2015 iDigBio Community Survey Summary of Results

Summary

IDigBio conducts an annual survey of its internal team and the collections community, broader scientific community, partners, stakeholders, and others interested in the national digitization effort. This year, nearly 250 individuals responded; among those representing the community about one-third were affiliated with a TCN or related network. The majority of respondents (over 70%) have engaged with iDigBio at a workshop, webinar, symposia, or similar event.

The 2015 survey included questions on iDigBio leadership and communication and project impacts. Those results are summarized in a <u>separate document</u>. The survey also focused heavily on the iDigBio portal as this topic has been largely neglected in prior surveys. Over 200 community respondents plus internal team members provided input on the portal. The number of respondents who have submitted data to the portal has doubled since 2014 and most report few or no problems with the process. Those who needed assistance were largely satisfied with the help they received. About one-half of the respondents who do have data have yet to submit it to the portal; they cited a number of barriers some of which iDigBio may be able to address. While respondents were generous with their feedback regarding the portal, most have favorable evaluations of its development.

Nearly one-half of community respondents visit the iDigBio website at least monthly, most often to access information related to workshops. This information on the website is highly valued by the community, although many observe that the sheer amount of material on the website makes it difficult to find what one is looking for.

Members of both the community and the internal team believe iDigBio is making effective use of the newsletter and social media to keep the community apprised of iDigBio activities. That said, few of those participating in the survey are active users of social media suggesting that it may be helpful to seek additional input from those who do use those tools.

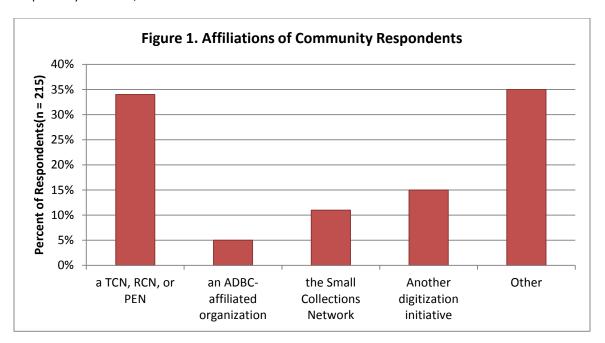
The community gives iDigBio high marks for its efforts toward achieving its major goals. When asked about priorities for the upcoming year, respondents most often cited continued training, software development, and bringing more data online. With respect to the most serious challenges facing the national digitization effort, respondents cited funding, time and staffing, and sustainability.

Approach

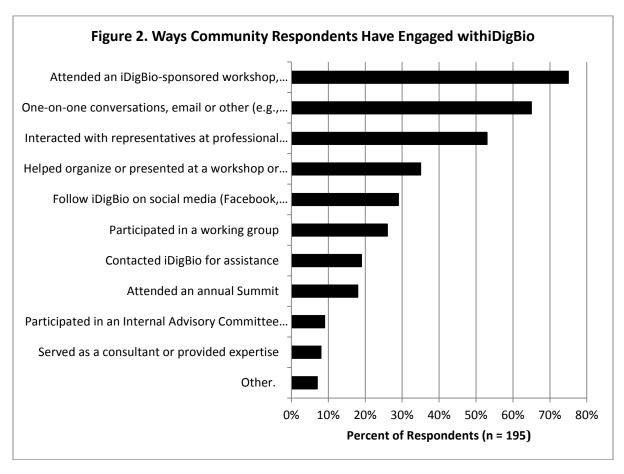
Invitations to participate in the 2015 Annual Community Survey were emailed to over 1,300 individuals who have attended iDigBio events, subscribed to the newsletter, are affiliated with a TCN, or are collaborating/partnering with iDigBio in some way. To encourage broad representation, anonymous links to the survey were also provided via the newsletter, Facebook, and Twitter. Nearly 250 individuals responded (including iDigBio team members), yielding an 18% response rate (not counting those reached through social media), which is well above the 10% standard for industry surveys.

Respondents

One-third (34%) of respondents identified themselves as current members of a TCN, RCN, or PEN. Respondents who identified themselves as "other" included representatives of NSF Bio Centers, SPNHC, Specify, VertNet, Symbiota/SEINet, Paleobiology Database, National Park Service, wildlife agencies, and unaffiliated museums and herbaria including several outside of the U.S., as well as university scientists, a high school AP Biology teacher, and members of former TCNs (see Figure 1). Forty-seven percent of respondents were male, 88% U.S. citizens, 82% not Hispanic/Latino, and 84% White; 13% and 14% of respondents selected "do not wish to provide" for the ethnicity and race questions, respectively.



