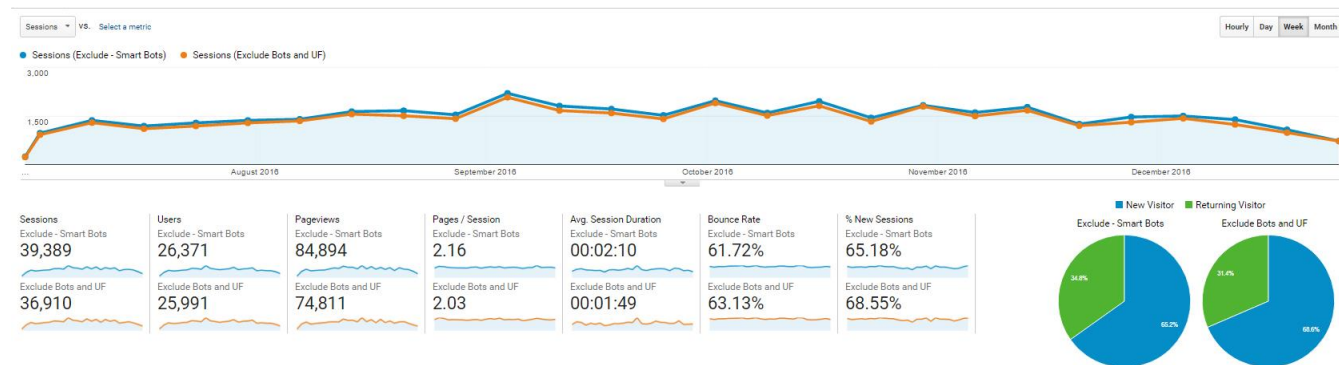
Table of Contents.....	1
iDigBio Online Presence (7/1/2011 thru 12/31/2016)	1
Website Usage Statistics (7/1/2016 thru 12/31/2016).....	2
Portal Usage Statistics (7/1/2016 thru 12/31/2016)	6
eNewsletter Statistics (7/1/2016 thru 12/31/2016).....	7
Adobe Connect Usage Statistics (7/1/2011 thru 12/31/2016)	9
Adobe Connect Recording Statistics (7/1/2016 thru 12/31/2016).....	10
Social Media Statistics (7/1/2016 thru 12/31/2016)	12
Data Ingestion Statistics (7/1/2016 thru 12/31/2016)	17
Portal Data Use Statistics (7/1/2016 thru 12/31/2016)	18

IDIGBIO ONLINE PRESENCE (7/1/2011 THRU 12/31/2016)

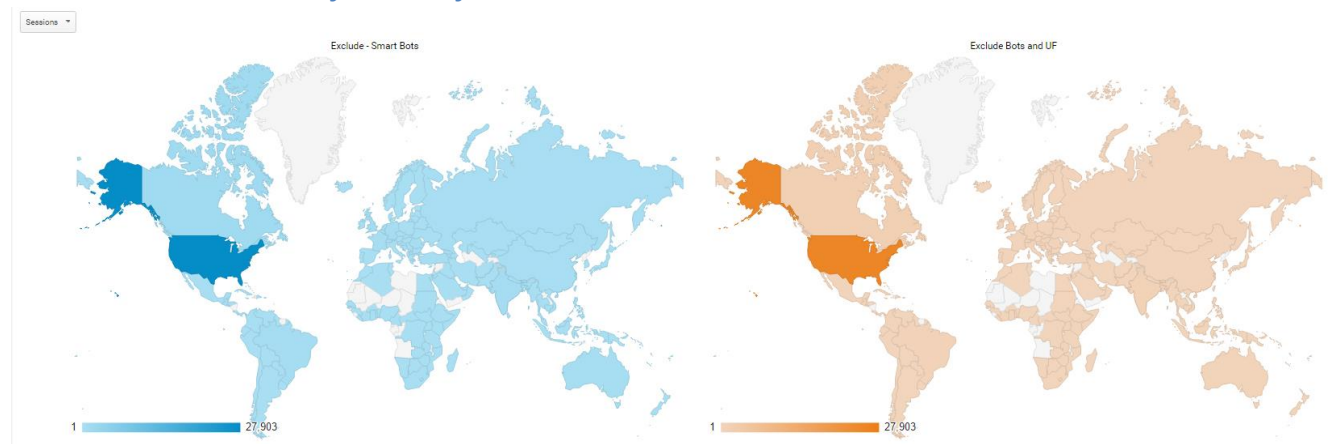
Timeframe	Total Registered Users	Total Facebook Likes	Total Twitter Followers	Total Newsletter Recipients
Fiscal Year 1 (7/1/2011 – 6/30/2012)	59	102	64	349
Fiscal Year 2 (7/1/2012 – 6/30/2013)	218	191	130	535
Fiscal Year 3 (7/1/2013 – 6/30/2014)	480	433	405	647
Fiscal Year 4 (7/1/2014 – 5/22/2015)	764	768	900	926
Fiscal Year 5 (7/1/2015 – 6/30/2016)	1,324	1,493	1,535	1,051
Fiscal Year 6 (7/1/2016 – 12/31/2016)	1,510	1,804	1,915	1,163



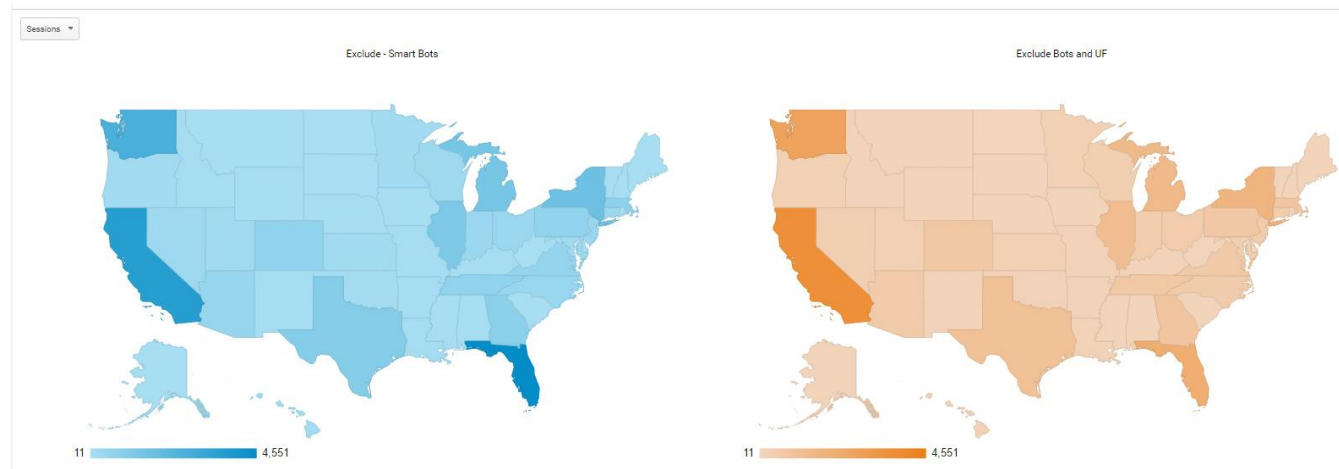
WEBSITE USAGE STATISTICS (7/1/2016 THRU 12/31/2016)



Total Website Visits by Country



Total Website Visits by State



Top Ten Page Views

Page Title	Pageviews	% Pageviews
1. iDigBio Home iDigBio		
Exclude - Smart Bots	15,370	18.10%
Exclude Bots and UF	12,346	16.50%
2. Wiki Home - iDigBio		
Exclude - Smart Bots	5,253	6.19%
Exclude Bots and UF	5,041	6.74%
3. Reset password - iDigBio		
Exclude - Smart Bots	5,028	5.92%
Exclude Bots and UF	5,028	6.72%
4. Log in - iDigBio		
Exclude - Smart Bots	4,865	5.73%
Exclude Bots and UF	4,865	6.50%
5. About iDigBio iDigBio		
Exclude - Smart Bots	1,944	2.29%
Exclude Bots and UF	1,695	2.27%
6. iDigBio The National Resource for Advancing Digitization of Biodiversity Collections		
Exclude - Smart Bots	1,848	2.18%
Exclude Bots and UF	1,186	1.59%
7. iDigBio Summit 2016 - iDigBio		
Exclude - Smart Bots	1,704	2.01%
Exclude Bots and UF	1,521	2.03%
8. Georeferencing for Research Use - iDigBio		
Exclude - Smart Bots	1,625	1.91%
Exclude Bots and UF	1,616	2.16%
9. Research iDigBio		
Exclude - Smart Bots	1,550	1.83%
Exclude Bots and UF	1,269	1.70%
10. Digitization Resources - iDigBio		

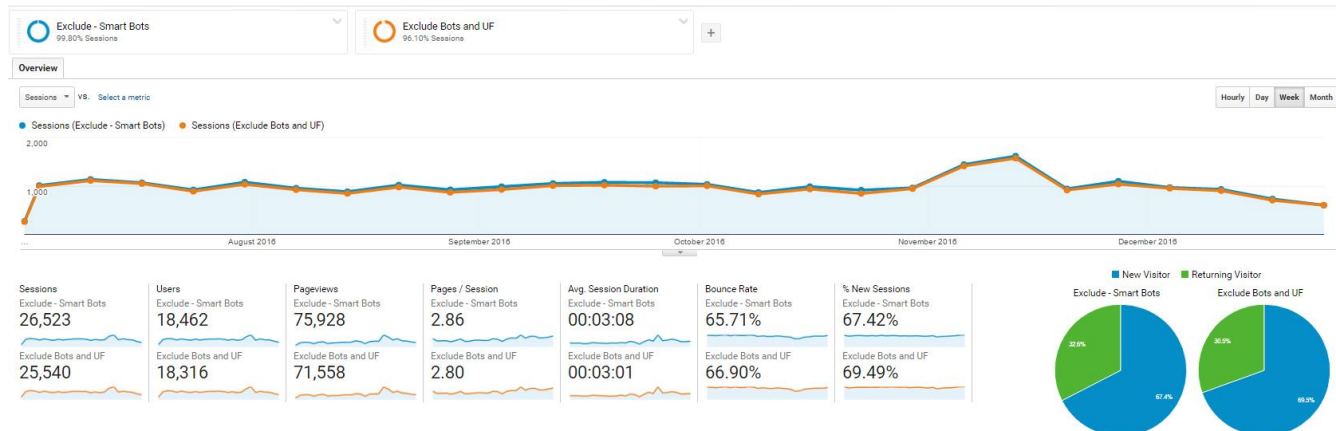
Wiki Statistics

Page statistics	
Content pages	201
Pages (All pages in the wiki, including talk pages, redirects, etc.)	1,980
Uploaded files	1,482
Edit statistics	
Page edits since iDigBio was set up	23,512
Average edits per page	11.87
User statistics	
Registered users	380
Active users (list of members) (Users who have performed an action in the last 30 days)	12
Bots (list of members)	0
Administrators (list of members)	14
Bureaucrats (list of members)	16
Manual_confirmed (list of members)	1
View statistics	
Views total (Views to non-existing pages and special pages are not included)	2,043,194
Views per edit	86.90
Most viewed pages	
Talk:iDigBio External Advisory Board	126,379
iDigBio Working Groups	113,682
Digitization Resources	66,536
GWG Train the Trainers Workshop	48,055
Wiki Home	35,238
iDigBio Workshops	31,558
GWG Second Train the Trainers Workshop	26,928
Data Ingestion Report	24,808
Dried Insect Digitization Workshop	23,391
Georeferencing	22,843

