Of People & Data

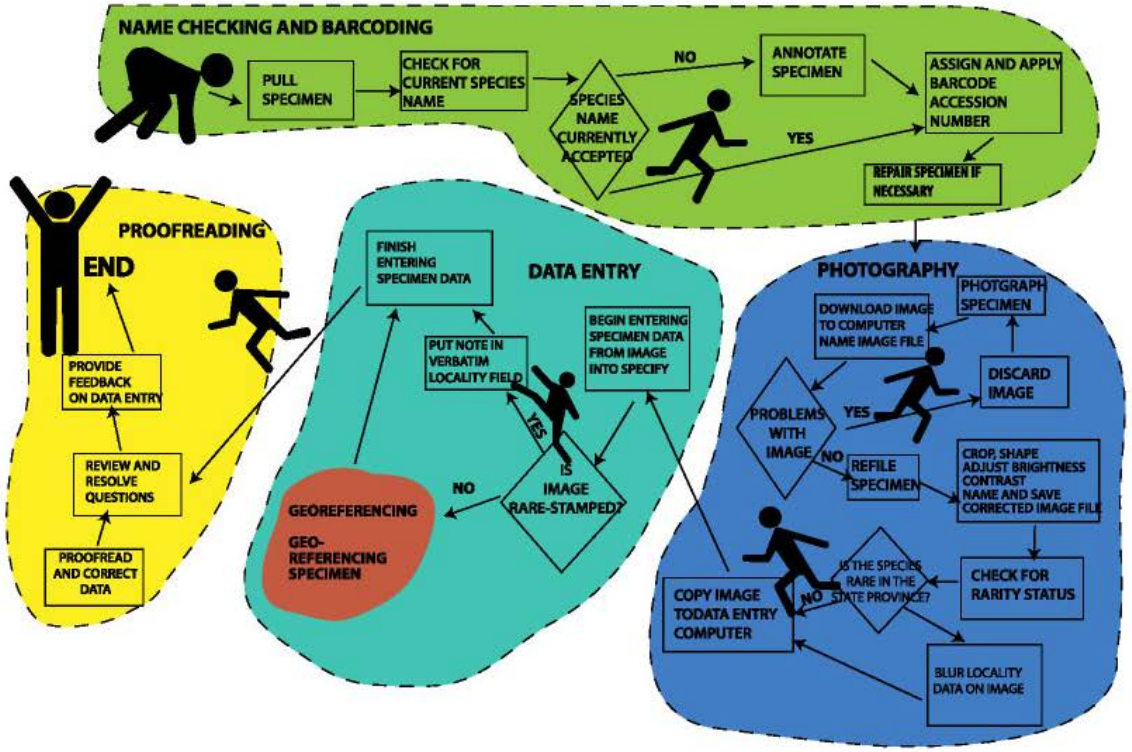
Association for Southeastern Biologists (ASB) 2013
Charleston West Virginia
iDigBio Digitization & Workflows Workshop
Deborah Paul, iDigBio
April 13, 2013

NSF

ASSOCIATION OF SOUTHEASTERN BIOLOGISTS • 1937 • ASB

iDigBio
Integrated Digitized Biocollections

PEOPLE IN THE LOOP



Of People & Data Overview

- Social issues (People)
- Specimen Identifiers
  - CollectionObject
- Share!
- Excel ≠ database
  - Open Refine

Social Issues

- Why discuss people issues?
  - People
  - Technology
- The invention of the wheel
- Use resources wisely, effectively
- Knowledge about social issues

Social Issues

- observing digitization workflows
- social issue layers
- business processes
- change, management
- motivation
  - examples
  - opportunities

