

Creating workflows for effective data migration and imaging in an invertebrate paleontology collection

Ann Molineux, Liath Appleton, Angella Thompson, Louis G. Zachos*
Non-vertebrate Paleontology Laboratory, Jackson School of Geosciences, The University of Texas at Austin;

* Department of Geology & Geological Engineering, University of Mississippi

Acknowledgements

- Volunteers and associates of NPL especially: Gareth Cross, Christopher Garvie, Cissy Geigerman, Judy Humer, Keith Minor
- Melissa Winans, Data analyst at Texas Natural Science Center
- Many of the protocols are based upon work supported by the National Science Foundation under grants:
 - DBI-1057396: *Open Access: Conservation, Digitization and interoperability of the Historic Non-vertebrate Collections of the Texas Natural Science Center.*
 - EF-1305070: *Digitization PEN: Targeted digitization to expand and enhance the Paleoniches TCN.*
 - Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.



Summary

9/14/2013

iDigBio PaleoDigWkshp

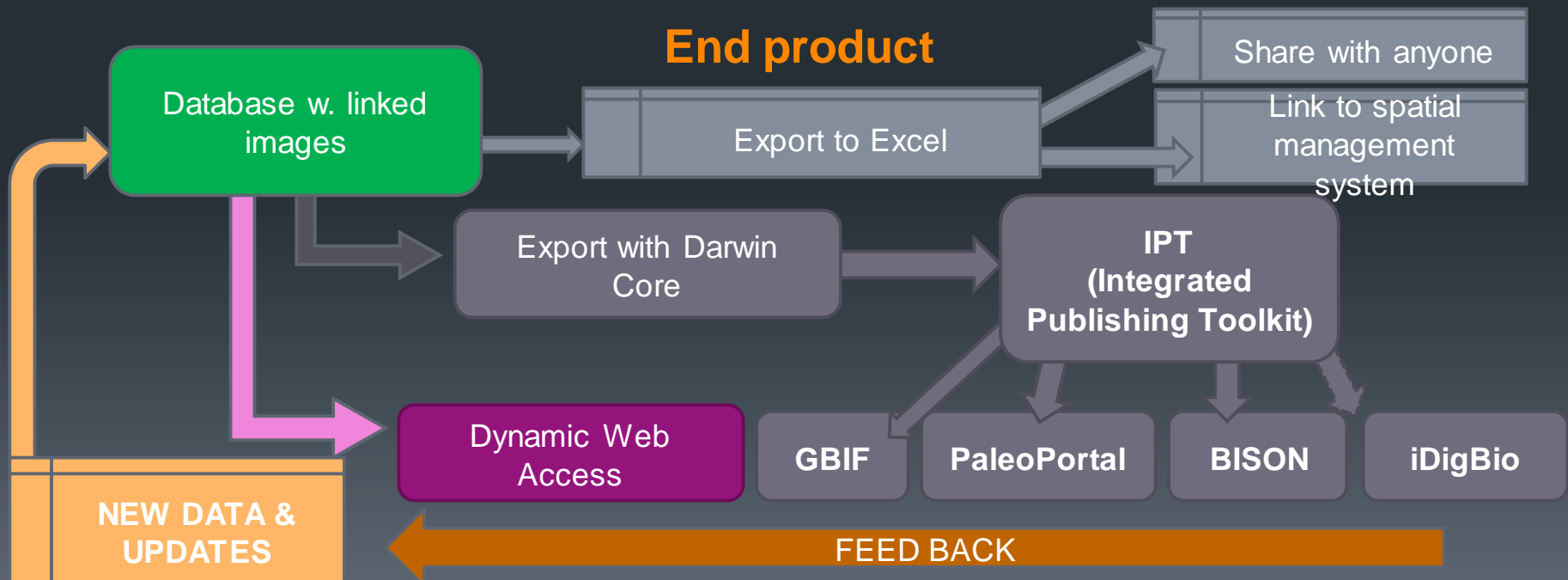
3

Goals

- Digital archive
- Useful data
- Accessible data
- 'Short' time frame
- 'Lowest' cost

Tasks

- Digitize
- Digitize accurately
- Digitize effectively
- Digitize efficiently



Major issues-

9/14/2013

4

iDigBio PaleoDigWkshp

Data migration

- Common relational database
 - Migrate from MS Access into Specify6
- Normalized data
 - Build reference trees
 - Translate data
 - Refine data
 - Upload data

Specimen imaging

- High quality images
 - Cope with multi-scale objects
 - Create excellent depth of field
 - Provide analytical ability
- Reference images
 - Basic specimen image
 - Capture label data
 - Group images (drawers)

Migration resources

9/14/2013

iDigBio PaleoDigWkshp

5

MAIN FOCUS

Translation tables

- Agent name
- BEG localities
- Chronostrat
- Lithostrat
- Storage [adding aisles]
- Taxon
- Titles [publications]

DATA
CLEAN-UP

Open Refine

- Collection number
- Locality name
- Locality [verbatim]
- Storage location
- Old storage location
- Type status
- Preparation type
- Horizon and Series
- All dates
- Duplicate fields
- Catalog remarks

MAIN FOCUS

DATA
STRUCTURE
FOR
WORKBENCH

Standardized [Normalized] data

- Import into Refine database as xls, csv,xml etc
- BUT export from Refine will be formatted in 97-2003 xls

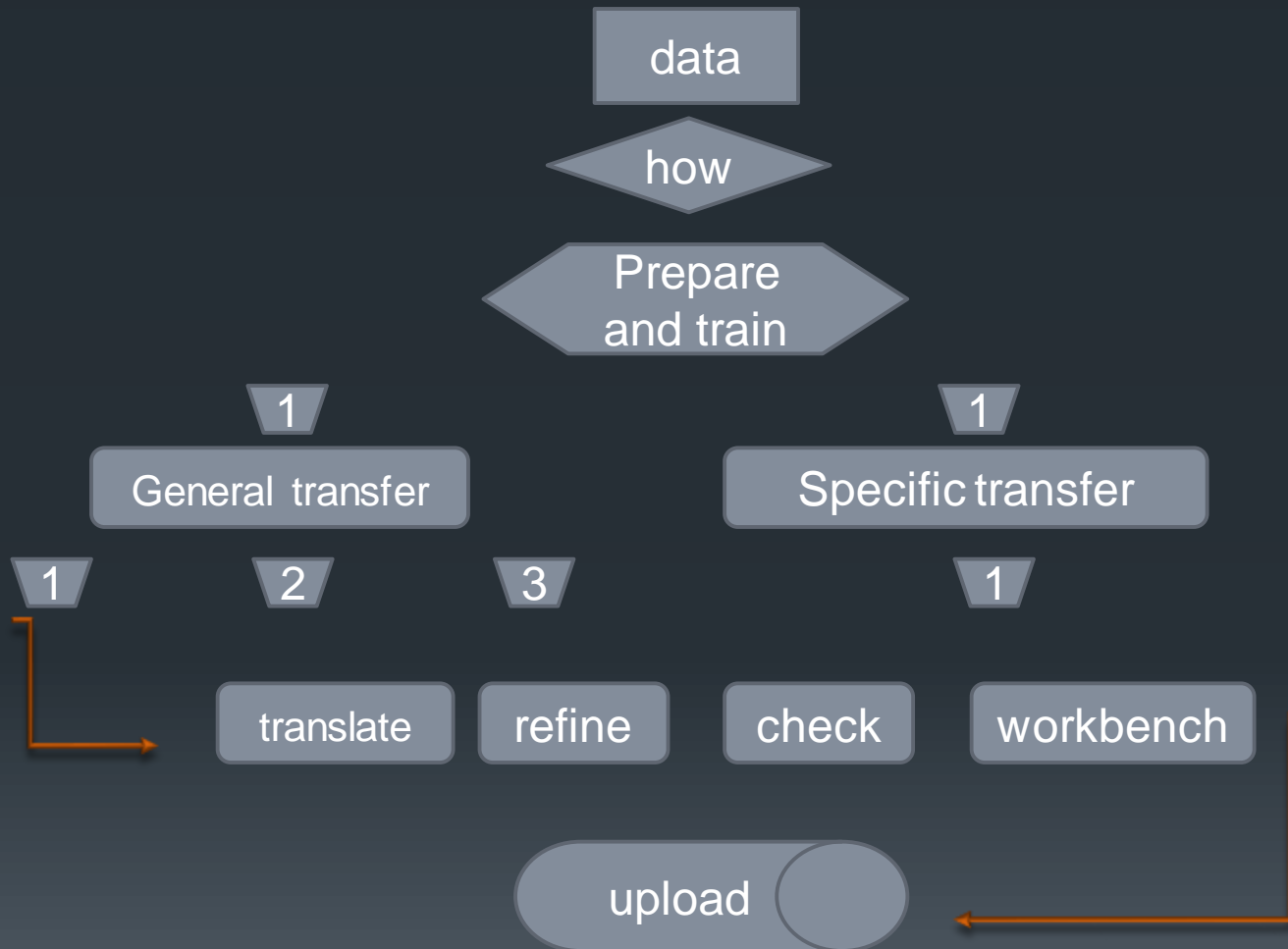
<https://github.com/OpenRefine/OpenRefine/wiki>

