S. Ellis, May 2014

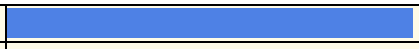
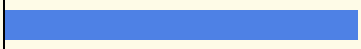
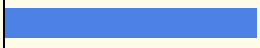
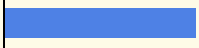
# 2014 iDigBio Evaluation Summary

## Overview

This report summarizes key findings from the 2014 iDigBio Community Survey and individual training workshop surveys. An invitation to the community survey was distributed electronically to over 800 email addresses of individuals who had interacted with iDigBio in some way—from participating in a working group to registering for the newsletter. Of the approximately 200 individual who responded, 32% are affiliated with a TCN or PEN, 21% with iDigBio, and 8% with an ADBC-affiliated organization. The 25% response rate is comparable to that of the 2013 community survey. Correspondence with several individuals suggest that some organizations felt it sufficient if just one member of their organization completed the survey; if this thinking is widespread, it would obviously depress the response rate.

The majority of respondents (85%) had visited the iDigBio website, three quarters had participated in at least one workshop or symposia, and over half had interacted with iDigBio staff via email or adobe connect and/or read the iDigBio newsletter. Only 7% had retrieved data from the portal, while 18% had submitted data. Of those who have engaged with the portal, most find it easy to use and are satisfied with the help and timeliness of assistance provided by iDigBio. It may be valuable over the course of the upcoming year to collect feedback about the portal (beyond data ingestion) from the community via surveys and/or interviews.

**Table 1. Ways that Respondents Interacted with iDigBio**

| | | |
|---|---|---|
| Visited the website |  | 85% |
| Attended one or more workshops or symposia |  | 74% |
| One-on-one email or other (e.g., adobe connect) exchanges |  | 54% |
| Interacted with representatives at professional meetings |  | 53% |
| Read the newsletter |  | 52% |
| Presented at a workshop or symposium |  | 40% |
| Participated in a working group |  | 29% |
| Contacted iDigBio for assistance |  | 18% |
| Submitted data to the portal |  | 18% |
| Attended an annual Summit |  | 16% |
| Organized or facilitated a workshop or symposium |  | 13% |
| Tweeted or created Facebook posts related to iDigBio |  | 13% |
| Contributed content to the website |  | 12% |
| Other |  | 8% |

| | | |
|---|---|---|
| Retrieved data from the portal | | 7% |
| Participated in IAC meetings | | 6% |
| Served as a consultant or provided expertise | | 6% |

One of the most successful of all the iDigBio efforts continues to be workshops, symposia, and webinars. Evaluations of training seminars reveal that, on average, 95% of participants report increased awareness, knowledge, or skill in key aspects of digitization. The most common responses to the query "What is iDigBio doing well" are workshops and training and community building.

Nearly 60% of respondents are satisfied with the iDigBio website and only 15% currently find the website "difficult" or "very difficult" to navigate. Resources on the website are valued by the community, although respondents did offer suggestions of ways to improve access as the amount of information continues to grow.

The newsletter is also popular among respondents with nearly 80% reading it. iDigBio social media efforts are appreciated, although fewer than have of respondents engage with iDigBio on either Facebook or twitter. Those who do have requested more frequent postings and posts with more original content.

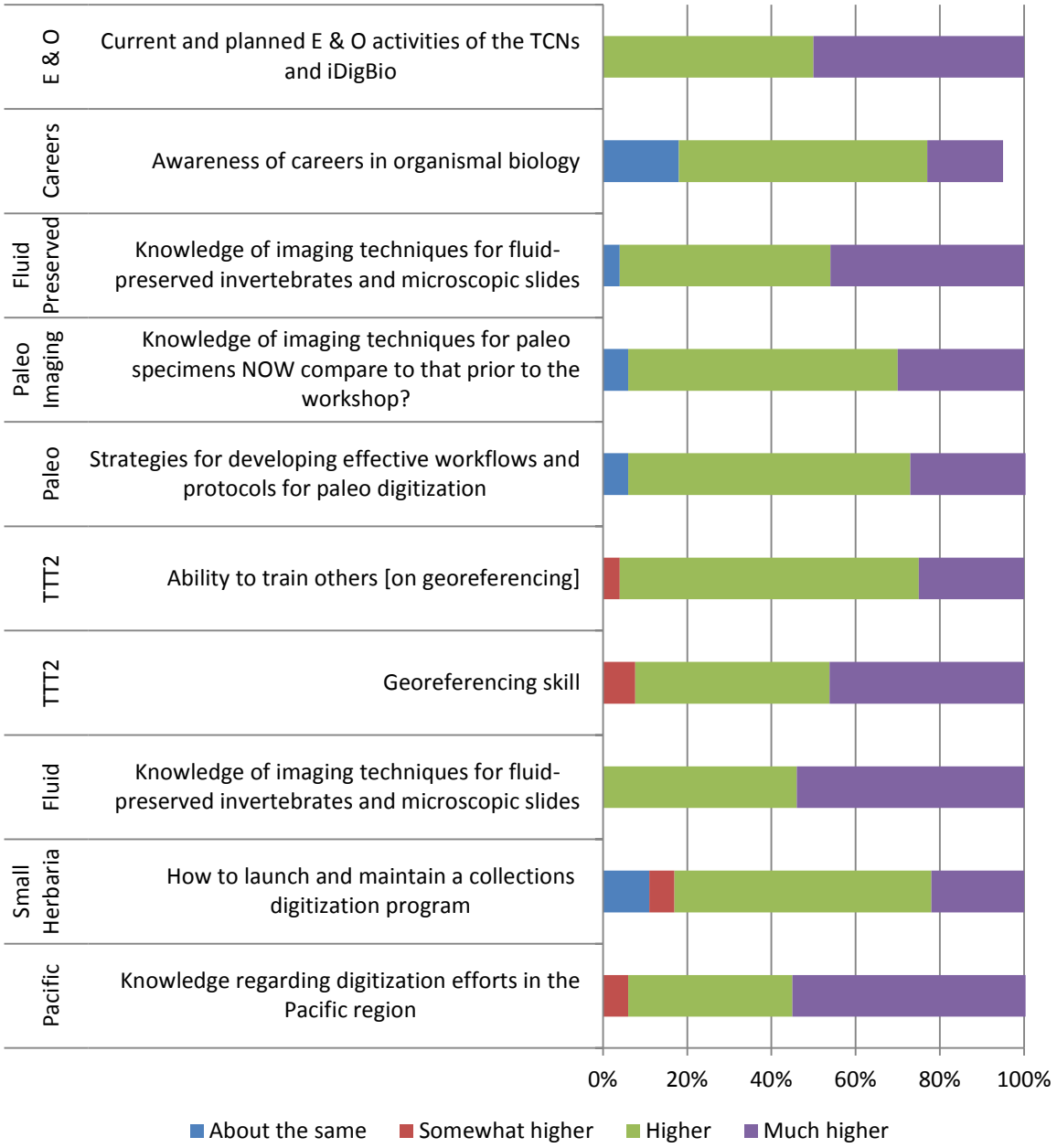
Although iDigBio is lauded by many for their community building efforts, the most common response to the question "What is iDigBio doing less well" is some variation of "reaching out"—to both small and large collections, different collection types, other initiatives, other experts, and the broader community.

When asked about the most important tasks iDigBio should address in the upcoming year, respondents suggested a list of challenging problems, followed by more reaching out and community building. The overwhelming response to a question about the biggest challenge faced by the national digitization effort was resources including alone or in combination—funding, time, and people followed by the related issue of sustainability.

**Workshops**

Of the 21 workshops and symposia sponsored by iDigBio during 2013-2014, 9 were training workshops. While the specific goals of each training workshop are unique, the overarching goal is to increase awareness, skill, or understanding within the realm of digitization. On average, 95% of workshop participants report increases (see Figure 1). Those who did not typically began the workshop with high levels of expertise and were often themselves presenters.

**Figure 1. Changes in Awareness, Knowledge, or Skill**



**Comments from workshop surveys**

***Did the workshop meet your expectations?***

- It was better than expected. The opportunity to meet with other collections managers and hear about their problems and solutions; finding out that there was a huge amount of help out there to solve our problems (and that we were not alone in having a particular set of problems); and learning that we may be ahead of the curve in dealing with some issues. I felt really lucky to have had the chance to participate. (Digitization in the Pacific)

- It exceeded them. I was so impressed with all the diverse solutions to imaging and now have a much better idea of what methods are most appropriate for different types of subjects and research and outreach needs. I hope I can synthesize and condense it down to be able to effectively communicate all the information to students in a shorter format, and will definitely be recommending that they look at the wiki content. (Paleo Imaging)

***Additional comments***

- Thank you for organizing this wonderful workshop! I developed the seeds for lots of ideas over the two days, and I am excited to figure out ways to make them grow. I look forward to continuing to build collaborations with folks at iDigBio, the TCNs, and local STEM educators. (E & O)

- This workshop is wonderful. I'm so excited to get started on georeferencing and help make a difference at my institution, both in in terms of helping educate others on georeferencing and speeding up access to this important data. (TTT2)

- All of the iDigBio workshops I've attended have been well-organized, tuned to the right level of detail for the particular audience, and prompted good engagement by participants. This particular workshop was even more so, and was the best so far, I think. Super-helpful. (Small Herbaria)

**What iDigBio is doing well**

The success of the workshops is further reflected in responses to the question "What is iDigBio doing well?" on the community survey (See Appendix A). Of the responses, the most common reference workshops, community building, providing resources, and reaching out and promoting best practices.

