Where do we go from here? 
The future of collections use, attribution and communications in a digital age through the lens of a data aggregator

Randy Singer, Kevin Love and Larry Page
Florida Museum of Natural History/iDigBio
Outline

▪ What types of data do collections produce?
▪ What are current data deficits?
▪ How are collection staff giving the data to end users?
▪ How are collections’ data being used in research?
▪ How does the broader community view collections’ role in research?
▪ How can we “future proof” collection data for best use in research?
What types of data do collections produce?
What types of data do collections produce?
What types of data do collections produce?

What types of data do collections produce?
What types of data do collections produce?
Giving Data to End Users

- Directly (via loans and database use)
- Indirectly (via data aggregators)
- Super indirectly (via publications by researchers who used collection data)
- Visits to the collection
Data Deficits

- Isotope data
- Media (Sound and Video)
- Ecological data (environmental data)
- Behavioral data
- Higher taxonomic classifications
- Habitat descriptions/images
- Field photos of specimens
- ????????????????????????????????
US Fish Collection Holdings
So how much collections’ data is digitally available?

Most of it?

Some of it?

We don’t 100% know but we are getting close!
Collection Holdings

Holmes et al. 2016
What are people doing with the data?

- Research
- Education and Outreach
- Conservation efforts
- Industry
- Government projects
- Art?
Data Uses in Research

Changes in stickleback armor over 50 years

Woodrat body size and historical temperature

The role of natural history collections in documenting species declines
Shaffera et al. (1998)
Data Uses in Research

Biofluorescence in *Scyliorhinus rotifer*
Gruber et al. (2016)

The Covert World of Fish Biofluorescence: A Phylogenetically Widespread and Phenotypically Variable Phenomenon.
Sparks et al. (2016)
Data Uses in Research

Testing for mercury vapor in herbarium cabinets
Hawks et al. (2004)

Seabirds as biomonitors of mercury inputs to epipelagic and mesopelagic marine food chains
Thompsona et al. (1998)
What types of researchers use collections?
Survey of Ichthyologists: The Who

Ecologists!
Management!
Conservation Biologists!
Survey of Ichthyologists Not Using Collections

No relevant data: ~30%........................Don’t know about data or how to access: 32%

Why?

Collections do not provide relevant data to my research. 29.8% 17
I am not familiar with how to access the specimen and/or data. 12.2% 7
I do not feel like I have access to the data. 7.0% 4
I do not value the data quality in biodiversity collections. 1.7% 1
The format that the data is presented is incompatible with my research. 0.7% 5
I was not aware that biodiversity collections had data that might be relevant to my research. 19.3% 11
I was trained by my superiors/mentors/advisors that biodiversity collections and digitized collections data is not useful for my type of research. 0.0% 0
The data was not as precise as I would have liked. 3.5% 2
I am morally or ethically opposed to the collection of specimens. 0.0% 0
My institution, department or supervisor does not allow the use of biodiversity collection specimens or data. 0.0% 0
Other 10.1% 9

My methods are destructive (e.g., lith sampling) and I need large sample sizes.

I have to go through a 3rd party to access the data.

I have not had the chance yet to.

I have only just begun my research. My undergrad research involved collecting fish to be deposited in a natural history collection.

I am involved in sport fish management, therefore these collections are not used.
Survey of Ichthyologists

Which of the following adequately describes your use of biodiversity collections? (ch...)

<table>
<thead>
<tr>
<th>Description</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>I currently use data from a collection at my institution.</td>
<td>100%</td>
</tr>
<tr>
<td>I worked a collection outside my institution to gather data.</td>
<td>93%</td>
</tr>
<tr>
<td>I borrowed specimens from a collection outside my institution.</td>
<td>87%</td>
</tr>
<tr>
<td>I downloaded data from an institutional collections database.</td>
<td>85%</td>
</tr>
<tr>
<td>I downloaded data from a data aggregator (BIRD, GBIF, etc.)</td>
<td>72%</td>
</tr>
<tr>
<td>I requested specimens via request to curator or collection manager.</td>
<td>69%</td>
</tr>
<tr>
<td>Other</td>
<td>2%</td>
</tr>
</tbody>
</table>

TL;DR: Most people download digital data
Survey of Ichthyologists

Q15 - Have you used a data aggregator (e.g., GBIF, iDigBio, FishNet2, etc.) for finding digit...