Observing Digitization & People

- iDigBio's Gil Nelson and Deb Paul visited
  - 28 Collections in 10 different museums
  - Collections varied in size, kind, staffing
    - entomology
    - invertebrate paleontology
    - invertebrate zoology
    - ornithology
    - botany
    - vertebrate paleontology
    - zoology
- Paper submitted to ZooKeys
  - Not about social issues

http://microecos.wordpress.com/2009/03/

Layers

specimens as boundary objects

Community

think *little science* to *eScience*

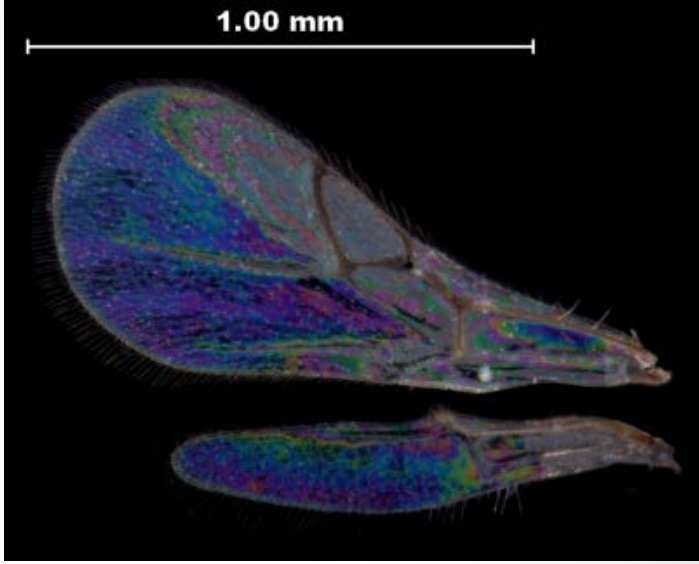
Organization

the museum

the individual collection

The individual digitizing

Change: the three r's



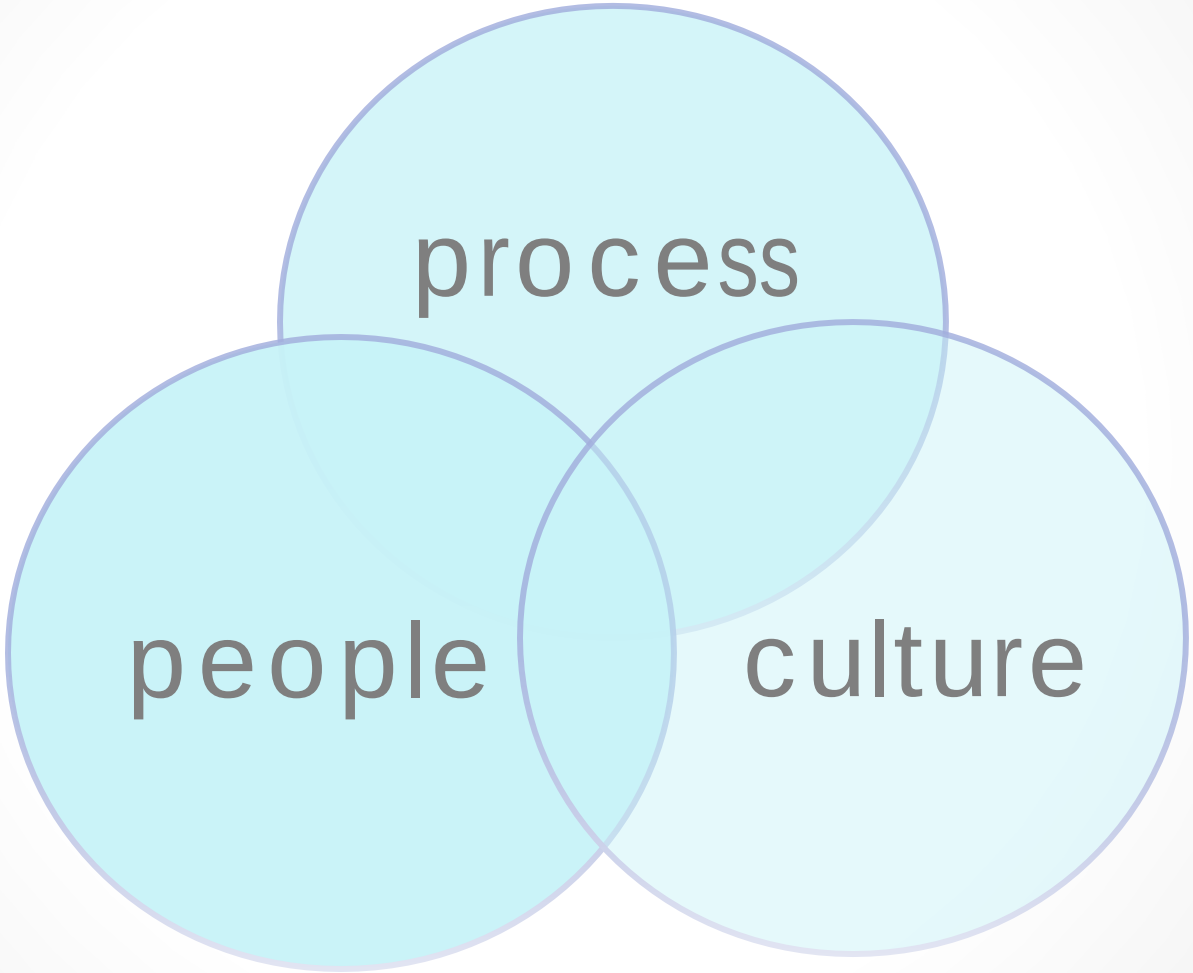
business processes applied to
workflow process improvement