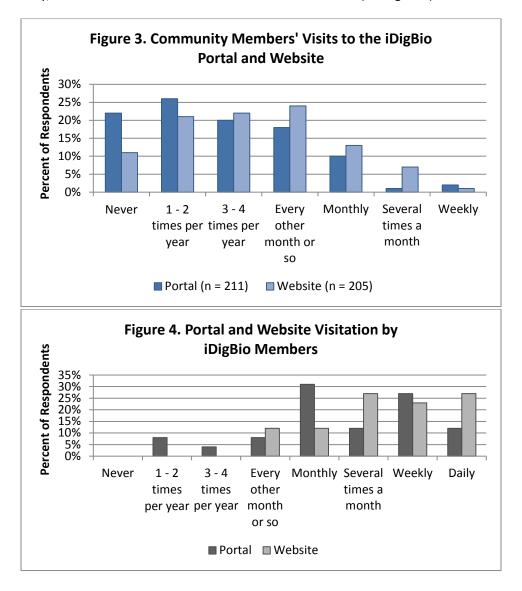
Community members who responded to the survey have engaged with iDigBio in a variety of ways, most often as a participant in workshop or webinar, hackathon, or symposium (see Figure 2).



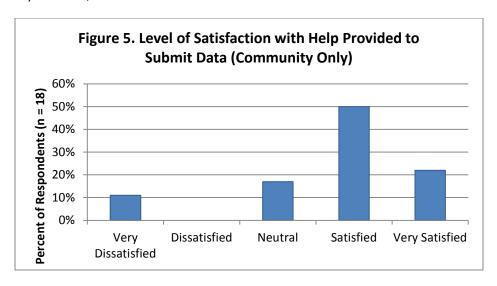
Portal

The 2015 Community Survey focused heavily on the data portal with 16 questions devoted to the topic. The questions were formulated with the assistance of IT team member Andréa Matsunaga.

<u>Visitation</u>. Community members were about evenly split between those who visited the portal two or fewer times per year and those who visited three of more times. Two-thirds visited the website three or more times per year, with 11% claiming to have never visited it (see Figure 3). In contrast, 80% of iDigBio members visit the portal at least monthly, and 76% visit the website at least several times a month (see Figure 4).



<u>Data submission</u>. Nearly half (49%) of community respondents with data reported submitting it to the portal. Of these, only 20% required more than minimal assistance to successfully submit data. Most respondents who required assistance were "satisfied" or "very satisfied" with the help received (see Figure 5), although one individual reported, "Sent multiple emails to help submit data. Received very short terse answers that did not solve the problem or answer the question."



"Our preference would have been to not do this step yet and just wait until we had our customized HUB interface built. But it didn't take too long to accomplish this - so that was fine."

"Staff were very helpful and worked to make sure the data transfer went smoothly. I dealt most directly with Joanna McCaffrey. I was very impressed with how responsive and helpful she was in getting all the kinks worked out regarding the transfer of our data. She also helped get some of the other participants in our TCN, who were slacking a bit on the data transfer to iDigBio front, to get things moving in the right direction. I found her to be very personable and efficient."

"Some of our data [are] very difficult as it does not fit into Darwin Core. It is a process to figure out what to do with those data, so sometimes it takes longer than I would hope. All of the persons involved have been very helpful."

"As a TCN, we were asked to contribute our data ASAP. We have a plan for providing data via a Hub developed as part of our TCN project, but in the meantime our institution has worked with iDigBio staff to make sure that specimen data can be directly sent to the iDigBio specimen portal via IPT directly from our collections management software system (Specify). To make sure that the data were feeding correctly, we needed to make a few adjustments to the IPT and this required communication with the iDigBio team."

"Typically our workflow involves submission to Joanna with an initial review, than a secondary review by someone in the ACIS group."

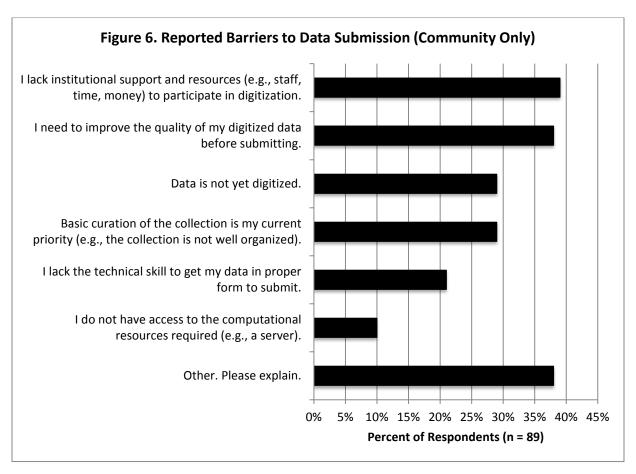
"We worked with the lead data person on our TCN. Since we've moved to using an IPT server things have gotten easier (for everyone)."

