TaxonWorks as a Collection Managment Tool

M. Yoder, R.E. DeWalt, D. Dmitriev INHS, University of Illinois

ECN 2015, Minneapolis





Outline

Motivation

Species File Group

TaxonWorks

Use cases

Frost Odonates

INHS Collection Data

Moving forward

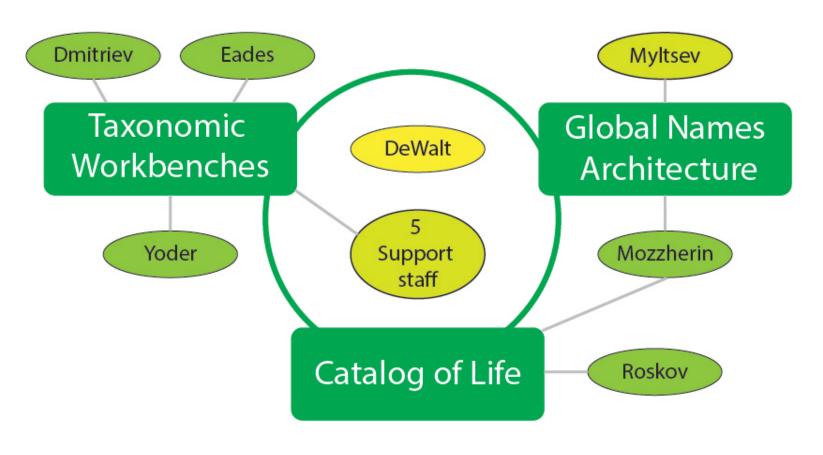
a FAQ for collection managers:

Now

Soon

In a while

Species File Group



endowed, long term funding

Taxonomic Workbenches

mx

SpeciesFiles

3i

deprecated languages, desktop constraints, loose semantics

TaxonWorks

TaxonWorks

from insight to archive comprehensive, hard core taxonomy web based data and tasks tokenized API open source

Core/basic data

taxonomists

Specimens

Taxon names

Collecting events

Descriptions

References

Images

Identifiers



collections

Loans

Collection profiles

Biocuration categories

Preparation types

Advanced data

Core Data

Georeferencing

Data Attributes

Labels

Notes, Tags

Interactive Keys

Matrices

Citations

Identifiers

Alternate Classifications

Data- import and exposure

Both Out ln **CSV** Scaffolded forms **JSON API** "Task" forms Console scripts **DwCA DB** Dumps **DwC A Symbiota JSON API IPT**

Coupling projects and developmnet

Securing Several Large Donations and Orphaned Collections of Alcohol Preserved Insects and Spiders and Improving Storage of Wet Types at the INHS

Simple, staged, digitization



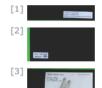
Simple, staged, digitization

tape off a stage position metadata position specimen, labels take picture specimen batch load all images image depiction transcribe verbatim depiction metadata match/parse verbatim

Task - transcribe depiction

Task: Transcribe depiction

sqed depiction: 8 | original image





Drag box to ocr.
Drag box to dail

Curator metadata Use [a]	Identifier Use [s]	Annotated specimen Use [d]
Created By: Emily Sandall Created On: 6 March 2015 Repository: Frost Entomological Museum (PSUC)	PSUC FEM """ 000042831	TEAMEA LACERATA Hagen ' chthvGIiBmvh HI Dallas, TEXAS 12 August,1959 O & R Sanders 0011'

Identifiers		Taxon Determinations
Namespace	Identifier	Name
Select a namespace	k	Select a otu
Buffered collecting event	Buffered determinations	Buffered other labels
Total 1 Note	Tag with 💠	
Save changes Save and next [n]	Skip to next	

Task - transcribe depiction

Identifiers

Task: Transcribe depiction

sqed depiction: 8 | original image



Curator metadata Use [a]	Identifier Use [s]	Annotated specimen Use [d]
Created By: Emily Sandall Created On: 6 March 2015 Repository: Frost Entomological Museum (PSUC)	PSUC FEM """ 000042831	TEAMEA LACERATA Hagen ' chthvGIiBmvh HI Dallas, TEXAS 12 August, 1959 O & R Sanders 0011'

Taxon Determinations

Namespace		Identifier	Name
Select a namespace	□ lock		Select a otu
Buffered collecting even	nt	Buffered determinations	Buffered other labels
_			
Total 1 Note		Tag with 📑	
Save changes Save and	next [n]	Skip to next	

Drag box to ocr.