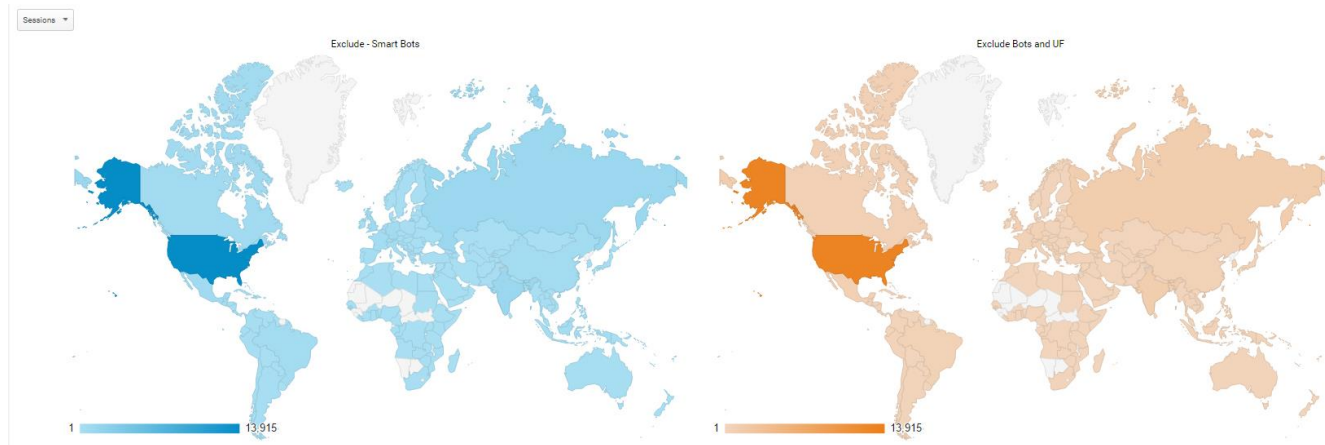
Popular Pages

1. [iDigBio Working Groups](#) (113,682 views)
2. [Digitization Resources](#) (66,536 views)
3. [GWG Train the Trainers Workshop](#) (48,055 views)
4. [Wiki Home](#) (35,238 views)
5. [iDigBio Workshops](#) (31,558 views)
6. [GWG Second Train the Trainers Workshop](#) (26,928 views)
7. [Data Ingestion Report](#) (24,808 views)
8. [Dried Insect Digitization Workshop](#) (23,391 views)
9. [Georeferencing](#) (22,843 views)
10. [2013 AOCR Hackathon Wiki](#) (18,937 views)
11. [iDigBio External Advisory Board](#) (17,962 views)
12. [Drawer Imaging Group](#) (15,336 views)
13. [Data Ingestion Guidance](#) (15,024 views)
14. [TCNs](#) (13,857 views)
15. [Paleo Digitization Workshop](#) (13,451 views)
16. [Paleo Digitization Working Group](#) (13,208 views)
17. [iDigBio Augmenting OCR Workshop](#) (12,307 views)
18. [Data Carpentry](#) (11,443 views)
19. [iDigBio API](#) (11,026 views)
20. [Glossary of Terms](#) (10,987 views)
21. [TCN Resources](#) (10,977 views)
22. [Fluid Preserved Invertebrate Imaging](#) (10,868 views)
23. [Transcription Hackathon](#) (10,764 views)
24. [OCR Resources](#) (10,361 views)
25. [Biodiversity Informatics Management](#) (10,349 views)

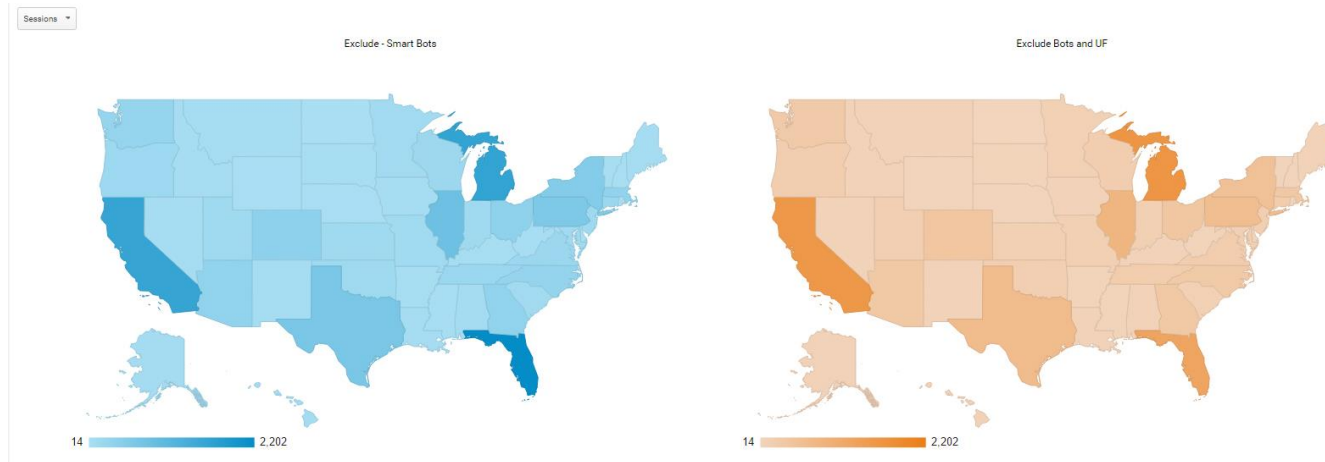
PORTAL USAGE STATISTICS (7/1/2016 THRU 12/31/2016)



Total Portal Visits by Country



Total Portal Visits by State



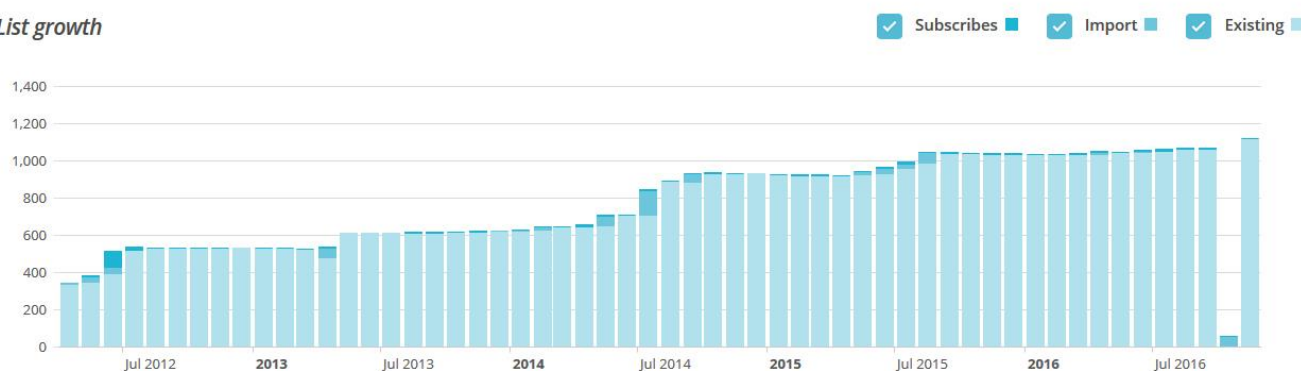
eNEWSLETTER STATISTICS (7/1/2016 THRU 12/31/2016)

eNewsletter Campaigns

Title	Send Date	Total Recipients	Successful Deliveries	Total Bounces	Times Forwarded	Unique Opens	Open Rate (%)	Total Opens	Unique Clicks	Click Rate (%)	Total Clicks	Unsubscribes
iDigBio e-newsletter July 2016	7/18/2016	1,066	1,063	3	0	253	23.8	659	60	5.6	111	2
iDigBio e-newsletter August 2016	8/19/2016	1,070	1,065	5	0	251	23.6	658	75	7.0	141	1
iDigBio e-newsletter September 2016	9/21/2016	1,070	1,066	4	0	270	25.3	702	95	8.9	170	1
iDigBio e-newsletter October 2016	10/21/2016	1,128	1,119	9	1	282	25.2	657	96	7.7	164	1
iDigBio e-newsletter November 2016	11/21/2015	1,168	1,163	5	0	293	25.2	630	84	7.2	166	2
iDigBio e-newsletter December 2016	12/21/2016	1,165	1,154	11	1	364	31.5	690	39	3.4	69	0

eNewsletter List Subscriber Statistics

List growth



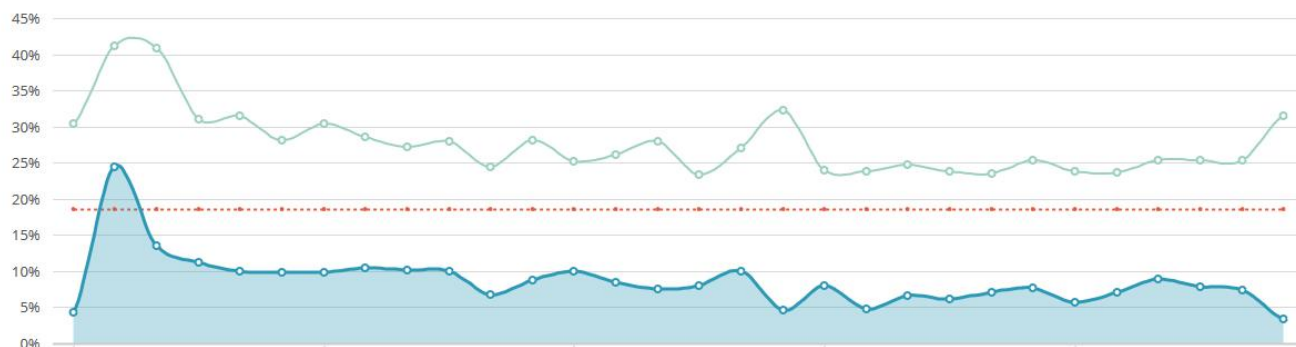
eNewsletter Open Rate & Click Rate Statistics

