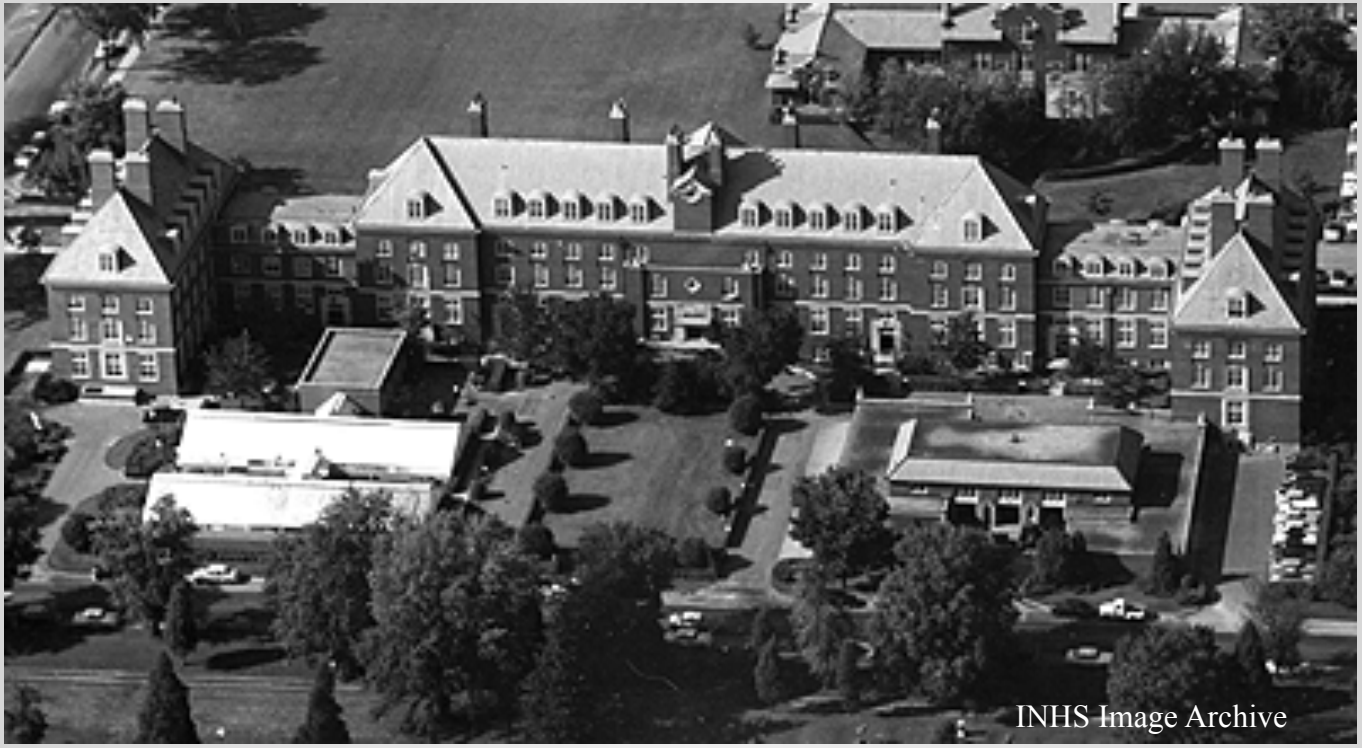
ILLINOIS NATURAL HISTORY SURVEY
PRAIRIE RESEARCH INSTITUTE

Getting an insect collection ready for digitization at the Illinois Natural History Survey's InvertNet project