Append [q], Replace [w]

Task - transcribe depiction

Task: Transcribe depiction

sqed depiction: 8 | original image







Dallas, IFEMS 12 August,1939 0 8: R Sanders c011./ A

Append [q], Replace [w]

Dallas, TEXAS
12 August, 1939
0 & R Sanders coll.

Curator metadata Use [a]	Identifier Use [s]	Annotated specimen Use [d]
Created By: Emily Sandall Created On: 6 March 2015 Repository: Frost Entomological Museum (PSUC)	PSUC FEM """ 000042831	TEAMEA LACERATA Hagen ' chthvGIiBmvh HI Dallas, TEXAS 12 August,1959 O & R Sanders 0011'

Identifiers

Namespace		Identifier	
Frost Entomology Collection	✓ lock	000042831	
Buffered collecting event		Buffered determinations	
Dallas, TEXAS 12 August,1939 O & R Sanders coll.		TRAMEA LACERATA Hagen [Female] det. by G.H.Beatty III	

ched Tag with :

Taxon Determinations

Name	
Select a otu	

Buffered other labels
Created By: Emily Sandall
Created On: 6 March 2015
Repository: Frost
Entomological Museum
(PSUC)

Total	1	Note	Head detached
Save	change	s Sav	e and next [n] Skip to next

INHS Insect Collection: Overview

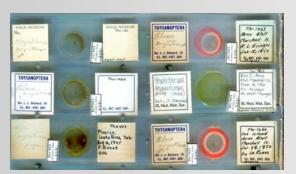
- One of the oldest and most comprehensive insect collections in North America (established in 1858)
- 9th largest collection in the USA (~7 million prepared specimens)
- >13,000 primary type specimens





Digitized material

- Specimen-level electronic database, since 1990's
- NSF funded projects
 - Digitize label data
- Retrospective georeferenced localities
 - Accessible via iDigBio and GBIF portals
 - Ephemeroptera, Plecoptera, Trichoptera
 - Hymenoptera
 - Central Asian grassland fauna
 - Hemiptera
 - Orthoptera



slide mounted specimens



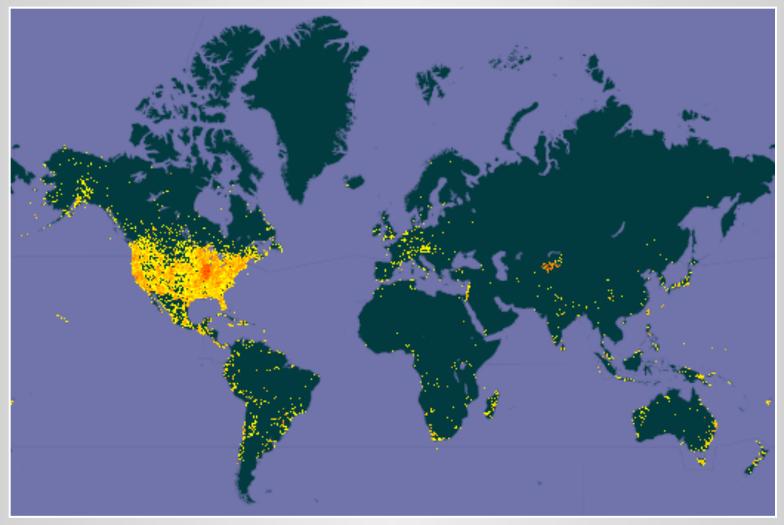
ethanol-preserved specimens



pinned specimens



Collection Digitization: Global Coverage



GBiF

INHS Insect Collection Database

Original sources

- •Insect collection database: File Maker Pro
- •TCN specimen records: SQL database
- •NEON specimen records: Excel file
- •Collection profile database: MS Access database