Overview

Revenue

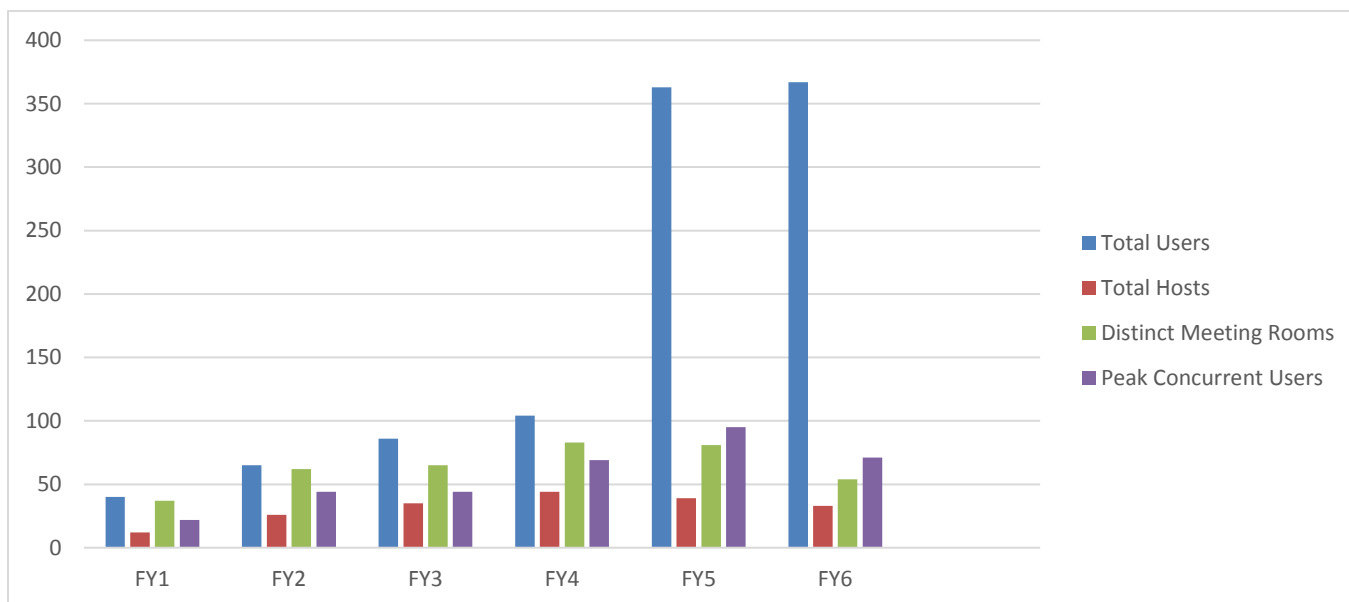
iDigBio e-Newsletter list ▼

Open rate Click rate Industry avg. open rate



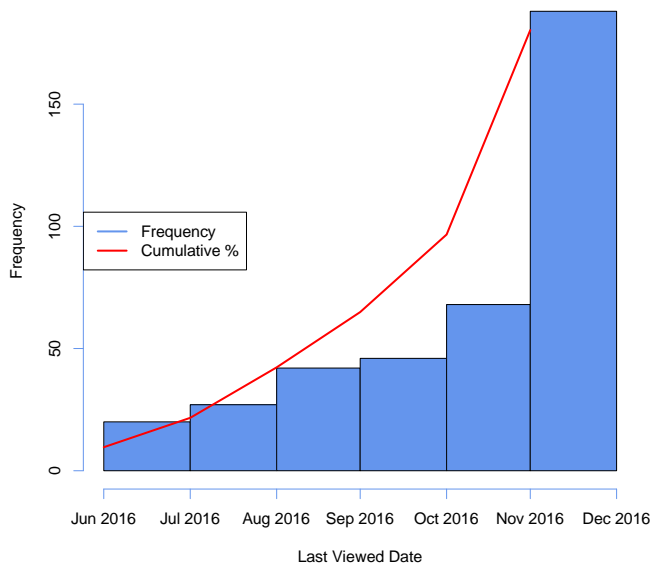
ADOBE CONNECT USAGE STATISTICS (7/1/2011 THRU 12/31/2016)

Metric	FY1	FY2	FY3	FY4	FY5	FY6
Total Users	40	65	86	104	363	367
Total Hosts	12	26	35	44	39	33
Distinct Meeting Rooms	37	62	65	83	81	54
Total Meeting Hours	768	4,429	5,932	8,273	5,744	3,592
Total Host Hours	304	2,182	2,268	2,711	1,897	1,044
Peak Concurrent Users	22	44	44	69	95	71
System Storage Consumption (GB)	0.8	2.6	5.2	52.6	12.3	18.1
Meeting Storage Consumption (GB)	6.8	8.1	484.3	27.8	32.5	5.8

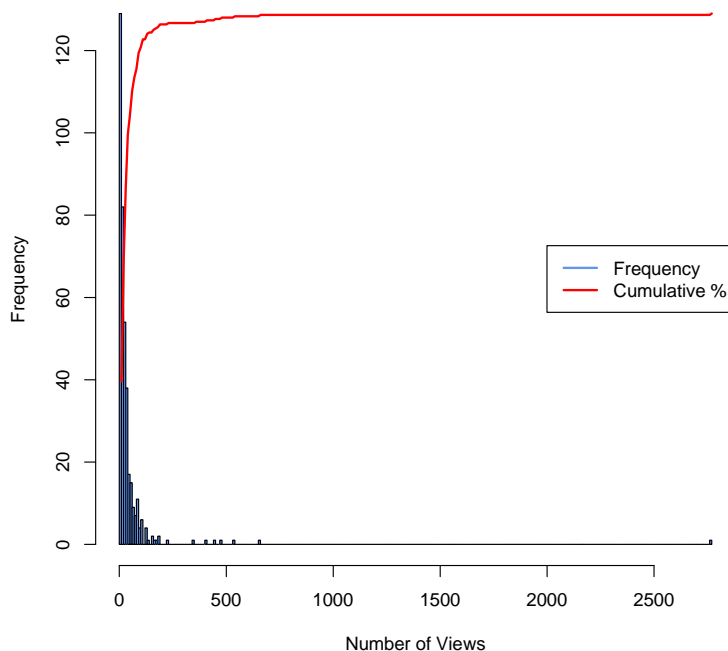


ADOBE CONNECT RECORDING STATISTICS (7/1/2016 THRU 12/31/2016)

Histogram of Adobe Connect Recordings
Last Viewed Date



Histogram of Adobe Connect Recording Views



Top Adobe Connect Recordings

Recording Name	Number of Views
----------------	-----------------

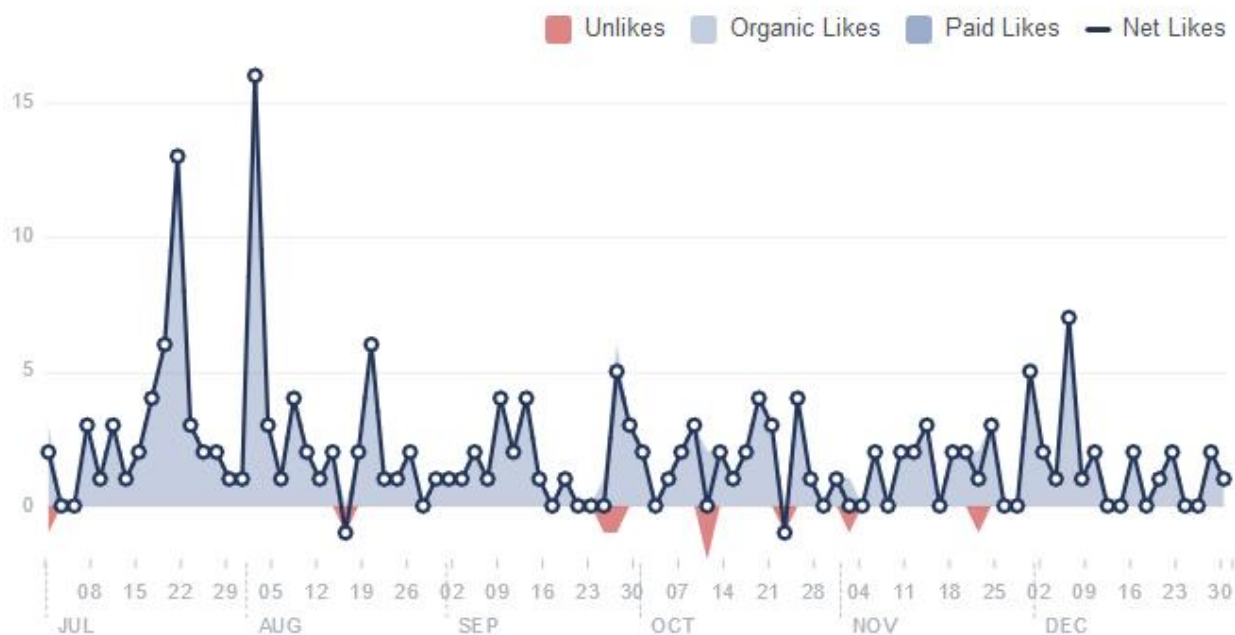
Education and Outreach Webinar Series: The BrainScoop	2,762
Symbiota: SCAN Bugs in my checklist	654
Education and Outreach Webinar Series: iPlant's DNA Subway	539
Webinar - Overview of GBIF IPT Installation and Set up	477
Symbiota: Skeletal Data Entry	443
ECN Annual Meeting 2015	409
FOSSIL Project Webinar Series 1 - Fossil Collecting: Where, When, How	350
FOSSIL Project Webinar Series 1 - Field Notes 101	224
TDWG Pre-Conference Training 2015: Introduction to Georeferencing	183
SWG - Coll. Management Features - EL (MiCC)	182
Webinar: Checklist Management - 2014-09-03	173
Roger Burkhalter Standard Views	161
Mini-webinar: Field Research - 2015-03-09_p1	154
Digitizing Paleo Collections Workshop (September 24 2013, A.M. Session 1)	153
Paleo Imaging Workshop (May 1, 2014)	131

SOCIAL MEDIA STATISTICS (7/1/2016 THRU 12/31/2016)

Facebook

Net Likes

Net likes shows the number of new likes minus the number of unlikes.