"We needed help with getting our specimen data and images uploaded. This was due to a variety of reasons, including the database we were using, figuring out the protocol for image uploading, and being sure that our fields matched."

"I mostly deal with iDigBio via VertNet."

"Some of our data flows into the portal from the Symbiota portal and I don't do anything. My problem right now is getting help with our CSBR data and images. We are working on a way to get the data to iDigBio as required by the grant, but so far haven't achieved that."

<u>Barriers to data submission</u>. Fifty-one percent of respondents with data have not yet submitted it to the iDigBio portal and 84% of these reported at least one barrier to submitting data (the remainder are submitting through other initiatives or are in the process of submitting data now). The most commonly cited barriers are lack of institutional support/resources and the belief that the quality of the digitized data needs to be improved prior to submission (see Figure 6). Some of the barriers cited are issues that iDigBio can potentially help with (e.g., lack of technical skill or computational resources). This suggests that iDigBio needs to continue widespread outreach to the collections community and possibly increase efforts to raise awareness of iDigBio resources.



Many of the "other" barriers cited are not in fact barriers (e.g., the data is submitted through another initiative, the data is nearly ready to be submitted). Descriptions of barriers offered by respondents:

"Data needs to be formatted for iDigBio ingestion and this is lower priority than other activities."

"Don't yet quite understand what the best order of operations is for us in getting data into iDigBio, VertNet, GBIF, other similar databases, our own online database, etc."

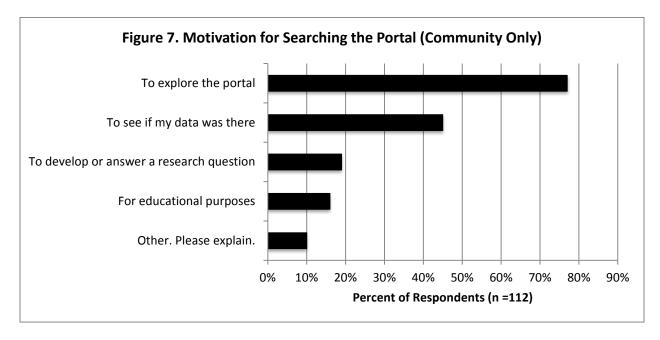
"I am unaware of how to submit data."

"Insufficient resources allocated by projects to submit separate iDigBio datasets for every TCN/PEN that we are involved in."

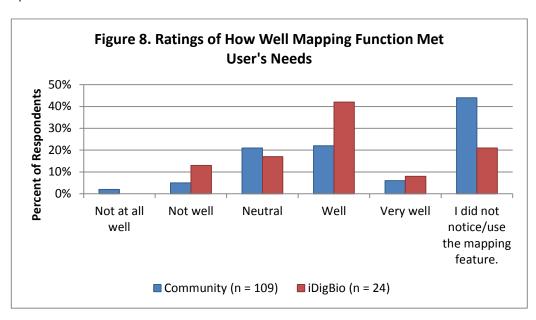
"Located in Canada (probably will submit to Canadensys)"

"Vertebrate fossil locality data restrictions - confidential data"

<u>Portal Search</u>. Fifty-three percent of community respondents who had visited the portal reported using the search function, mostly to explore.



Over 40% of the community members who searched the portal either did not use or notice the mapping function. Of those who did, 50% thought the function worked "well" or "very well," 38% were "neutral," with the remaining thinking it "did not work well/work well at all." Among the iDigBio team who used the mapping function, 63% thought it worked "well" or "very well" while 16% rated it as working "not well" (see Figure 8). Comments suggest that the map is "buggy," can be difficult to center, and sometimes drops the user to a point off the map. One respondent suggested that the team focus less on the mapping feature and more on getting data (because good mapping alternatives are available elsewhere), while others would like additional mapping features to allow them to "get deeper into the data."



Comments:

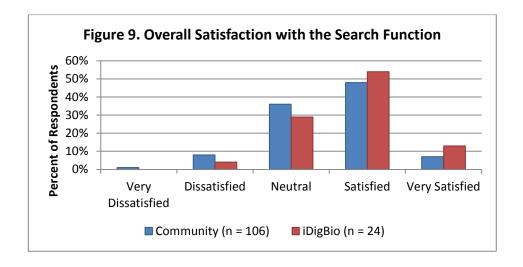
"I recognize that this is a work in progress. I was impressed with what was developed thus far and I feel that Andréa M. and staff have done an excellent job with this and I'm sure further improvements will be made. If I had to point out specific problems some times I did have a bit of trouble getting the map to always center and some times I would be dropped down in a point outside of the field of view of the map. This is not meant as a critique of what was developed except in a constructive way."

"It would be better if the map where zoomed in to the area that was covered by the specimens that had come up on the search."

