iDigBio Top Priorities

What do you think iDigBio's top priorities should be over the next three years? Do you think there are areas where we don't have sufficient staff or resources to accomplish these priorities?

- Ingest as much data as possible from natural history collections
- Develop computational cloud and user interface to take advantage of new research opportunities using biodiversity data (with an emphasis on ecosystem services)
- Work with NIBA to develop a viable economic model for sustainability of national digitization effort
- Professional development and workforce training
- Education and outreach to broaden public appreciation and stewardship of biodiversity along with improving public awareness and participation with critical environmental issues
- As a data aggregator, ingest data from data providers
- iDigBio may need additional professional education staff in the future to grow our efforts in this area.
- From where I sit, the program still feels very technical and introverted. More information for the public, and more content about results and achievements.
- Data Cleaning and Mobilization tool development (Out of scope, I know... I know). Somebody has to do it though.
- Ramping up E&O and the resources required to do this.
- Improving the usability of iDigBio should be the top priority over the next three years. Different organizations have different standards and iDigBio users will want iDigBio to be more user-friendly through time. The code development of iDigBio virtual appliance may not have sufficient staff when we head to the next steps, such as integrating the ingestion tools and other software with the virtual appliance.
- I think that iDigBio should continue to improve its outreach activities to the general public over the next three years (camps, high school/undergrad interns, etc.).
- Unfortunately, I am not well suited to answer this question. It seems to be a managerial level inquiry.
- Over the next three years iDigBio should motivate a large number of scientists and educators to use and talk about the data in iDigBio. Build tools that are helpful to the community to explore and use the data and find ways to keep this data available for everyone for decades to come.

iDigBio is funded by a grant from the National Science Foundation's Advancing Digitization of Biodiversity Collections Program (#EF1115210)

- In general, the top priorities are: to mobilize more data and improve its quality in order to be able to demonstrate the value of it, to grow organically, and to start looking for venues and processes that would ensure sustainability of iDigBio beyond the existing 10-year roadmap. With respect to IT, the top priorities are to develop according to the existing plan: data feedback workflow, data linking, and advanced data visualizations.
- Building a user-friendly and science-relevant portal that serves a large amount of new digital content (the collection of which was facilitated at least in part by iDigBio). It is my view that engaging the public in digitization has a lot to offer iDigBio and the public.
- Continue to ingest specimen data in a timely way, i.e., when the TCNs tell us they are ready, and going forward with their updates.
- Develop broader community participation in digitization as a whole: small herbaria, non-TCN, education & outreach, research.
- Develop non-mediated input mechanisms for data ingestion to the portal.
- Develop relationships with community groups to use our API to present their own view of our data.
- Develop material that reaches a wider community of interest
- I think iDigBio probably has the resources but might need stronger leadership for these topics. A working group and regular meetings would be useful.
    - small collections training
    - sustainability plan
    - education outreach
- Get the database portal up and fully functioning, hopefully with images. Until there is really a product out there to use, it seems premature to do much in the way of engaging the K12 community, the broader public, etc.
- More or different outreach to the collections community with an overall goal of community building. I get a sense from some that there is an “in group” of large, well-funded institutions that could possibly accomplish digitization without iDigBio and an “out group” of smaller and/or less prestigious collections that may be feeling left out. Also, continue to try to build community among the TCNs. It seems like if this is happening, it is largely because of efforts among the TCNs themselves.
- Take steps to increase involvement of under-represented groups (and this still includes females).
- Continue to work on digitization efficiency and standardization.
- At some point, increase education and outreach once there is something to point to.

- Having a large enough database of digitized collections to be useful.
- Engaging the collections and research community so that participation in submitting and using iDigBio data is the norm.
- Beginning to serve the downstream users. The NSF Site Visit Report suggests the need "to develop education modules and mechanisms to facilitate dissemination of outreach activities by TCNs."
- The NSF Site Visit Report suggests the need to develop "potential future options for sustainability."
- Increasing diversity.
- Education and outreach!! I think that if you can involve the public/citizen scientists in the goals of iDigBio (probably through engaged citizen science groups out of the different TCNs) then they could help to move us more rapidly towards the imaging/labelling/media organization goals of iDigBio
- Broadly stated, from a services perspective, we need to do data ingestion, enable data access and feedback *in the large*, as well as enable certain forms of data processing. From an operational perspective, the IT component needs enough resources to cover each of the design, programming and testing step of the IT workflows. IT staff is never sufficient. In particular, we need someone specializing on data-related IT issues and testing. Resources (storage and computers) will be needed but not immediately.
- Build a portal that provides a sufficient quantity of data in a usable fashion that someone can answer an actual research question. The shortest path to doing this is to find someone with a question that can't be answered with an existing tool and construct something around it.
- Settle the mechanical ends of data identification so information can be moved and copied between locations without relying on internal inspection of the semantics of the data (DWC triple, collection name, genus & species etc.)
- Bundle the infrastructure into a reproducible software or software and hardware package that can be passed to others. In the worst case, this may offer way for the ADBC effort to move forward without iDigBio if we're non-renewed. In the best case this may provide other sites with a way to contribute storage and compute resources without simply handing iDigBio hard-to-come-by IT dollars.
- I think our top priorities should be to address those issues that NSF raised in their review. For me personally it would be to address our web presence and how the portal can be better integrated with our CMS.