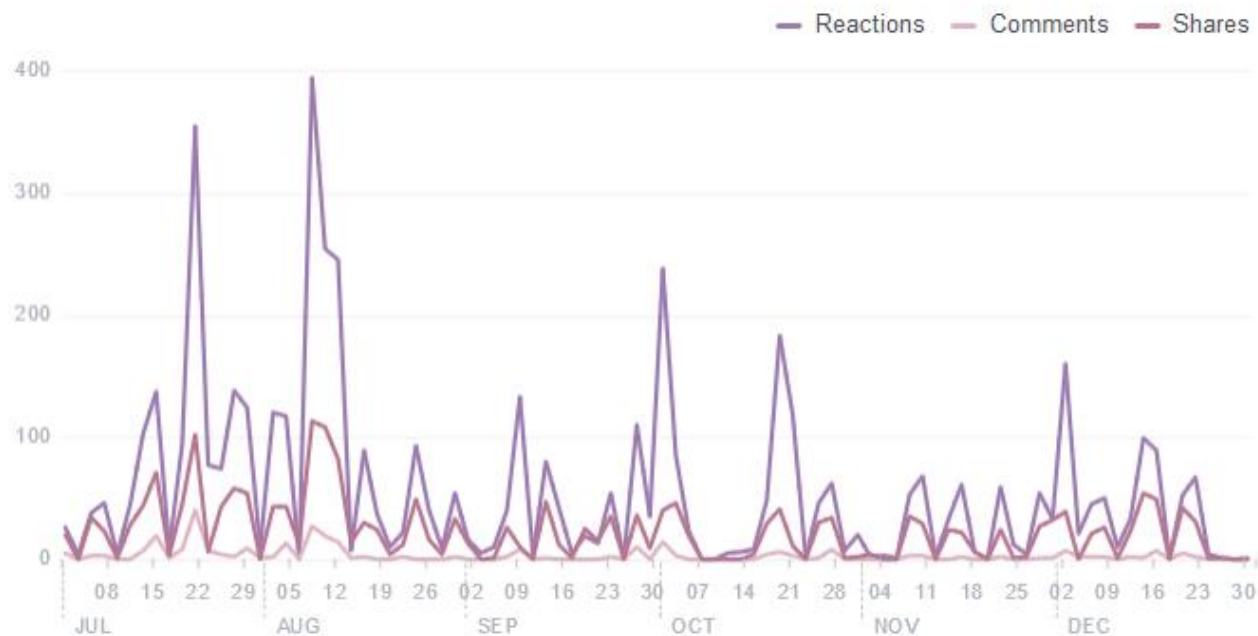


Total Page Likes as of Today: 1,810



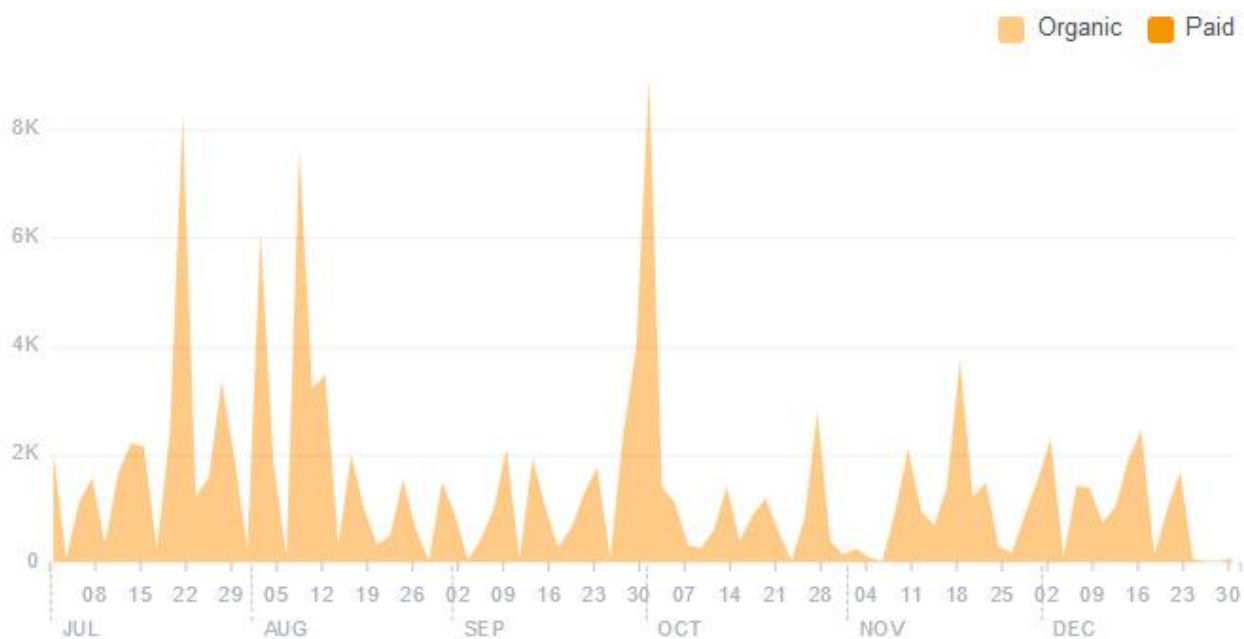
Reactions, Comments, and Shares

These actions will help you reach more people.



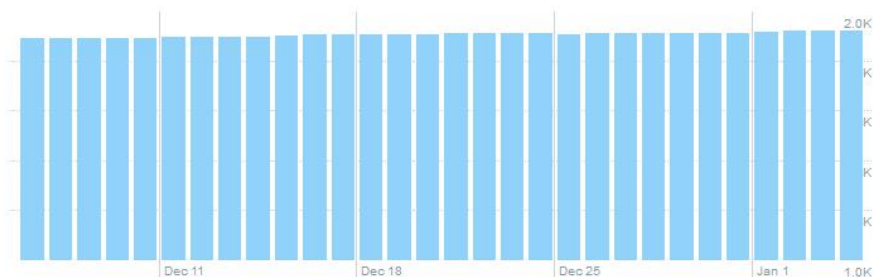
Post Reach

The number of people your posts were served to.



Twitter

OVERVIEW	DEMOGRAPHICS	LIFESTYLE	CONSUMER BEHAVIOR	MOBILE FOOTPRINT
Top interest Science news	Top language English	Top lifestyle type Online buyers	Top buying style Premium brands	Top wireless carrier AT&T

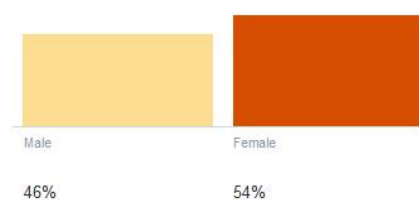


Your current follower audience size is 1,922
That's 31 more than the same time 30 days ago. You've gained around 1 new followers per day

Interests

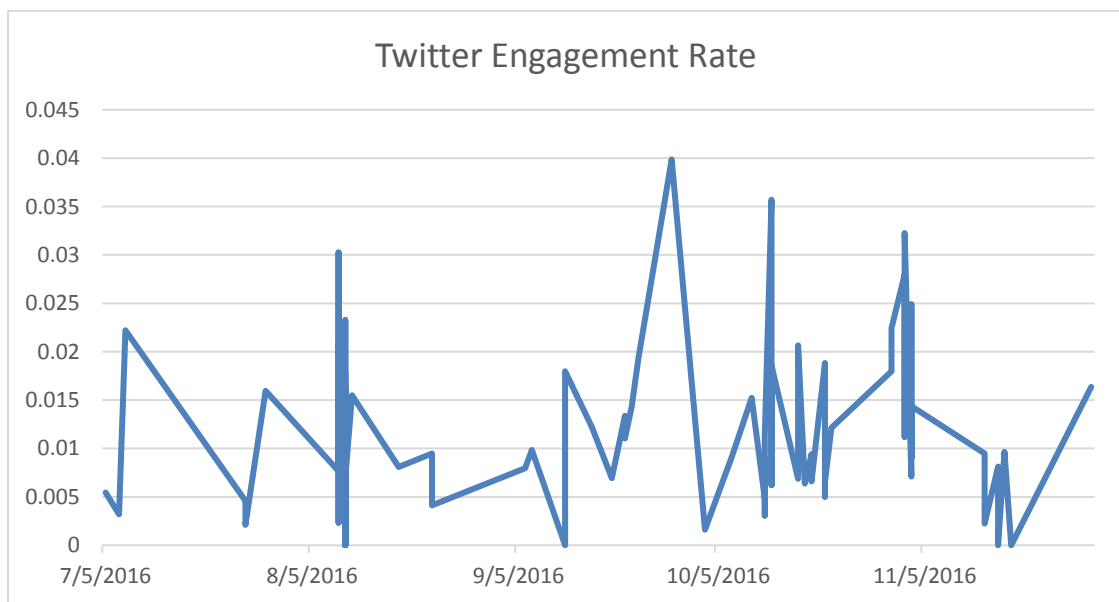
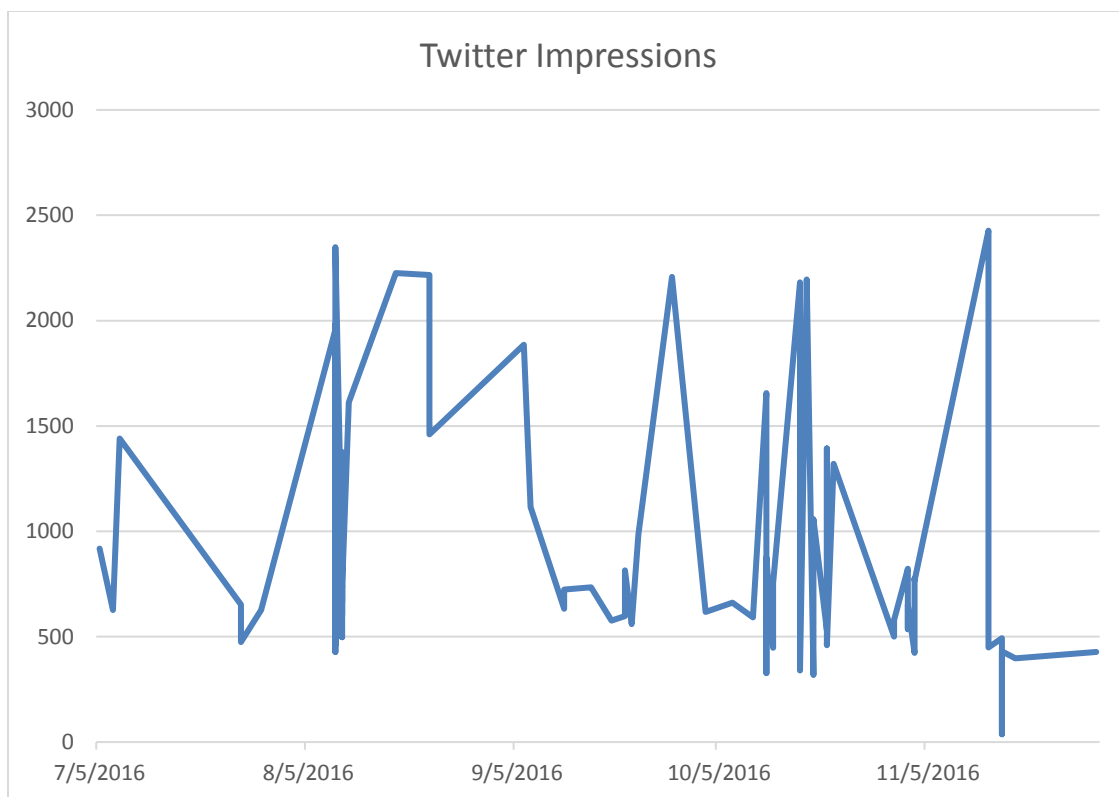
Interest name	% of audience
Science news	93%
Biology	89%
Documentary	71%
Tech news	70%
Books news and general info	67%
Physics	66%
Business and news	63%
Space and astronomy	62%
Politics and current events	61%
Technology	57%