Migration flow

9/14/2013

iDigBio PaleoDigWkshp

6



Training tool-wiki

9/14/2013

7

iDigBio PaleoDigWkshp

Dashboard › Working with Specify 6 › Working with Specify 6 › How to prepare an Access table for Specify Upload

Search

- Detailed reference
- How to prepare an Access table for Specify Upload
- How to Write a Locality
- Latitude and Longitude
- Specify Quick Guide
- Uploading data via Workbench
- Using Google Refine to prepare a spreadsheet for upload

How to prepare an Access table for Specify Upload

Added by [Angella C. Thompson](#) last edited by [Angella C. Thompson](#) on Feb 19, 2013 (view change)

How to prepare an access table for Specify Upload

Isolate the dataset you want to work with:

"the following series of actions should take place within your copy of TAGRSURV, or any other database you create. DO NOT act on the main database in the NPL- Databases file."

First, run a Make Table query in the TAGRSURV database. Set the criteria to match what you are isolating, be it a cabinet, a certain formation or time period. You will be prompted to name your new table. If you are isolating specific catalog numbers, it is easiest to make a new table with just these numbers (via import from access, or import from a select query).

Field:	Collection catalog inventory	Specimen # catalog inventory	Suffix catalog inventory	Epoch catalog inventory	Period catalog inventory	Notes catalog inventory
Table:	catalog inventory	catalog inventory	catalog inventory	catalog inventory	catalog inventory	catalog inventory
Sort:						
Criteria:	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
or:					Jurassic	

Powered by Atlassian Confluence 3.5.17, the Enterprise Wiki | Report a bug | Atlassian News
Confluence Documentation | Web Privacy Policy | Web Accessibility

- ⊕ Detailed reference
- ⊕ How to prepare an Access table for Specify Upload
- How to Write a Locality
- Latitude and Longitude
- Specify Quick Guide
- Uploading data via Workbench
- Using Google Refine to prepare a spreadsheet for upload

Standardize

9/14/2013

8

iDigBio PaleoDigWkshp

When finished, RUN update

Field:	Family	Genus	Species	Qualifier
Table:	Jurassic	Jurassic	Jurassic	Jurassic
Update To:	[NPLTaxa]![Family]	[NPLTaxa]![Genus]	[NPLTaxa]![Species]	[NPLTaxa]![Qualifier]
Criteria:				

Refine (Google/Open)

9/14/2013

iDigBio PaleoDigWkshp

9

Google Refine recipe order.txt - Notepad

File Edit Format View Help

Run Translation Table steps 1-4 before uploading to Refine

1. delete the two exported to specify columns as well as any Access created index columns
2. Do a text facet clean up of suffix (remove decimals, set nulls to "0" and make letters uppercase)
Do a text facet clean up of collection as well.
3. -lumped1 (convertCatNum, combineOtherNums, combineRemarks, addColumns1...it also should be moving columns to the correct place)
4. Find and add accession number
5. Add columns: Qualifier 1 and Determination Remarks1
6. Check taxonomy, and add any necessary determination remarks from the Taxon column (see Liath for explanation) text filter taxon for "?" , "aff" , "cf" and "var"--many of the varieties (subspecies) are not in the taxonomy translate table.

For any taxon names that are not translated, you can just move all of that data over to determination remarks to be resolved later (or you can look up info and resolve the data right away)

7. check type columns and, then remove column Fossil Inventory catalog inventory

8. check accuracy column and Method for any useful info. In TA listed as "low", "med" or "high". (if accuracy is really long,

ADD METHOD AND ACCURACY TO LONGITUDE

value + " " + cells["Method"].value + " - Accuracy: " + cells[

9. -lumped2 (renameMoveTypeStatus, combineLatLong, combineLoca

« Start Over

Configure Parsing Options

	Loan_ID	Collection	Specimen	Suffix	CatalogNumber
1.	1	TMM	822	96	TMM00000822.096
2.	2	BEG	19163	.	BEG00019163.000
3.	2	BEG	20928	.	BEG00020928.000
4.	2	BEG	19168	.	BEG00019168.000
5.	2	BEG	19166	.	BEG00019166.000
6.	2	BEG	19167	.	BEG00019167.000
7.	2	BEG	20938	.	BEG00020938.000
8.	2	BEG	19164	.	BEG00019164.000
9.	2	BEG	20926	.	BEG00020926.000

Workbench upload

9/14/2013

10

iDigBio PaleoDigWkshp

Data Set Mapping Editor

Data Types

- Acc Accession
- Agnt Agent
- CE Collecting Event Attribute
- CI Collecting Information
- CT Collecting Trip
- CO Collection Object
- COA Collection Object Attribute
- DNA DNA Sequence
- Det Determination
- FN Field Notebook
- FNp Field Notebook Page
- FNs Field Notebook Page Set
- Gcd Geo Coord Detail
- Geo Geography
- CS Geologic Time Period
- Jou Journal
- LS Litho Strat
- Loc Locality
- OI Other Identifier
- PC Paleo Context
- Pre Preparation
- RW Reference Work**
- Tax Taxon

Specify Fields

- Format
- Article/Book Title**
- Publisher
- Place of publication
- Date published
- Volume
- Pages
- URL
- Library number
- Author First Name1
- Author Last Name1
- Author Middle1
- Author Title1
- Author First Name2
- Author Last Name2
- Author Middle2
- Author Title2
- Author First Name3
- Author Last Name3
- Author Middle3
- Author Title3
- Author First Name4
- Author Last Name4

Specify Fields

- Acc Accession Number
- CO Catalog number
- Det Type status 1
- To Be Discarded**
- Det Genus1
- Det Species1
- Det Subspecies1
- Det Variety1
- Det Qualifier 1
- CI Collector First Name1
- CI Collector Last Name1
- CI Verbatim date
- CI Field number
- PC Horizon
- PC Series
- LS Litho Group
- LS Formation
- LS Member
- CS Era
- CS Period
- CS Subperiod
- CS Epoch
- CS Age

Data Set Columns

- accession
- Catalog Number
- Type status
- Title 1**
- Genus
- Species
- Subspecies
- Variety
- Qualifier
- CollectorFirstName
- CollectorLastName
- Collection date
- Collectors number
- Horizon
- Series
- Litho Group
- Formation
- Member
- Era
- Period
- Sub-System/Period
- Epoch
- Stage