- Publicity: What is iDigBio and why should people care (or not be scared/wary on yet another big NSF juggernaut)
- Outreach: Development of mechanisms by which the image hub will be utilized by more than just museum professions - specifically the public, educators, and citizen scientists.
- Dedicated help to re-design (and perhaps separate) the Drupal site from the Wiki site or just use one (and not the other).
    - Latest hire, Greg, may be doing this?
    - Templating would be useful now to help organize and keep up with linking all the material together.
    - Usability testing of site/s – a few people at a time > re-design > repeat.
- Need staff to help with data organization and versioning.
    - possibly, in conjunction with E & O, work on our eLearning materials
    - enhance our workflow documentation
    - enhance our presentations (prezi?)
        - possible help from Robert Hanner (from Univ of Guelph) or someone from UF or FSU Distance Learning
- Development of Web Services for the community
    - Need to make it possible for community to implement the digitization strategies we are discovering that help make process more efficient and faster
        - Example OCR services (See SaaS)
            - https://www.idigbio.org/wiki/index.php/OCR_SaaS
    - Alex has said we do need staff for this (maybe a Grad student)
        - Could be at UF or FSU
        - A position may make this happen a lot faster than a graduate student?
    - More web services for Georeferencing.
    - Setting up easy system for community to utilize crowd-sourcing to manage data input into their databases.
        - Note just learned from visit to Smithsonian – they are planning to develop a crowd-sourcing tool.
            - Another example of how we're all re-inventing the wheel…
            - Need to share our crowd-sourcing tools
- I think iDigBio should focus its efforts and resources on the ingestion of data, collaborations with existing projects and sustainability (turning iDigBio into a revenue generator). I feel that resources, both human and physical, will be necessary for iDigBio to operate at the

scale and scope suggested by the collaborative agreement. These areas include: user services, systems administration, marketing/branding, and physical space for meetings and workshops.

- We should strive to become the source of specimen information and related media objects for the US biodiversity community. To do so requires acquisition of information, discovery and searching capabilities, and packaging of information for export. We need to be finding and acquiring many sets of digitized data. We will also need to become data managers and help collections managers and users improve data accessibility and quality.
- We likely need more staff who are actively finding new sources and assisting collection managers with export issues. In addition, we need staff to manage the sources and data that we have.
- We also need to help the users to take best advantage of the resources we have.
- We've listened to the community, we've built the infrastructure, we've practiced training data providers (and I'm sure there is a lot more that I'm not mentioning). It seems like the priorities should now begin to shift towards marketing: e.g., who is using iDigBio and who could be using iDigBio, how are we going to engage them or increase levels of engagement, and how can we increase the momentum of digitization? We also will need to continue to assess the needs and experience of our end users, as well as continue to refine the iDigBio portal to meet the needs and expectations of all users. I think we do have sufficient staff to accomplish most of the things I've mentioned, but it seems iDigBio will need to assemble a marketing team. I'm not sure we have all the expertise required to do this kind of work--some of it, definitely.
- Get portal fully functional for basic data accessibility; THEN add images; add/link to georeferencing and mapping functions; links to other metadata, GenBank #s, DNA collections, etc.
- Work on integration with federal collections
- Work on integration with similar resources internationally
- Integrate with other 'big data' sources in different relevant disciplines
- Enable research thru development of appropriate computational environment
- Enable education and outreach
- Continue to integrate the collections community into ADBC
- We need to begin a series of targeted digitization workshops that provide in-depth training for specific digitization tasks and digitization personnel, to include specific types of imaging (similar to the fluid-preserved arthropod imaging workshop we are sponsoring at the U.

Michigan); discipline-specific "getting started" training targeted only to those who have not yet begun digitization activities; in-depth training in specific open source database systems (primarily Specify and Symbiota) targeted specifically to biodiversity informatics managers who have the responsibility for assisting curators and collections managers with database management; co-sponsored Specify and Symbiota workshops for end users (which will secondarily demonstrate our support for and collaboration with other successful NSF initiatives); increase our presence at national meetings by coordinating symposia and one-day workshops in association with them; increase our international cooperation through international working and interest groups; organize and cooperate with database and imaging equipment vendors in an effort to create digitization packages targeted at specific collection types, collection sizes, and price points; offer direct, onsite consultative services to assist small collections or groups of small collections embark on collections digitization (this may include organizing or supporting statewide collections consortia as we are currently attempting in West Virginia and Georgia). With more staff (support and direct service), we could enhance workshops and training offerings, involve a large percentage of our constituents in working groups and training, and realize the goal of providing onsite consultation, the latter of which will be impossible without additional funding and staff. I also think we should be able to fund and help install digitization stations for digitization projects too small for direct NSF support.

- Making iDigBio data and relevant to the research community and coordinating resources for TCNs are core priorities that I think have not changed. Having sufficient digital data to address research questions and having the right kind of access are necessary for iDigBio to be relevant on the long-term. GBIF, for whatever reasons, continues to be a difficult resource to data mine. The data set won't be complete enough and relevant unless there is some integration with federal collections and international efforts. The EAB report made that point.
- The noSQL/big tables approach to data ingest has been the right one, but we also need to have capability that allow users to make sense of that data. My sense is the latter, again, for the sake of relevance. How those data relate to other biological and environmental data is also a critical factor that will affect usability of the data. This all relates to data indexing, linking via GUIDs, and addressing what level of structured query support is needed (e.g., ontologies, data modeling)? Some of this could be done via external tools (e.g., BiSciCol), but we need to assess what's needed via internal resources. Current staffing levels in IT are probably insufficient.

- I think priorities revolve around the ability to seamlessly ingest, link, annotate, and feed back data, in a manner that addresses the needs of the bio-collections community so the iDigBio cloud is recognized as a resource that provides unique capabilities, is easy to use, is scalable and has high availability. From an IT perspective, I think our "throughput" is limited by the number of technical staff members in the project - who are extremely capable, but there are only so many hours in a day - so we can deliver on the priorities, but could accomplish more with additional programmers and staff that could work well in the interface between the cloud and bio-collection users. I don't think we are limited (yet) in terms of computer resources.