- review
  - observe
  - document
  - evaluate
- redesign
  - observe
  - document
  - evaluate
- reengineer

= layers of change

http://www.jiscinfonet.ac.uk/infokits/process-improvement

Change Management



- http://www.jiscinfonet.ac.uk/infokits/change-management/aspects-of-change

The *irrational side* of change management

- what motivates you ≠ what motivates others
- we're better off **listening**, rather than telling
- we need the + and the –
- leaders mistakenly believe that they already "are the change"
- it's not just about the leader, it's about how **receptive** "society" is to the ideas
- money is the most **expensive** way to motivate people.
  - satisfaction equals perception minus expectation.
  - **small unexpected rewards** can have disproportionate effects.
- process and outcome have to be fair to get buy-in

Aiken & Keller, 2009 in The McKinsey Quarterly

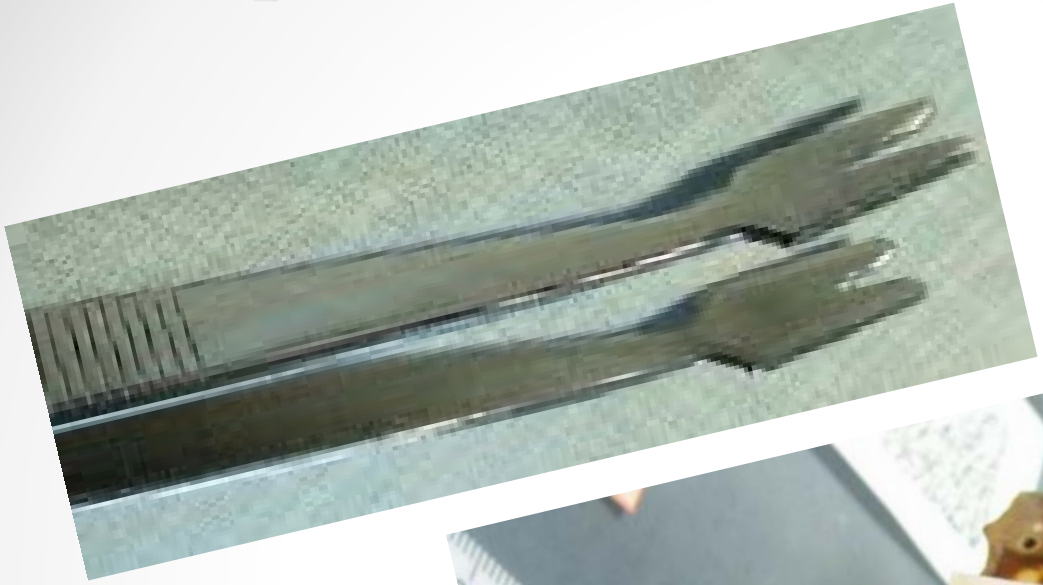
Social Issues

- observing digitization workflows
- social issues layers
- business processes
- change, management
- motivation
  - examples
  - opportunities

Motivation

- What is known about what motivates us?
- Where can we integrate what is known about motivation into our workflows?