Gender



Household income categories

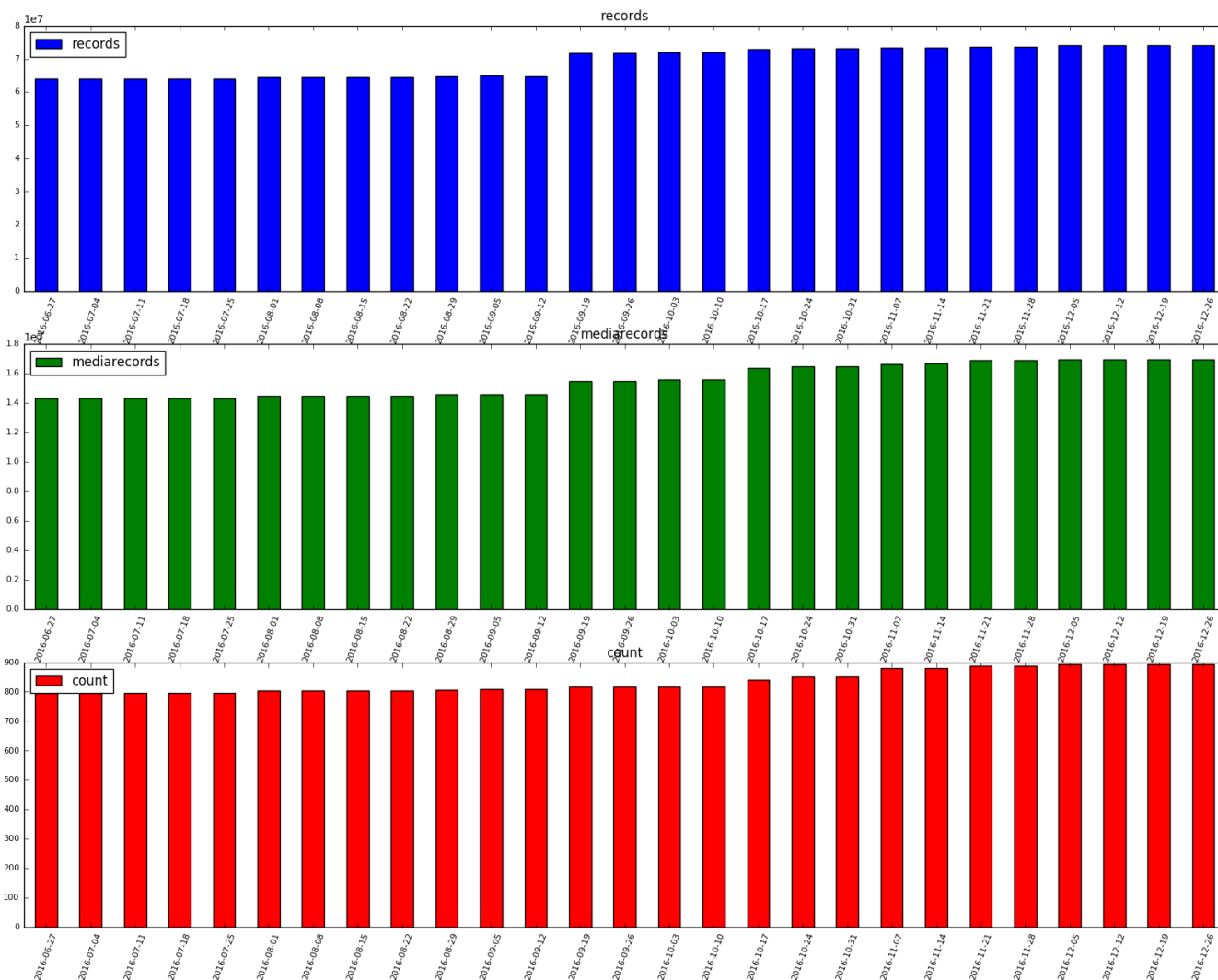
Income category	% of audience
\$75,000 - \$99,999	17%
\$100,000 - \$124,999	12%
\$150,000 - \$199,999	12%
\$125,000 - \$149,999	9%



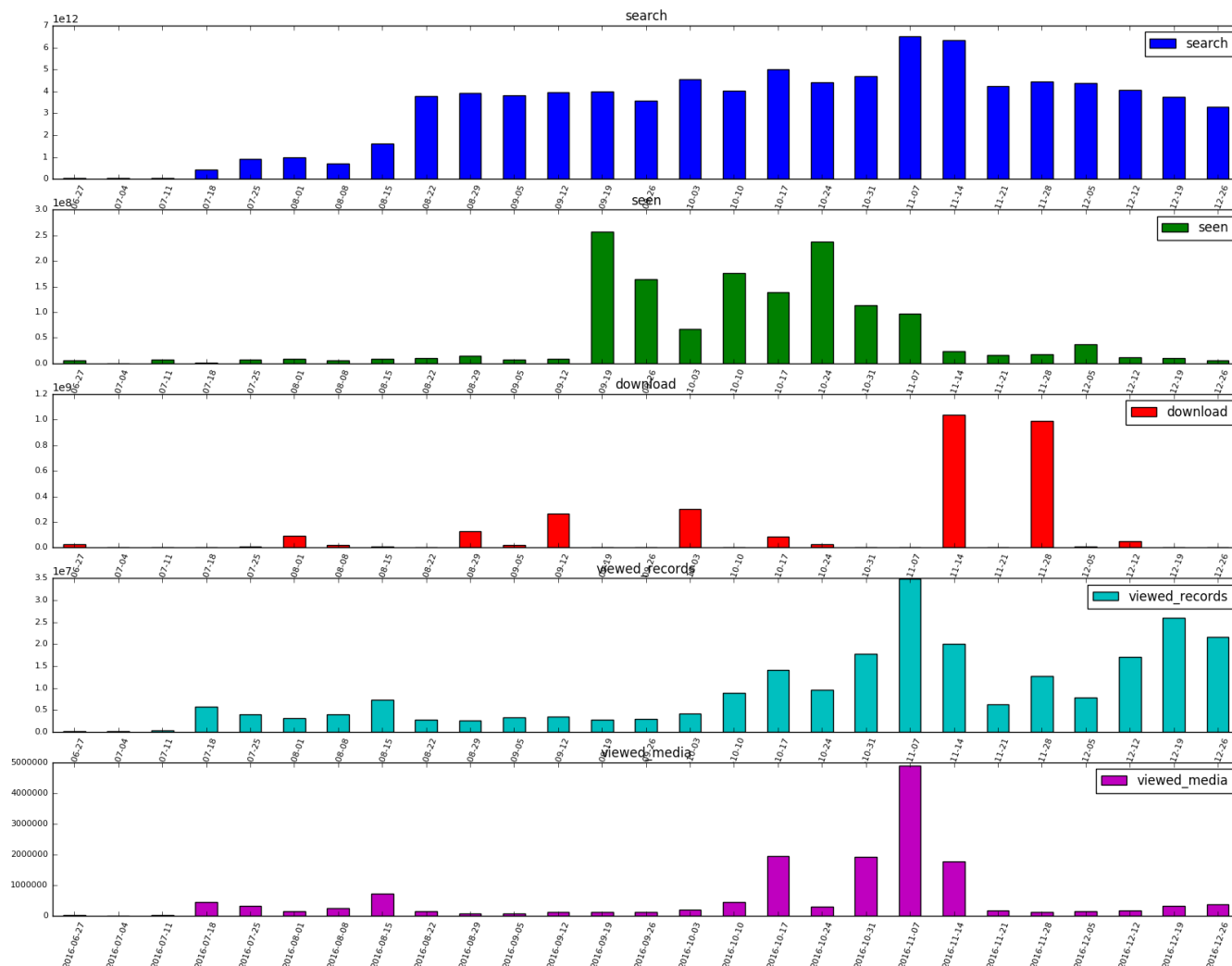
Vimeo



DATA INGESTION STATISTICS (7/1/2016 THRU 12/31/2016)



PORTAL DATA USE STATISTICS (7/1/2016 THRU 12/31/2016)



Strategic Plan for Integrated Digitized Biocollections (iDigBio)

Prepared By:

David Jennings, iDigBio Project Manager

Dr. Larry Page, iDigBio Project Director

Dr. Bruce MacFadden, iDigBio Principal Investigator – Education & Outreach

Dr. Pam Soltis, iDigBio Principal Investigator – Serving the Research Community

Dr. Greg Riccardi, iDigBio Principal Investigator – Digitization

Dr. Jose Fortes, iDigBio Principal Investigator – Cyberinfrastructure

Revision Date:

6/1/2015

Table of Contents

1. Project Overview	3
1.1 Vision	3
1.2 Mission	3
1.3 Horizon	3
2. Project Organization	5
2.1 Project Oversight	5
2.1.1 Executive Committee	5
2.1.2 Steering Committee	5
2.1.3 Internal Advisory Committee	5
2.1.4 External Advisory Board	5
3. Strategies and Objectives	6
3.1 Project Administration & Management	6
3.2 Education and Outreach	6
3.3 Serving the Research Community	6
3.4 Digitization	7
3.5 Cyberinfrastructure	8
4. Assessment and Evaluation	9
5. Environmental Scan	10
6. Sustainability	11
6.1 Maintaining Strengths	11
6.2 Addressing Weaknesses	11
6.3 Capitalizing on Opportunities	13
6.4 Mitigating Threats	13

1. Project Overview

Integrated Digitized Biocollections ([iDigBio](#)) is the national coordinating center for the Advancing Digitization of Biodiversity Collections ([ADBC](#)) program funded by the U.S. National Science Foundation ([NSF](#)). iDigBio is enabling digitization of data and media for millions of biodiversity specimens from U.S. collections and is integrating those data to make them available online for the research community, government agencies, students, educators, citizen scientists, and the general public to promote understanding of biodiversity and societal consequences of environmental issues.