Intermediate processing

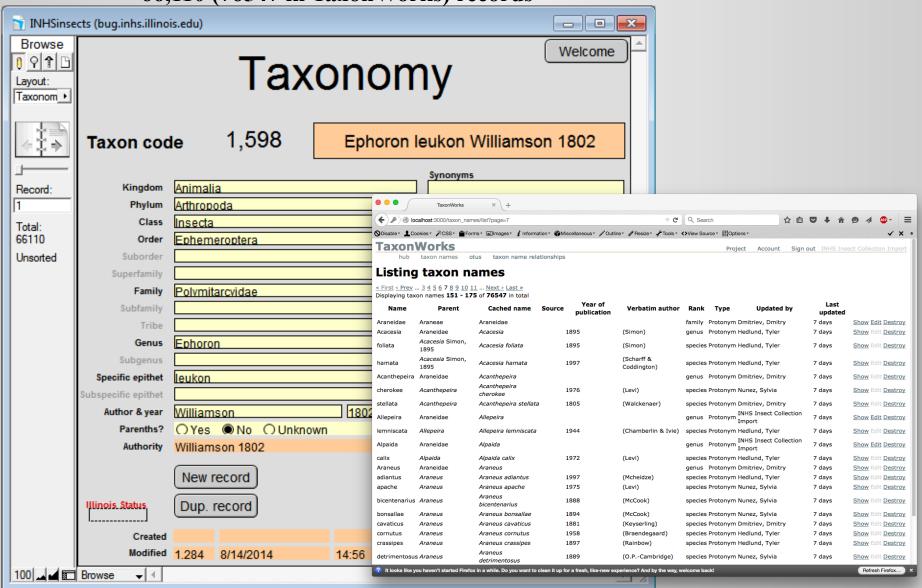
•MS Access database, Tab delimited txt files

Imported to

•TaxonWorks PostgreSQL database using Ruby script

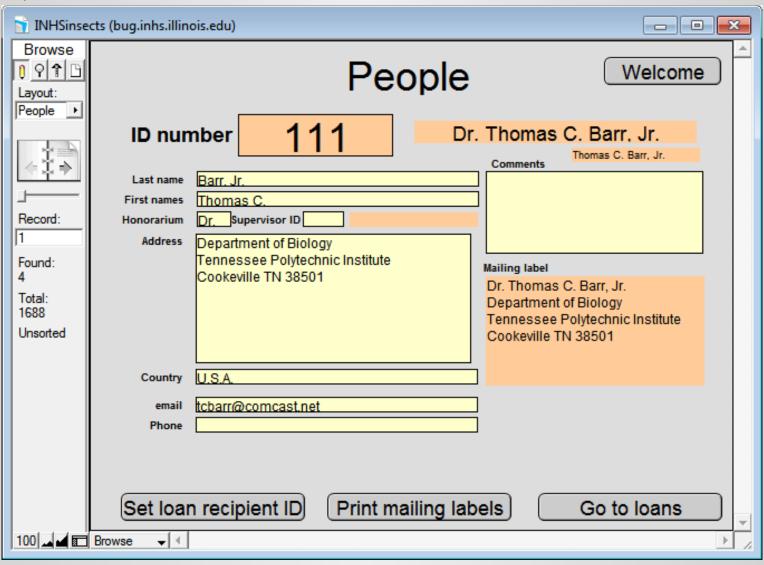
INHS Insect Collection Database: Taxonomy