**Website**

We asked a series of questions related to the iDigBio website—partly in response to earlier concerns that the website was difficult to navigate which led to significant revisions. Among our respondents, 58% reported being satisfied with the website, with 59% visiting the website at least monthly. The most common reasons for visiting the website was for information on workshops, upcoming events, and news from the digitization community (see Appendix B).

The website includes a significant number of technical resources. The most frequently used of those are materials related to workflows, imaging, databases, and georeferencing. Slightly more than half of the respondents (55%) report using the wikis associated with workshops and working groups. Nearly 60% find them “helpful” or “very helpful.”

One-third of respondents find it “easy” or “very easy” to find information on the website with 52% feeling neutral on the topic.

***Additional comments***

- My answer to this is directed to more of a heads up than a doing less well answer. The amassing of such an abundance of information on collections issues will eventually lead to a point where it will become more difficult for the user to quickly access needed information from the web site. Since the site is currently not overwhelmed with too much information it may now be the time to evaluate the web site format and consider the changes needed before it becomes difficult to manage. I would hate to see the iDigBio web site not being used because of too much information that is difficult to find. This concern may have already have been pointed out is currently being addressed.
- Presenting synthesized information on processes / techniques (what was learned and presented from various workshops) to those the general community (those that couldn't attend the workshops). It appears to me currently results and information are provided on a workshop by workshop basis, so someone would have to go back to a certain workshop to find out information on certain processes. I think iDigBio has developed some great information on processes, methods, etc. but I think these need to be organized and synthesized somewhat differently from how they are presented currently in the website to be more accessible to the wider community. In other ways, those resources are provided right now under a certain working group or a certain workshop, but your average person not familiar with the structure of iDigBio will never be able to find these resources as they are not organized by themes such as Digitizing Specimens that talks about a specific action someone at a museum would do.

**Newsletter, Facebook, Twitter, and Blogs**

The iDigBio newsletter is the most popular of our communication efforts with 79% of respondents reporting reading it and a subscriber list exceeding 600 individuals. Forty percent of respondents engage with iDigBio on Facebook, 23% follow iDigBio on Twitter, and 45% follow blogs (see Appendix C). Respondents are about equally divided between those who are satisfied or very satisfied with the communication efforts and those who are neutral or dissatisfied. Suggestions on ways to improve communication include: Enlist guest authors/bloggers from outside of the core iDigBio Staff, TCNs & PENs so as to pull in more interest from those in the greater community; have more content for the Newsletter, it's digital after all so length should be less of an issue; include more information on outreach opportunities and ways to involve those not in TCNs & PENs.

**Portal**

As noted above, only a minority of respondents have worked closely with the portal. However, among those, satisfaction is fairly high. Only 4% report it difficult to submit data to iDigBio, and no one was dissatisfied with the time frame in which they received assistance with two-thirds being "satisfied" or "very satisfied" and equal percentage reporting that those providing assistance usually communicated with them in a way they could understand. General comments about the portal are provided in Appendix D. A few additional comments about the portal can be found in Appendix E—What iDigBio is doing less well.

**What is iDigBio doing less well?**

Although respondents cited iDigBio's efforts at community building and reaching out as some of what the team is doing exceptionally well, reaching out tops the list of what iDigBio is doing less well, too. Specifically, respondents noted that iDigBio should continue efforts reaching small collections, but that the team should also reach out more to large collections, different collection types, other initiatives, other experts, and the broader community (see Appendix E).

**Most important tasks for iDigBio in the upcoming year**

Respondents were asked what they thought iDigBio should focus on during the upcoming year. Topping the list was an assortment of challenging problems to solve (e.g., extracting label data) followed by reaching out to the community and other projects (see Appendix F).

**Biggest challenge facing the national digitization effort**

By far, the most frequently cited challenge was resources including funding, time, and people. This was followed by the related issue of sustainability. Other issues include data quality, duplication of efforts, and getting everyone on the same page (see Appendix G).

## Appendix A. What iDigBio is Doing Well

### Workshops and training (n = 19)

- Education, I believe the workshops are very helpful. In particular the recent small collections workshop was very informative. The digitization of paleo workshop was less so.
- The workshops are an amazing way to engage a lot of people in the broader scientific community who may not be at places with funded projects. I think that is likely to be one of the main ways in which iDigBio has a long term impact, in that it has raised the standards through diverse training opportunities.
- Webinars, workshops
- Workshops and webinars!
- Workshops and training
- I have participated in three iDigBio workshops as an organizer and presenter. They were very well organized and very productive in terms of stimulating meaningful interaction and open discussion relevant to the topics at hand. I am a member of a working group (NANSH) and that has been a very productive experience too.
- The workshops and symposia are very well organized and are an excellent way to disseminate information on digitization processes / techniques, and other topics to the community. In the ones I've attended, I'm always impressed about the caliber of the information and how useful they are for assisting members of the collections community in conducting certain activities.
- Offering extremely productive and useful workshops for a variety of user groups towards improving digitization efforts and facilitating communications among and across disciplines.
- Education
- Webinars
- Education about workflows
- Stimulating meetings and symposia.
- Making workshops more visible and accessible to students who are interested in this field and who want to learn new skills.
- Workshops
- The workshops are well done, and they seem to provide useful content on the above area.
- Teaching others; offer numerous opportunities to learn about digitization
- Teaching and sharing data
- Nice workshop topics
- Training

### Community building (n = 18)

- In my experience IDigBio is doing particularly well at helping to organize the collections community to collaborate and share information and in general discipline-wide outreach and assistance. It has helped people realize that existing wheels do not need to be reinvented, and that all of the necessary software already exists to digitize collections.
- Training, updating the community, bringing together multiple entities, networking, providing positive support and motivation
- Bringing together a community of like-minded individuals to further the digitisation cause.

- Great communication and community building. Gil Nelson and Deb Paul are great at building consensus through workshops, Working Groups, etc.
- Connecting communities of collection managers and curators across the country. Responding to the needs of community members.
- Connecting different collections together.
- Enabling personal connections between potential collaborators that may have otherwise never met. Facilitating continued interactions by hosting them on Adobe Connect.
- Providing an incredibly useful forum for communication among the museum collections community,
- Bringing collection managers together.
- Community building
- Community engagement, trying to bring in people not normally at meetings. Very good at listening often.
- Keeping the group organized. Getting information to everyone. Keeping everyone interacting with workshops.
- Bringing the community together around specimen digitization; getting small collections mobilized
- iDigBio is really bringing together the different institutions charged with holding and storing scientific material. They are helping share ideas of what works and doesn't work between institutions so that we aren't all reinventing the wheel.
- Bringing together a group of people with a broad range of skills and knowledge in order to teach and share with those who are new to the world of digitization. As far as workshops go, I am thrilled to see more breakout sessions occurring. I feel much more gets accomplished during those and it gives everyone a chance to process the information they just heard from the talks prior. They are working hard to hear small collections with their wide variety of problems and trying to develop networks to unite everyone in this ambitious digitization effort. It is especially helpful to have the workshops and webinars and working group meeting available to everyone who is interest and can participate via Adobe Connect. LOVE that!
- Providing a sense of community.
- iDigBio is bringing together people who wish to develop the skills and knowledge to share information about their specimens.
- Interacting with the collections community, the people that do the digitisation at the ground level.

**Providing resources (n = 13)**

- Excellent at providing information on various aspects of digitizing natural collections. They are very responsive to questions about the process. One major advantage is that they provide multiple methods not just a single avenue to achieve your end goals.
- Providing a source of workflows and information to save other collections having to "reinvent the wheel".
- Being a central clearing house and coordinator
- Trying to organize best practices documents.
- Being a centralized resource for digitization.
- Providing concrete resources via the wikis.

- Providing a repository for information and resources on digitization that is useful across a broad spectrum of users.
- Great source of information for institutions at all stages of implementing a digitization program;
- Assembling a set of resources that are useful for those in digitization efforts
- I think iDigBio is a fantastic resource for people just starting in the digitization effort.
- Providing a resource that pulls together the experience and technology of a diverse group of digitization projects.
- I like the emphasis on workflows and the idea of sharing info so we in the community are not all having to recreate the wheel.
- I think iDigBio is doing an excellent job of presenting materials and expertise in digitizing collections to a wide variety of disciplines in the Natural Sciences.

**Raising awareness/communicating (n = 8)**

- Helping the museum community realize they need to digitize their specimen data.
- Raising the general level of awareness about digitization techniques
- Spreading the word about digitization
- iDigBio is reaching out to the digital community and making it aware that various kinds of resources, for various types of collections, at various levels of organization, are available to help improve the accuracy of data capture and dissemination. As the year has progressed, the details of managing the digital and the physical aspects of collections continue to be refined (and defined) by the iDigBio web site.
- And being highly visible!!
- Working with social media
- Nice newsletter. Good frequency of emails/newsletters (not too much, not too little)
- Communication

**Reaching out (n = 7)**

- Reaching out, communicating with prospective contributors. You all seem very active and present at relevant meetings, etc.
- Reaching out to a diverse (many senses of the word) constituency.
- Reaching out to the public; I've not used the portal as I've been told by our project manager that it really isn't up and running yet...I will have to try it out soon.
- Community outreach - workshops, working groups, etc are doing a phenomenal job of engaging with the collections community
- Outreach to the collections community.
- Reaching out to as many folks as possible
- Outreach.