1.1 Vision

The vision for ADBC is a permanent repository of digitized information from all U.S. biodiversity collections that leads to new discoveries through research and a better understanding and appreciation of biodiversity through improved outreach, which then leads to improved environmental and economic policies. Creation of the permanent digitized repository is occurring in four stages:

1. An initial stage where the effort to digitize U.S. biodiversity collections is catalyzed by funding from NSF and enabled by iDigBio activities that foster collaborations, identify priorities, demonstrate the value of biodiversity and collections, and generate information on best practices related to standards, workflows, and data management.
2. An intermediate stage where digitization at Thematic Collections Networks ([TCNs](#)), Partners to Existing Networks ([PENs](#)), and other participating institutions/networks improves methods and strategies and demonstrates the scientific and societal benefits of validated and readily accessible data.
3. A third stage in which the vision for ADBC is realized through participation by all U.S. institutions with biodiversity collections.
4. A fourth stage in which digitization is a routine and sustained practice in all institutions with biodiversity collections, and the national database is easily accessible as an up-to-date source of information on biodiversity.

1.2 Mission

The mission of iDigBio is to develop a national infrastructure that supports the vision of ADBC by overseeing implementation of standards and best practices for digitization; building and deploying a customized cloud computing environment for collections; recruiting and training personnel, including underserved groups; engaging the research community, collections community, citizen scientists, and the general public through outreach activities; and planning for long-term sustainability of the national digitization effort.

1.3 Horizon

iDigBio is enabling digitization of data from all U.S. biodiversity collections and is integrating those data to make them broadly available and useful with shared standards and formats. Ultimately, ADBC is

furthering the discovery and understanding of biodiversity, and iDigBio is engaging the research and collections communities in a spirit of collaboration in an effort to open biodiversity research collections to new downstream user communities.

iDigBio involves the development of a permanent and powerful cloud computing infrastructure to link biodiversity data from collections across the U.S. into a single unified web interface, overcoming the “data silos” that exist across the country. Search and analytical tools enable users to mine diverse data including taxonomy, geographic location, 2- and 3-dimensional images, vocalizations, and molecular resources tied to specimens in collections. These data promote integrative biodiversity research on living and fossil species and provide an immense resource for agricultural science and land use management, human health, and assessing the impacts of climate change, invasive species, and other natural resource management issues.

Key partners in this effort are the Thematic Collections Networks ([TCNs](#)), which form a national grid of institutions that are digitizing specimens and associated resources. Integration with the greater community of biocollections resources, tools and organizations is critical to accomplishing the grand challenge of digitizing and integrating data from all U.S. collections, large and small. For more details regarding the larger community that encapsulates ADBC, please refer to the Network Integrated Biocollections Alliance ([NIBA](#)) [strategic plan](#) and [implementation plan](#).

2. Project Organization

The iDigBio project is organized into five domains to achieve its goals: (1) Project Administration & Management, (2) Education & Outreach, (3) Serving the Research Community, (4) Digitization, and (5) Cyberinfrastructure. The [Principal Investigators](#) (PIs) provide leadership within each domain, and the Project Director and Project Manager provide overall leadership for the project.

2.1 Project Oversight

The project incorporates several oversight committees to provide governance, advice, and leadership concerning goals, strategies, implementation, activities, and progress. The Project Director, in consultation with the Project Manager, Executive Committee, and Steering Committee, makes project decisions that affect scope, budget, and/or risk. The Project Manager effects day-to-day planning, execution, and decision-making.

2.1.1 Executive Committee

The Executive Committee (EC) is composed of the iDigBio Project Director, Principal Investigators, Project Manager, and key Senior Personnel. The Executive Committee is responsible for overall project management, general oversight of iDigBio activities, managing conflicts of interest, and implementation of the strategic plan, including assurance that the digitization, research, training, and outreach missions of iDigBio are integrated and accomplished.

2.1.2 Steering Committee

The Steering Committee (SC) is composed of the iDigBio Project Director, Principal Investigators, Project Manager, and Senior Personnel. The Steering Committee is responsible for reviewing progress and coordinating activities in digitization, research, training, and outreach at iDigBio, and for advising others in iDigBio on overall resource allocation, strategic directions, and management policies.

2.1.3 Internal Advisory Committee

The Internal Advisory Committee (IAC) is composed of the Project Manager, Biodiversity Informatics Manager, and representatives from the Thematic Collections Networks (TCNs) and Partners to Existing Networks (PENs). The IAC meets regularly to report on progress in digitization efforts, share and identify best practices and standards, identify gaps in digitization areas and technology, and enhance training efforts.

2.1.4 External Advisory Board

The External Advisory Board (EAB) is composed of members selected by iDigBio and approved by the NSF program officer. The EAB meets annually and is responsible for advising iDigBio on its strategic directions, management policies, and activities, including progress and integration of digitization projects, research, training, and outreach activities among all funded institutions.

3. Strategies and Objectives

3.1 Project Administration & Management

iDigBio Project Administration & Management decisions and activities are executed under the leadership of Dr. Larry Page, iDigBio Project Director, who is based at the University of Florida's Florida Museum of Natural History ([FLMNH](#)). The Project Administration & Management domain provides administrative and logistical support to the other project domains and ensures that activities are coordinated within and across iDigBio domains and with collaborative organizations as appropriate. The key objectives of the Project Administration & Management domain are:

- a. Coordinate project oversight committees and other key meetings;
- b. Plan for long-term sustainability, including participation with NIBA;
- c. Comply with NSF cooperative agreement and reporting requirements;
- d. Build and maintain strategic partnerships;
- e. Plan and monitor project budget and expenditures;
- f. Implement a structure to measure progress against goals;
- g. Manage internal and external communications, including broad dissemination of outcomes;
- h. Develop and maintain collaboration and communication capabilities; and
- i. Serve as the central resource for ADBC and the collections community, including promotion of cohesion and interconnectivity.

3.2 Education and Outreach

iDigBio Education and Outreach (E&O) decisions and activities are executed under the leadership of Dr. Bruce MacFadden, iDigBio co-PI, based at who is based at the University of Florida's Florida Museum of Natural History ([FLMNH](#)). iDigBio E&O activities are focused on digitization curricula development, stakeholder identification, E&O materials and protocols, and public speaking engagements, which are recorded and published for national impact. The key objectives of the iDigBio E&O domain are:

- a. Foster project awareness within the scientific and collections communities;
- b. Engage the public and collections community through resources and opportunities that highlight the importance of biodiversity collections and digitization;
- c. Identify and assess the needs of target audiences, downstream user groups, and other stakeholders, and execute outreach activities to meet those needs;
- d. Identify and assess the needs of downstream partners and stakeholders, and execute outreach activities to meet those needs;
- e. Develop outreach resources related to digitization and biodiversity; and
- f. Measure and track the impacts and outcomes of outreach efforts.

3.3 Serving the Research Community

iDigBio decisions and activities related to serving the research community are executed under the leadership of Dr. Pamela Soltis, iDigBio co-PI, based at who is based at the University of Florida's Florida Museum of Natural History ([FLMNH](#)). The Research domain leverages existing relationships to deliver

branding and messaging related to iDigBio to foster collections/research community adoption of iDigBio's services, infrastructure, tools, resources, and data including: [specimen portal](#), [website](#), [Wiki](#), [Listservs](#), [Adobe Connect](#) teleconferencing, [Workshop](#) coordination/funding, [Working Group](#) coordination/funding, [Appliance](#) development, and Hosting of services. In addition, the Research domain highlights developments and opportunities for research using specimen data alone or in conjunction with other types of data. The key objectives of the Research domain are:

- a. Engage the research community to promote community adoption of iDigBio services, infrastructure, tools, resources, and data;
- b. Promote and facilitate both traditional and novel uses of specimen data;
- c. Produce use cases of research applications of specimen data, provide them to the Cyberinfrastructure team, and help validate the effectiveness of implementation;
- d. Seek opportunities for integration of iDigBio specimen data and API services with key data and research services from other projects and organizations;
- e. Identify strategic partners in the research and collections community and develop synergistic relationships with those partners; and
- f. Measure and track the outcomes of research use efforts.

3.4 Digitization

iDigBio decisions and activities related to development and optimization of digitization workflows/processes, digitization documentation, and efforts to share/improve digitization tools within the collections community are executed under the leadership of Dr. Greg Riccardi, iDigBio co-PI, based at Florida State University's Institute for Digital Information and Scientific Communication ([iDigInfo](#)). iDigBio digitization experts work with the community to understand gaps in workflows, processes, practices, and tools that prevent effective and efficient specimen digitization. These digitization-related activities are catalyzed by site visits, virtual and on-site workshops, working groups, contact with tool developers, and user contact to foster a high-degree of community involvement. The observations and subsequent analyses from these activities yield documentation, papers, training materials, and presentations that promote effective digitization practices and workflows, including qualitative and quantitative measures of success. The key objectives of the Digitization domain are:

- a. Engage the collections community to market and build interest in adopting iDigBio services, infrastructure, tools, resources, and data;
- b. Establish minimum information standards and data fitness for use parameters;
- c. Develop, optimize, and disseminate digitization workflows;
- d. Conduct digitization training and produce online training materials;
- e. Increase awareness and utility of digitization tools and resources that can improve efficiency and scalability of digitization efforts;
- f. Evaluate, document, and publish analyses related to digitization hardware and software tools; and
- g. Identify gaps, bottlenecks, and challenges in digitization efforts, standards, and best practices.