~70% Use aggregators
Survey of Ichthyologists

Q17 - How often did you cite the catalog number, tissue number, or some other institutional...
The Future!
Human-Robot Alliance
Future Proofing Collections

What lasting role will biodiversity collections play in the future of research?

What do we do once we have fully digitized biodiversity collections?

How can collections better facilitate research?

How can researchers help collections sustainability?
Future Proofing Collections

What last role will biodiversity collections play in the future of research?

“As natural history museums [prepare to] enter the twenty-first century, much of their core still sits in the 1800s.”

(Krishtalka and Humphrey 2000)
**Future Proofing Collections**

What lasting role will biodiversity collections play in the future of research?

**Library/Storage of Biological Specimens**

“Biodiversity collections should function as a library of biological specimens, field notes, photos, and associated documentation for collections. Their function should be to preserve this material in perpetuity so that they are available for future use.”

**Public Access to Specimens and Data**

“Public access to all available data needs to be a high priority to all collections. Currently, there are whole states who's data is locked away on a few computers and are not accessible, except by a few individuals that dictate who can get the data. This includes public institutions that are tax payer funded.”

**Training Center**

“Training to agencies, individuals, and institutions that require to identify specimens in the field, and maybe collect specimens. If collection of specimens is required by such agencies, museums could train them also in the collection and preservation of specimens and establish collaborations for depositing material at their institutions.”

**Research Facilitator**

“Biodiversity collections, as a priority, are a library of specimens. Hence, irreplaceable specimens and associated metadata (e.g. locality, environmental data) are the original data that future scientists can build upon.”
Future Proofing Collections

What do we do once we have fully digitized biodiversity collections?

“As the major issues in exploration and systematics are resolved and society’s interest in biodiversity wavers, museums need to embrace important new uses for natural history collections and, with new partners, begin laying new foundations for a postbiodiversity future.” (Winkler 2004)

Long term storage of data/organization to make it very useful = #1
What do we do once we have fully digitized biodiversity collections?

- Improve the data?
- Generate new types of data?
- Function more like a library?
- Offer more services?
- Produce new products?
- Get good at solitaire?
Questions are changing

Figure 18. The phylogenetic kinship relations between the species of a monophyletic group, represented in two different ways.

Hennig 1966

Letunic and Bork 2015
Future Proofing Collections

How can collections better facilitate research?

“Data aggregators are ok, but I really like to be able to check on a specimen at the original source.”

Diversity of aggregation!

Formatting will be key!

Need good UIs!
Future Proofing Collections

How can collections better facilitate research?

- Formatting data better for end use
- Talking with researchers to find out their needs
- Show off the cool stuff we have and the tools to utilize it
- “Preach to the choir” less and interact with potential new users
- Not just serving up the data “as-is” and assuming its good enough
- Not assuming we know what the data needs to the research community are and serve up all the data types we can
Future Proofing Collections

How can researchers help collections sustainability?

“Biological collections are the most important repositories of our tree of life. The most important service is to provide access to the specimens/vouchers and associate data to conduct different kinds of studies. Although not all the collections have to deal with it, it’s important that collections promote open activities to the general public and to the stakeholders.”

Citations!
Sharing data
Data formatting
Contributing holistic data
Citations!
Future Proofing Collections

How can researchers help collections sustainability?

Keep using collections!

Collect and record data in a method where it has the most use to others

Format data in such a way where it is easy interpreted by others

Consider open source publishing options to increase exposure

Build a “specimen deposition” plan with funding for curation in your grants that involve specimen collection
Future Proofing Collections: Summary

• We need to keep collaborating
• We need to be continuously thinking about the future (sustainability focused vs project focused)
• Look inward for new data from old specimens
• Imagine we are using our own data in 50, 100, 200 years and how we can make it as useful as possible
• Don’t make assumptions that we know who, what, when and where collections data are being used
• Data aggregators are your friend
Thank You!

- iDigBio
- University of Florida
- Dave Blackburn
- Molly Phillips
- Shari Ellis

@Randchovy
@iDigBio