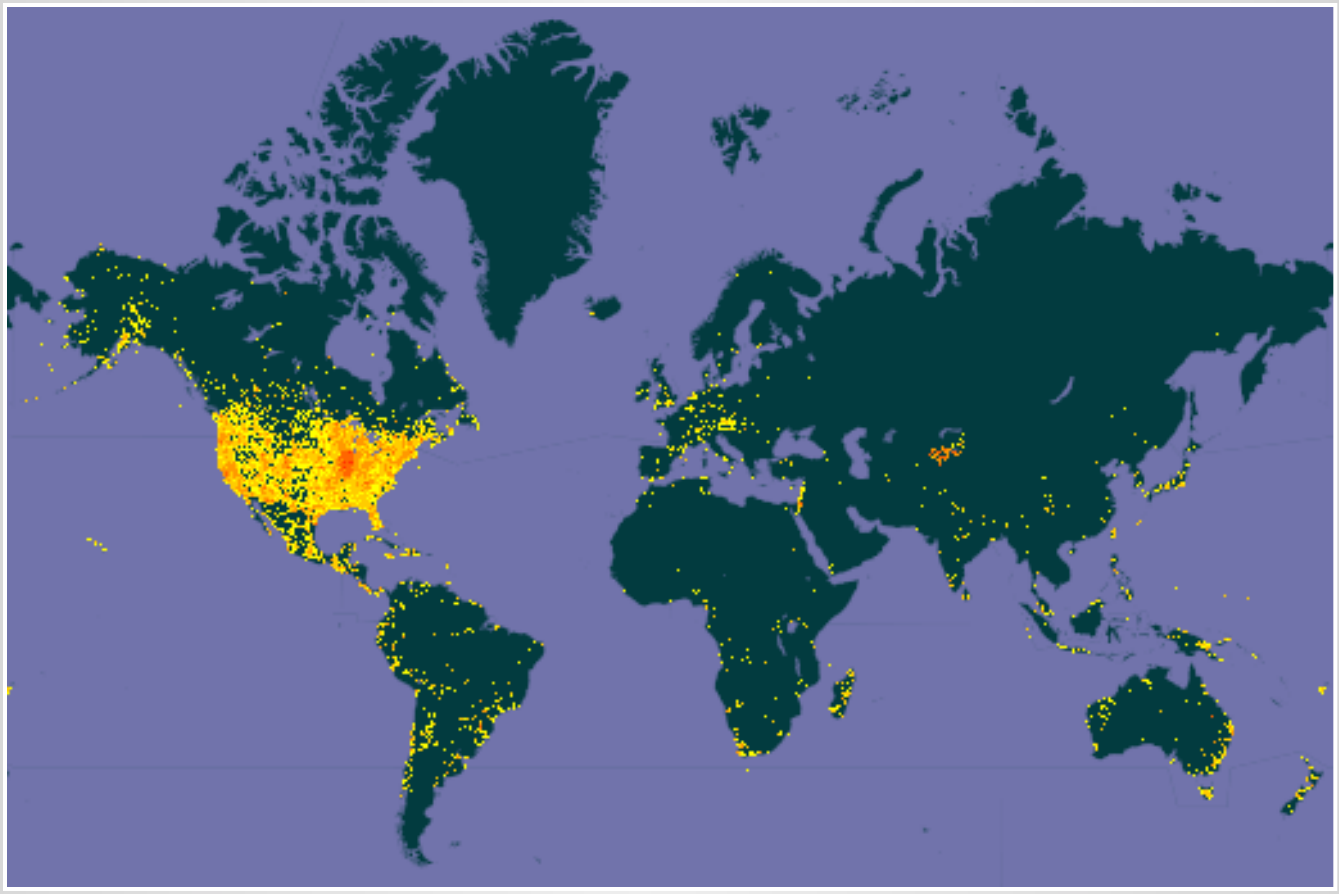
Dmitry Dmitriev



Illinois Natural History Survey collection

> 7,000,000 specimens

> 14,000 primary types



Today, ~900,000 records (2.1 million specimens)