"Some aspects are confusing, such as the "missing' or 'present' buttons and the mapping features."

"Rather than just plotting generic points on a map, it would be useful to be able to have different colored points representing values in other data fields (e.g., collection date range)."

Nine percent of community members of 4% of iDigBio team are "dissatisfied" or "very dissatisfied" with the search function, while 55% and 67% of the respective groups are "satisfied" or "very satisfied" (see Figure 9).



Comments suggest that users find the interface attractive, appreciate how quickly results are returned, and like the Encyclopedia of Life synonyms feature:

"I like the ability to sort by clicking on column headers. Advanced search is excellent and powerful while still being intuitive. Love the tab design on specimen records, and the ability to view query results as a list, labels, or images."

"Nice to have a general simple search to begin. If that doesn't yield results, then the advanced search pops up. Very nice to keep a search going, without seeing the null set only. The simple search should search all available fields, not just the genus/species."

"The search is very quick."

"Specimen search worked quickly when correct parameters were entered (genus, country)."

"I really like the search function. It is very easy to use, intuitive and based on Darwin Core."

"Really like the EOL integration to add synonyms. Like that you can add fields for searching."

Those who are satisfied with the search function find it easy to use and intuitive, while others users sometimes struggled. Users are divided on the issues of fields and taxa with some apparently wanting more of both and some wanting less. Respondents offered a variety of suggestions for improvement—some quite specific and perhaps idiosyncratic. Users would like the option of searching by institution, and note that additional features are needed to improve the usefulness for paleontological searches. The most serious criticism is that the searches did not always return all the information that the user expected based on their experiences with other data portals; this obviously threatens their confidence or trust in the data. Comments:

"It can be very tricky to figure out what the different fields mean. Is 'collection code' an institution or a group of animals? Also, where does one find an 'institution code'? Ultimately I was able to find my data, but it took several trials to find the 'magic' combination."

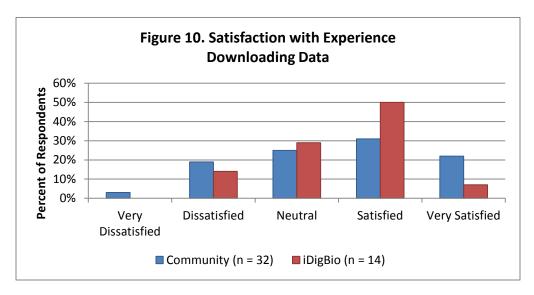
"Once your initial search is built, things get confusing. Sometimes I run the same search twice, but get different results, because something was still checked that I thought I had cleared. The user interface isn't super-intuitive."

"The search itself is good, but the ability to add fields to the search and results is not obvious--I think this could be a barrier to first time users accepting and appreciating the usefulness of the portal."

"Needed higher taxonomy field names. Not sure I returned all specimens of group I was interested in."

"Search functions are more difficult to use than the average Symbiota portal. The information that is returned is, for the most part less useful (and comprehensive) than that returned from a search of most Symbiota or herbarium-aggregated (e.g., Consortia of Pacific Northwest Herbaria) portals."

<u>Downloading specimen data experience</u>. Only fifteen percent (n = 32) of those who have visited the portal have downloaded data. Slightly more than half were "satisfied" or "very satisfied" with the experience, which reflects that users' experiences varied—some had trouble downloading the data, some did not, and for some it took a long time (see Figures 10 & 11). Several users question the format of the output, noting that csv files are easier or that a FAQ sheet or other assistance is needed to help the user know how to view the data once they have it.



"Downloading a massive XML files is not very useful to people uncomfortable with IT methods. How about a simple csv file or spreadsheet that can be opened in Excel."

"Data was downloaded, but then what do I do with it? Need an FAQ on the download area on how to view/use the data."

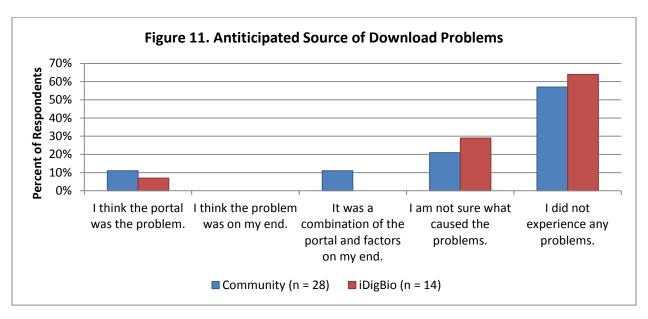
"To be perfectly honest, I usually use the CNALH pages to access our data and the info on specimens from other institutions. Have had a few format problems (with how the page appears, things being missing, etc.) We communicated with someone and they fixed it."

