Getting Started!

Task Cluster 1 - Pre-digitization curation and staging: decisions / opportunities / options

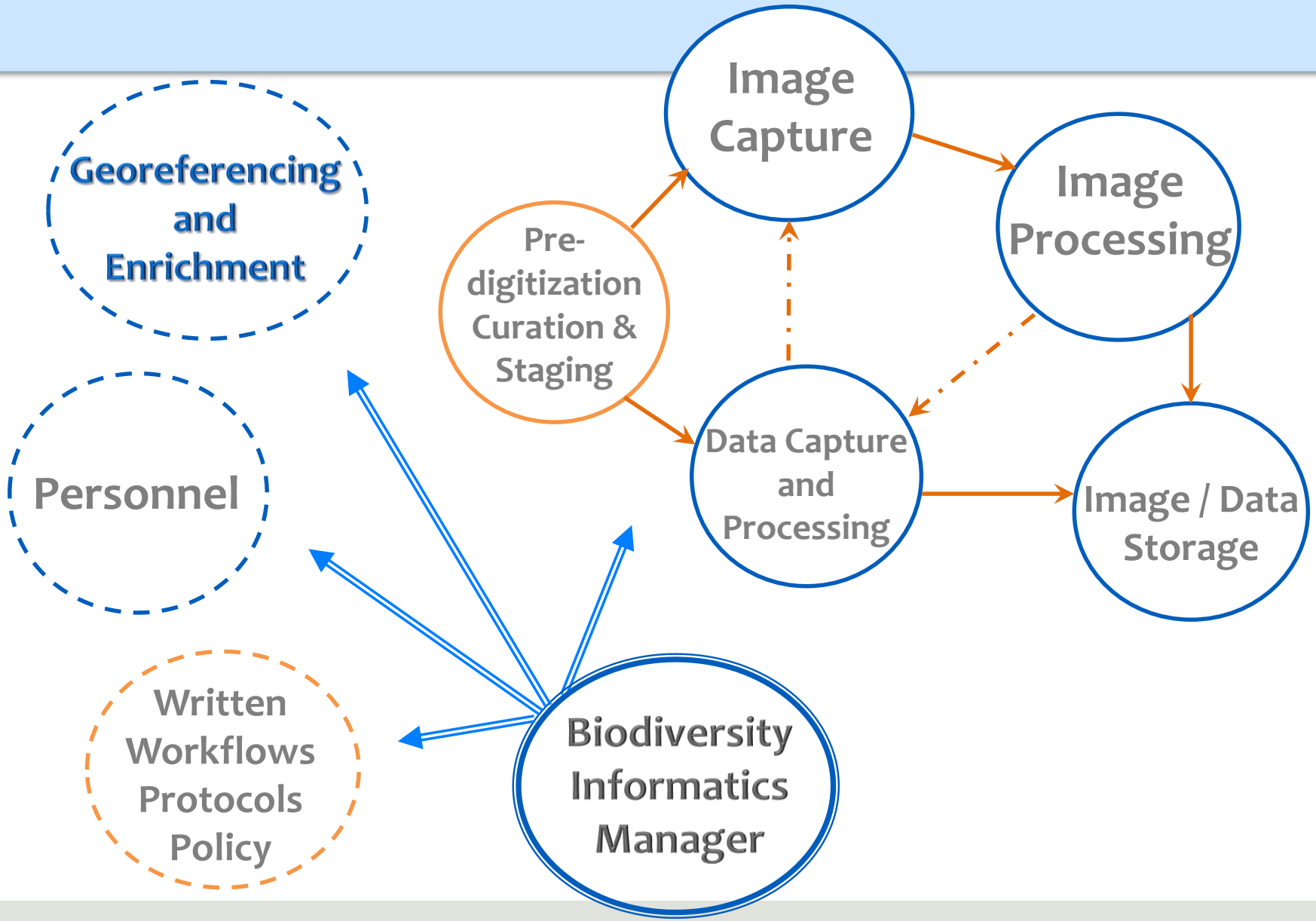
Presenter: Deborah Paul, @idbdeb
Florida State University
Integrated Digitized Biocollections (iDigBio)

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

#SiBBr

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.

PRE-DIGITIZATION CURATION



Pre-digitization Curation

DROID Workflows Workshop

- Developing Robust Object to Image to Data Workflows
  - Workflows by storage type
  - DROID1 – flat sheets
  - Module 1 – Pre-digitization Curation
- https://www.idigbio.org/content/workflow-modules-and-task-lists

Pre-Digitization Module Tasks (Part 1)

- T1 – apply storage locator barcodes
- T2 – selecting what to digitize
- T3 – apply machine readable barcodes at collection level
- T4 – locate specimens (flag cabinets)
- T5 – pull specimens from cabinet*
  - *(optional) sort by collector, date, geography
- T6 – curate collection in place (check nomenclature and annotations)

Pre-Digitization Module Tasks (Part 2)

- T7 – transport specimen to imaging station
- T8 – placeholder to flag pulled specimens
- T9 – sort to remove any already imaged / barcoded
- T10 – separate specimens needing conservation work before imaging
- T11 – apply barcodes
- T12 – create skeletal database record

Deciding to Digitize

- What will you digitize?
  - All or part of your collections
- How do you decide?
  - Researcher-based needs
  - Appeal to public, outreach, preservation, conservation
  - Fragility
  - Access
  - Cost
  - Staffing
- Will you be taking images? If yes, more decisions!