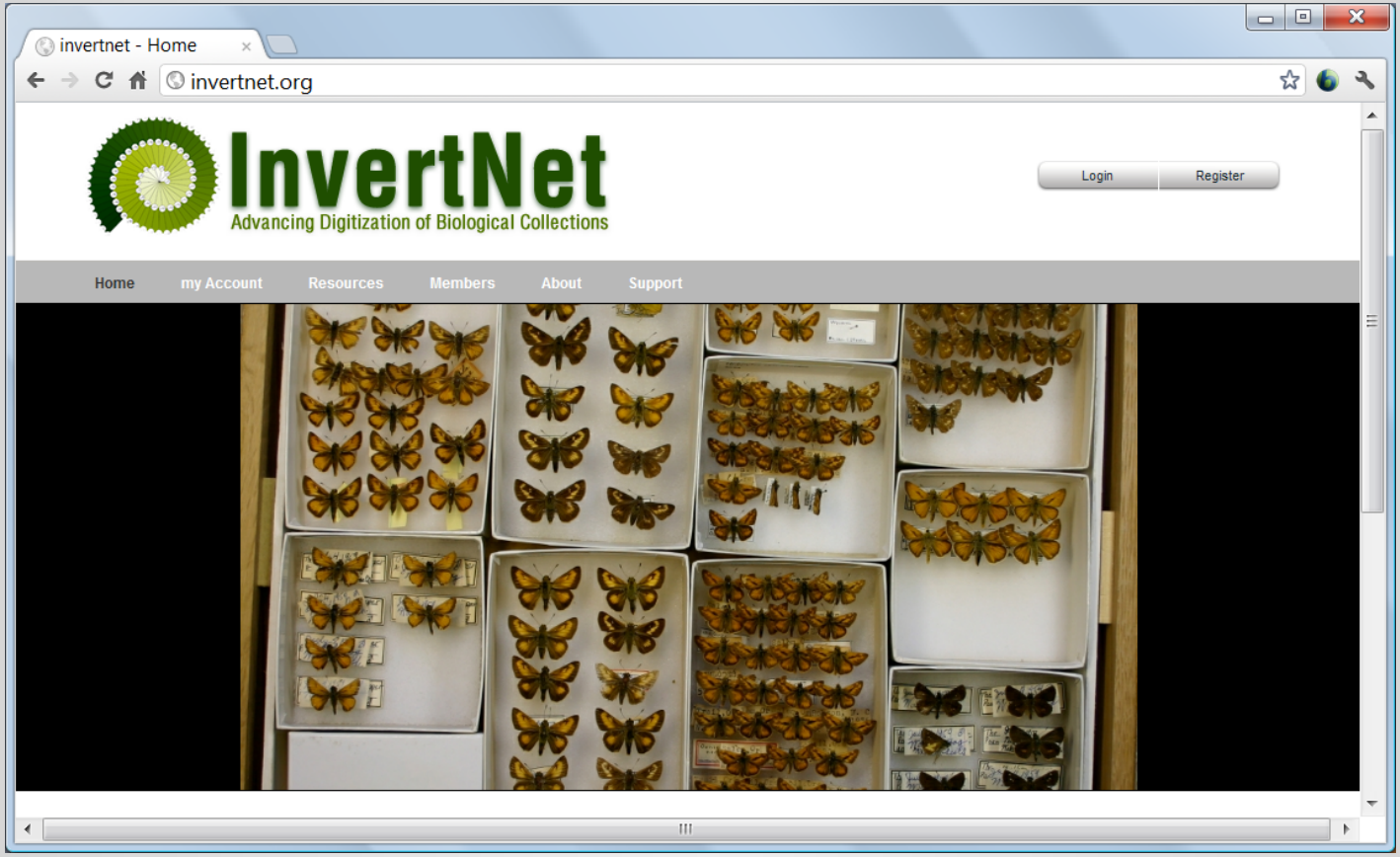
GBiF: http://data.gbif.org/datasets/provider/75

InvertNet Rationale

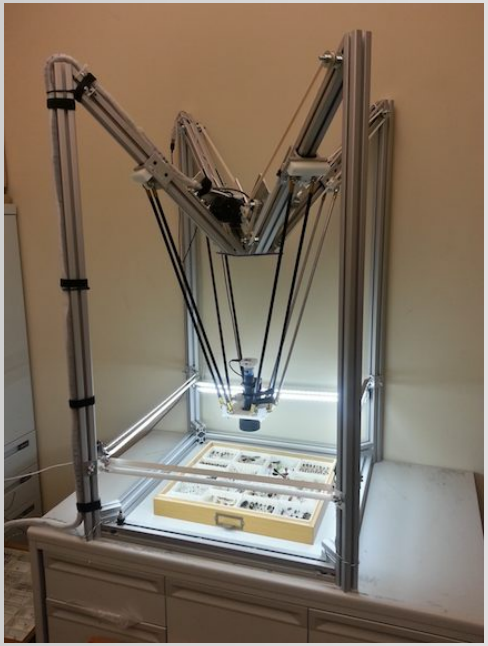
- vast majority of specimens the collections are invertebrates
  - primarily insects and related arthropods
  - less than 5% available online
  - only label data usually provided
- existing digitization methods are inadequate
  - slow and expensive ($1+ per specimen)
  - risk of damage to specimens from handling