3.5 Cyberinfrastructure

iDigBio Cyberinfrastructure decisions and activities are executed under the leadership of Dr. José Fortes, iDigBio co-PI, based at the University of Florida's Advanced Computing and Information Systems ([ACIS](#)) laboratory. Cyberinfrastructure decisions are informed by the experience and expertise of ACIS personnel, key stakeholders in iDigBio, biodiversity community input, and information technology community input. The Cyberinfrastructure implementation process follows a formula that balances strategic planning with the agility to meet new challenges, short-term project needs, and enhanced/clarified specifications. The cyberinfrastructure team releases new features and upgrades to the specimen portal on a semi-annual basis. The key objectives of the Cyberinfrastructure domain are:

- a. Research, implement, and maintain a scalable cloud infrastructure for text (data/metadata) and object (media) storage;
- b. Implement infrastructure to enable hosting of web services and/or websites for strategic partners, such as TCNs;
- c. Deploy iDigBio appliances and services via multiple channels (e.g., web services, locally-run virtual machines, infrastructure-as-a-service cloud implementations) to enhance, simplify, and/or improve activities of data providers and data consumers;
- d. Implement a comprehensive authentication and access control system to enable data tracking and a cohesive user experience among iDigBio systems;
- e. Develop, implement, and maintain iDigBio APIs to access text and media data stores;
- f. Develop, implement, and maintain a Graphical User Interface (GUI) to provide end-users, including data contributors and data consumers, access to search, visualize, and download text and media data from the cloud infrastructure (i.e., the [Portal](#));
- g. Integrate iDigBio services and user portals with key strategic partners and other collaborators;
- h. Secure infrastructure resources to maintain adequate performance and capacity;
- i. Serve as a central site for aggregation of digitized collections data; and
- j. Plan for the management and long-term preservation of iDigBio's digital data.

4. Assessment and Evaluation

iDigBio's Project Evaluator, a professional with training in social science and expertise in evaluation, oversees assessment and evaluation of iDigBio-supported activities. The evaluator employs a multi-method approach including participant observation, surveys, interviews, analysis of project records, and tracking of data use in publications. The evaluator participates as an observer in annual Summits and Retreats, meetings of the Steering Committee, Internal Advisory Committee, External Advisory Board, periodically in IT meetings, and as an advisor in Core Team meetings and Working Groups (e.g., E & O and Georeferencing). The evaluator collaborates with iDigBio personnel and partners (e.g., Data Carpentry, Small Collections Network, and WeDigBio) on the development of instruments for needs assessment and evaluation and provides summary reports. The evaluator also conducts evaluations of project management, leadership, and communication.

The purposes of iDigBio's assessment and evaluation efforts are to:

- a. Document progress against goals in the five domains: (a) project management, (b) education and outreach, (c) serving the research community, (d) digitization, and (e) cyberinfrastructure;
- b. Perform assessments to determine the needs of iDigBio's various constituencies, stakeholders, and audiences in order to develop effective tools, services, training, outreach, and communication;
- c. Conduct formative evaluation of iDigBio efforts spanning the five domains listed above; and
- d. Measure the impact of iDigBio on the national effort to digitize natural history collections.

Key metrics used to evaluate the impact of iDigBio include (but are not limited to):

- a. Effectiveness in building cohesion among the digitization community;
- b. Awareness of the collections digitization effort among the collections and scientific community;
- c. Engagement of the public and collections community in digitization efforts related to biodiversity;
- d. Engagement with resources related to digitization and biodiversity;
- e. Understanding the value of digitization of biodiversity collections;
- f. Use of iDigBio services, infrastructure, tools, resources, and data by the research community;
- g. Rates of digitization;
- h. Efficiency of digitization;
- i. Use of digitization best practices; and
- j. Use of the specimen portal by down-stream users.

5. Environmental Scan

The following is a summary of the major strengths, weaknesses, opportunities, and threats that have been highlighted by the External Advisory Board, NSF Site Review team, and surveys of the collections community:

	POSITIVE	NEGATIVE
INTERNAL	<p><u>Strengths</u> – iDigBio has...</p> <ul style="list-style-type: none"> • Successfully engaged with the collections community, particularly through workshops and working groups; • Developed a vast repository of resources and information; • Developed and optimized digitization protocols and workflows; • Actively reduced barriers to digitization; • Demonstrated strong leadership and management; and • Fostered partnerships, networks, and collaborations within the community and has developed significant synergy with TCNs. 	<p><u>Weaknesses</u> – iDigBio has...</p> <ul style="list-style-type: none"> • Non-intuitive website navigation and inconsistent arrangement of resources; • Minimal data in specimen portal; • Inconsistent internal communication; • Insufficient involvement of smaller collections and institutions; • Insufficient effort to promote diversity and include under-represented groups; • Lack of explicitly defined of project leadership evaluation plans; and • Poor coordination of outreach activities across ADBC.
EXTERNAL	<p><u>Opportunities</u> – iDigBio should...</p> <ul style="list-style-type: none"> • Highlight research uses of data; • Develop digitization standards and encourage compliance; • Advocate for greater digitization funding and formation of cooperative groups across institutions and projects; • Coordinate efforts across institutions, disciplines, and agencies to bridge initiatives and facilitate networking; • Integrate directly with collection management software and tools; • Capture “dark data” from both large and small institutions; and • Collaborate with key existing projects, such as Global Biodiversity Information Facility (GBIF) and Biodiversity Information Serving Our Nation (BISON). 	<p><u>Threats</u> – iDigBio should...</p> <ul style="list-style-type: none"> • Develop a strategy for financial sustainability to reduce dependence on federal funding; • Develop a strategy for long-term storage and maintenance of its cyberinfrastructure, metadata and images; • Be mindful of its extremely broad project scope, which includes a vast array of collaborations and partnerships; and • Be mindful that iDigBio depends on a number of external projects, tools, and initiatives whose long-term sustainability could be questionable.

6. Sustainability

To ensure long-term sustainability of the national digitization effort, iDigBio is implementing strategies to maintain its strengths, address its weaknesses, capitalize on available opportunities, and mitigate potential threats.

6.1 Maintaining Strengths

The Integrated Digitized Biocollections ([iDigBio](#)) project has made significant progress since the initiation of funding in 2011. iDigBio's innovations include both sociological and technological accomplishments with wide-ranging benefits to the collections community.

First, iDigBio has established successful **communication** between the Information Technology (IT) and biodiversity collections communities. Having bridged this “cultural” barrier, iDigBio personnel are working together to identify challenges and to design appropriate solutions. This communication extends beyond the personnel specifically working on iDigBio to other partners, such as the TCNs, which allows for collaboration, synergy, and effective training throughout the community.

Second, perhaps the most successful innovation of iDigBio is the series of **training workshops** that have been organized and sponsored by iDigBio personnel. These workshops have delivered effective training on digitization-related methods and practices, as well as on other topics contributed by the workshop participants. These workshops and training materials, publically available on the [iDigBio Wiki](#), have provided a wealth of new resources and have secured iDigBio's leadership role in **workforce development** within the collections community. During its first four years, iDigBio has sponsored 58 workshops, summits, symposia, and other events and that have been attended by 2,094 participants (1,160 of which are unique) from 485 unique institutions to its. In addition, iDigBio has held 31 webinars that have been attended by 1,103 participants.

Finally, the forthcoming **availability of massive amounts of specimen data** has energized the collections community about the use of specimen data for a variety of big research questions that have been intractable to this point. This renewed energy within the community has fostered iDigBio to produce significant innovations in IT design and implementation, including:

- Creating the practice of introducing identifiers in the data stream to enable data linking
- Development of emerging data models for ingestion and integration of data sets from diverse collections
- Cloud architecture for data storage, retrieval, and management
- Open schema infrastructure to offer flexibility and agility in handling an evolving data model
- Appliance framework to respond to the needs of biocollections informatics

6.2 Addressing Weaknesses

iDigBio has recognized that its website and specimen portal are the primary interfaces between the project and its audiences. Completely revised versions of the iDigBio website and specimen portal were released to the community in December 2013, and minor changes have continually been made since

then. The website was redesigned to focus on making it easier to understand and to use and to be more approachable to a lay visitor. The specimen portal and APIs were redesigned to create the foundation for a system that will serve the community for years to come, including user interface improvements and improved stability and flexibility.

Availability of specimen data in the portal is a top priority, and iDigBio has worked diligently to improve the efficiency of the data mobilization and ingestion process. iDigBio has the goal of ingesting data sets within 2 weeks of data mobilization completion. Data mobilization is dependent on the provider, although iDigBio staff continually track the process from initiation to closure using [Redmine](#). From December 2013 thru May 2015, the number of collections in the portal has increased from 121 to 438 (↑262%), the number of specimen records has increased from 4,410,237 to 28,036,830 (↑536%), and the number of media records has increased from 1,005,679 to 4,511,503 (↑349%).

iDigBio senior personnel have acknowledged that communication, both within iDigBio and with our clients, could be improved. iDigBio believes that less-than-ideal communication is the result of having a small staff charged with multiple tasks who are unable to find the time to report on activities in a meaningful way and in a reasonable timeframe. In response, iDigBio hired a Communications Coordinator who was charged with tracking iDigBio activities and broadly disseminating information about those activities. The Communications Coordinator improved communications by defining categories of participating audiences and then focusing high-impact messages tailored for those audiences using a variety of media outlets. iDigBio personnel are making a concerted effort to write and post concise reports on our website in a timely fashion, and the Communication Coordinator announces their release using Twitter, Facebook, and other more traditional means.

iDigBio has initiated efforts to provide national leadership in developing methods to enhance the cultural and ethnic diversity of participation in the biodiversity sciences. iDigBio is hosting a series of workshops targeted at undergraduate students in underrepresented populations focusing on opportunities for careers and graduate study in environmental biology, biodiversity, ecology, and evolution.

iDigBio has initiated efforts to annually evaluate the project and project leadership. The project is evaluated by the collections community to gauge interest and involvement with iDigBio and to solicit input on iDigBio's progress. The project leadership is evaluated by iDigBio staff and students, giving them an opportunity to share their views regarding their own roles and responsibilities and to solicit input on project leadership and management.

iDigBio is actively recruiting an Education & Outreach Coordinator who will be responsible for coordinating and implementing the Education & Outreach (E&O) activities of iDigBio. In addition, the Education & Outreach Coordinator will be responsible for communicating, coordinating, and networking across ADBC (i.e., iDigBio, the TCNs, and other stakeholders) to promote, encourage, develop, and implement E&O activities that achieve broad reach and high impact.