66,110 (76547 in TaxonWorks) records



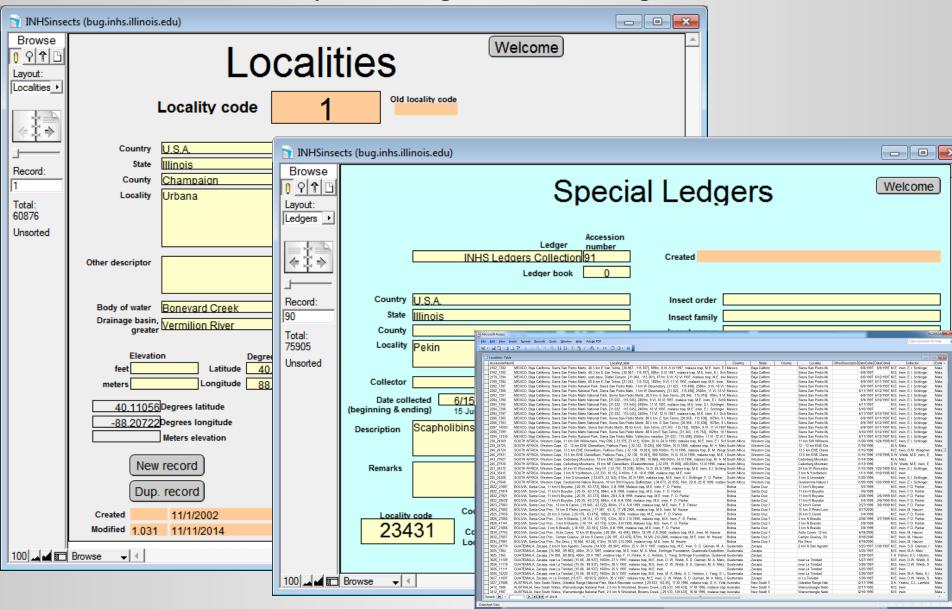
INHS Insect Collection Database: People

1,688 records



INHS Insect Collection Database: Localities

60,876 locality, 75,905 ledger, 20,115 collecting event



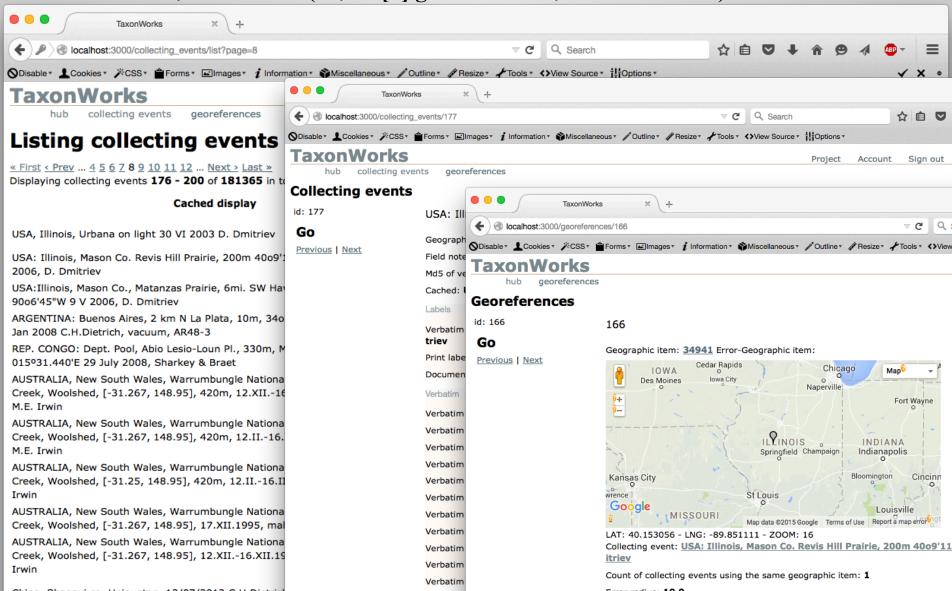
INHS Insect Collection Database: Index cards