InvertNet

- Digitize all holdings of 22 arthropod collections (>50 million specimens)
- Provide access to images and other data via online virtual museum
- Provide platform for research and development of additional tools and resources



Drawer Imaging



- Delta Robot, digital camera, telecentric lens captures grid of single, close-up images at 40-60 x/y coordinates and 5 perspectives

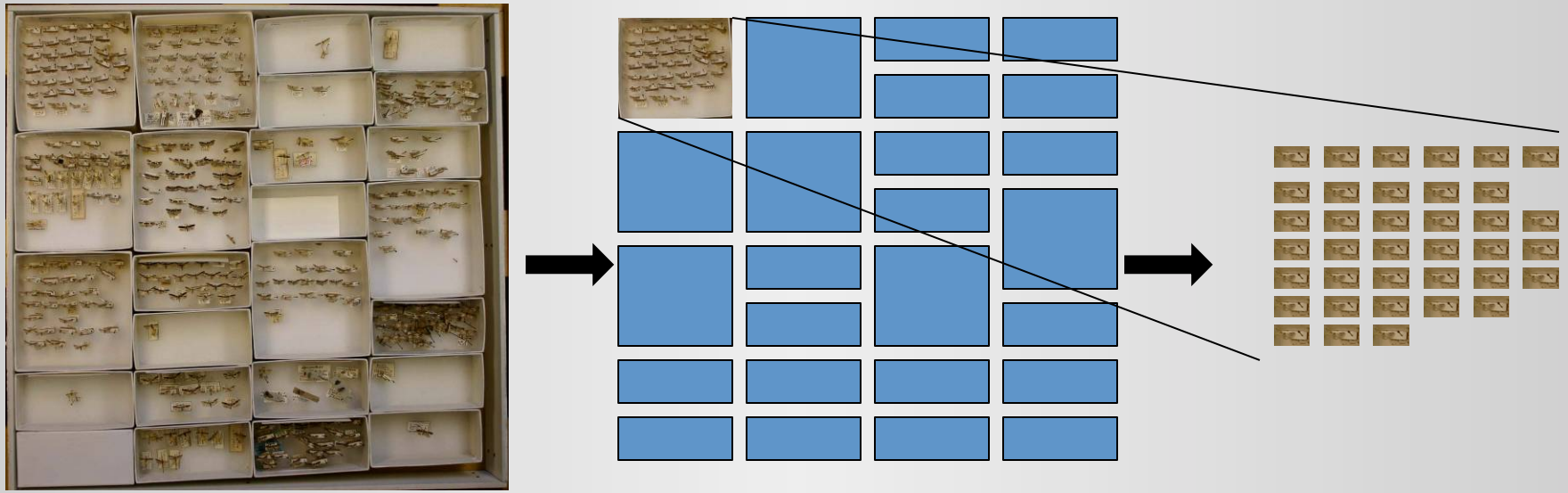
- Single images stitched to yield Gigapixel images from multiple viewpoints

