Accelerating the Discovery of Biocollections Data

Siro Masinde, Programme Officer for Content Mobilization

IDIGBIO SUMMIT V, WASHINGTON DC, 4-6 NOVEMBER 2015
1. Document best practices in data mobilization
2. Document successful business models for resource mobilization
3. Consult with other capacity building and content mobilization initiatives. e.g., iDigBio, ALA
4. Guidance in training and outreach materials

GBIF’S TASK FORCE MEMBERS

1. Leonard Krishtalka, USA, Chair
2. Barbara Thiers, USA
3. Deborah Paul, USA
4. Eduardo Dalcin, Brazil
5. Masanori Nakae, Japan
6. Ian Owens, UK
7. Jean Ganglo, Benin
8. Marc Pignal, France

Siro Masinde, GBIFS task force Coordinator

Photo by S. Masinde: GBIF task force meeting, 3 Nov 2015.
ACCELERATING DISCOVERY OF BIOCOLLECTIONS DATA

Metadata approach

- c. 90% of world’s c. 3 billion specimens of NHCs are not digitized
- Metadata is critical to advancing digitization efforts, priorities and opportunities, and enables the community of users to search and find relevant collections
- Opportunity to participate whether institution is digitizing or not
Purpose is to determine:

• Digital readiness of the world’s biocollections and their institutions
• Benefits to a collection / institution that digitization engenders
• Impediments to collection data digitization
• Make appropriate recommendations

https://ufl.qualtrics.com/SE/?SID=SV_byg2mgbtYdeF3bT

Please participate in survey, it takes 10 - 30 minutes
PRELIMINARY SURVEY RESULTS

- c. 600 responses
- USA among best response rate
- Survey closes Nov 20, 2015, hence responses still coming in
NORTH AMERICA RESPONSES

ADBC PROJECTS

SURVEY RESPONSES

*Image taken from iDigBio Summit V program*
<table>
<thead>
<tr>
<th>Collection Type</th>
<th>Vascular Plants</th>
<th>Bryophytes</th>
<th>Fungi &amp; lichens</th>
<th>Algae</th>
<th>Arthropods</th>
<th>Mammalogy</th>
<th>Ornithology</th>
<th>Herpetology</th>
</tr>
</thead>
<tbody>
<tr>
<td>Respondents</td>
<td>337</td>
<td>187</td>
<td>182</td>
<td>143</td>
<td>116</td>
<td>67</td>
<td>68</td>
<td>63</td>
</tr>
<tr>
<td>Percent</td>
<td>63%</td>
<td>35%</td>
<td>34%</td>
<td>27%</td>
<td>22%</td>
<td>13%</td>
<td>13%</td>
<td>12%</td>
</tr>
<tr>
<td>Respondents in Group</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Vascular Plants</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ichthyology</td>
<td>63</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Malacology</td>
<td></td>
<td>58</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Invertebrates, terrestrial</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Marine Invertebrates</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Paleobotany &amp; Palynology</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Vertebrate Paleontology</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Other</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Invertebrate Paleontology</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Other</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

GBIF
<table>
<thead>
<tr>
<th>Collection</th>
<th>#</th>
<th>Databased</th>
<th>#</th>
<th>Shared</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vascular Plants</td>
<td>n = 239</td>
<td>46%</td>
<td>n = 231</td>
<td>37%</td>
</tr>
<tr>
<td>Bryophytes</td>
<td>n = 141</td>
<td>37%</td>
<td>n = 132</td>
<td>28%</td>
</tr>
<tr>
<td>Fungi &amp; lichens</td>
<td>n = 141</td>
<td>47%</td>
<td>n = 123</td>
<td>38%</td>
</tr>
<tr>
<td>Algae</td>
<td>n = 112</td>
<td>36%</td>
<td>n = 101</td>
<td>37%</td>
</tr>
<tr>
<td>Arthropods</td>
<td>n = 87</td>
<td>31%</td>
<td>n = 77</td>
<td>28%</td>
</tr>
<tr>
<td>Mammalogy</td>
<td>n = 58</td>
<td>67%</td>
<td>n = 27</td>
<td>47%</td>
</tr>
<tr>
<td>Ornithology</td>
<td>n = 54</td>
<td>65%</td>
<td>n = 51</td>
<td>30%</td>
</tr>
<tr>
<td>Herpetology</td>
<td>n = 52</td>
<td>65%</td>
<td>n = 49</td>
<td>34%</td>
</tr>
</tbody>
</table>
MOTIVATION FOR DIGITIZING

- Funding/grant priority and opportunity
- Research priority
- Taxonomic priority
- Geographic priority
- Partnership in a larger community effort
- Other. Please specify.
- No target priority, just opportunistic.
- Health and human service needs

Percent of Respondents
IMPORTANCE OF METADATA

We are currently providing this information.  
I think it is critical we provide this information.
WHERE METADATA IS SHARED – 60 LOCATIONS

- Other: 55%
- Index Herbariorum: 40%
- GBIF Registered Organization: 35%
- Print publication: 25%
- Not applicable: 20%
- GRBio: 15%

Percent of Respondents
SOURCES OF FUNDING

<table>
<thead>
<tr>
<th>Source of Funding</th>
<th>Percentage of Respondents</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ad hoc funding (e.g., opportunistic gifts)</td>
<td>20%</td>
</tr>
<tr>
<td>Part of the regular institutional budget</td>
<td>60%</td>
</tr>
<tr>
<td>External funding</td>
<td>70%</td>
</tr>
</tbody>
</table>
LONG TERM PLANS

We do not plan to digitize our collections.

We will digitize only new additions to the collections.

We will digitize only components of collections unique to our institution (e.g., type specimens, endemic species).

We will digitize components of collections in response to research-based needs.

We will gradually digitize our entire collection.
TOP 5 REASONS FOR NOT DIGITIZING

1. Funding/resources not available
2. Lack of time among personnel
3. Limited expertise among personnel
4. Not a priority of the individual in charge of the collection Insufficient information on digitization process and protocols
5. Size of task is overwhelming
NEXT STEPS

Task Force recommends that the DATA → KNOWLEDGE → APPLICATION value chain framework be used to mobilize resources and prioritize digitization.