**Facilitating digitization, raising standards and promoting best practices (n = 7)**

- iDigBio is doing an amazing job of engaging a diverse representation of the biodiversity communication. They engage individuals from across the workflow spectrum to identify best practices, define new best practices where none exist, identify challenges and successes. They

facilitate discussion in a way that enables not just individual institutions to maximize the efficiency of their digitization efforts but also for the community as a whole to grow.
- Establishing standards/best practices/guidelines.
- Synthesizing information on best practices
- Creating forums, workshops, and other venues to increase collaboration among the scientific community to facilitate more efficient methods for digitizing biocollections nationwide.
- Getting people together to talk about digitization efforts is great idea and iDigBio does this well. I think the most important part is learning how other people/museums are digitizing their collection. I like specific examples. I like being able to see what works for other people and then applying it to my own collection.
- Facilitating digitization efforts across different collection disciplines and expertise areas
- Assisting with development of best practices for digitization.

**Data ingestion/portal (n = 6)**

- Data ingestion, working with the community.
- The data portal seems to be going well. The team has done well in dealing what turns out to be a complex problem
- Accepting data
- Ingestion and integration of specimen data
- Providing collections information and images in a robust and easily accessible form. Having the collections available in this format is greatly helpful and much appreciated.
- General IT infrastructure; visuals of portal.

**Support TCNs (general) (n = 2)**

- Supporting TCNs
- Providing guidance, news, and support for TCN.

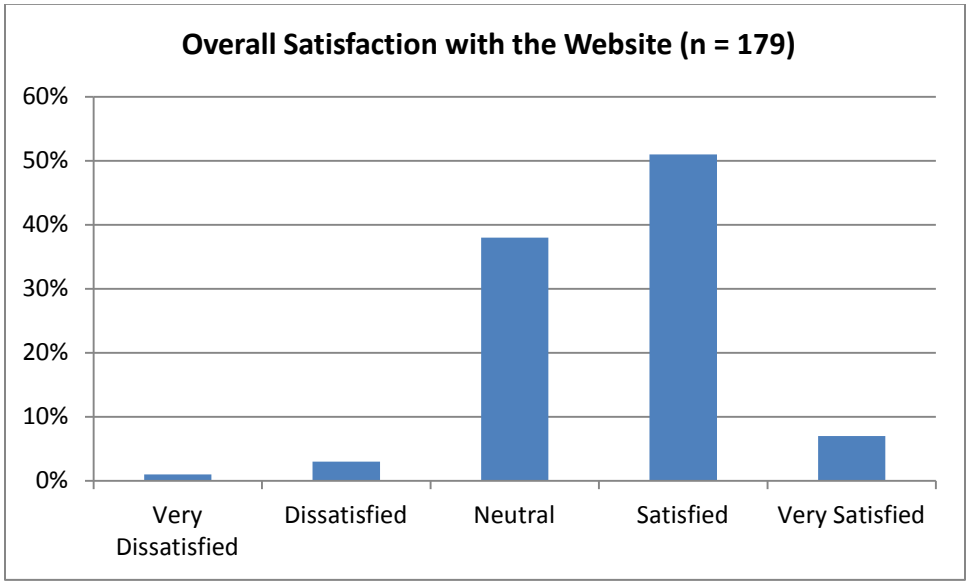
**Supporting small collections (n = 3)**

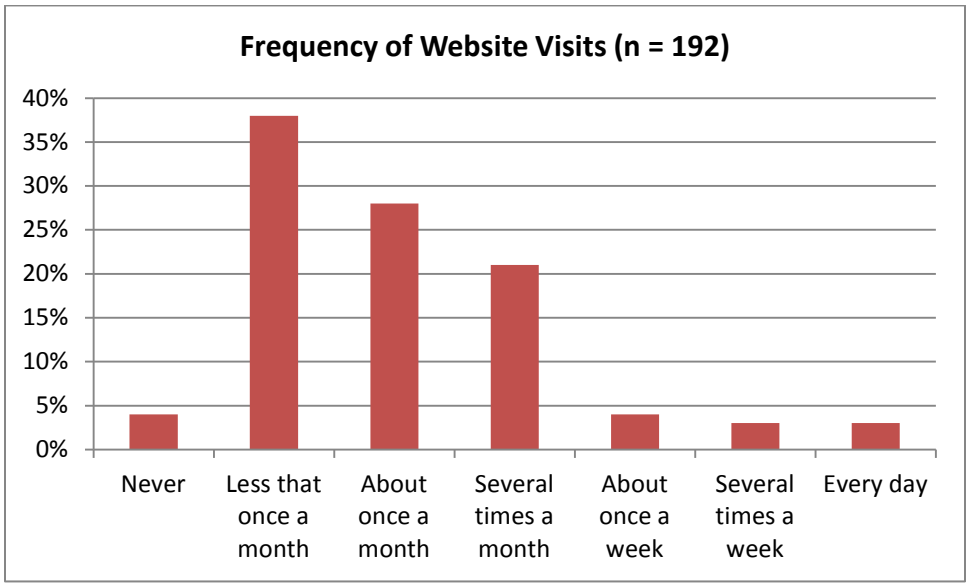
- I believe that iDigBio has done a great job of helping small institutions and in fact an amazingly wide variety of collection types to get going on digitization. I hear from folks that the workshops are quite helpful for many people. There is certainly a huge increase in the number of institutions tackling digitization since iDigBio ramped up.
- Cooperating with smaller institutions who are currently active
- The workshops and workflows seem extremely important to smaller collections getting their data online

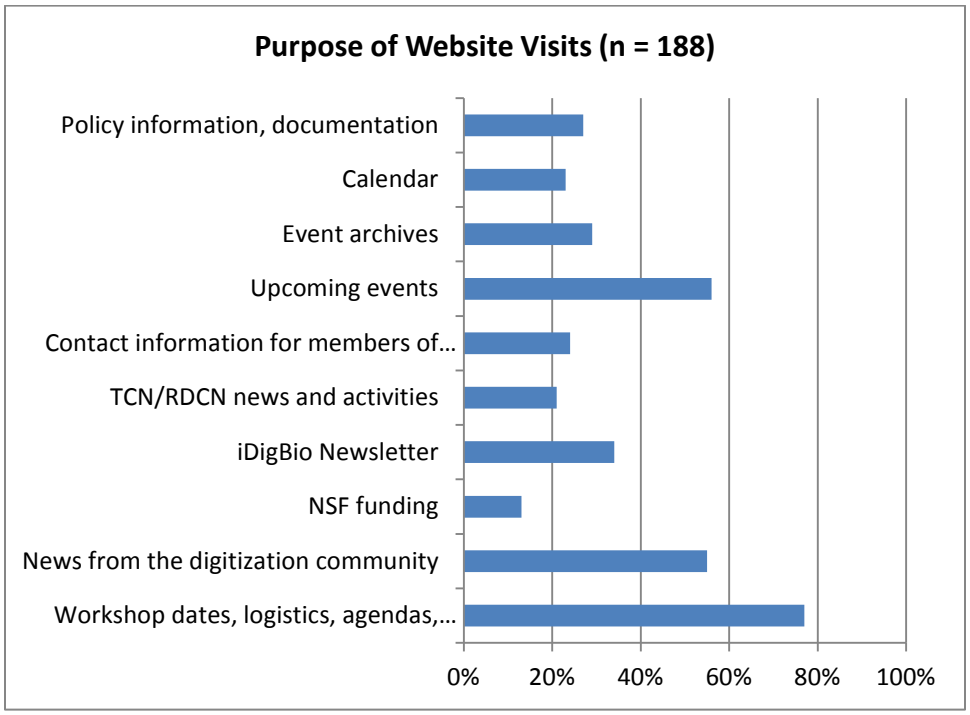
**Other**

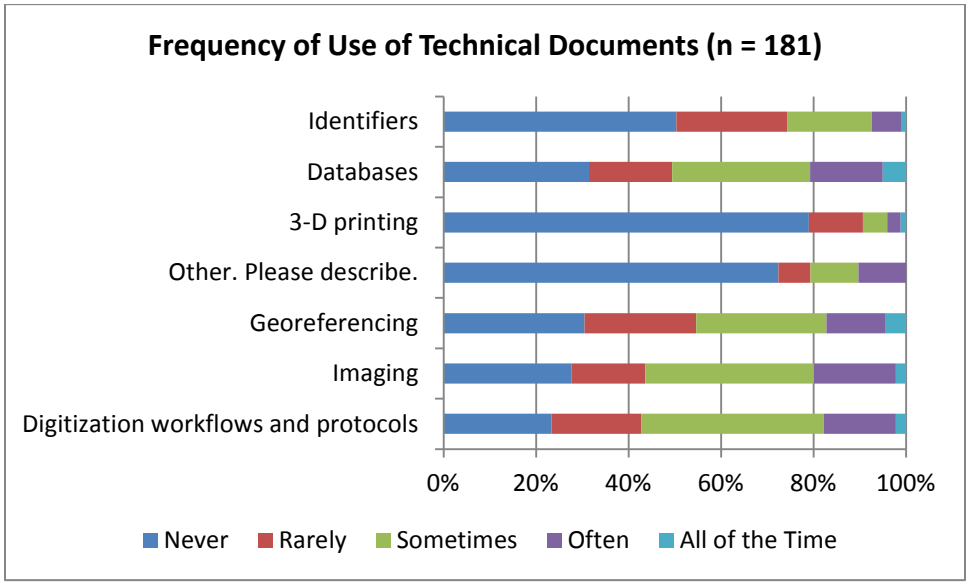
- Very active, helpful, organized, and purpose-driven
- Yes because it has spun a number of successful projects which have secured NSF support
- Gathering the information is the first step. Taking into account the size of the project, the first phase has been done quite well
- Based on my answers it seem that iDigBio does not advertise very well