"Downloading wasn't the problem. The lack of organization of the fields in the download was. Cleaning up / arranging the download took 2 weeks. It would be nice to pick the fields and be able to order the fields the way you want them in the download. Or at least arrange them alphabetically or group them by like information (ALL types of numbers, IRN, USI, Catalog should be grouped together). Taxon names - together (Order, Family, Genus species, common name) at least, then unique specimen info, then collecting info."

"The download works in a timely manner. You can download large datasets with no problem. The data are provided in a standard format, and the provided citation is well formatted and a very nice bonus. Data was difficult to use outside of the portal."

"Occasionally, the download did not work; I had to contact ACIS. The 'time to download' did not seem accurate. It would be nice to be able to select the fields I'm interested in to download - as it is, I get all of them, and then I have to edit my file to clean it up for my purposes."

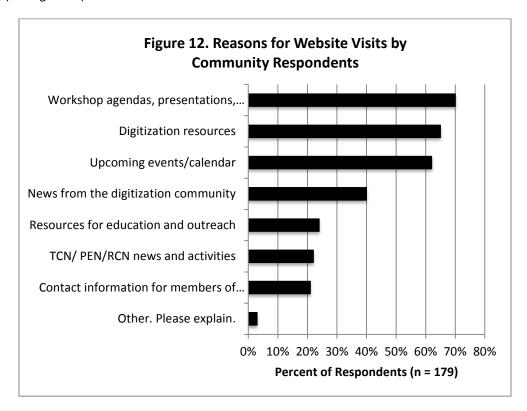
"I do see an issue with the estimator of time to download a data set. The portal software estimated 10 seconds for a 60 record data set. It took over a minute. A 4 record data set (also estimated as 10 sec) took over 30 sec. Note from restaurant management: overestimating your wait, and seating you before your overestimated time to a table -- is much better received by restaurant patrons, than telling them it will be 10 minutes -- and turns out to be 30."



Website

The survey included four questions about the iDigBio website. (A more in depth analysis of website usage as well as workforce training provided in a separate document—Year 4 Impact Evaluation).

As shown in Figures 3 and 4 above, 89% of the iDigBio team visits the website at least monthly compared to 45% of the community respondents. The most common reasons the community visits the website are to access workshop agendas, presentations, and reports (70%), digitization resources (65%), and information about upcoming events (62%) (see Figure 12).



When asked where they might find information currently available on the website if the website did not exist, 19% of community respondents reported they did not think they could. As one individual observed, "I have no idea. I don't think I could. One of the most valuable aspects of the website is all the wikis and being able to see past workshop presentations and other such documentation that may never be published but are essential resources." Another noted that some of the resources would not even exist without iDigBio: "There's no one place where this information would be available, and without iDigBio none of the workshop and working group resources would exist."

While the website, and the information it contains, is highly valued by the community (e.g., "It is s an outstanding resource! It's packed with information, and has frequently saved me from wasting time figuring out solutions to problems already solved by someone else."), comments suggest the website remains difficult to navigate:

"It is often difficult to find specific information about digitization when I have looked for it. This could be partly due to the vast amounts of information on the iDigBio website. However, most of the information is in the form of adobe connect videos which makes it difficult to find anything specific without wading through an entire talk."

"As iDigBio has grown so has to website. It has gotten pretty overwhelming and this makes it hard to find what you need at times. I wish I had a suggestion as to how to deal with this situation."

"It is not easy to determine from looking at the home page where to find information and results from past iDigBio workshops. I guess one can do a search, but if you don't remember the exact title of the workshop it may be hard to determine where that information is. It would be nice to have an easily accessible link in the home page to a page where we could browse for results and other materials of past workshops."

Several respondents noted the lack of resources related to education and outreach, while others suggested there is a need to make the information provided more accessible to broader audiences.

"There's a distinct lack of education and outreach resources."

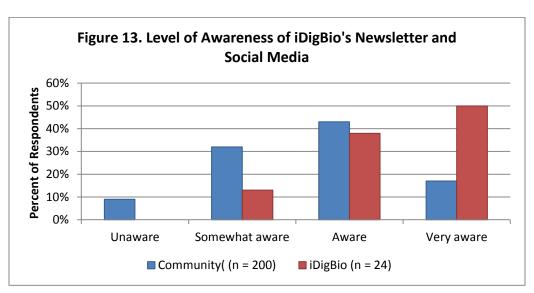
"Link more publications that exemplify the use of digitized data for education and research."

"I understand it's a new modern digital age, but I, along with many others did not grow up with this and some of the terms are ambiguous to me. For example—digitization community. What, really, does that even mean?"

"A 'For Dummies' section, most of the information presumes knowledge of certain terms that create extra hurdles for collections managers with less knowledge in technical areas or CMs that have less experience with natural sciences. For example I have both humanities and natural science collections and am much more knowledgeable in the humanities. Having to learn the technical jargon for both the science and data at the same time is slow and difficult work."