A Unique Tool for the Job: Forceps



http://www.ecnweb.org/dev/files/12_Eastwood_2010.pdf

Task Order / Task Method

- Task order didn't matter
  - o imaging ← →data entry
  - o update taxon names ← → imaging
- Task method didn't matter
  - o sort by collecting event
  - o apply barcodes
- Social effects
  - o autonomy
  - o alleviates boredom
  - o engages worker → giving them a choice
  - o novel strategies (mastery and purpose)

And the winner is

V

S

- Social effects
  - o seems to engage workers
  - o steps reduced → less tedious
  - o novel strategies

Using Skill Sets

- opportunities to use individual skills
  - typing
  - imaging
  - language
  - writing
  - software / hardware
- Русский
- student is OCD

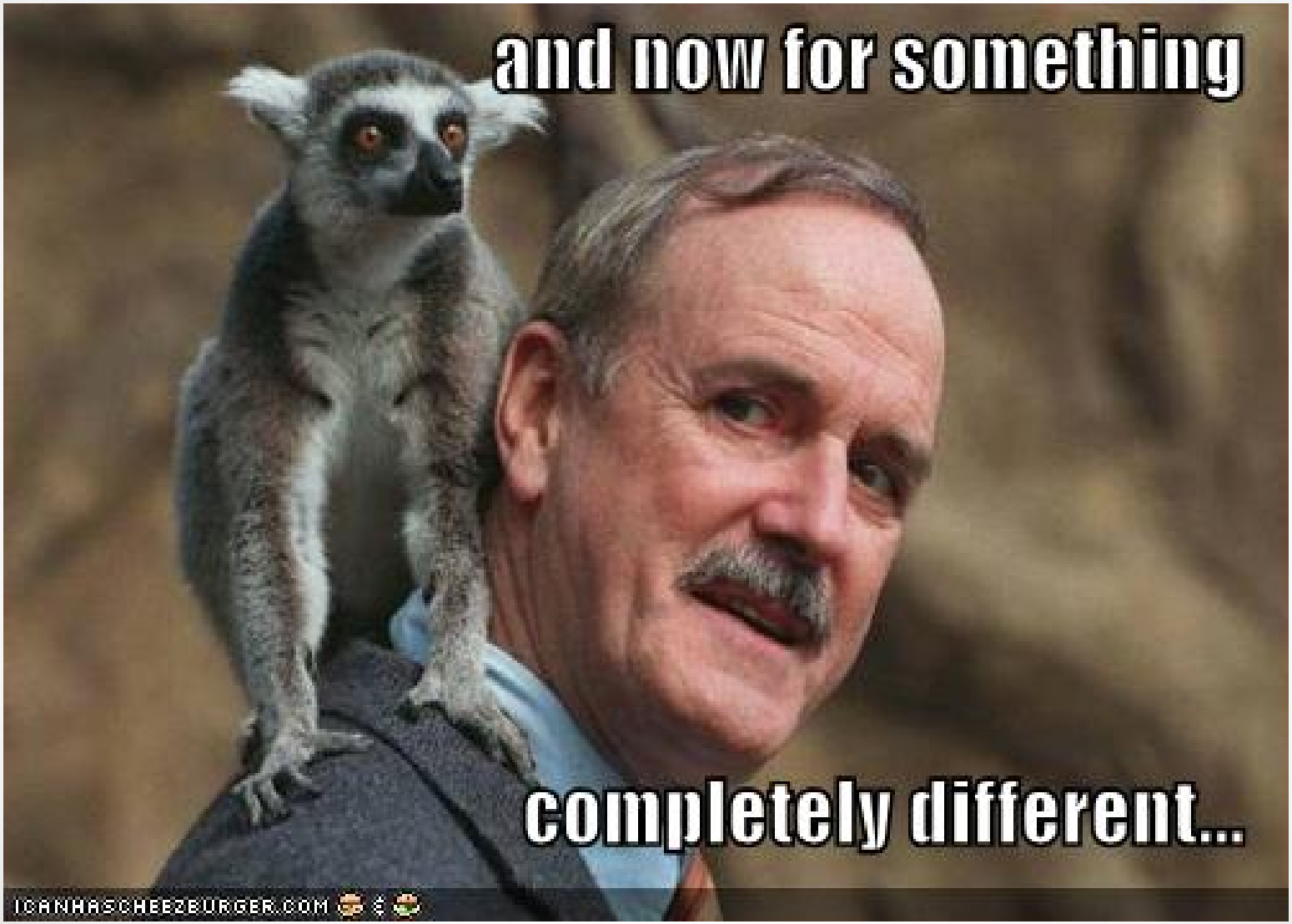
Using Skill Sets

- opportunities to use individual skills
  - typing
  - imaging
  - language
  - writing
  - software / hardware
- Русский
- student is meticulous
  - taxonomy
- students / staff share what works with each other
- effects
  - efficient use of talent (resources)
  - engaging someone's sense of purpose and usefulness allowing them to utilize mastery

Motivation: 3 keys

1. autonomy
   - possibilities
2. mastery
   - software skills
   - hardware skills
   - people skills
   - training (giving and receiving)