1,499 cards

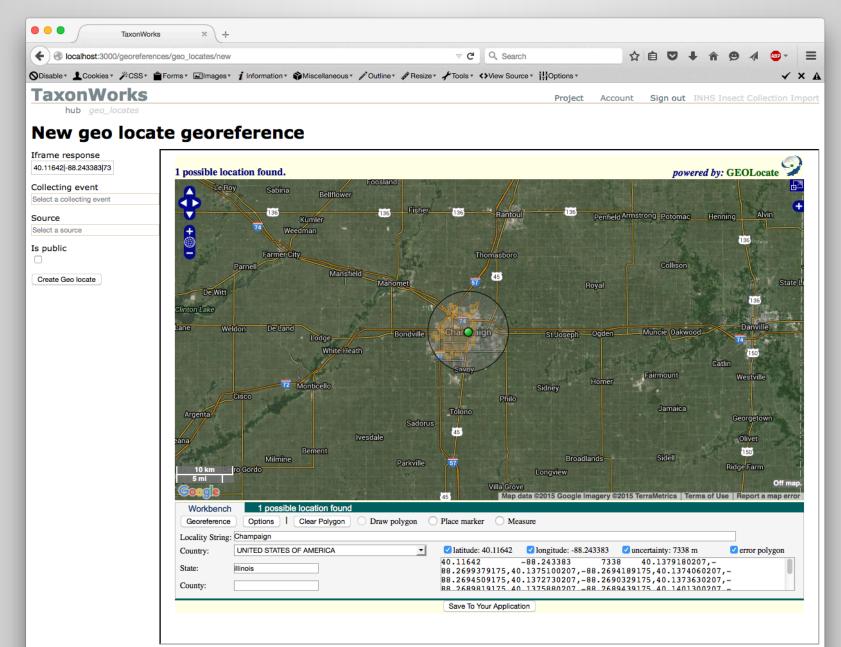
GL 13 KANSAS Clayton, June 27, 1955, Ross & Ross						
Grasses on open 1 never ploughed.	Grasses on open limestone slopesgrazed but never ploughed. Bunchy grass dominant.					
Flexamia pedinata GL 11 COLORADO F. albida Hartzel, June 7, 1954, Ross & Ross F. abbreviata Grasses in open flat floodplain, including much Eleocharis?						
889 Exitianus exitios	2N Lemellus bima IN, 30 5 & Psamir	GL 12 COLORADO Sapinero, June 8, 1954, Ross & Ross				
		Grasses along open marsh around small pondbadly grazed.				
		3N,28 Lemellus bimazulatus 289 Hebe caphalus oxcidentalis				
		3N, 60 9 Ambly selves dorsti				
		707 Psaumolettix akeny 607 Amblysellus grex				

TaxonWorks: "Localities", georeferencing

181,365 records (28,967[?] georeferences, should be 80%)