Specify Field Description: Title of reference

OK Cancel Help

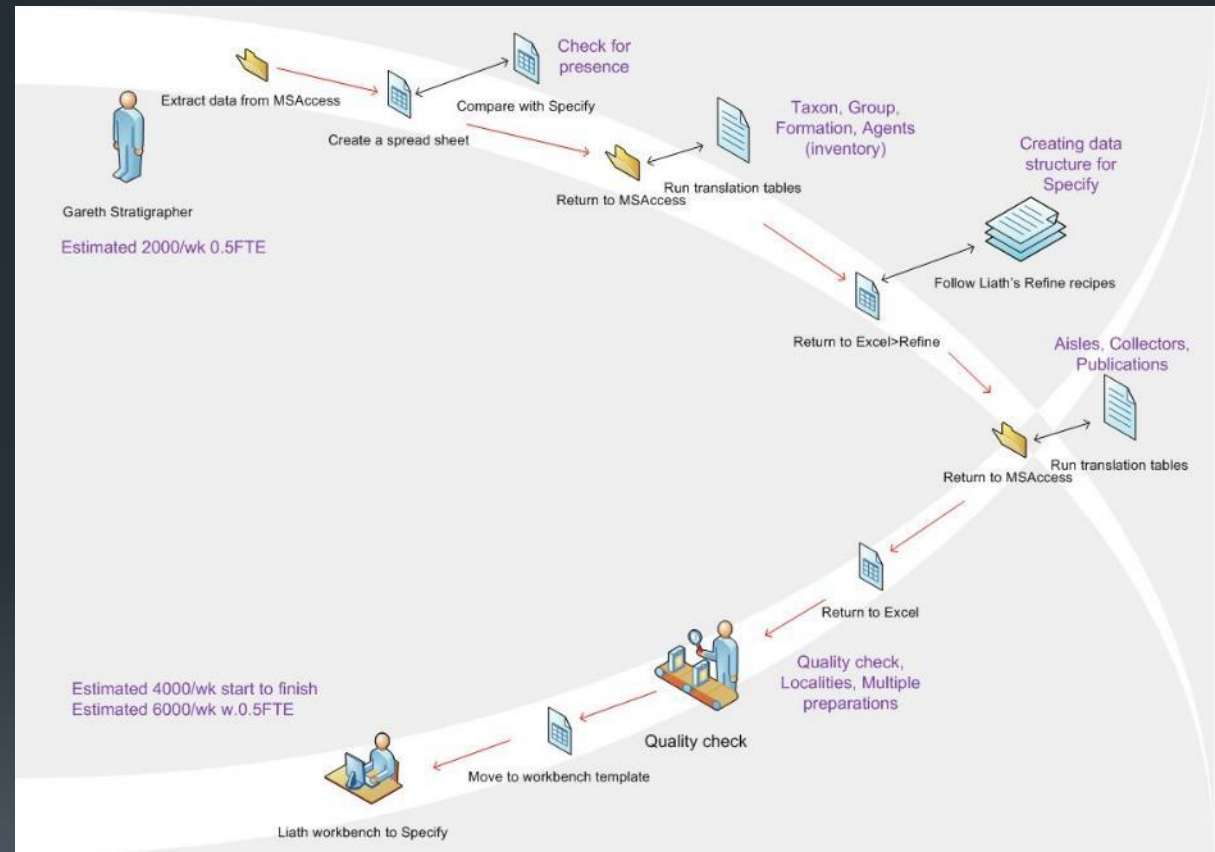
Example workflow

9/14/2013

11

iDigBio PaleoDigWkshp

- Content 'expert' to database 'expert'
- ADD time to develop protocols
 - Translation tables from legacy data
 - Refine recipes to structure data



Workflow rates

9/14/2013

iDigBio PaleoDigWkshp

12

Data migration

- 4000 records per week-
refine and migrate to
Specify via workbench
(1FTE)
- 6000 per week, with
1.5FTE
- 24000/month, 300,000/year

How long will it take?

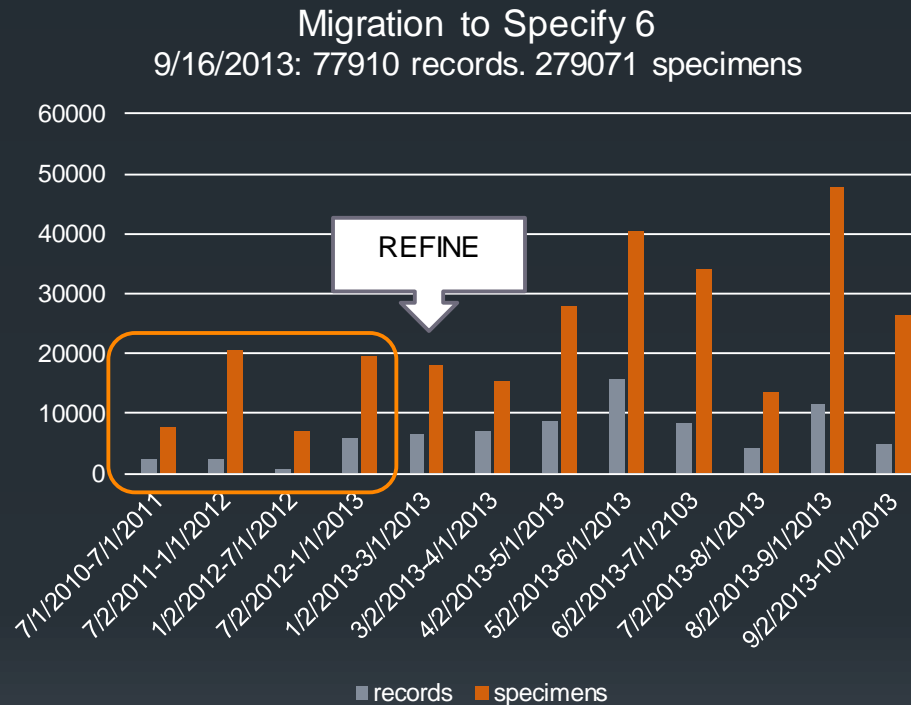
- Current record count from
all databases 500,000
- TWO years
- Assume no major glitches

Data migration rate

9/14/2013

13

iDigBio PaleoDigWkshp



NB: Graph x-axis variable time scale

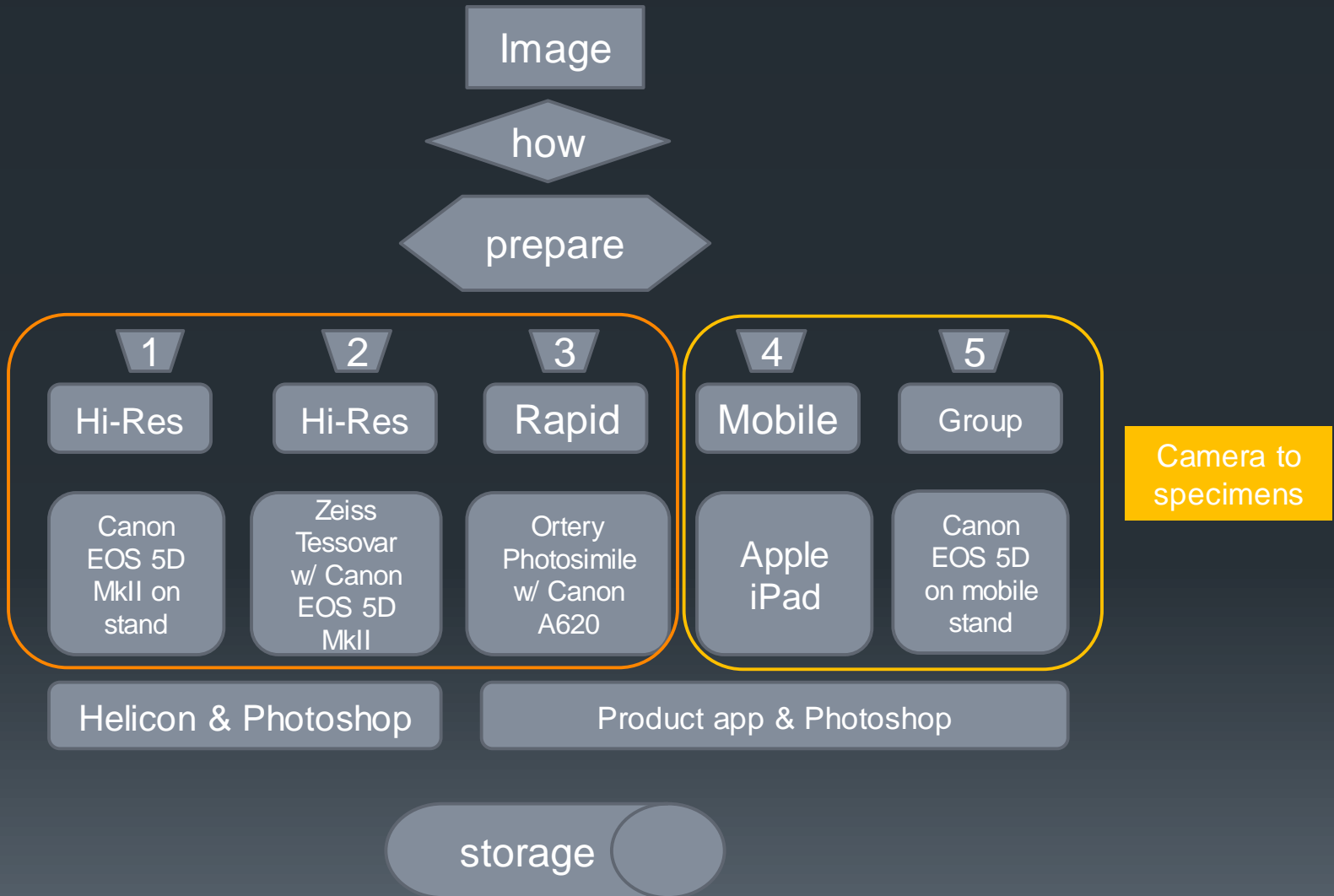
- iDigBio 'Train the trainers' georeferencing workshop 10/8/2012
- In-house transfer of training knowledge 11/2012
- Rate is FTE constrained and FTE knowledge base