   - knowledge (collections, curation, content)
3. purpose
   - engagement & inspiration
   - the ADBC initiative is rich with purpose

http://www.**danpink**.com/archives/2010/06/whiteboard-magic
http://youtu.be/u6XAPnuFjJc


and now for something
completely different...
ICANHASCHEEZBURGER.COM

Of People & Data Overview

- Social issues (People)
- Specimen Identifiers
  - CollectionObject
- Share!
- Excel ≠ database
  - Open Refine

Specimen Identifiers (Collection Object)

- What good is identification?
- Examples and Use
- Specimen identifiers
- Providing IDs
- Annotations and Feedback

What good is identification? I

- Aggregation
  - If you get info from 2 sources that are about the same object, you can combine the info
- Resolution (finding information about object)
  - Types of resolution
    - Determine where to get information
    - Determine how to get information

What good is identification? II

- Aggregate
- Resolve
- Cite
- Find locally
- Feedback

Identifier examples

- record id 112234
- uuid urn:uuid:954c8760-e1a6-4b4b-ab82-6bf7311c25f3
- lsid urn:lsid:zoobank.org:act:8BDC0735-FEA4-4298-83FA-D04F67C3FBEC
- http - uri http://ids.flmnh.ufl.edu/herb/abcd12345678
- ezid ark:/87286/B2954c8760-e1a6-4b4b-ab82-6bf7311c25f3
- doi doi:10.1038/ng0609-637
- catalog urn urn:catalog:FSU:Herbarium:000023905