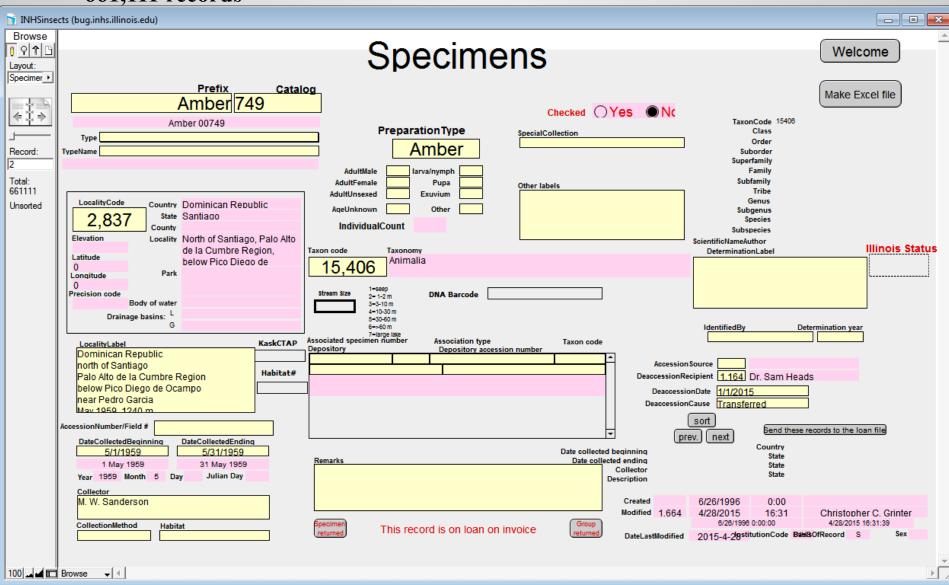


TaxonWorks: GeoLocate



INHS Insect Collection Database: Specimens

661,111 records

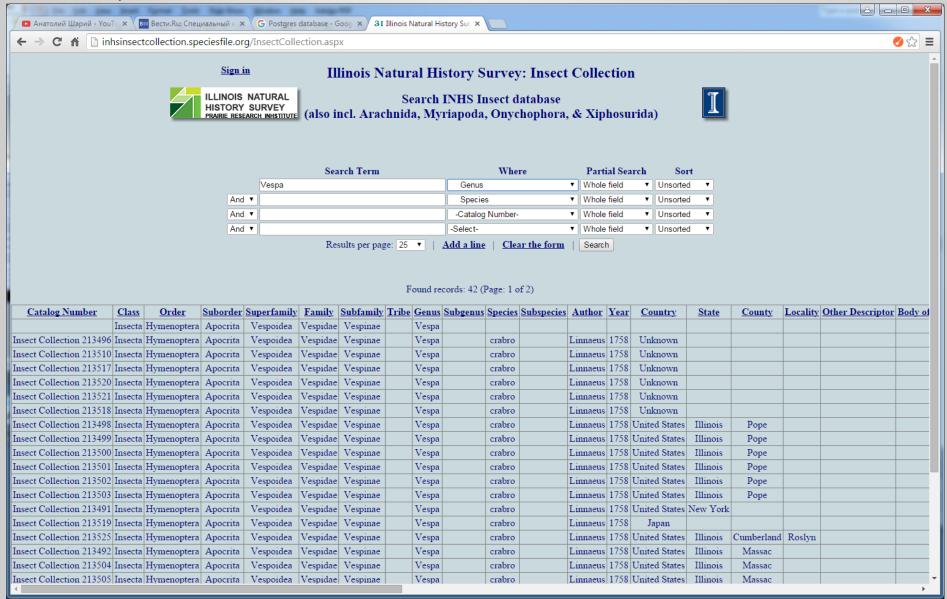


Parsing verbatim labels

- •Matching with existing parsed records (~40% match)
- •RegEx date extraction (~80% records)
- •RegEx coordinate extraction (~95% records with coordinates)
- •Finding collector (~70% records)
- Student parsing
- •Crowd-sourcing (have not been tested yet).

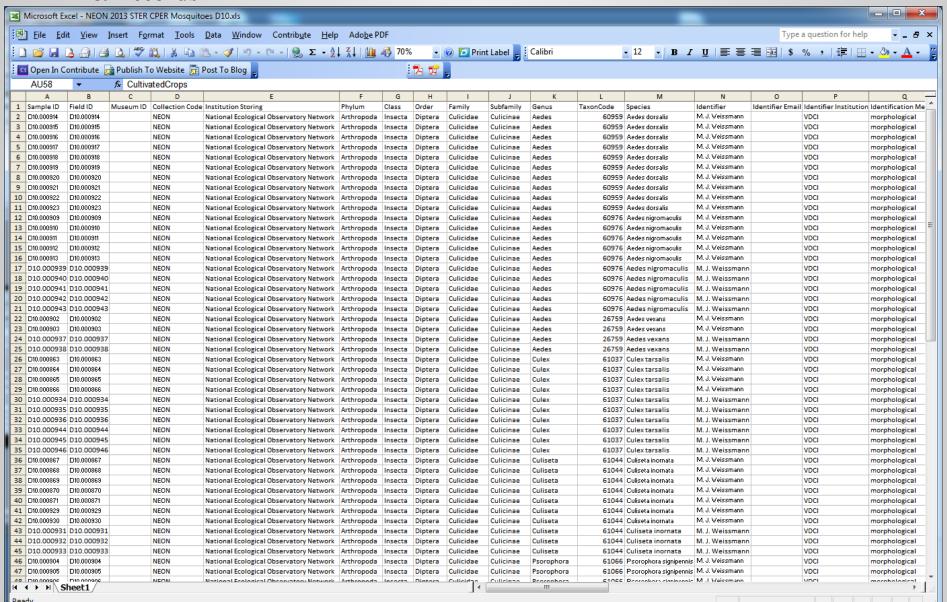
INHS Insect Collection Database: SQL database