Top-down view

Angled view

- Enables virtual tilting

Image segmentation/annotation



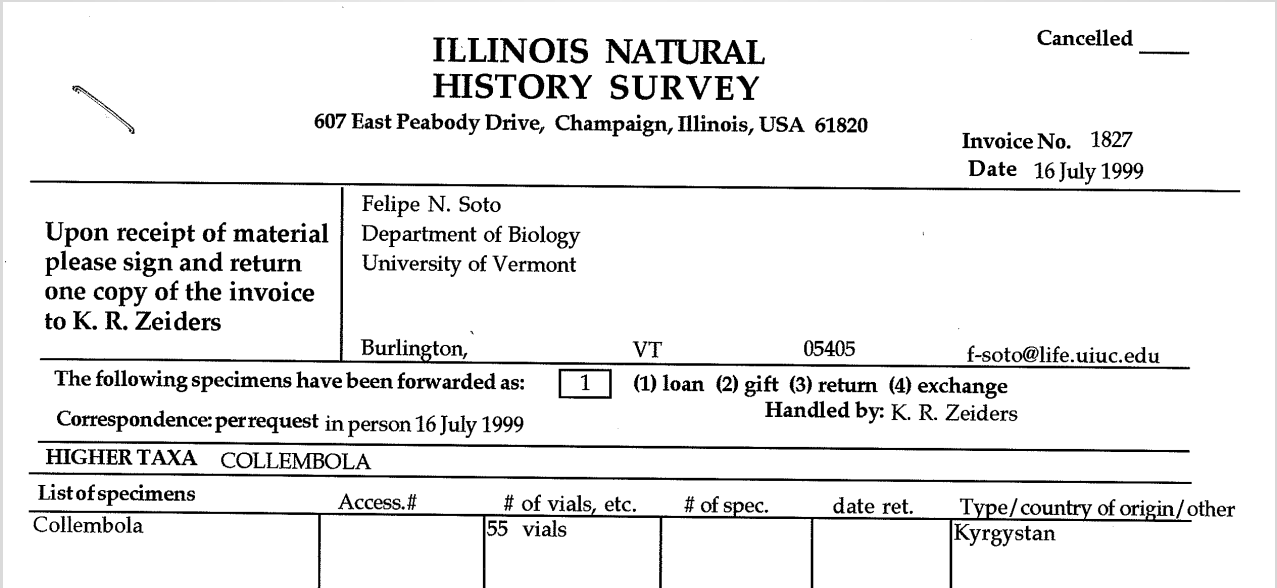
1. capture image of drawer + metadata (location, contents)
2. segment unit trays (image analysis software)
3. segment specimens
4. capture label data

Preparation of the collection for digitization

- Loan returns
- Merging large donations
- Updating nomenclature
- Unifying labels
- Collection profiling

Loan returns

- In 2011: 876 loans in the collection database, 223 of those cancelled.
- Frustrations: missing records, inaccurate records, human factor.
- Now: 3068 records in the database, 2565 cancelled, 82 renewed. All the invoices scanned.



ILLINOIS NATURAL HISTORY SURVEY
607 East Peabody Drive, Champaign, Illinois, USA 61820

Cancelled ____

Invoice No. 1827
Date 16 July 1999

Upon receipt of material please sign and return one copy of the invoice to K. R. Zeiders

Felipe N. Soto
Department of Biology
University of Vermont

Burlington, VT 05405 f-soto@life.uiuc.edu

The following specimens have been forwarded as: 1 (1) loan (2) gift (3) return (4) exchange

Correspondence: per request in person 16 July 1999

Handled by: K. R. Zeiders

HIGHER TAXA COLLEMBOLA

List of specimens	Access.#	# of vials, etc.	# of spec.	date ret.	Type/country of origin/other
Collembola		55 vials			Kyrgystan

Large donations

- The material was sitting in temporary storage area, and has not been incorporated in main collection.

Updating nomenclature

CoL match: Diptera – 60%, Hemiptera – 60%, Lepidoptera – 80%, Coleoptera – 35%.