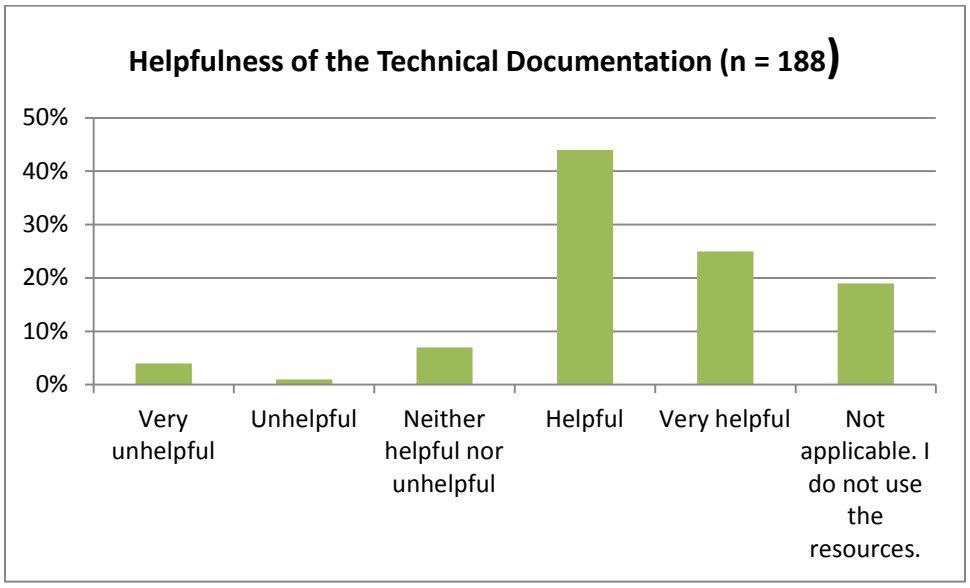
## Appendix B. Findings Related to the Website

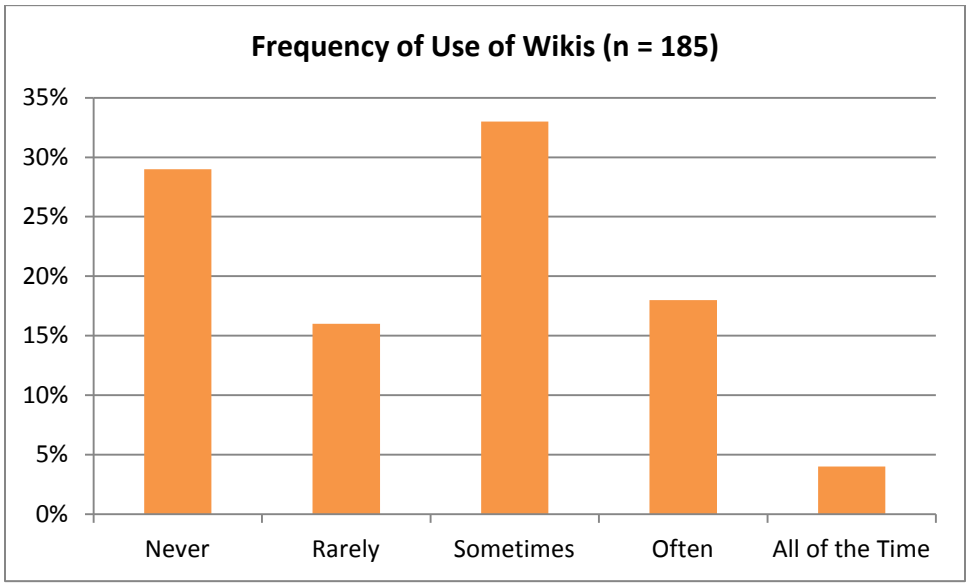


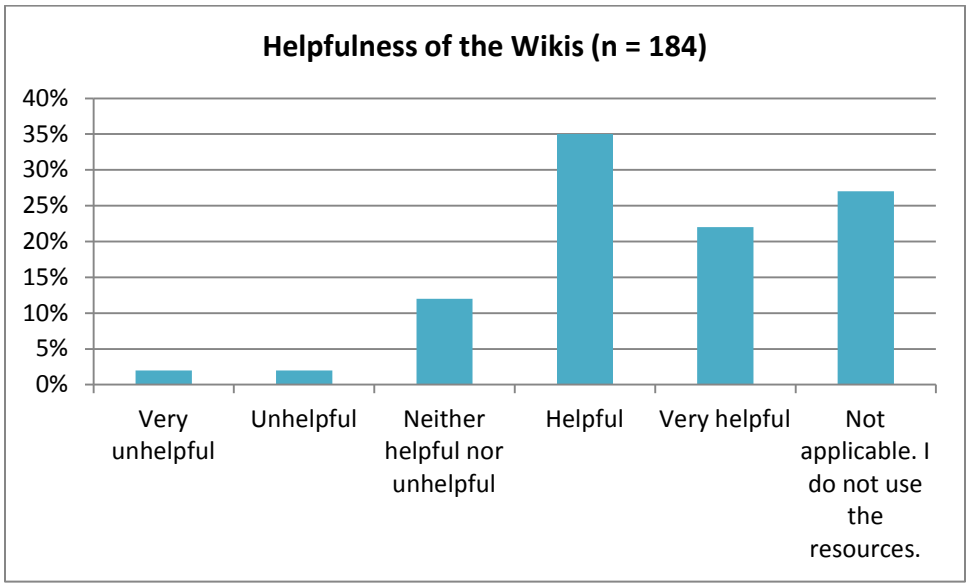



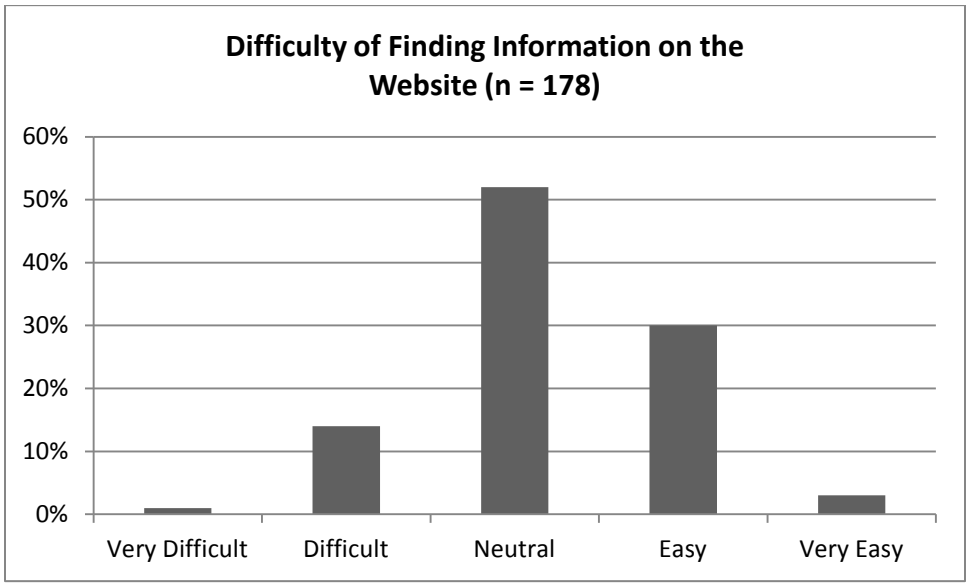

Purpose of Website Visits (n = 188)
Policy information, documentation
Calendar
Event archives
Upcoming events
Contact information for members of…
TCN/RDCN news and activities
iDigBio Newsletter
NSF funding
News from the digitization community
Workshop dates, logistics, agendas,…
0%
20%
40%
60%
80%
100%


Frequency of Use of Technical Documents (n = 181)
Identifiers
Databases
3-D printing
Other. Please describe.
Georeferencing
Imaging
Digitization workflows and protocols
0%
20%
40%
60%
80%
100%
Never
Rarely
Sometimes
Often
All of the Time