231,109 records



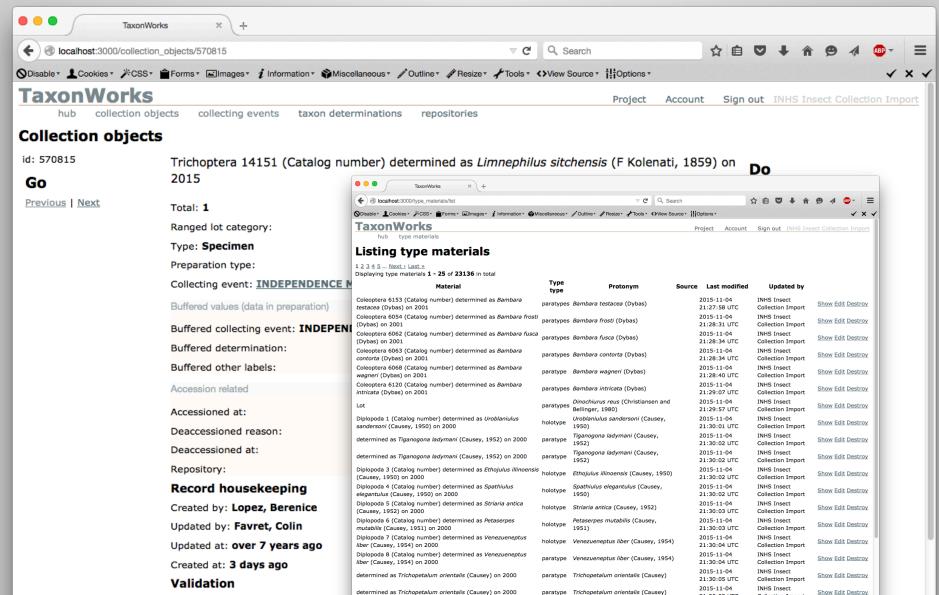
INHS Insect Collection Database: NEON

189 records



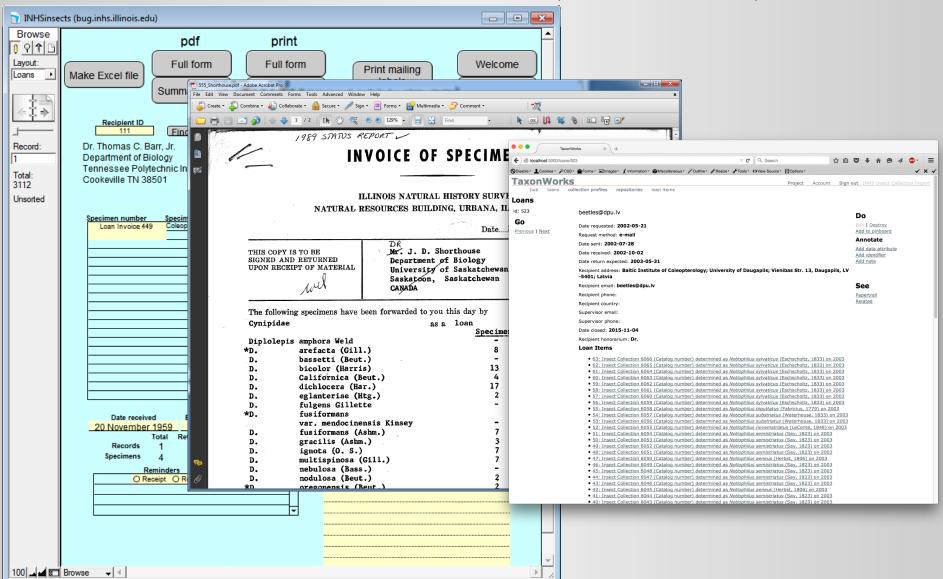
TaxonWorks: Collection Objects

912,593 records, with 23,136 type material records



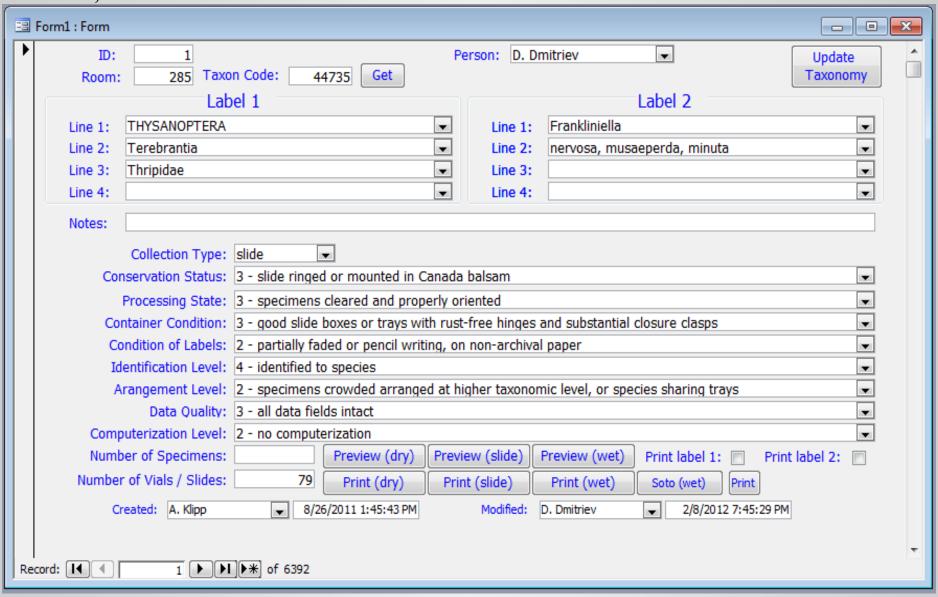
INHS Insect Collection Database: Loans

3,112 records, loan items: 56,689 records (1:1 in TaxonWorks)

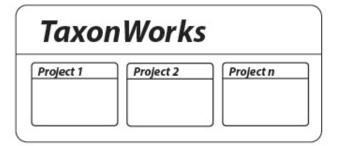


Collection profiling (following Favret et al. 2007)

16,449 records



Accessing a TaxonWorks project



Self managed

Desktop (development)

Local Server

Remotely managed

Institutional Server

Cloud

iPlant?

TaxonWorks as a basic CM tool: TODO

TW Deployment Strategy