Image flow

9/14/2013

14

iDigBio PaleoDigWkshp



Common

- File naming convention
- One image views
- Standard lighting
- Scale bar
- JPEG and PDF

Specific to HiRes

- Multiple views
- Multi-focus composites
- Variable lighting
- Standard scale bars
- Embedded scaling
- TIFF and JPEG

Variable

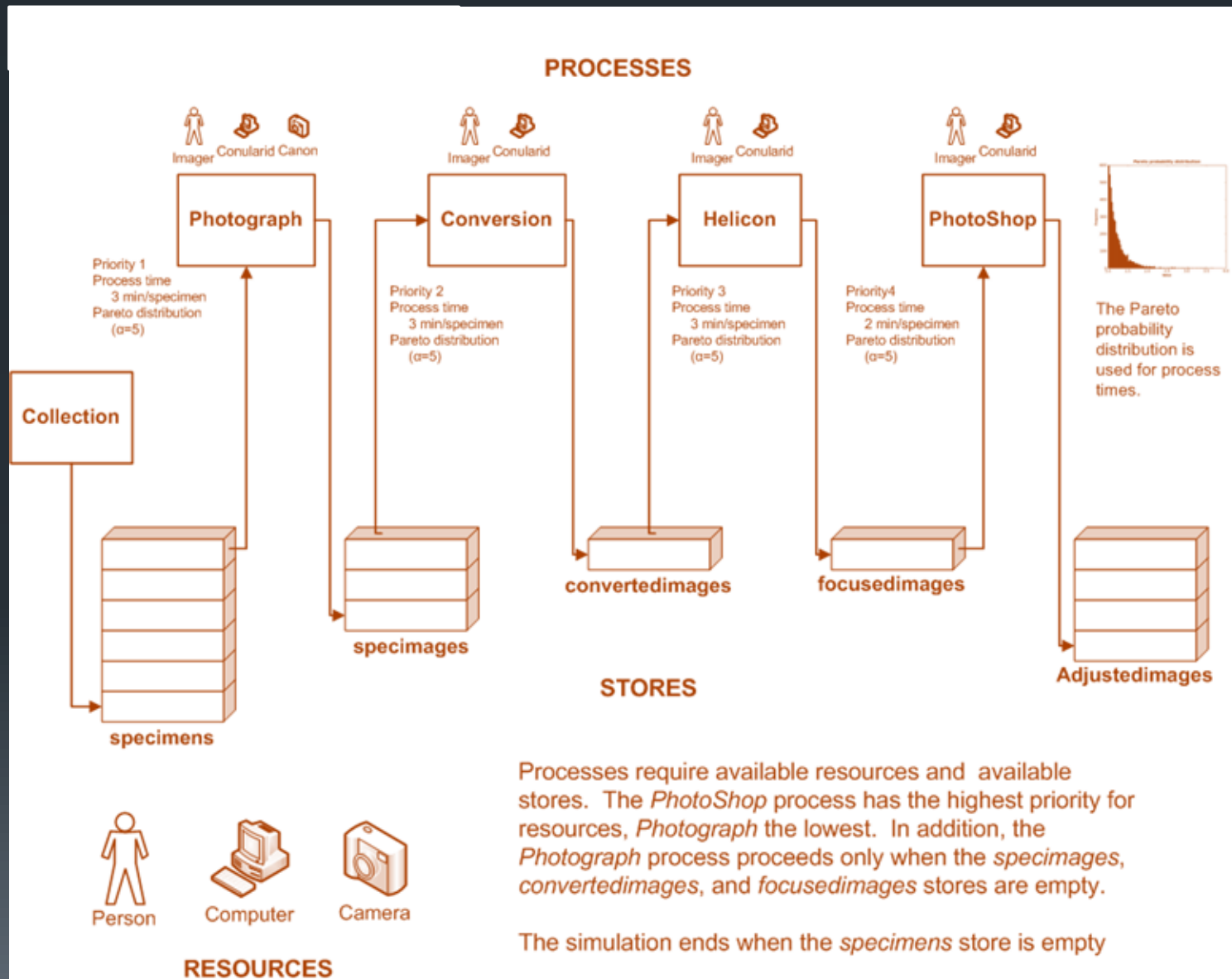
- Ammonium chloride smoked
- India ink or less permanent stain
- Emersion