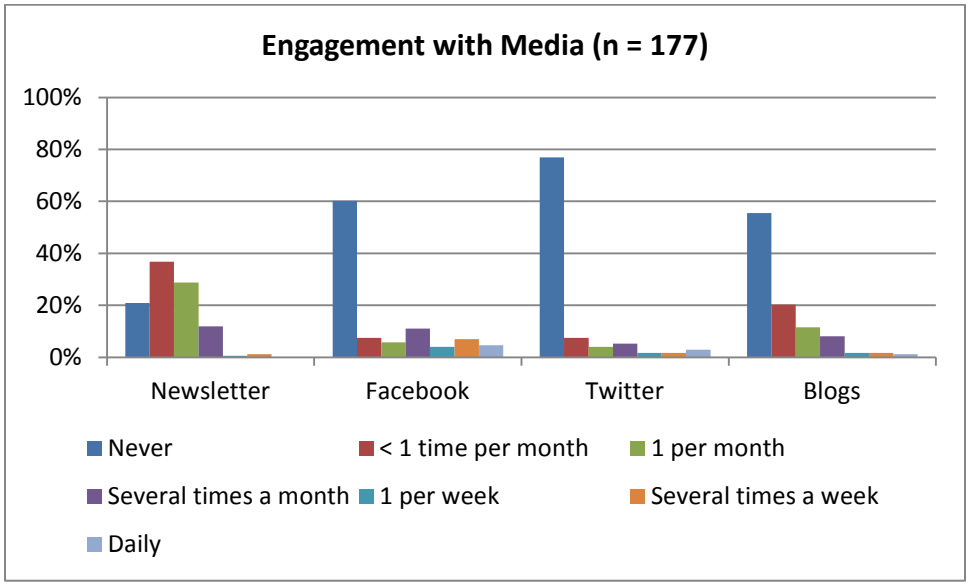

Helpfulness of the Technical Documentation (n = 188)
50%
40%
30%
20%
10%
0%
Very unhelpful
Unhelpful
Neither helpful nor unhelpful
Helpful
Very helpful
Not applicable. I do not use the resources.

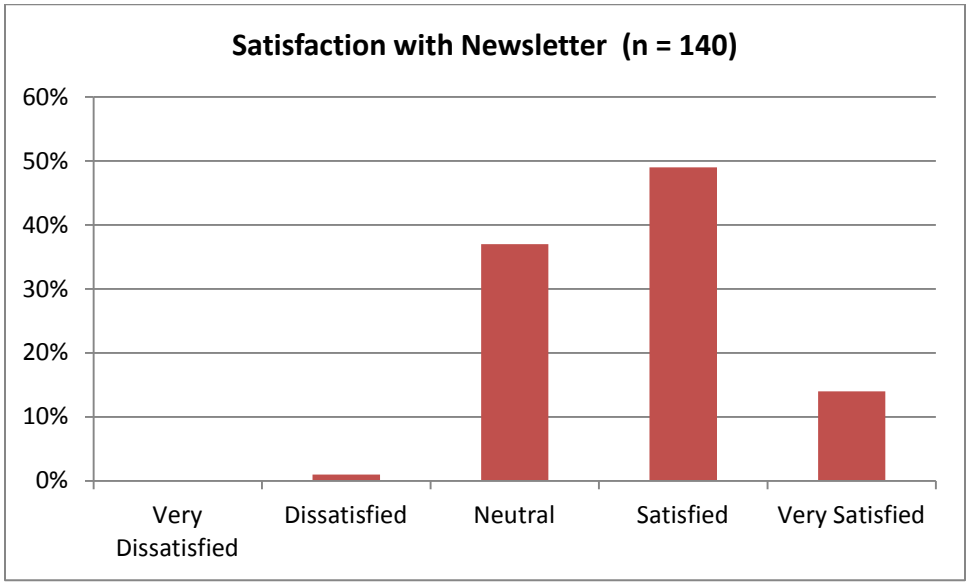

Frequency of Use of Wikis (n = 185)
35%
30%
25%
20%
15%
10%
5%
0%
Never
Rarely
Sometimes
Often
All of the Time

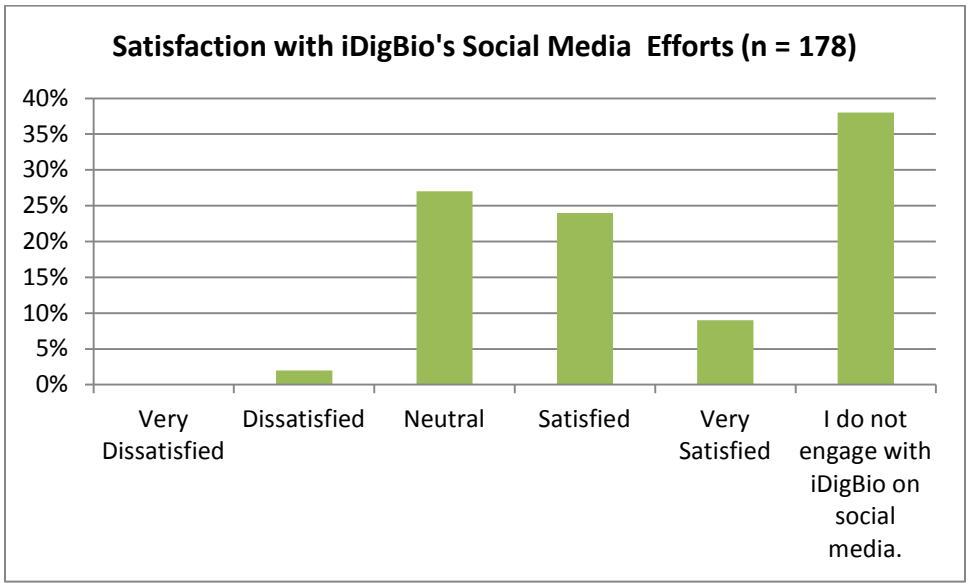

Helpfulness of the Wikis (n = 184)
40%
35%
30%
25%
20%
15%
10%
5%
0%
Very unhelpful
Unhelpful
Neither helpful nor unhelpful
Helpful
Very helpful
Not applicable. I do not use the resources.


Difficulty of Finding Information on the Website (n = 178)
60%
50%
40%
30%
20%
10%
0%
Very Difficult
Difficult
Neutral
Easy
Very Easy

## Appendix C. Communication Efforts




Satisfaction with iDigBio's Social Media Efforts (n = 178)
40%
35%
30%
25%
20%
15%
10%
5%
0%
Very Dissatisfied
Dissatisfied
Neutral
Satisfied
Very Satisfied
I do not engage with iDigBio on social media.

## Appendix D. General comments about the portal, data ingestion process, or service from iDigBio personnel

- Our dissatisfaction has to do with **image uploads**, which have been challenging, partly because the informatics infrastructure is under development

- **Data Field Display** order matters putting them in alphabetical order is inappropriate they should be grouped in standard orders Taxonomy hierarchy for example; **Standard search fields** are not as encompassing as they should be (especially with paleo specimens also being in the data) with unclear documentation about how to add additional fields that some disciplines need to have to have meaningful searches

- **Data sometimes lost**. **Lag** makes it difficult to know the fate of what is submitted.

- We portal our data to GBIF, so it seems a bit redundant to portal it to iDigBio too (I)

- I'd like to hear how the iDigBio portal will work with the VertNet portal and the various smaller portals like Ornis, Manis, etc. It seems like there is some duplication of effort here and I'd like to know more about where this is going in the future. Will all that data eventually be ingested into iDigBio or will I need to do something myself if I want it to happen? Is the portal only for TCNs?

- This is big topic -- does iDigBio have a working/evolving statement to what extent data can be directly edited (improved/enhanced) IN the iDigBio portal? What are the plans here to coordinate with other data environments in which data cleaning occurs directly?

- I think the data mobilization team is doing a great job making the community feel valued. (I)

- Well my one thought is that they have a LOT to work on, and many more tasks ahead to do everything being asked of them. Perhaps it's already happening, but I you should try to clone Alex :)

- Some on the cyberinfrastructure team have very little respect for the scientists for whom the portal is serving. I never know what they are working on, and often they are working on features that they have been told to wait for further input from the biologists (e.g., mapping). (I)

## Appendix E. What is iDigBio Doing Less Well? (I = iDigBio respondent)

### Reaching Out

#### Broader Community

- Soon iDigBio is going to have to turn their attention to building the user base for digitized specimen data from users outside the collections and systematics community -- the problem of how to make these data available for the wide range of uses we believe that can be useful has still not been addressed head on in a very effective manner. To be sure, this is probably the most difficult aspect of what iDigBio needs to accomplish.
- Appealing to non-scientists (downstream users) [I]

#### Other Initiatives

- Integrating with other initiatives [I]
- Connecting to GBIF

#### Other Experts

- From what I have seen, many of the meetings have the same people attending. I think recruitment of more collections and different experts would enhance the program.
- Reaching out to PIs for perhaps a bit different perspective on database issues. It seems the contact is primarily with the project manager (which IS good), but the PIs may have stuff to offer that doesn't necessarily reach iDigBio that I think might be important.

#### Small Collections

- Engaging the smaller collections/museums and those collections with limited funding; Engaging those not already in TCNs or PENs;
- Engagement with small museums - there are some efforts, but it could be better. (not a hopeless issue, as small museum participants are in many of the workshops, but the resources often seem to be tailored for places with big budgets and staff)
- Bringing additional small museums into the fold

#### Large collections

- I'm not sure that iDigBio is connecting effectively with larger museum staff. I think there could be a feeling in a big museum of "we're already there" or "we already know how to do this" and those users might not connect with iDigBio well enough.
- Getting larger collections mobilized

#### Collection Types

- I see some concern about different level of enthusiasm seen depending on the collection types: zoology (especially vertebrates) are not as excited about digitization as entomology or botany or paleontology but ideally the entire community should make progress in increasing accessibility.