Proxy request

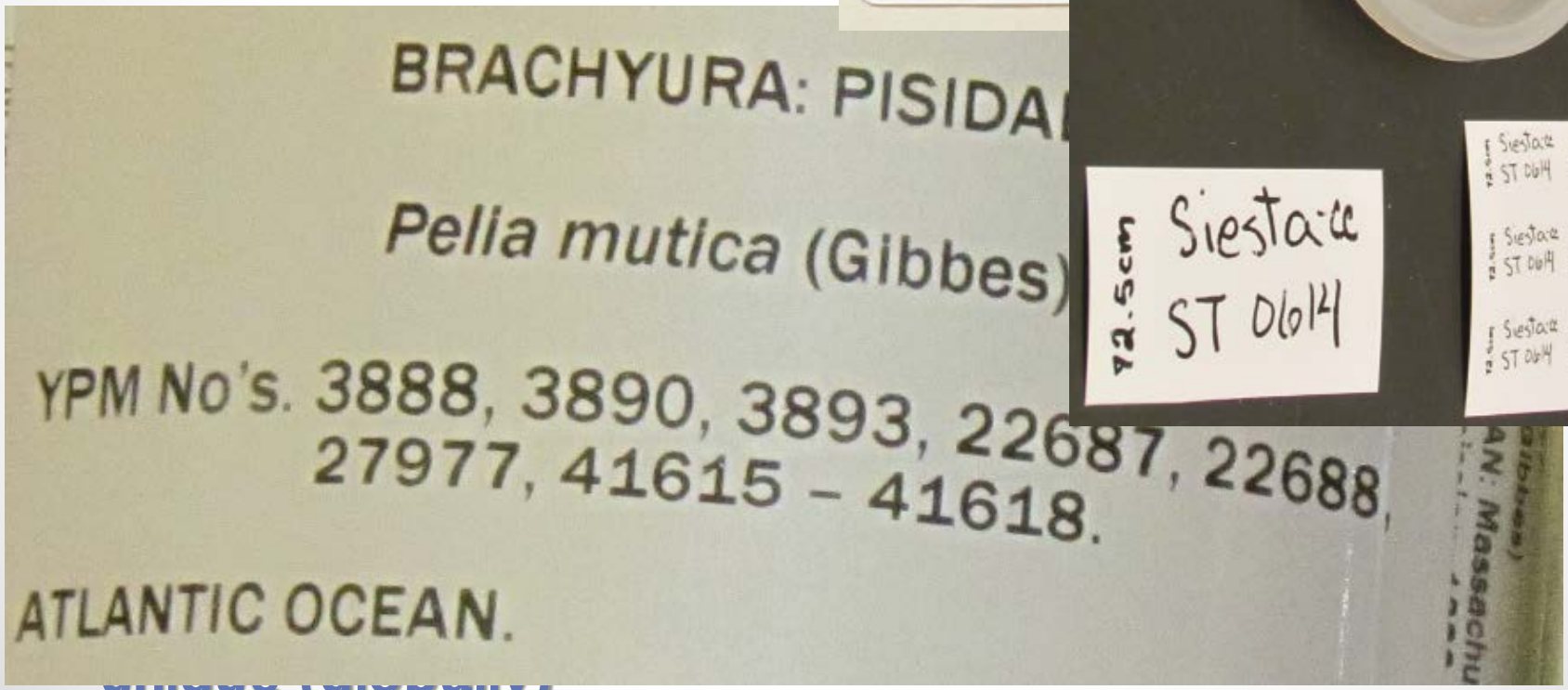
http://dx.doi.org/10.1038/ng0609-637

Zoobank Uses Identifiers

- Consider web page
  - http://zoobank.org/NomenclaturalActs/4DFD6D95-C287-4AFF-8473-E073D8960EC6
- Look for identifiers
  - urn:lsid:zoobank.org:act:4DFD6D95-C287-4AFF-8473-E073D8960EC6
  - HOLOTYPE: Deposited as No. 7113, Hancock Parasitology Collection, University of Southern California
- Search in google for these

Specimen identifiers?

- Identifier on the specimen?
  - readable text



- unique (globally)
- **actionable (resolvable) maybe**

Identifiers

- Opaque identifiers
  - specimenId = **urn:uuid:**954c8760-e1a6-4b4b-ab82-6bf7311c25f3
  - o No where, which or how
  - o Smith No. 1153
- Importance of Registration
  - o Institution Code, Collection Code, Owner Institution ID
- Finding Specimens
  - o Globally
  - o Locally
- Darwin Core OccurrenceID

Identifiers: Curating, Providing, Using

- Store the ones you know about
  - Database changes to accommodate identifiers (global or not)
  - Analogous to determination history
  - Analogous to ownership history
  - How are identifiers stored in your database?
- Sharing is Key
  - Facilitate finding / BiSciCol / relationship illumination
  - Filtered PUSH
  - SGR
  - Standards (Darwin Core) to share identifiers
- Cite
  - What do you cite in your papers?
  - Metrics for administrative justification?
  - Metrics for tenure credit?