What Database?

- what database, what other software (optical character recognition (OCR), voice, touch screen)? How to decide?
- preparing the database
- taxonomic trees / tables
- getting all localities done beforehand
- what kind of identifiers?
- how will data be shared / exported / re-integrated?

Pre-Digitization Opportunities

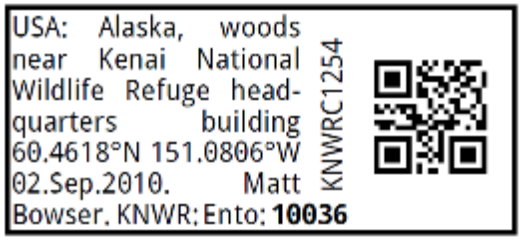
- evaluate collection health
  - Profiling Natural History Collections: A Method for Quantitative and Comparative Health Assessment
- hard data for museum directors & administrators
- "an important tool in reinvigorating collection management and in particular providing data to support funding requests."
- finding unknown unknowns and lost material
- experts or non-experts?
- high-hanging fruit (or tasks perhaps long put off)
- cabinet reorganization
- equipment updates
- loan returns
- specimen repair

Not all steps require a professional

Curation is a potential bottleneck

Bar Codes

Types



Use and Content of Barcodes

- What suits your collection type/s?
- Is all printed text on the label in the encoded part?
- What ought to be in the encoding?
- Will you need to re-print?
- Can the barcode be seen easily (insect / wet collection issues)?
- Is your barcode identifier globally unique or only unique inside your collection?
- Must you use one?
  - They speed up processing and tracking of loans
  - They make automation possible for some digitization processes
  - Accountability

Preparing the collection – Curatorial tasks

- Updating the taxonomic identifications
  - Or not
- Updating nomenclature in the database
- Tracking loans
  - What's been digitized, what has not
  - Updating loan records
- Label updates / standardization
  - Cabinets, drawers, trays, jars, slides
- Collection health
  - Curation pipeline
  - Conservation status
  - Condition of labels
  - Data quality
  - Computerization level
  - Container condition
- Incorporating new materials (gifts) waiting to be accessioned

Who is going to digitize?

Tasks

- Preparation
  - Cabinet organization
  - Re-pinning
- Bar code application
- Data Transcription
- Imaging
- Data Validation
- Georeferencing
- Determination Annotation
- Enhancement

Potential Resources for these Tasks

- Staff
- Volunteers
- Public

- Using Optical Character Recognition software
- Voice recognition software
- Touch screen technology
- Light Box
- Conveyor belt

Predigitization Unanticipated Benefits

- inspect / repair / specimen damage (ipm)
- collection health,
- inventory collection
- re-pin / remount specimens
- replenish / replace preservatives
- attach a unique identifier
  - (most often a 1- or 2-D barcode)
  - to a specimen, container, or cabinet
- discover important but
  - unknown, lost, or dislocated holdings
  - (e.g. those owned by other institutions or the federal government)
- update nomenclature and taxonomic interpretation
- reorganize the cabinets, cases, trays, and containers
- vet type specimens, and
- select exemplars for digitization / imaging*

Assessing Digitization Tasks

- Reed Beaman, James Macklin, Michael Donoghue, James Hanken. 2007. Overcoming the Digitization Bottleneck in Natural History Collections: A summary report on a workshop held 7 – 9 September 2006 at Harvard University.
- Íñigo Granzow-de la Cerda and James H. Beach. December 2010. Semi-automated workflows for acquiring specimen data from label images in herbarium collections. Taxon 59 (6): 1830-1842
- Bryan Kalms. Digitisation: A strategic approach for natural history collections. Canberra, Australia, CSIRO, 2012.
- John Tann & Paul Flemons. 2008. Report: Data capture of specimen labels using volunteers. Australian Museum
- Ana Vollmar, James Alexander Macklin, Linda Ford. 2010. Natural History Specimen Digitization: Challenges and Concerns. Biodiversity Informatics 7 (1): 93 – 112
- Favret C, Cummings KS, McGinley RJ, Heske EJ, Johnson KP, Phillips CA, Phillippe LR, Retzer ME, Taylor CA, Wetzel MJ. 2007. Profiling Natural History Collections: A Method for Quantitative and Comparative Health Assessment. Collection Forum 22(1–2): 53 - 65
- Nelson G, Paul D, Riccardi G, Mast AR 2012. Five task clusters that enable efficient and effective digitization of biological collections. In: Blagoderov V, Smith VS (Ed) No specimen left behind: mass digitization of natural history collections. ZooKeys 209: 19–45. doi: 10.3897/zookeys.209.3135
- iDigBio Developing Robust Object to Image to Data (iDigBio DROID) Workshop – May 30 – 31, 2012
- https://www.idigbio.org/content/workflow-modules-and-task-lists

About iDigBio | Research | Technical Information | Education

iDigBio
Integrated Digitized Biocollections

Google™ Custom Search Search My account Log out

Making data and images of millions of biological specimens available on the web

24,705,794
Specimen Records

4,046,837
Media Records

357
Recordsets

Search the Portal

Why digitization matters
More about what we do and why

Digitization
Learn, share and develop best practices

Sharing Collections
Documentation on data ingestion

Working Groups
Join in, contribute, be part of the community

Proposals
New tool and workshop ideas

Citizen Scientists
How can you help biological collections?

Obrigada SiBBr! 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

dpaul@fsu.edu
@idbdeb

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.