Optimized Loan Form

Loan
Completion Form

Loan Letter

DwC A Export

Things to do!

Use a TaxonWorks VM

Provide Feedback Poke at Sandbox

SFG Hackathons

Masters Student!

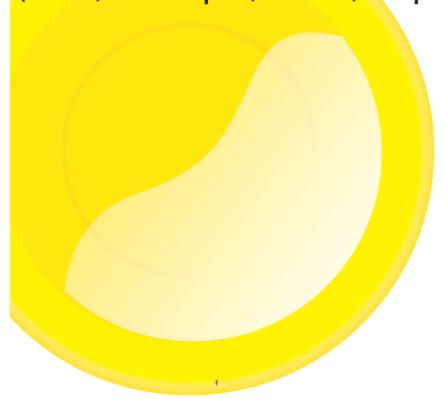
Summer Internships

Draw a Wireframe

Hack on TaxonWorks!

Wanted: YPTs

(bowls, not samples, well OK, samples too, but for now, bowls)



How? Where?

Combined, custom order?

Thank you

NSF-CSBR-1458285, and prior support INHS collorators and museum support

C. Dietrich, T. Schuh, C. Bartlet, S. Heads,

B. Morris, A. Bader, J. Zahniser, many undergrads!

Frost currators and staff

Species File Staff

Jim Tucker, Marilyn Beckman, Rich Flood, Mike Maehr, Beth Frank

NSF-DBI-1356515