Feedback with IDs

- Annotations
  - Target of annotation
    - http://www.morphbank.net/818505

Related Annotations

Taxonomic Name	Taxon Author	Prefix	Suffix		
Opuntia humifusa	(Raf.) Raf.	none	none	1	0

  - filtered PUSH
  - SGR
  - BiSciCol
- linked data, aka the semantic web
- updating the database
  - be(a)ware
  - store and share other IDs


Specify 6
FilteredPUSH
iDigBio
Integrated Digitized Biocollections
EMu
Museum
Management
System
BioPortal
eol
Encyclopedia of Life
GBIF
ARCTOS
Identifier
morphbank
morphbank.net
Kepler Kurator
DISCOVER LIFE
NCBI
Systematic Biology
Symbiota
ZooKeys
VerSeNet
Cladistics
Silver
COLLECTION
BMC
Bioinformatics
PhytoKeys
vizzuality
ZOONIVERSE
REAL SCIENCE ONLINE
PLoS one
www.plosone.org

Of People & Data Overview

- Social issues (People)
- Specimen Identifiers
  - CollectionObject
- Share!
  - Collaborating with TCNs and PENs
  - Contributing to iDigBio
  - Data Aggregators
- Excel ≠ database
  - Open Refine

Ways to Share Data

- Thematic Collection Networks (TCNs)
  - have data ready to share?
  - fits a current TCN theme?
- Partners to Existing Networks (PENs)
  - join the effort
- Does my collection have to be part of a TCN or PEN to contribute data?
  - No.
- Through an existing portal or repository
  - Symbiota
  - VertNet
  - Morphbank
  - iDigBio
  - GBIF
- Help is everywhere!

Sharing data with iDigBio

- All specimen data welcome at iDigBio
- Custom export
- CSV files
- DwC-A files
  - Integrated Publishing Toolkit (IPT)
  - DwC-Extensions
    - MeasurementOrFact
    - ResourceRelationship
    - AudubonCore
- Specimen Identifiers
- Record Identifiers
- *Expect Feedback*

Social Issues in Collaborative Digitization

- Barton, J. 2005. "Digital libraries, virtual museums: same difference?", Library Review, Vol. 54 No. 3, pp. 149-54
- 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. [Accessed January 2012]
- 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. 2012. Digitisation: A strategic approach for natural history collections. Canberra, Australia, CSIRO
- Liew, Chern Li. 2009. Digital library research1997-2007:Organisational and people issues. Journal of Documentation 65(2): 245 – 266 http://dx.doi.org/10.1108/00220410910937606
- Eric T. Meyer. 2009. Moving from small science to big science: Social and organizational impediments to large scale data sharing. In: E-Research: Transformation in Scholarly Practice. Jankowski, Nicholas W. (Ed), Routledge. 350 pp.
- Susan Leigh Star and James R. Griesemer. 1989. Institutional Ecology, 'Translations' and Boundary Objects: Amateurs and Professionals in Berkeley's Museum of Vertebrate Zoology, 1907 – 1939. Social Studies of Science 19: 387 – 420
- 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
- **iDigBio Workshop: Developing Robust Object to Image to Data Workflows** May 30 – 31, 2012

Acknowledgment & Thanks

American Museum of Natural History (AMNH)
Botanical Research Institute of Texas (BRIT)
Florida Museum of Natural History (FLMNH)
Florida State University (FSU)
Harvard Herbarium (HUH)
Museum of Comparative Zoology (Harvard)
New York Botanical Garden (NYBG)
Southeast Regional Network of Expertise and Collections (SERNEC)
Specify Software Project (University of Kansas)
Symbiota Software Project (Arizona State University)
Tall Timbers Research Station and Land Conservancy (TTRS)
Tulane University Museum of Natural History
University of Kansas Biodiversity Institute
Valdosta State University (VSU)
Yale Peabody Museum (YPM)
and ***all participants, presenters and organizers of the iDigBio Workshop:***
Workflows and Challenges in Digitization of Museum Specimens
Host: – Association of Southeastern Biologists
Marshall University
April 14, 2013 Charleston West Virginia