3D object imaging

9/14/2013

16

iDigBio PaleoDigWkshp

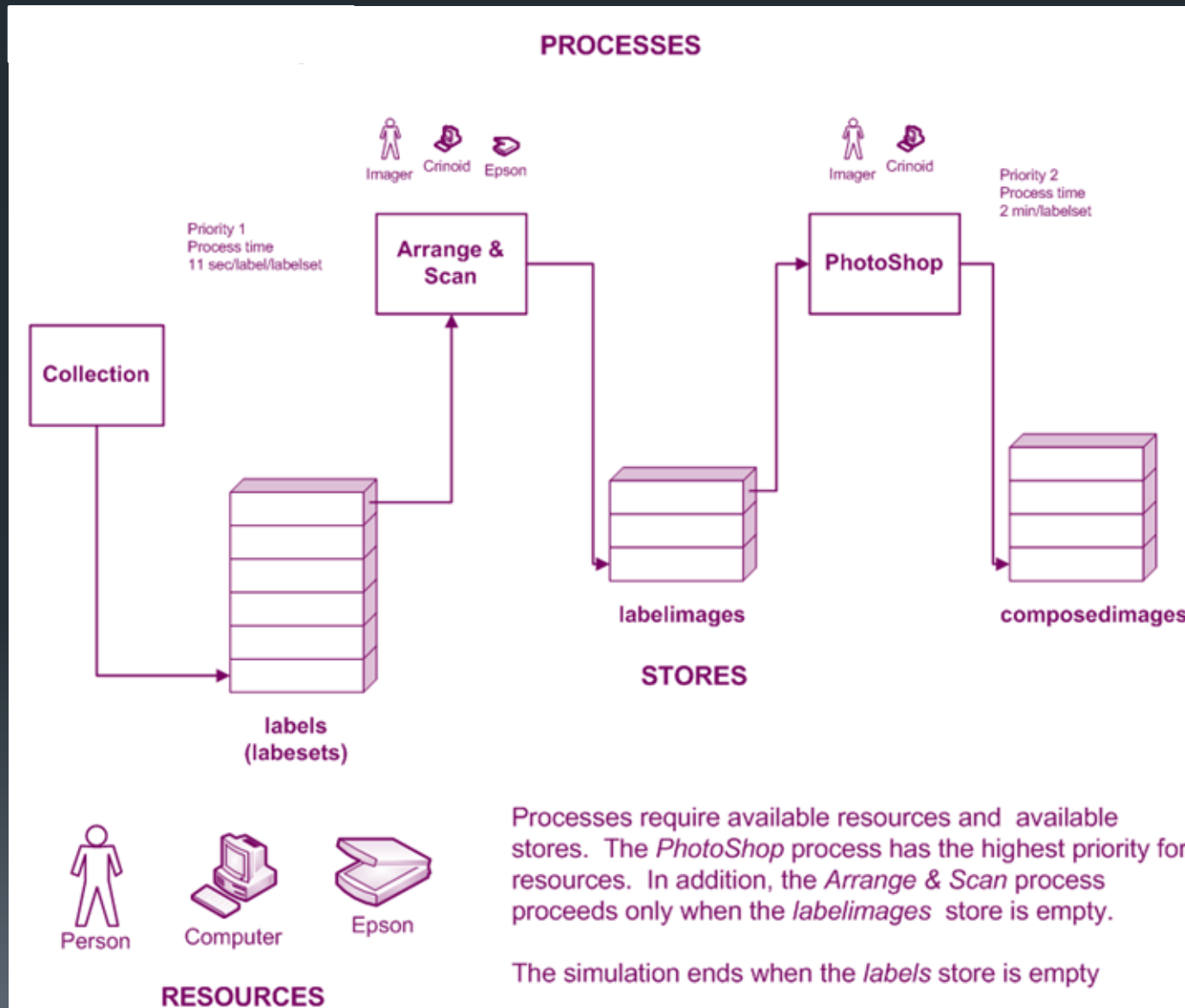


2D object scanning

9/14/2013

17

iDigBio PaleoDigWkshp



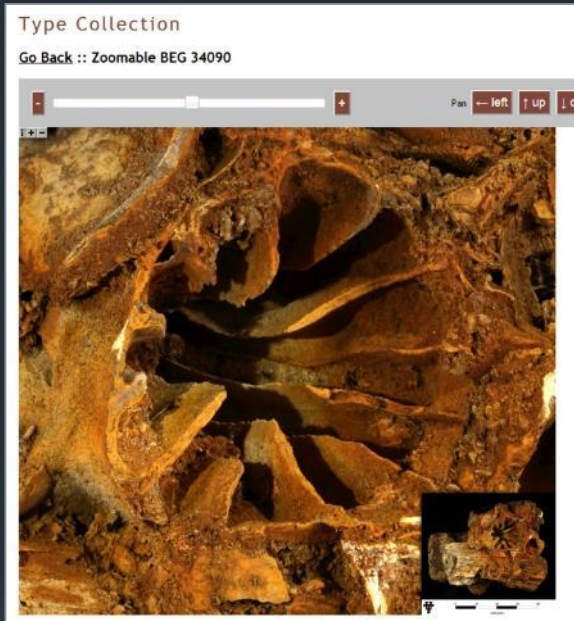
Hi-Res product

9/14/2013

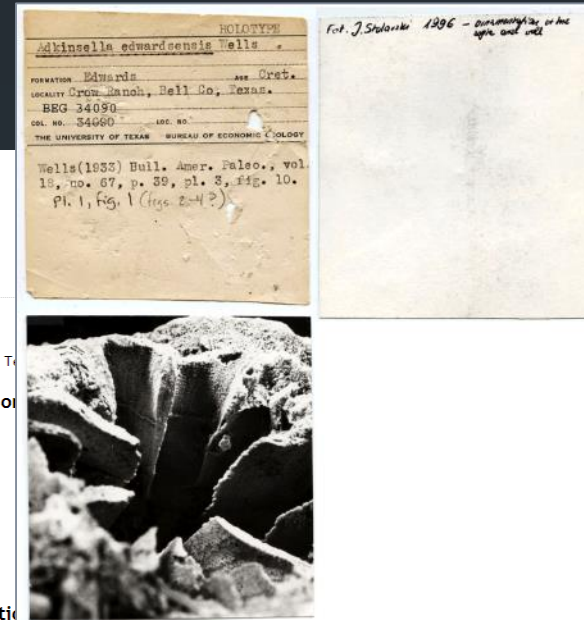
18

iDigBio PaleoDigWkshp

3D- Specimen image



2D-Labels and photos



ion

Locality Verbal

Crow Ranch, Bell County, T

Geographic Location

Bell County, Texas, USA

Geologic Age

Era: Mesozoic
Period: Cretaceous
Epoch: Early
Age: Albian

Stratigraphic Position

Group: Fredericksburg
Formation: Edwards
Member:

Storage Location

PRC122, Types, 7, 182

Collector

Collector: WS Adkins
Collection Date:

Publication

Wells (1933) Bull. Amer. Paleol., v. 18, no. 67, p. 39, pl. 3, fig. 10, pl. 1, figs. 1-4

Images

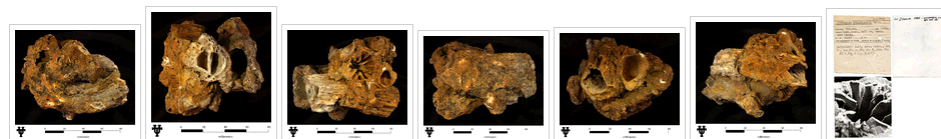


Image workflow

9/14/2013

19

iDigBio PaleoDigWkshp


A



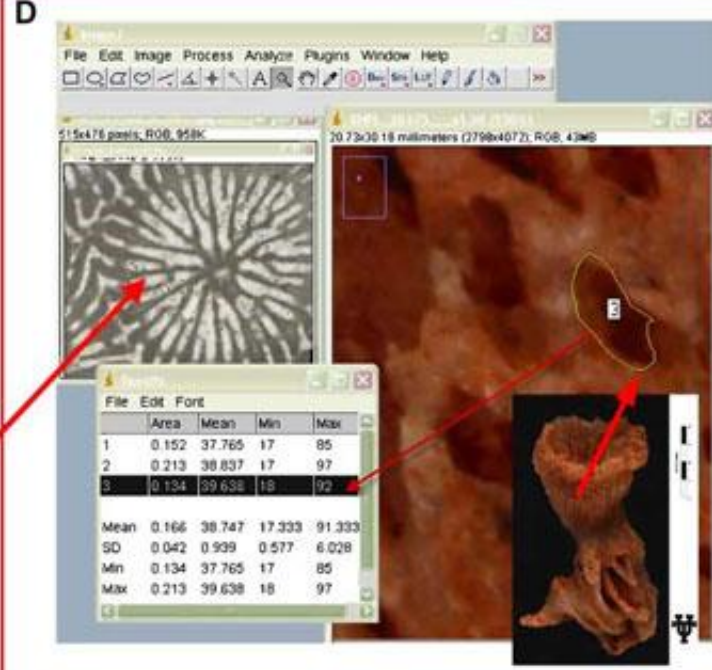
B



C




D



	Area	Mean	Min	Max
1	0.152	37.765	17	85
2	0.213	38.837	17	97
3	0.134	39.638	18	92
Mean	0.166	38.747	17.333	91.333
SD	0.042	0.939	0.577	6.028
Min	0.134	37.765	17	85
Max	0.213	39.638	18	97

E



PaleoCentral

HOME MAP RECORDS REFERENCES TOOLS LOGIN

Types Collection

Records » TX 1121.66

« prev rec

Scientific Name

Eopinnacrinus pinnulatus Brower & Veinus n.gen., n

Common Name

crinoid

Taxonomic Hierarchy

Phylum: Echinodermata

Class: Crinoidea

Order:

Family:

Collector

Collector: Graffham, A.

Collection Date:

Publication

Brower & Veinus. 1982. Sprinkle, J. T (Ed.), University of Kansas

Paleontological contributions, Monograph 1, pl. 11 fig. 3, 7, fig. 42a, fig. 42b, p. 141.

Locality Verbal

West Branch Sycamore Creek, Daube Ranch, Jo

Geographic Location

Johnston County, Oklahoma, USA

Geologic Age

Era: Paleozoic

Period: Ordovician

Epoch: Late

Age: Katian (Mohawkian)

Stratigraphic Position

Group: Simpson

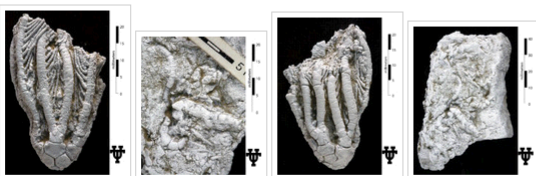
Formation: Bromide

Member: Mountain Lake

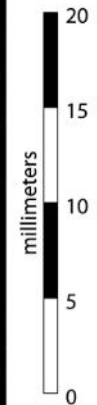
Storage Location

PRC122, Types, 19, 50

Images



Eopinnacrinus pinnulatus Brower & Veinus n.gen., n



view zoomable image

1/4

CLOSE

HiRes imaging

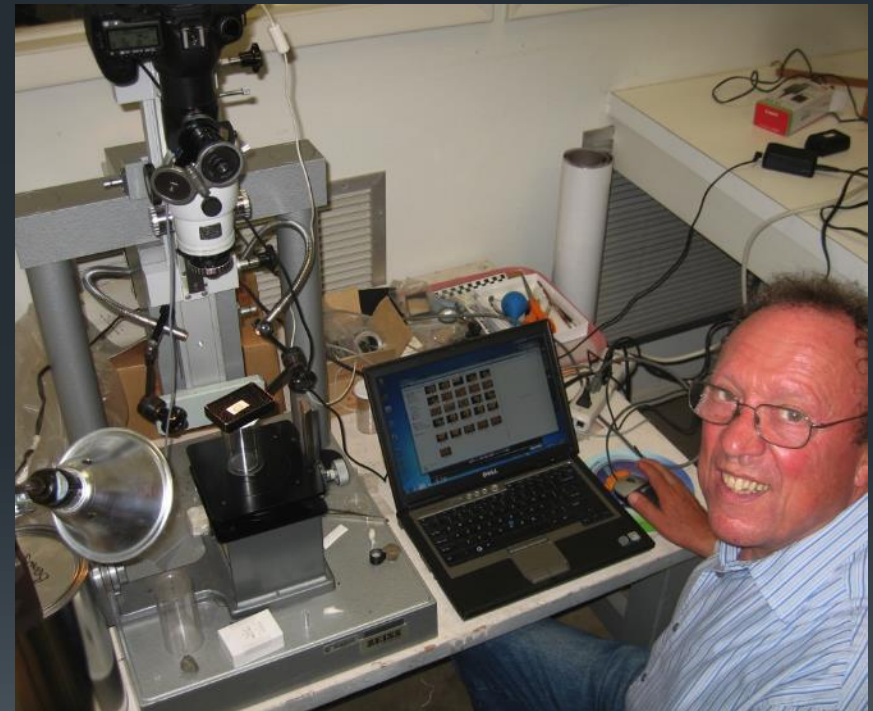
9/14/2013

21

iDigBio PaleoDigWkshp

Zeiss Tessovar

- Found adaptor on line for Canon EOS 5 Mark II
- Using the same software protocols
- Especially useful for small specimens