- They have a hard time addressing the areas in the most need with the less representation like marine invertebrate collections and terrestrial vertebrate collections but they do a great job addressing the areas with the most success like botany and entomology
- Travel funds are often limiting, making it hard to participate in workshops. Work on expanding remote participation and making recordings so we can view later, I hear a lot about TCNs- what about the rest of us who don't have one of these? Where/how/when do we fit in? I'd love to have some data checking tools to use on my database I heard a lot of "let us get up and running" at first but it's been a couple of years and it seems like there should be a plan- I'd love to see it
- Maybe reach out outside the [end of response]

**Website**

- As mentioned earlier, I would love to see the iDigBio website become more of a central point of access for collections. So, continuing to ingest data, developing the forum maybe in a Stack Overflow kind of format, and organizing the site to be easier to use. However - the site is one of the best I've seen in the natural history community!
- The website and wikis have become a little too dense with content that it is hard to navigate it all and find direct answers. Some simplified summaries and purging might be helpful.
- Sometimes information gets buried on the website and can be difficult to get to. For example, if you didn't know the information was in a wiki, you could get lost for a bit.
- My answer to this is directed to more of a heads up than a doing less well answer. The amassing of such an abundance of information on collections issues will eventually lead to a point where it will become more difficult for the user to quickly access needed information from the web site. Since the site is currently not overwhelmed with too much information it may now be the time to evaluate the web site format and consider the changes needed before it becomes difficult to manage. I would hate to see the iDigBio web site not being used because of too much information that is difficult to find. This concern may have already have been pointed out is currently being addressed.
- Presenting synthesized information on processes / techniques (what was learned and presented from various workshops) to those the general community (those that couldn't attend the workshops). It appears to me currently results and information are provided on a workshop by workshop basis, so someone would have to go back to a certain workshop to find out information on certain processes. I think iDigBio has developed some great information on processes, methods, etc. but I think these need to be organized and synthesized somewhat differently from how they are presented currently in the website to be more accessible to the wider community. In other ways, those resources are provided right now under a certain working group or a certain workshop, but your average person not familiar with the structure of iDigBio will never be able to find these resources as they are not organized by themes such as Digitizing Specimens that talks about a specific action someone at a museum would do.

**Portal**

- The portal remains very basic, compared to, for example, Specify, GBIF, ALA. No mapping is provided, and the image search interface is especially basic - providing a dump of all images, specimens through labels

- Portal search result page not so easy to work with. CSV download from Portal will be nice. Would be nice to standardize collector names (it's a bit of a mess since not all have first name or initials, there can be 5+ versions of a name for a single person). Link to Tutorial on the Portal Home page is broken.
- Making the iDigBio portal the "one stop for data" that is should become.
- Managing the data that we upload. Getting the data into a database through advancing the label data transcription in some way.
- Developing a wonderful portal that appeals to biologists. So what that it is technically advanced? [I]

**Supporting research**

- Defining interactions with more advanced, and more research-y, biodiversity informatics projects. How much of iDigBio is "just store, display, extract what is submitted" (as is); versus "let's have users enhance data/"own" those enhancements, somehow.
- Engagement in the research process and research community. How will researchers contribute? Improvement of status of collection managers, or collections people as researchers.
- Outreach about how the portal of the database is helping collections and research [I]

**Help with funding**

- I think it would be great if they helped museums find funds to implement new digitization practices.
- Offering assistance and advice to the museum community regarding accessing external funding sources (NSF/ADBC, but mostly otherwise) for new and ongoing digitization initiatives. Often resources are the primary stumbling block towards advancing digitization at the institutional level. Providing guidance once funding is secured, however, is a crucial role that iDigBio continues to play.
- The real problem that most of us - particularly those of us with large collections have is 2-fold: 1) convincing the upper levels of administration to dedicate the resources necessary to digitize these very large collections, and I believe that somehow iDigBio should be able to help with this more; and 2) funding.....the real problem is that we need funding beyond what is available through research question driven TCN's. I don't know what the solution to the funding issue is, but I am worried that lots of money has been spent on workshops that could have been spent on data entry.
- Groups not successful in securing funding have disengaged. There has to be some person who would be willing to serve as a consultant for groups preparing grants for database projects without experienced iDigBio participants. With that being said, the iDigBio team has a huge charge and has made excellent progress in creating a synergistic databasing network.
- I think we can all agree that digitization is important, and can certainly work out best standards in the long run. But the main impediment with digitization continues to be the lack of funding and human resources. I would like to see more effort placed in providing and helping researchers, curators, and managers successfully obtain funding so they can implement best practices, training and digitization of their data and collections. [internal]

**Communicating**

- Some parts of iDigBio have been a bit tone deaf, but in all cases have come around. I personally think that we could use more leadership in helping TCNs understand and implement their software choices.
- Hard to communicate clearly all the activities and the relationships between them
- Not so clear how iDigBio is going to get to the presumed end target of having all data in one portal- I'd like to know more about what's coming up and what I as a curator need to do to participate.
- I don't really understand much about iDigBio as a portal or app developer; I would like to know more about this, rather than just hearing that the resource is there.
- Communicating the work being done and the reason it is useful to a general public.
- Creating concise and clear best-practice manuals
- Also, other than attending webinars and workshops I'm still not really familiar with iDigBio
- Following up with workshop attendees to see if the workshop help their digitization efforts or if they are still struggling with something;
- Communication in general (I)
- Broader penetration into related life sciences might bring about greater awareness of what iDigBio is doing, and the resources that are being developed. Having a presence at meetings that are related but not directly tied to collections will open up more awareness and participation by potential affiliates. [I]

**Other**

- Ignoring the elephant in the room called Intellectual Property and Copyright
- Fluid preserved sample data capture
- Perhaps there are a few too many workshops and working groups.
- I think where iDigBio drops the ball is in understanding that a lot of collections need basic curation before they can even start to digitize.

**Things are good**

- They are already improving on the areas I was previously concerned about. Keep up the good job iDigBio.
- My experiences are all positive.
- None. I think everything is going pretty well (I)
- Perhaps the functionality of the hub is not as great as it could be, but it will get better. [internal]

**Appendix F. Tasks iDigBio should focus on during the upcoming year. (I indicates iDigBio respondent)**

**Figure out solutions to hard problems (n = 12)**

- Although I do not know of all the new initiatives, the low-hanging-fruits have been addressed for the most part. Now need to attack more difficult issues such as **extracting data from labels in small vials (i.e., fluid-preserved)**.
- Figuring out a solution to **OCR and transcription**. Pick one and move forward either OCR or Citizen Science rather than this state of limbo.
- Getting **data from the labels** into a database.
- More tools to **improve the efficiency of digitization** and make them **more easily accessible** (maybe more publicized) for general (non-TCN) users.
- There are **support data**, such as collector lists, locality lists, etc., that all digitization projects must build on their own (the smaller institutions do not have the luxury to do this). I don't think it is necessarily iDigBio's responsibility to build these resources, but they could provide a framework for some working groups to start constructing these. Everyone would benefit.
- Developing tools for **seamless cross referencing** of collection records to derived biodiversity data products (images, sequences, articles, etc)
- **Resolve the GUID question** in regards to submitting data to iDigBio.
- More support for other **techniques for creating text records from images** (beyond OCR, which for many specimens is a dead end).
- More attention to **how existing digitized data can be leveraged** for populating data records for newly digitized records.
- **Ensuring no uploaded image data are lost**.
- Developing **tools/protocols for data quality control** and error correction.
- Address the linking of taxonomic information to specimen data to assess the reliability of specimen identifications (I)

**Reach out to other communities/projects (n = 9)**

- Work to expand the digitization effort beyond the relatively small number of non-federal collections supported through ADBC
- Linking with other nodes/regions to develop regional (i.e. outside the US) nodes of iDigBio. I work closely with the Atlas of Living Australia and I'd like to see them more engaged with iDigBio.
- How do the rest of us without TCNs fit in to the effort?
- Interfacing with various user communities (professional societies, collections societies/groups).
- Focus on filling in the voids. Identify the groups of extant/fossil organisms that are not currently being databased and plugged into the iDigBio project.
- Digitizing marine invertebrate collections
- Collaboration with other large scale international digitization projects.
- Data sharing with other exiting digitization projects (I)

- Rounding up tools and people to make a more 'integrated into the community' offering. NSF should be using a business model instead of an academic model for organizing efforts like this. What about some mergers and acquisitions, e.g., VertNet? (I)

**Continue training and other support (n = 7)**

- Keep being available in case any of us "out there" need help.
- How to sustain the great efforts you have started. Stay the course with old initiatives. If you keep adding new initiatives when others aren't completed then no substantive progress will be achieved.
- Help facilitate for TCNs.
- Continue to provide georeferencing training
- Continued georeferencing training
- Workshops
- Inadequate technical expertise with data storage and data management is a very big problem for many in the community. In a brief conversation over lunch at a recent iDigBio workshop, I learned about the enormous potential for inexpensive data storage in the cloud. From the perspective of a curator of a small collection, the cloud option would seem to be a viable one.
- Further development of teaching and training tools. Especially the continuing promotion of virtual access to workshops and symposia. Focus needs to be placed on **production of videos and converting some of the power points and videos in to more finished products** as well as organization of them on line with easy access. (I)

**Facilitate funding (n = 7)**

- Provide more resources or links to funding agencies.
- Poll contributors and assist with compiling information regarding creative project funding sources.
- Continue to encourage NSF to select a wide diversity of project types (taxonomically and otherwise) for TCN funding.
- Defining "boundaries/opportunities" for derivative biodiversity informatics research. Where does iDigBio stand, e.g. regarding standing/projected NSF-ABI projects?
- Funding is a huge problem for all. iDigBio has done an excellent job exploring and promoting alternative funding sources at workshops. I'm sure additional local funding would be made available if administrators of academic institutions were made more aware of the importance of digitization of collections.
- Funding efforts to support digitization (I)
- Assistance with funding for staff to complete digitization efforts (I)

**Improve data portal (n = 6)**

- Make the data portal easier to use.
- Getting the portal to be user friendly and useful.
- Fix the portal and get the ability to download data.
- User interface for the data.
- Better visualization tools for portal, especially mapping and images

- Must share data with GBIF! Data portal and services up! with methods for accessing portal for research (not CSV download) Accept data from publication (i.e. Biodiversity Data Journal IPT) Better documentation for 'one off' sharing of data. How to share if a small collection or individual researcher.