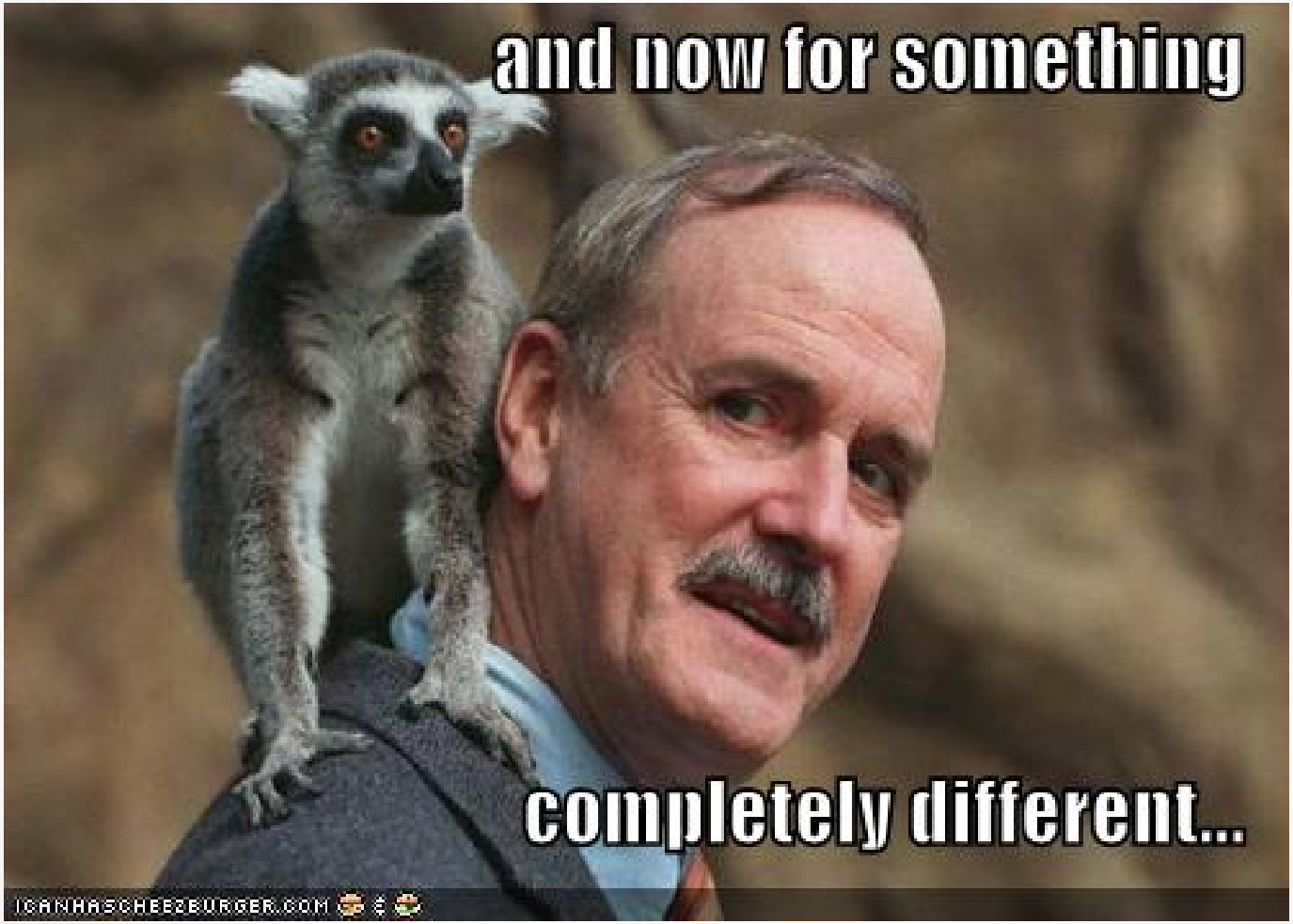
iDigBio
Integrated Digitized Biocollections

NSF

Thank you

iDigBio
Integrated Digitized Biocollections

iDigBio is funded by a grant from the National Science Foundation's Advancing Digitization of Biodiversity Collections Program (#EF1115210). Views and opinions expressed are those of the author not necessarily those of the NSF.


and now for something
completely different...
ICANHASCHEEZBURGER.COM

Of People & Data Overview

- Social issues (People)
- Specimen Identifiers
  - CollectionObject
- Share!
  - Collaborating with TCNs and PENs
  - Contributing to iDigBio
  - Data Aggregators
- Excel ≠ database
  - Open Refine