Improving throughput

9/14/2013

22

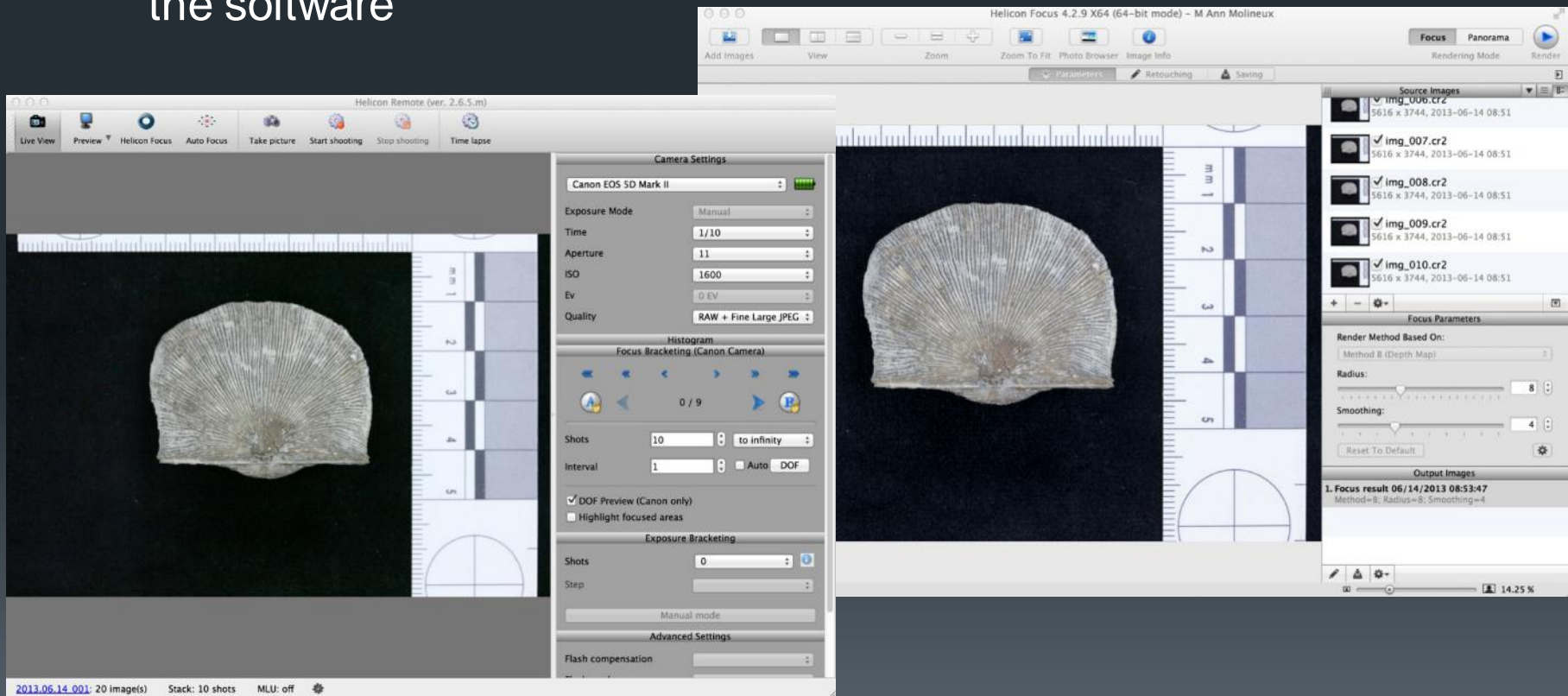
iDigBio PaleoDigWkshp

Remote shooting

- Helicon remote
- Camera lens is driven by the software

Multi-focus rendering

- Economic multi-focus shooting



Batch processing

- Final PhotoShop processing includes several JavaScripts, batch routines to generate scaled images, image metadata, standard backgrounds, scale bars, and trademarks.
- Scripts are available at <http://www.utexas.edu/tmm/npl/projects/imaging>



Analytical quality

9/14/2013

24

iDigBio PaleoDigWkshp

JavaScripts applied to rendered image

The screenshot displays two software windows. The left window, titled 'Helicon Focus', shows a fossil shell with a scale bar at the bottom ranging from 0 to 40 millimeters. The right window, titled 'ImageJ', shows a 'Results' table with the following data:

	Area	Mean	Min	Max	Angle	Length
1	0.198	65274.940	0	65535	0	10

Below the 'Results' table, there is another table with the following data:

Scale Factor	Count	Length	Angle
1.000000	1	511.318300	-0.895751
1.000000	1	930.019323	-0.369346

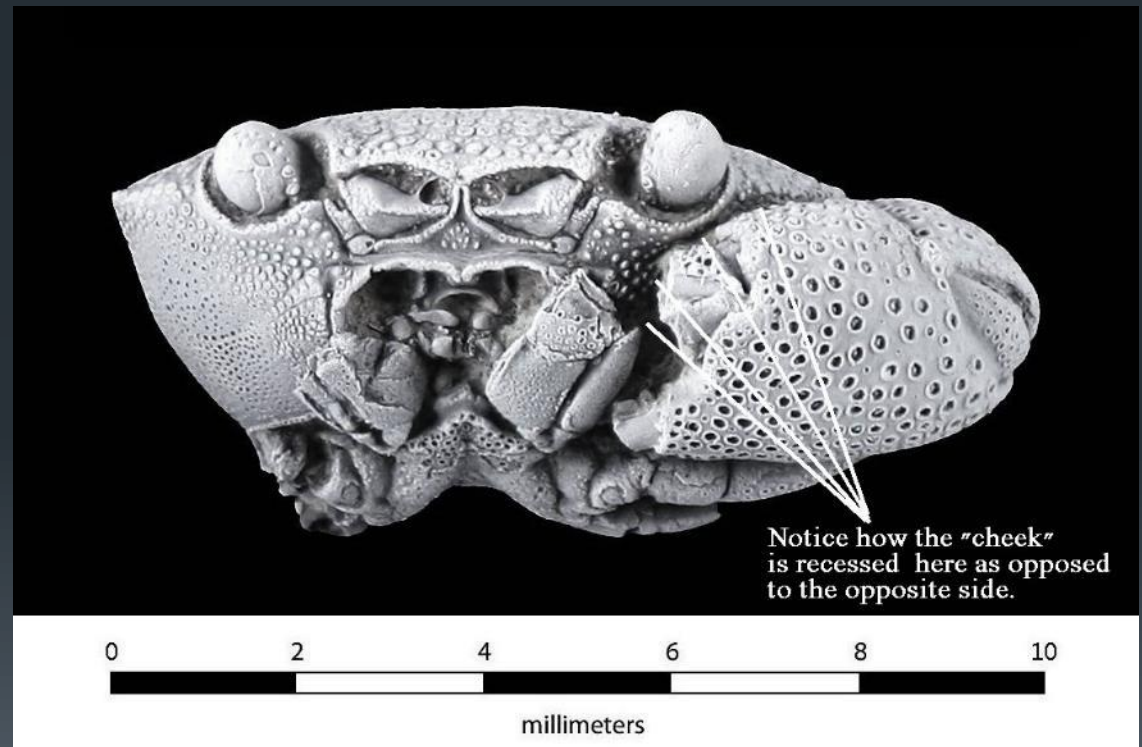
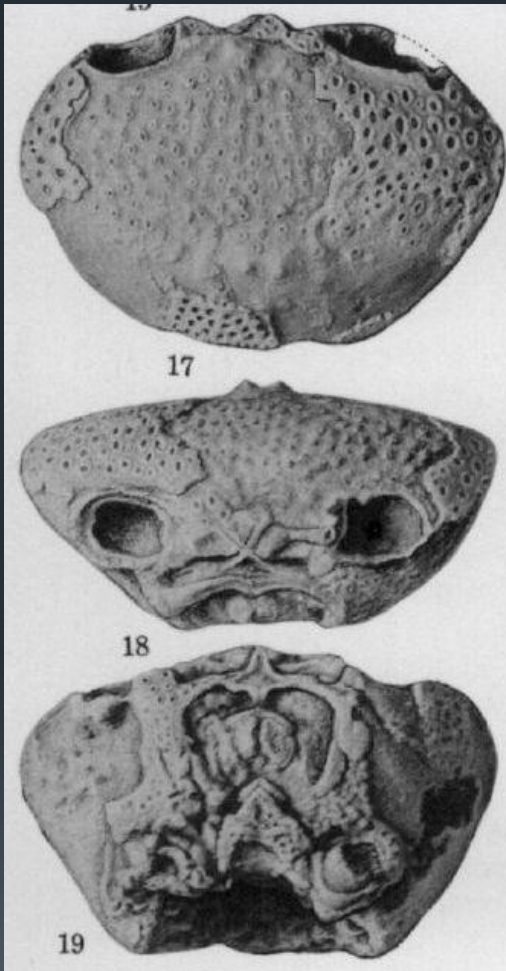
Example using ImageJ