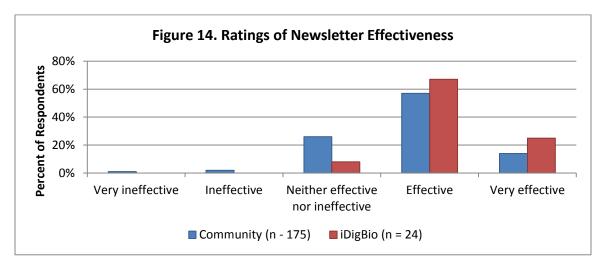
iDigBio Communication Efforts

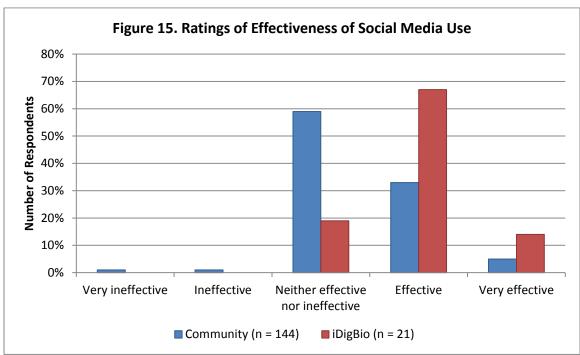
The survey included six questions about iDigBio's communication efforts with an emphasis on the use of social media. The majority (>90%) of the community and iDigBio team are at least "somewhat aware" of the newsletter (iDigBio Spotlight) and social media efforts (see Figure 13).

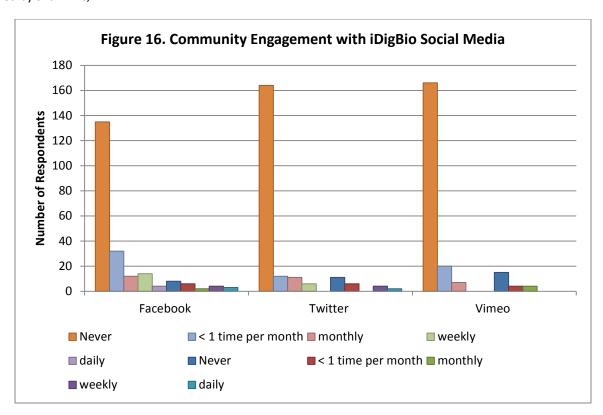


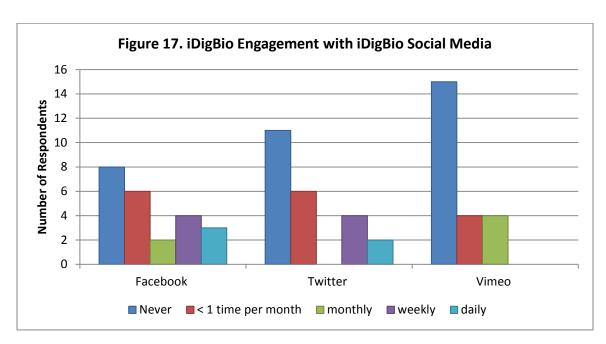
Over 90% of the iDigBio team and 71% of the community respondents rate the newsletter as "effective" or "very effective" at keeping the community informed about iDigBio activities, while 81% of iDigBio and 38% of the community respondents rate the social media efforts as "effective" or "very effective" (see Figures 14 and 15). The lower rating of social media use, particularly among the community respondents, likely reflects their active dislike of these communication channels. While the number of individuals following iDigBio on Facebook and Twitter has

increased steadily to over 700 and 800, respectively, the vast majority of respondents to the survey are not members of those groups (see Figures 16 and 17).









"While having the Facebook page share relevant articles is interesting & does lead me to sometimes find out about new things I hadn't heard of before, it would be good to also see the Facebook page talk more about what iDigBio is doing as far as events/workshops and updates from these events."

"Don't just have a Facebook page. If you want to engage the community you have to go to their forums, they won't come to yours without a reason. Get involved in the identification forums for mushrooms and plants and look for opportunities to throw in a plug for iDigBio and show people how they can access the data themselves to answer their questions."

"I only use FB, but I think they're doing a super job with this. I regularly share these posts with my FB community, and it's certainly sparked interest, even in my non-biology contacts!"

"Educators at the secondary level are, IMO, an untapped user base for the resources you provide."

"The broader community consists of many people not in college. I suggest 'dumbing it down' to the level of the everyday broader community. Just my thoughts. However, if the intent of IDigBio is to involve academicians, mostly, then keep doing what you're doing."

"Get the word out outside of the museum community so that it is more visible in the scientific community outside of TCNs, PENs, & iDiqBio Workshop participants."