**Assist small collections (n = 6)**

Helping smaller collections get their data into national/international databases and portals.

- Assisting smaller collections
- Developing a network of professionals who know digitizing strategies that can share their expertise/train others to carry out digitizing (regional or state contacts/mentors that smaller collections or those less versed in digitization can contact for assistance - sort of like the train the trainers on georeferencing but on digitizing best practices techniques) especially as it could potentially allow smaller collections to get involved.
- Establish a more cohesive network that helps smaller collections get to know the larger collections in their area so that they can partner to share digitizing knowledge/resources and or learn how to digitize (we are greater together). Thereby allowing bigger collections to help smaller collections find a voice.
- iDigBio did an excellent job of bringing university administrators and curators together at a recent workshop on sustaining digitization at small collections. I think such efforts will be productive and should be continued.
- Has iDigBio ever discussed with SPNCH or like programs the possibility of creating "most improved" or other kinds of awards for collections? Anything (!) that help raise the profile of smaller collection that can help highlight their importance and give academic administrators something to crow about (I'm not saying this facetiously) is a good thing. Collections need to be on administrators' radar screens as active, important facilities. How could administrators consider closing a facility (or shipping it to the nearest Division 1 school) if the facility had recently received a national award of some type? (Most improved digitally, most improved website, best record of scholarship from a small institution, highest level of outreach, greatest participation in curation by undergrads, etc..... let's all put on our thinking caps and come up with some other possible ways to recognize facilities.)

**Data ingestion (n = 7)**

- Keep aggregating data...the more the better.
- Probably data ingestion and making folks aware that these data are available.
- Data ingestion is huge!
- Continuing to ingest data. (I)
- Encouraging additional collections to provide data for ingestion (I)
- Again showing how those datasets and the movement of the data to iDigBio is especially helpful versus the other options for database portals already available like GBIF, BISON, etc. (I)
- Continue to expand the amount of data and collections digitizing.

**Promote use of the data for research (n =5)**

- Coming up with research projects for the data that has been made available. Outreaching to researchers to let them know about data availability and how it can be applied.
- A workshop focused on digitization for research purposes
- Data use cases, research applications, more community involvement and interest (using the data) (I)
- Develop cyberinfrastructure and other methods to use the data in research (I)
- Increasing use of data in the portal. (I)

**Crowdsourcing (n = 4)**

- Developing models for crowdsourcing like some kind of open source template that institutions could install to connect the public with the database and records to finish
- More support for crowdsourcing for specimen label transcription (I know they are working on this)
- More crowd-sourcing options/resources/information
- How to start your own crowd-sourcing initiative would be a great workshop/symposium topic. This is one way to really help places without a TCN or other digitization grant (which is most places).

**Standards and best practices (n = 4)**

- Personally I would like to see emerging or potential legal issues associated with collections digitization: intellectual property and associated ethical dilemma pertaining to biodiversity data. iDigBio can serve as a clearing house to develop best practices
- Reach agreement and provide additional best practices documents whenever possible.
- Continued development of standards and workflows.
- Dissemination of best practices

**Raise awareness about digitization (n = 4)**

- Spreading the word about the national digitization effort. There are a lot of institutions that have never heard of this effort or iDigBio.
- Increase awareness further throughout the Pacific especially Papua New Guinea.
- Reminding everyone of the big picture
- I think collection work, use and innovation within both needs to be promoted. Collections might face scrutiny with recent criticisms about collections and wildlife conservation.

**Clearinghouse for documentation (n = 4)**

- Continuing to coordinate and serve as the place to get documentation for digitization projects.
- Continue as a repository and facilitator of technological advances, workflows, and new ideas.
- Synthesizing and providing some of the great information and resources that iDigBio has in a more user friendly way.

- Create a directory of U.S. collections (I)

**Facilitate collaboration (n = 3)**

- More resources to collaborate online with people at other conferences (ex. a Linked In of iDigBio people), more opportunities to network
- Some of the digitization projects are in their last year, so perhaps developing or encouraging the continuation of communities of practice so that those that have the most knowledge from their digitization experience don't totally leave the group but are participants that can help share their expertise and provide guidance to those that have not embarked in those processes yet.
- Continue to bring us together to share ideas and form collaborations.

**Education and outreach (n = 3)**

- More education and outreach (I)
- E & O (I)
- Develop cyberinfrastructure and other methods to use the data in education and outreach (I)

**Public outreach (n = 3)**

- Make communications to the lay audience of greater importance and continue posting on social media, namely Facebook.
- The outreach to public participation is crucial to the success of national digitization efforts (I)
- Developing ways for downstream users to engage with iDigBio. (I)

**Sustainability (n = 3)**

- Sustainability of digitization efforts and related software. (I)
- Plans for sustainability (I)
- Strategic planning and distilling strategy down into objectives all the way to tasks and goals for sub-projects and staff. (I)

**Document use of specimen data (n = 1)**

- People who publish using specimen data should be required to submit those data to an online database prior to publishing just like we have to do with genetic data + GenBank etc. There are far more people using specimen data than staff at the museums that curate the specimens so these people's efforts should be tapped to help digitize the world's specimen data.

## Appendix G. Biggest challenge facing the national digitization effort

### Resources (n = 33)

#### General

- Biggest challenge – resources. Need to find a way to show how digitisation complements research not competes against research support staff.

#### Time

- Time to get everything done electronically.
- Some of our biggest challenges institutionally are just having the staff time to work on our data. We have such a small staff that demands on our time for other things, such as research requests and visits, exhibits, programs, volunteers, end up eating all our time.

#### Funding

- Increased **funding** that will affect a broader range of stakeholders. The present TCN's only benefit a minority of individual collections across the country. While I am supportive of the goals of the TCN program, there is a need to raise the bar and accessibility to digitization funding across all collections. (I)
- **Funding (I)**
- **Funding**
- Probably **funding**. By advocating for the continuation of the ADBC program into the future for several more cycles.
- **Funding** cuts to collections. Digitisation is often seen as a "bonus" that shouldn't be prioritised over core curation and taxonomy (I don't agree but it's hard to fight this view).
- **$$$$$$$**
- Acquiring digitization **$$$** support for digitization large Museum collections. iDigBio can continue to contribute workflows that help with proposal writing.
- **Funding**. Institutional resources are frankly quite often pulled in too many directions to allow for major institutional contributions towards individual digitization initiatives (at the department level). We continue to work towards convince our administrations that we require additional support and assistance for our endeavors, however perhaps iDigBio might facilitate the collections community as we explore creative/alternative options for funding/ramping up our individual efforts towards enhanced digitization.
- **Funding**
- Obtaining **funding** for people and equipment
- There are lots of organizations and Institutions throughout Pacific that work on preserving biodiversities physically in their respective localities but having them recorded electronically for

all to see is not very good because of lack of infrastructure and **funding** being the major constraint.
- Prioritizing and **balancing funding** initiatives effectively and equitably to achieve the most efficient use of limited resources is very important and will be very difficult.
- **Funding.**
- **Funding!**

**Time and funding**
- **Funding and time** are always the biggest constraints. Museum budgets keep getting slashed. The amount of **time and money** necessary to digitize the collections.
- **Funding and time**. Digitization has actually increased our workloads, and not in only trivial ways. Even a one-semester buyout of one course would help me a lot. But iDigBio does not exist to disburse funds.
- **MONEY and TIME**!

