Task Cluster 4 - **Electronic Data Capture:** collection data management and data capture

Presenter: Deborah Paul  
Florida State University  
Integrated Digitized Biocollections (iDigBio)

Brazilian Biodiversity Information System (SiBBr) Launching Event  
25 November 2014  
Brasilia, DF, Brazil  

iDigBio is funded by a grant from the National Science Foundation’s Advancing Digitization of Biodiversity Collections Program (Cooperative Agreement EF-1115210). Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.
DATA CAPTURE

- Image Capture
- Image Processing
- Data Capture
- Image / Data Storage

- Pre-digitization Curation or Staging

- Georeferencing and Enrichment
- Personnel
- Written Workflows
- Biodiversity Informatics Manager
Staffing

Data Entry, Imaging, Data Validation

- In-house volunteers
- Paid staff
- Citizen scientists

Considerations

- Training
- Roles
- Autonomy
- Managing Staff and Data
Labels
Notebooks
Card Files
Vials, ...
Minimal Data Capture

- “filed as” name
- higher geography
- barcode
- image

- all sheets in folder get the same initial data
- only the barcode differs

How do we get these minimal records completed, or more complete?
Note darwin core / georeferencing **standards**

[Image: http://www.britishmuseum.org/images/rosettawriting384.jpg]
Inside the 1899 Harriman Expedition

New York Botanical Garden (NY)

Collector: Wm. Trelease
Associated Collectors: De Alton Saunders
Exsiccat Title
Scientific Name: Wahlenbergia mucosa

Country: United States
State/Province: Alaska
County: Juneau
Locality: Taku Inlet

Latitude: 58.255678
Longitude: -134.078102
Uncertainty: 17128

Elevation in Meters

Verbatim Elevation

Habitat
Substrate
Notes

Save Edits
Status Auto-Set: Pending Review

OCR Image
Options:
- OCR whole image
- OCR w/ analysis

Notes:
Source:
ABBYY:2013-02-09

Save OCR Edits
Parse OCR (LBCC)
Delete OCR
Specimen Data Capture

• Extracting label data
  – Before, during, after imaging (a choice)

• Entering data from label images
  – reduces specimen handling
  – can facilitate ability to read labels
  – creates a voucher for the label

• Database interface often customized
  – speeds data transcription and enhances accuracy

• Data often imported from spreadsheets (Specify)

• Online data entry (in-the-cloud)
Data Capture Options

• Data capture with voice* (shhhhhhh)
• Using OCR software and OCR output parsed into database
  – vetted by a person
• OCR output is searchable!
• Records multi-keyed
• Crowd-sourcing
A few more key thoughts about data capture ...

• database software and data-entry issues
  – ditto,
  – drop-downs,
  – automated scripts for validation
• error catching / data validation strategies
• tracking what has / has not been entered / imaged
• protocols / workflows continuously evaluated
• **data quality / integrity**
Obrigada SiBBBr! Find out more at …
https://www.idigbio.org/content/workflow-modules-and-task-lists

www.idigbio.org

facebook.com/iDigBio
twitter.com/iDigBio
vimeo.com/idigbio
idigbio.org/rss-feed.xml
webcal://www.idigbio.org/events-calendar/export.ics

iDigBio is funded by a grant from the National Science Foundation’s Advancing Digitization of Biodiversity Collections Program (Cooperative Agreement EF-1115210). Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.