"Social Media is NOT all it's cracked up to be. Personally, I was on the FB quite frequently and saw many iDigBio posts, HOWEVER, I have very little time to piddle around with Social Media now that our TCN is up and running so I hardly ever seen anything on the FB (because I just don't check my FB account). Please don't ever consider opting for social media over a newsletter or vice versa...diversity of communication media is key."

"Not everyone uses social media and many people refuse to use social media. It should not be a primary means of raising awareness and engaging the community."

"I love the newsletter, but do not typically engage in social media such as Facebook and Twitter."

"I like the newsletter. More news is always better."

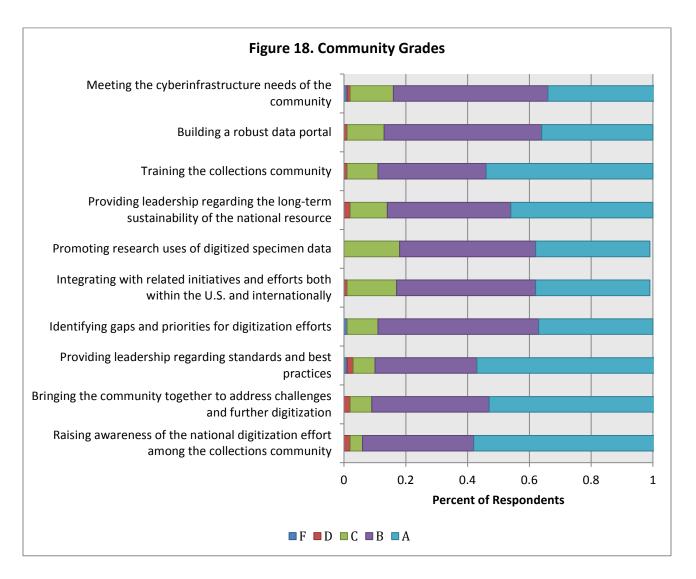
"Focus on developing an active web page with current news and information regarding iDigBio activities."

"The newsletter format is very attractive. The Biodiversity Spotlight was a great addition."

"I am old school (prefer not to use social media) so rely on the newsletter but younger staff tend to use social media more."

Overall Ratings of iDigBio

To gather an overall sense of how the community perceives iDigBio is performing, we asked the community to grade iDigBio in ten areas of impact. Respondents were allowed to assign plusses/minuses, but the grades are collapsed here for ease of presentation. As shown in Figure 18, the community clearly believes that iDigBio is performing well and making progress across all areas. For several reasons, however, the results should be interpreted with caution. On average, only about one-half of the respondents were comfortable assigning grades/answering these questions. Some who did nonetheless commented that they really had no basis to judge but did not realize that they had the option of skipping the question. (We did not include a "no basis to judge" option here in a misguided effort to allow respondents to skip ahead.) Comments provided by respondents also remind us that individuals vary widely in how they assign grades. As one respondent observed, "Some are hard to judge. If you didn't get an A it doesn't mean you aren't doing well at it, but rather that that area might be the next to focus on." Despite these limitations, we can conclude that overall the community believes iDigBio is doing well achieving its goals.



"While iDigBio does seem to enhance and facilitate digitization activities, it merely provides the resources and connections to <u>promote research</u> involving digitized specimen data. It would be great if there were more of a research focus, but that seems to be a priority that is to be sought after some extensive digitization efforts have been achieved nationally. Additionally, while iDigBio does provide leadership, it does cause me concern to see how this model will be <u>sustainable</u> or even persist beyond the funding regime outlined by the NSF ADBC. Sustainability of this model does seem to be questionable, in my opinion."

<u>"Identify gaps and priorities</u>: I think iDigBio could provide more leadership in this area. It is a fine line between guiding and letting the community naturally develop."

"I think you have, so far, fallen short in the area of <u>international data sharing</u>. I think you may be changing that soon, however."

"iDigBio is great in connecting people from widely separated collections and/or small collections to show the importance of our aggregated data. I have been greatly helped by workshops and webinars, and have been able to increase the numbers of online specimen records from my collection."

"I gave you a B in <u>research</u> because I've only seen workshops focused at the collections community and none focused on how to do research with the collections. I'm a faculty member and collections director, so I want a workshop on research for undergrads, grads, and myself using the collections/iDigBio portal."

"Overall iDigBio has provided leadership in many of these venues, or promoted many of these topics among the collections community. I think the workshops have been one of the best things about iDigBio and bringing the community together."

"I fully support what you have done so far, and appreciate the training that has been made available (staff from my institution have attended a number of workshops and share the info). I think the effort is driving my institution to digitize the collections at a faster rate than would have been done without the federal support. That being said, we are still a couple of years behind in being ready to upload our collections data (e.g. still cleaning up and converting old databases to KE, still digitizing specimen data especially for invertebrate collections, etc."