Images 'loaned'

9/14/2013

25

iDigBio PaleoDigWkshp



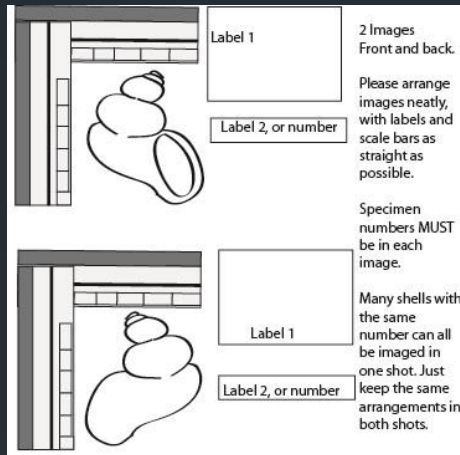
Improving throughput

9/14/2013

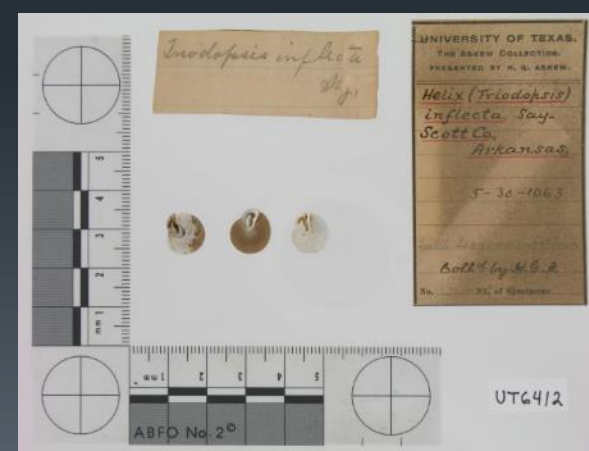
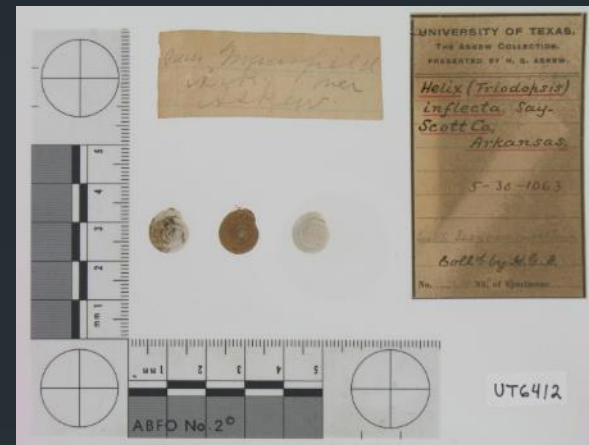
iDigBio PaleoDigWkshp

26

Basic reference + label



Minimal processing



Label text is legible: Specimen detail adequate: Thumbnails for drawers

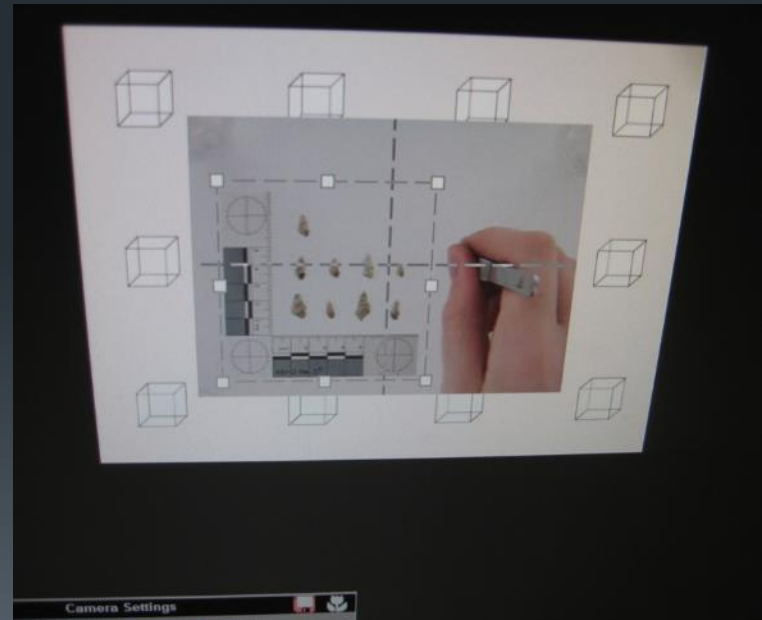
Rapid imaging

9/14/2013

27

iDigBio PaleoDigWkshp

Stationary Photosimile light box



In situ imaging

Individual and drawer imaging

9/14/2013

28

iDigBio PaleoDigWkshp

Basic inventory in situ

- Select drawer
- Clean and conserve specimens
- Count specimens
- Complete inventory sheet
- Make specimen # visible

Basic drawer imaging

- Image whole drawer
- Link to ArcOnline for browsing

Basic specimen imaging

- Select drawer
- Remove specimen from bag
- Flatten labels beneath glass
- Add specimens
- Image using iPad
- Bag historic labels
- Return drawer
- Attach image files to Specify record
- Create thumbnails to form 'drawer'
- Link to ArcOnline for browsing