6.3 Capitalizing on Opportunities

iDigBio is working to highlight research uses of the data in the specimen portal. iDigBio has begun implementing features in the specimen portal that are based directly upon input from biologists and how they anticipate using the data. For example, iDigBio is working to implement advanced mapping features in the specimen portal and is working with [GenBank](#) to create links between specimen data and genetic data. In addition, iDigBio is actively recruiting a Data Management Coordinator to facilitate the use of natural history collections data in addressing big-science questions by integrating tools and services into a computational environment for data integration, analysis, and visualization. The Data Management Coordinator will act as a liaison between museum collections staff/researchers and IT/cyberinfrastructure staff/developers; develop and document use cases for the use of natural history collections data, including translating these into cyberinfrastructure requirements; prioritize requirements for iDigBio's cyberinfrastructure and data products; provide advice on cyberinfrastructure recommendations and direction; provide advice on cyberinfrastructure data standards and linked data; and serve on the iDigBio Steering Committee to provide strategic advice and direction.

iDigBio is continuing to advocate for, coordinate with, and facilitate networking among the collections community through its involvement in professional meetings and societies. iDigBio is also working to continue direct representation in key organizations, such as The Society for the Preservation of Natural History Collections ([SPNHC](#)), Natural Science Collections Alliance ([NSCA](#)), and the American Institute of Biological Sciences ([AIBS](#)).

iDigBio is continuing to support a variety of [Working Groups and Interest Groups](#). iDigBio is allowing these groups to form organically to dynamically address the needs of the collections community. Some of iDigBio's groups are focused on delivery of short-term objectives, while others are tasked with ongoing research, development, and improvement activities.

iDigBio is continuing to develop its partnerships with [Specify](#), [Symbiota](#), and [KE EMu](#) to introduce globally unique identifiers and to create export features specifically tailored to iDigBio. The goal of these partnerships is to streamline the path to data ingestion with iDigBio.

iDigBio has started to mobilize data sets beyond the TCNs by working with the Global Registry of Biorepositories ([GRBio](#)) to make contacts with U.S. institutions outside the TCNs. In addition, iDigBio is advancing its existing partnerships with the Global Biodiversity Information Facility ([GBIF](#)) and Biodiversity Information Serving Our Nation ([BISON](#)).

6.4 Mitigating Threats

iDigBio, in concert with the Network Integrated Biological Alliance (NIBA), is examining and developing business models that create an environment for long-term economic sustainability. Examples of such business models include institutional commitments, cost-recovery, grants, endowments through benefactors, and engaging in public/private partnerships. The goal is to establish a multifaceted funding base for long-term sustainability through community engagement and partnerships with government agencies, private foundations, and international stakeholders.

Building a community of use around iDigBio is paramount to sustaining the infrastructure. As the success of ADBC continues, collections-based data are becoming much more generally available, used, and appreciated, not only for their scientific value, but also for understanding the importance of biodiversity to the economic stability and health of the planet. Coupled with tools from informatics and computer science, collections data are being made available for a host of applications related to biodiversity research, natural resource management, and public policy development. iDigBio has developed a data mobilization and ingestion process to ensure the vast number of collections being harvested are tracked from initiation to closure using [Redmine](#).

iDigBio remains vigilant in managing its scope. To reduce uncertainty and to prevent scope creep, iDigBio has published the [Project Scope](#) along with a series of agreed upon in-scope and out-of-scope activities on the iDigBio website. The iDigBio Steering and Executive Committees regularly review tasks, objectives, and progress to insure they are within scope.

iDigBio is establishing strategic partnerships, including international, with external projects, tools, and initiatives. These strategic partnerships not only promote integration of tools and services into iDigBio but also promote long-term sustainability of both iDigBio. Ultimately, these strategic partnerships will contribute towards a robust comprehensive infrastructure.

Sustainability Plan for Integrated Digitized Biocollections (iDigBio)

Prepared By:

David Jennings, iDigBio Project Manager

Dr. Larry Page, iDigBio Project Director

Dr. Bruce MacFadden, iDigBio Principal Investigator – Education & Outreach

Dr. Pam Soltis, iDigBio Principal Investigator – Serving the Research Community

Dr. Greg Riccardi, iDigBio Principal Investigator – Digitization

Dr. Jose Fortes, iDigBio Principal Investigator – Cyberinfrastructure

Revision Date:

5/31/2015

Table of Contents

Sustainability Plan for iDigBio during and beyond the 10-year funding from NSF	3
Key components of Sustainability	3
1. Institutional Support.....	3
2. Business Plan	4
3. Cost Recovery	5
4. Grants and Contracts.....	5
5. Globalization.....	5
6. Endowments.....	6

Sustainability Plan for iDigBio during and beyond the 10-year funding from NSF

Integrated Digitized Biocollections (iDigBio) at UF serves as the national coordinating center for the NSF-funded 10-year \$100 million program to digitize information in natural history collections housed in U.S. institutions in response to the collections community's *Network Integrated Biocollections Alliance (NIBA) Strategic Plan* (2012). Currently, iDigBio is working with 439 collections in 268 institutions in the U.S. to reach this goal, and those numbers will increase over the next few years.

iDigBio is in the fourth year of a five-year cooperative agreement with NSF, funded at \$13 million, and is preparing a proposal for renewal for another five years with an anticipated budget of \$15 million. During the second five years, iDigBio will develop and transition into a program for sustainability beyond the 10 years of funding from NSF. This will be done through a combination of institutional support from UF and FSU, cost-recovery and revenue-generating mechanisms, and grants that capitalize on digitized data and software tools that facilitate use of biodiversity data. Plans for sustainability of all aspects of the national digitization effort, including iDigBio, will be generated by the NIBA Research Coordinating Network, funded by NSF in 2015 (RCN Award #1441785).

Funding from NSF/ADBC to iDigBio will decrease beginning in Year 7 and will end in Year 10. Funding from UF and FSU in support of iDigBio will remain constant through Year 10 and will then diminish starting in Year 11 to a level that is critical for long-term support of staff and infrastructure. Support from other sources will increase during Years 6-10 and will continue indefinitely.

Key components of Sustainability

1. Institutional Support

Substantial contributions from the University of Florida (UF) and Florida State University (FSU) guarantee permanence of the infrastructure and operation of iDigBio. These commitments, coupled with the resources available through iDigBio, will secure the ability to generate external funding.

A. Support from UF

- Physical facilities including office and meeting space, equipment, and data storage facilities.
- Key staff and operational funding:
 - Currently, UF is providing 2 staff positions—a project assistant and a project director—plus funding for expenses not covered by NSF.
 - For the renewal, the UF Office of Research and the Provost are providing critical resources to ensure long-term sustainability: three staff positions—a project director, a program assistant, and a data portal manager—plus funding for expenses not covered by NSF. These resources will be provided during the 5-year renewal period, but the spending period will be up to 7 years to facilitate sustainability beyond NSF funding.
 - The Office of the Provost will commit a Preeminence line in Biodiversity as permanently allocated to the project for the Project Director of iDigBio. This line remains beyond funding from NSF.
- Over time, iDigBio will become linked with UF's Biodiversity Institute and Informatics Institute.

B. Support from FSU

The FSU Office of Research and the Provost are providing half the cost of a software developer.

2. Business Plan

A business plan will be developed and launched over the next 2-3 years as the amount of data in iDigBio increases and software that enables use of the data is enhanced and increased. The plan will be designed to generate revenue based on direct-billed cost-recovery for consulting, support, storage, and computing resources at rates approved by federal OMB CASB Circular A-21.

Clients of iDigBio include providers of data, who benefit from the standards and workflows developed for digitization of specimens and related information as well as the increased visibility and use of their data; scientists, educators, conservationists, and others who use biodiversity data for research, outreach, and policy-related activities; and individuals using ‘cloud-computing’ services and cyberinfrastructure resources, such as online storage space and access to software tools developed for data analysis. Channels through which data and services will be provided, and revenue potentially generated, include:

A. Workshops

iDigBio has hosted an average of 1-2 workshops/month over the 4 years of the project involving a total of 1,245 unique participants from 511 institutions. Workshops and training will continue, but over time registration fees and travel costs will be paid by some participants.

B. Consulting Services

iDigBio will charge for consultation and assistance in performing analyses that require biodiversity data and software developed by iDigBio and collaborators (for niche modeling, global-change predictions, phylogenetic analyses using specimen-based data, etc.) when users are unable to perform analyses without assistance.

C. Downstream User Accounts

Establishing links and revenue-generating accounts with downstream users of biodiversity data, including those in industry, conservation organizations, and private, state, and federal agencies that require data and analyses for environmental assessments. [PI Page directed a program like this at the University of Illinois that generated \$1.5 million/yr.]

D. Education & Training Materials

New education and training materials and programs may be made available for classroom use for a fee.

E. Industry & Government

Information of economic importance (agricultural pests, invasive species, etc.) will be packaged and sold as requested by clients in industry, agriculture, health services, USDA, DOE, etc.

F. Technologies & Tools

New software technologies and tools developed by iDigBio may be applicable to industrial use for a fee, especially to entities working with ‘big data’ issues.

G. Long-term Data Storage

Long-term storage of data may be made available for a fee to institutions that lack facilities.

H. Mimetic Products

Working with industry and other potential users, economically valuable mimetic products—those simulating the action or effect based on structure and function of living organisms—may be developed using images and other data provided by iDigBio.

3. Cost Recovery

Cost recovery will occur through the use of vouchers to sustain cyberinfrastructure. Dryad Digital Repository (<http://datadryad.org>) uses a system in which resources are used by clients, and vouchers for these resources are submitted to NSF. This enables NSF to contribute to institutionally supported scientific centers at rates approved by OMB CASB Circular A-21.

4. Grants and Contracts

Grants and contracts will be sought from federal and state agencies and NGOs that rely on biodiversity data for research. Examples include:

- The huge amount of validated data, APIs, other software, and tools available through iDigBio enable previously intractable basic and applied research. iDigBio, UF, and FSU, with the database, search portal, and tool developers, are in strategic positions to partner with collaborators on grant proposals.
- The 268 institutions in TCNs that have received funding from NSF to digitize data around particular research themes will be looking to iDigBio for assistance with data use and management as well as research guidance, leading to proposals submitted with UF PIs.

5. Globalization

iDigBio is increasingly involved in digitization programs in foreign countries as investigators become interested in modeling their programs after those of iDigBio. International programs with which iDigBio is developing relationships include: the Atlas of Living Australia (ALA) at CSIRO National Research Collections Australia (ala.org.au); CRIA (Centro de Referencia em Informacao Ambiental; cria.org.br) and SiBBR (Sistema de Informaçao sobre a Biodiversidade Brasileira, www.sibbr.gov.br) in Brazil; NSII (National Specimen Information Infrastructure; nsii.org.cn) at the Chinese Academy of Sciences; NHM Digital Collections Programme at the British Natural History Museum (nhm.ac.uk); and Digitalarium (Digitisation Centre of the Finnish Museum of Natural History (digitalarium.fi/en)). These collaborations

create opportunities for International Programs as well as pay-for-services, including consulting on cyberinfrastructure and research.

6. Endowments

Donations for long-term support of iDigBio will be solicited and will become increasingly likely as the scientific and societal value of digitized collections-based data is realized. Development opportunities with professionals at the UF Foundation will be pursued.