Priorities for the upcoming year

Both the community and members of the iDigBio team were asked to share their thoughts about the most important tasks iDigBio should focus on during the upcoming year and any ideas for new initiatives and activities. The most frequent response was workshops and training, followed by software development and bringing more data online.

Representative comments:

"Continue hosting regional workshops and helping support attendance by interested participants.

Continue to foster interaction among curators through workshops, especially interactions between large and small herbaria."

"Doing better at opening up workshops/seminars/etc. to remote participation either live or viewable after the fact. I'm sure like many other places we don't have the funding to travel to FL."

"Continued support for the development of software and infrastructure to assist the collections management community."

"Continuing to bring more data online"

"...need for shared, central taxonomy webservice. It's insane that I'm plugging all these names into my own database and everyone else is doing the same. If a revision comes out with lots of name changes for a taxon how many different curators/collections managers are going to edit their own databases to update them to all those new names - this happens all the time, enormous waste of effort. Those edits should happen in only one place, and iDigBio taxonomy name server that everyone can benefit from sharing & linking to."

"Data management for collections managers who keep specimen data/images in many different databases. How do we get these databases to talk to each other, so that collection managers don't have to update data twice (or three times, if you have data in three different databases)."

"Tools and infrastructure for using digitized collections/data for research."

"At this point, for my particular ADBC-TCN, I think iDigBio's most important function this coming year will be to facilitate data uploading and sharing. Many of our collaborators have begun to image specimens this year, but have not yet begun (or are just beginning) to share those images and data. So we're in new territory once again, and iDigBio will be critical to us maintaining momentum."

"Focus on integrating existing efforts to provide one-stop-shopping for digital collections data. Make all efforts to aggregate existing digital data resources."

"Getting as actively involved as possible in the ongoing NIBA/B.CON sustainability initiative"

"Work to guarantee long-term sustainability of the digital archive and portal."

"It would be nice to help the larger institutions, as their digitization efforts mature, find a way to continue the digitization of accessions beyond their grants."

"Increase collaboration within/outside of ADBC"

"Providing ways smaller institutions (or institutions with smaller budgets, anyway!) can undertake digitization projects. Perhaps create a venue for sharing manuals or DIY ideas created in the museum community."

"Focus needs to turn to improving the quality of data and improving fitness for use Funding of digitization proposals."

"Networking users and educating administrators on ideas of where to look for funds to help in their efforts."

"Collections 'not in peril' but still un-digitized have an impossible time getting funding for digitization. What other funding sources are available for digitization projects? Get the word out."

"Figure out ways to get researchers to use the data."

"I think creating more specimen-based resources, e.g. educational tools, and research products from the already digitized data would provide the community with tangible reasons to continue digitizing legacy collections."

"Increase dissemination of best practices and standards resulting from the various workshops."

Biggest challenges

Both groups were also asked what they regard as the biggest challenges faced by the national digitization effort at this point in time and how iDigBio can help address these challenges. As in past years, the most common responses were sustainability, funding, and other resources (time and people). However, the community provided a wide variety of unique responses that have not been shared on prior surveys, perhaps reflecting the more diverse sample that completed the survey this year.

Representative comments:

"A better plan for long-term sustainability of digitization efforts. Collections are not static; they will always need to be updated and it does not seem like enough thought is being put into how all of this data and infrastructure will be maintained and updated in a post ADBC world."

"The biggest problem nationally is funding in museums. Nationwide museum are having difficulty finding money to hire people, pay for space improvements, or even upgrade data systems. Without more funding

for museums, it make little difference if the data is available. Without funding at the museum level, georeferecing specimens, imaging of specimens or catalogs, or other digital info, cannot be generated let alone be made available. Collections are at risk everywhere. iDigBio has to press government at all levels to fund museums and collections."

"'Mothballing' of collections. Administrators still just don't get it. They tend to see systematic collections and systematic in general, as dinosaurs that take time, money and space away from more important endeavors. Collections based in university academic departments are particularly in jeopardy."

"My greatest challenge is lack of resources (and especially personnel), both to capture the data and to update it and maintain quality control."

"Lack of time at an institution where teaching absorbs nearly all time."

"Clean data, check and update taxonomic identifications"

"The biggest challenge is providing skills or training for understaffed and overwhelmed collection managers."

"There is also the problem of varying speed in digitizing specimens - some taxa are easier than others. The working groups help organize people with similar imaging problems."

"Administrators often fail to recognize the importance of collections and I hope the digitization effort can help increase the utilization of collections in ways that administrators can appreciate."

"Making the data ubiquitous and available."

"Provide stable, long-term access to the data."