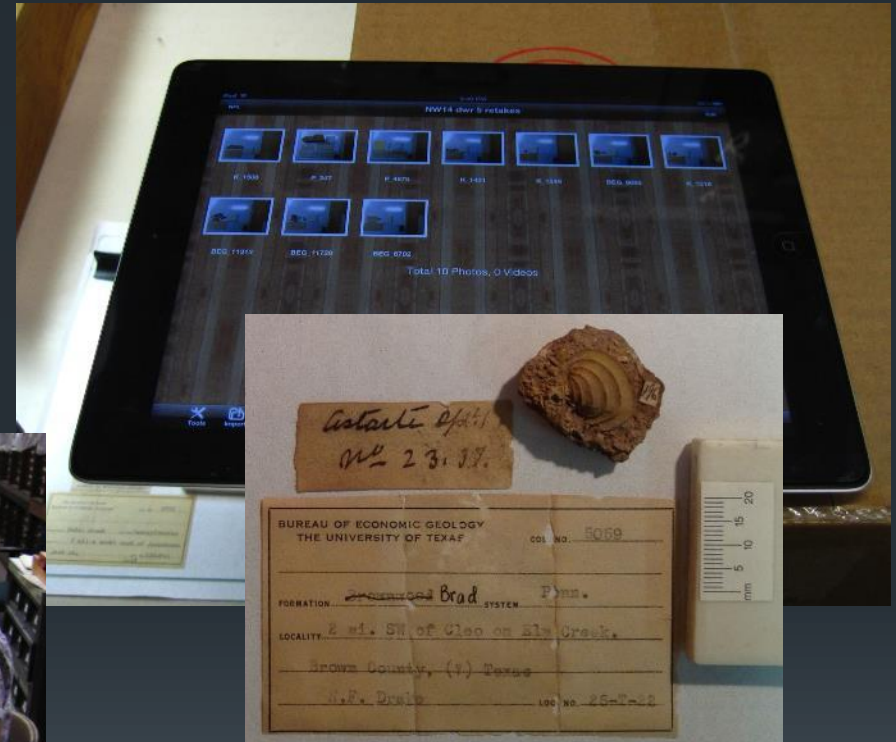
Rapid imaging flow

9/14/2013

29

iDigBio PaleoDigWkshp

Inventory-conserve-image



Results

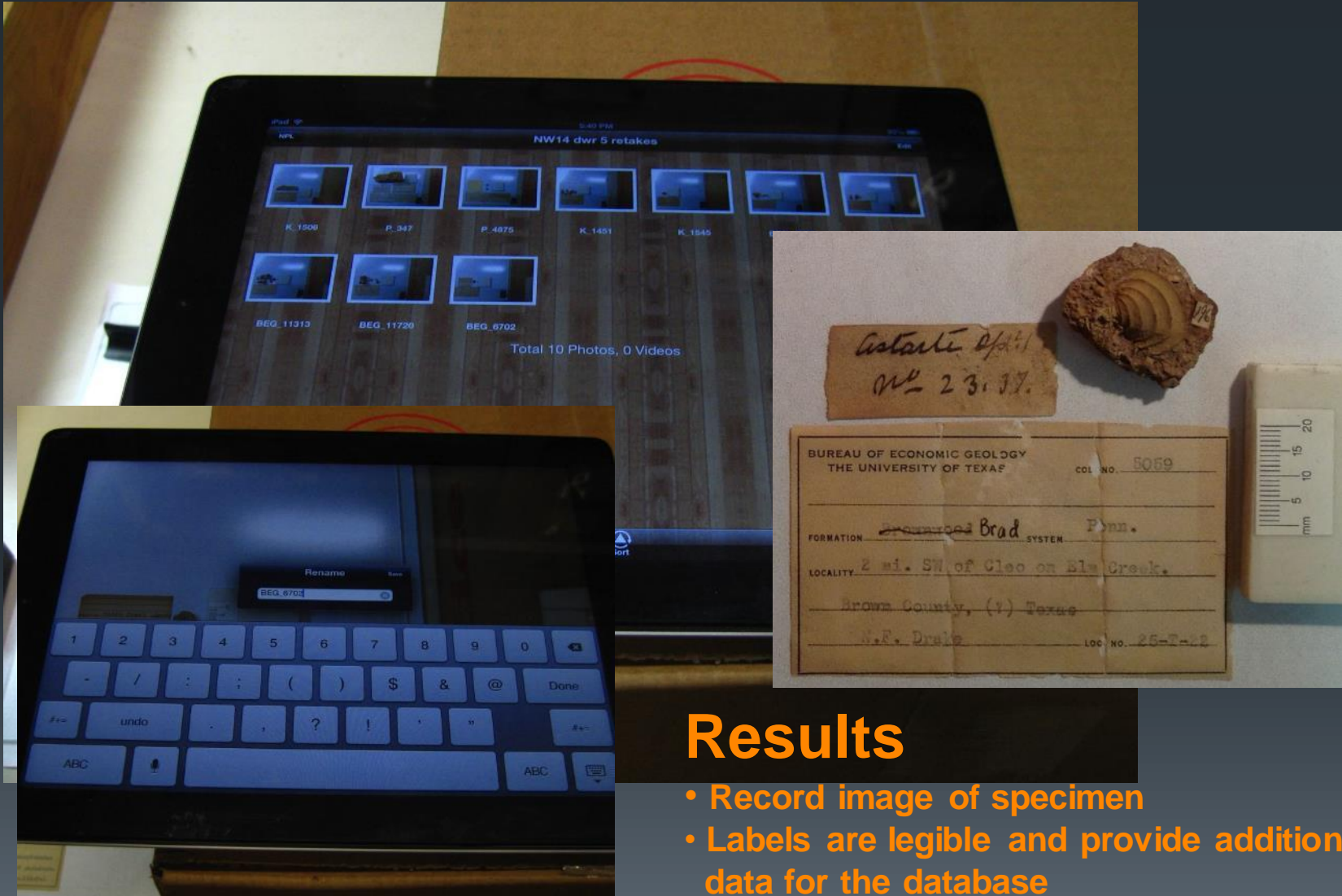
- Record image of specimen
- Labels are legible and provide additional data for the database
- Can be linked to ArcOnline for browsing

In situ imaging

9/14/2013

30

iDigBio PaleoDigWkshp



Results

- Record image of specimen
- Labels are legible and provide additional data for the database
- Can be linked to ArcOnline for browsing

- Imaging
 - High resolution- 5 per hour [on both Tessovar and Stand]
 - Processing-file numbering, composites, scaling
 - Basic imaging- 9-17 lots per hour
 - In situ (Mobile) imaging-17 lots per hour
- Trade-offs
 - Hi-Res- research quality- higher training curve
 - Rapid- inventory- low training curve
 - In situ- inventory- flexibility
- Results
 - All capture label data
 - All can be linked to the database as attachments
 - Complete drawer imaging provides a feasible way to virtually view via the GIS system

Online assets

9/14/2013

32

iDigBio PaleoDigWkshp

- Main campus access
 - Teaching
 - Research
- Global access
 - Research
- Query access
 - Research
 - Public enquiry
- Inventory
- Conservation
- Space assets

ArcGIS ▾ NPL_PRC33

Details Add Basemap Save Share Print Measure Bookmarks Find

NPL_PRC33

NPL Collection in PRC33
Web Map by RBurroughs
Last Modified: August 21, 2013
☆☆☆☆☆ (0 ratings, 0 comments, 233 views)
More Details...

Make your own map
Add to this map
Make a new map

Esri.com · Help · Terms of Use · Privacy · Contact Us · Report Abuse

University Of Texas At Austin

Online accessibility

9/14/2013

33

iDigBio PaleoDigWkshp



[Home-NPL](#) | [Home-JSG](#)

[About NPL](#)

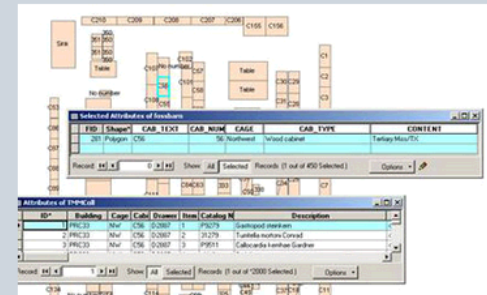
[Practices & Protocols](#)

GIS-Geographic Information Systems

The databases are linked to a geographic information system (GIS) map of both repository buildings. This allows us to find information about each specimen when in the repository, and also to know where it is located within the buildings of the repository.

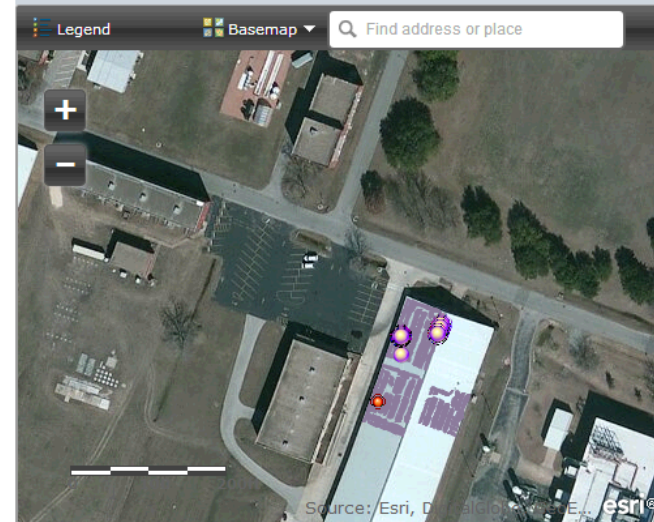
Current developments include on-line access to enable virtual browsing of the specimens with whole drawer images.

The GIS linked system improves our emergency management and will enhance our pest management. For example, we can locate zones susceptible to concrete moisture 'weeping' and ensure that any cabinets in those areas are raised off the floor surface. We can also monitor areas with noticeable pests and target mitigation procedures more effectively.



Part of the GIS map of the NW cage in the main repository. The tabular details refer to the highlighted cabinet on the map. Other aspects of the collection environment are tied to this system.

Visually browse our collections below or access more details via this [link](#)



Online browsing

9/14/2013

34

iDigBio PaleoDigWkshp



ArcGIS - NPL_PRC33 NEW MAP SIGN IN

Details Add Basemap Save Share Print Measure Bookmarks

312 - Drawer 3

Contents

- TAGSurv
- Cabinet 312
- PRC122
- PRC33 Layout (No Specimen Data)
- Imagery with Labels



University Of Texas-JJ Pickle Ctr

Microsoft, DigitalGlobe | Copyright: © 2013 Esri, DeLorme, NAVTEQ, TomTom **esri**

TAGSurv (8,179 features, 0 selected) Table Options X

ID	Autonumber	Specimen #	Cabinet #	Drawer #	Taxon	Locality	County	State	Collector	Period
82,631	97,668	7,225	9,921	15	Gryphaea sp.	Villers-Sur-Mer, cliff of Vaches-Noire, Channel Coast of Calvados, west of Deauville			Young	Jurassic

Esri.com | Help | Terms of Use | Privacy | Contact Us | Report Abuse

Goal

9/14/2013

35

iDigBio PaleoDigWkshp