**People**
- Still are **human resources**, even with all the technical improvements.
- **Resources to hire staff**…**lack of willingness of many collection staff to participate** because many institutional cultures do not value and reward curation
- Many collections lack the **staff** and resources to digitize their collections, and even those collections that have received TCN/PEN grants are only able to digitize a small proportion of their records/specimens. Digitizing all ~1 billion specimens in the US is beyond ADBC, and yet getting all of the collections digitized and ingested is crucial. By continuing to support training, iDigBio can help make digitization happen. (I).
- Again, there are a lot of collections that due to lack of funding or **staff** that can't get above water with basic curation, let alone think about digitization. Perhaps iDigBio could help funding institutions to recognize this as one of the other bottlenecks to digitization.
- **Dollars and ultimately staffing**. I think that iDigBio should do more to publicize the efforts, beyond the collections/natural history community.
- The same ones as always: **time, person-power, money**!
- Per item **cost** of digitization
- Without **trained people** to do the work it cannot be done.
- Funding, human resources. Training discussions. Successful funding initiatives. (I)
- Lack of expertise and coordination. But iDigBio has made a great progress on addressing these challenges

**Small collections and others not affiliated with TCNs**
- This is tricky since the biggest obstacle for digitization is funding. Funding for staff and equipment. iDigBio helps people who want to digitize but if those people do not have reliable outlets for funding, the training doesn't do much. iDigBio is already trying to rally together

small collections and see who can work together. Collaboration of small collections will be a huge help. While large institutions have a broad variety of specimens and localities, small collections are significant and iDigBio should continue working on developing an online community of small collections personnel so we can figure out a strategy to move digitization forward with minimal funds.

- Funding for digitization efforts. Currently, I am not enamored of the funding structure. Unless you can be integrated into an existing TCN or have the IT infrastructure to start your own, you are out of luck. Make it easier to fund basic digitization and you will get a lot more data. (I)
- Getting those who are not part of TCNs or PENs digitizing when money is tight. Help establish/setup equipment that can be loaned (e.g. scanners, cameras). Establish a network of individuals that can be technical advisors or mentors to collections without or with limited technology background/ability (not everyone who works with collections are comfortable with computers and some only know enough to get by - especially in the smaller less well funded collections).
- Among others, finding the resources to digitize small collections, which won't ever get to it otherwise, because in many cases they have no staff at all. Digitization has to be done to these collections by those at larger collections or resources have to be provided by granting agencies.

**Sustainability (n = 10)**

- Sustainability (I)
- Sustaining extant efforts post-funding.
- Sustainability (I)
- Sustainability and how to ensure expertise is sustained through various cycles of funding.
- Sustainability, given that so little new money is forthcoming
- Sustainability
- Financial sustainability (I)
- Sustainability for the digitization effort that iDigBio now leads is the biggest challenge, since it's pretty clear that we will not digitize all the natural history collections in this country by 2020. I know there is a committee working on sustainability -- this effort needs a concentrated approach.
- I am also concerned about what happens to these ADBC activities once the 10 years are up. iDigBio should continue to lead discussions about sustainability of the efforts. (I)
- In the long run we need models for sustainability, both of the digitization effort and the technology that facilitates it. Like many things the NSF does, both the foundation and iDigBio have said little about these aspects of collection digitization. In short, we need a business model for this kind of work, and few viable suggestions have come forward. Institutional collectives might offer an option, wherein a few institutions have direct responsibility for server, software, and data maintenance, although the financial responsibility is shared broadly.

**Crowd sourcing**

- I think the only way we can get the bulk of our collection digitized is through crowd sourcing. It would be great to have some easy to use tools to set up crowd sourcing or a site where we could just submit photos of our labels to have them digitized.
- Lack of participation in crowdsourcing efforts.

**Duplication**

- There are so many portals for our data, but which ones are actually effective and utilized is still a question. (I)
- There are many digitization efforts afoot, employing diverse methods, technologies, and standards. Pulling all of this together is daunting but essential. iDigBio has done much toward disseminating information about best practices and standards, and this should continue to be a high priority.
- Buy-in from people already happy with GBIF. Why switch? iDigBio needs to demonstrate its value to me. (I)
- Too many different portals, databases, and data collection platforms. Simplification would be good especially for those of us who want to quickly get data out without having IT expertise.
- Project duplication refers to the fact that the general structure of digitization funding encourages massive duplication of effort: it's easier to get a grant to build an entirely new portal than it is to fund data entry into existing portals. The result is a mess of local portals and various levels of aggregators, which would be less of a problem if it weren't for inevitable errors introduced at each new data-ingestion event -- not to mention the difficulty of propagating corrections up, down, and sideways across every single portal. It's insane.

**Data quality**

- I guess one of the possible challenges would be that some of digitized resources including images turned out to be actually of little use in scientific research communities due to data quality issues or research methodological constrains even if they may be useful for educational and general reference purposes.
- Making sure data is clean. I think that iDigBio is already doing a lot to address data quality - digitizing supplemental documents, and so on.
- Reliability of specimen identifications (I)
- Data quality: more resources focused on reviewing submitted data for consistency; creation of semantics to improve searching. Sustainability: make it so important that it must survive. This means proven for research.
- The two biggest problems for digitization in general are dirty-bucket tactics and project duplication. I'll explain both sequentially. Dirty-bucket tactics refer to the practice of getting as much data online as fast as possible, with little effective proofreading. It makes for databases

that are unreliable and barely usable. I work on a small-scale digitization project, and my inability to proofread most of my data is terribly frustrating. We could be so much more useful if we could be more certain of our accuracy! Unfortunately, the funding just isn't available to "clean up our bucket" by hiring better-educated data-entry personnel and proofreading more thoroughly. We have neither the time nor the money.

- Clean useful data with reproducible results.(I)

**Getting everyone on the same page**

- Getting all the players to have the same daily priorities and to move in the same direction. (I)
- Getting people on the same page. Setting basic standards so that the data can be useful. Doesn't have to be too overly-complicated.
- I think people need to be able to answer the question of: why should I put my limited resources into sharing, what's in it for me?
- Getting people more on the same page as far as workflow, data flow, portals available. Narrow down the choices and help point people in the right direction.
- Coordination with all partners and getting them to participate would be the biggest challenge, however, it seems you are on the right track and overcoming this well.
- We need to develop community support for use of iDigBio - possibly through a service like Dryad or various publishers (like GenBank was able to do).

**Software and other tools**

- The lack of adequate software to digitize specialized collections is going to continue to be a problem and will mean that a lot of data just aren't going to be mobilized.
- Lack of user-friendly tools for digitization and related information management seem to be one set of challenges. Awareness of efforts, working groups, and collaboration and training opportunities are ways in which iDigBio combats this. Possibly more tool and resource development in coordination with other projects could also help. (I)
- Simplifying both the process and software tools in order to promote the adaptation of digitization efforts, and encourage efficient workflow practices.
- Moving from images to values in text fields (I)
- Improving OCR accuracy so it will be a major time-saving measure.

**Supporting small collections and/or those not part of a TCN, etc.**

- Lots of talk about importance of small collections to digitization effort, but absolutely no understanding of the real challenges faces by curators of small collections and the types of support that are actually needed.
- Dealing with all the un-digitized specimens that aren't covered by a TCN. In reality, most collections won't get a TCN- does this mean they're left out?

- Reaching obscure or small collections (I)

**Improving the user experience**

- User interface is the biggest challenge confronting digitization efforts. Think Apple - before the Apple II computers were for the very technically inclined. That is where we are today with natural history collections data. Someone needs to make interacting with the data easier, more graphically rich, and more intuitive. There is so much happening with data analytics today, and those principles need to be incorporated into the user interface.

**Policy**

- One of the biggest challenges is inconsistent policy on things like imaging rights, etc., across museums. Although museum scientists often make a big deal on "collections being held in the public trust," it is a little disconcerting to see multiple layers of copyright slapped even on imaging data captured by outside workers. E.g., the Harvard policy of any images of specimens being copyright Harvard (as one example--not an uncommon policy), even if they are taken by an external researcher. Right now, many policies seem to serve institutional legal consultants (whose job it is to be risk averse) rather than the public and the research community (who may want to make it easier to share data). A conversation on this issue is an important one, but hasn't really happened in a serious way at the national level.
- Globally Unique Identifiers
- The biggest challenges in my opinion would be to establish guidelines that can be utilized in making sure that the accumulated data is preserved (it's integrity and safety) and maintaining access to this information (continued hosting of the information by the institution - What would happen to a collections information (and the collection) if it becomes "orphaned"?).

**Research using the data**

- Building smarter connections among data; making implicit connections explicit.

**Measuring impact**

- The biggest challenge I see is that all of this great information is being gathered and disseminated throughout the workshops, but the follow through with seeing how these workshops have resulted in better digitization efforts is unclear

**Reaching the public and downstream users**

- Making the project accessible to downstream users (I)
- Collection managers have to manage both specimens and electronic databases. iDigBio could possibly focus on ways to show how relevant and valuable digitization is to general public. Another way to put this would be to help collection managers find ways to express and demonstrate the value of these resources to universities and museums.

- Get the message out! Need to move beyond talking to yourselves and reach out to customers and investors.

**Other**

- The speed of change in technology. How can museums continue to afford buying new cameras and other equipment at the rate required to continue increasing the speed of digitization?
- One of the biggest challenges is connecting related resources within and across institutional boundaries. Within an institution, specimens, literature, and archival materials may all describe the same species and habitats but are often managed through separate information systems without clear connections to one another. Each type of resource may include information not available in the others and can fill in information gaps.
- Keeping up momentum. With fewer NSF grants being funded in this area (or funded at reduced rates) this may lower the support from institutions' administrators.
- Standarization and outreach. (I)
- Continue training and report and share information on data portal usage by people
- Aggregating data from a multitude of sources.