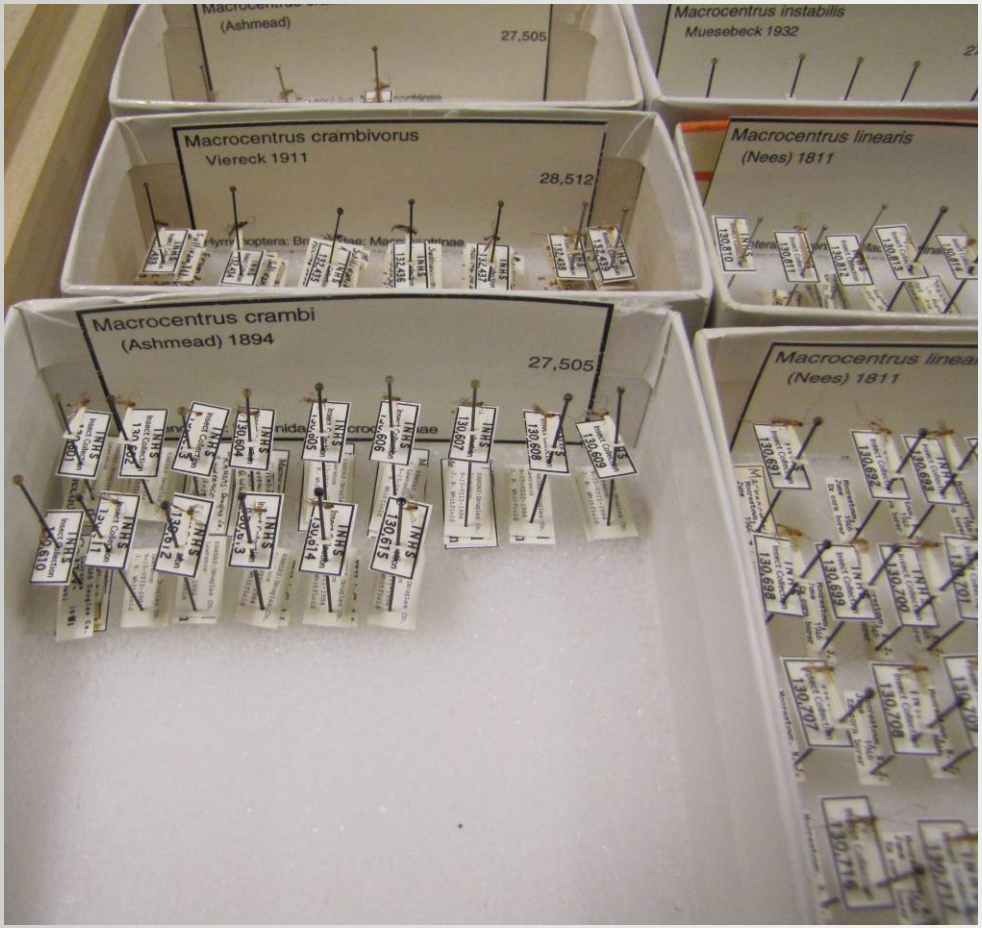
Now: 63,545 names in the database.

Coleoptera : Table

ID	Order	Family	Genus	Subgenus	Species	Supspecies	Author	year	# of labels
267	Coleoptera	Bruchidae	Algarobius		bottimeri		Kingslover		
268	Coleoptera	Bruchidae	Algarobius		johnsoni		Kinkslover		
269	Coleoptera	Bruchidae	Algarobius		nicoya		Kingslover		
270	Coleoptera	Bruchidae	Algarobius		prosopis		(LeConte)		
271	Coleoptera	Bruchidae	Algarobius		riochama		Kingslover		
272	Coleoptera	Bruchidae	Algarobius		sp.1				
273	Coleoptera	Bruchidae	Althaeus		folkertsi		Kingslover		2
274	Coleoptera	Bruchidae	Althaeus		hibisci				2
275	Coleoptera	Bruchidae	Althaeus		steineri		Kingslover		
276	Coleoptera	Bruchidae	Althaeus						2
277	Coleoptera	Bruchidae	Amblycerus		robiniae		(Fabricius)		3
278	Coleoptera	Bruchidae	Bruchidius		atar				
279	Coleoptera	Bruchidae	Bruchus		brachialis		Fahr.		2
280	Coleoptera	Bruchidae	Bruchus		kisoruni		(L.)		
281	Coleoptera	Bruchidae	Bruchus		pisorum		L.		
282	Coleoptera	Bruchidae	Bruchus						
283	Coleoptera	Bruchidae	Callosobruchus		chinensis		(L.)		
284	Coleoptera	Bruchidae	Callosobruchus		maculatus		(F.)		
285	Coleoptera	Bruchidae	Caryobruchus		gleditsiae		(L.)		
286	Coleoptera	Bruchidae	Cercidiestes		ulkei		(Horn)		
287	Coleoptera	Bruchidae	Cercidiestes						
288	Coleoptera	Bruchidae	Eubaptus		palliatus		Lacord.		

Record: 1 of 8703

Unifying labels



Cistalia signoretii (Guerin-Meneville 1857)	49,288
Heteroptera: Lygaeoidea: Rhyparochromidae	

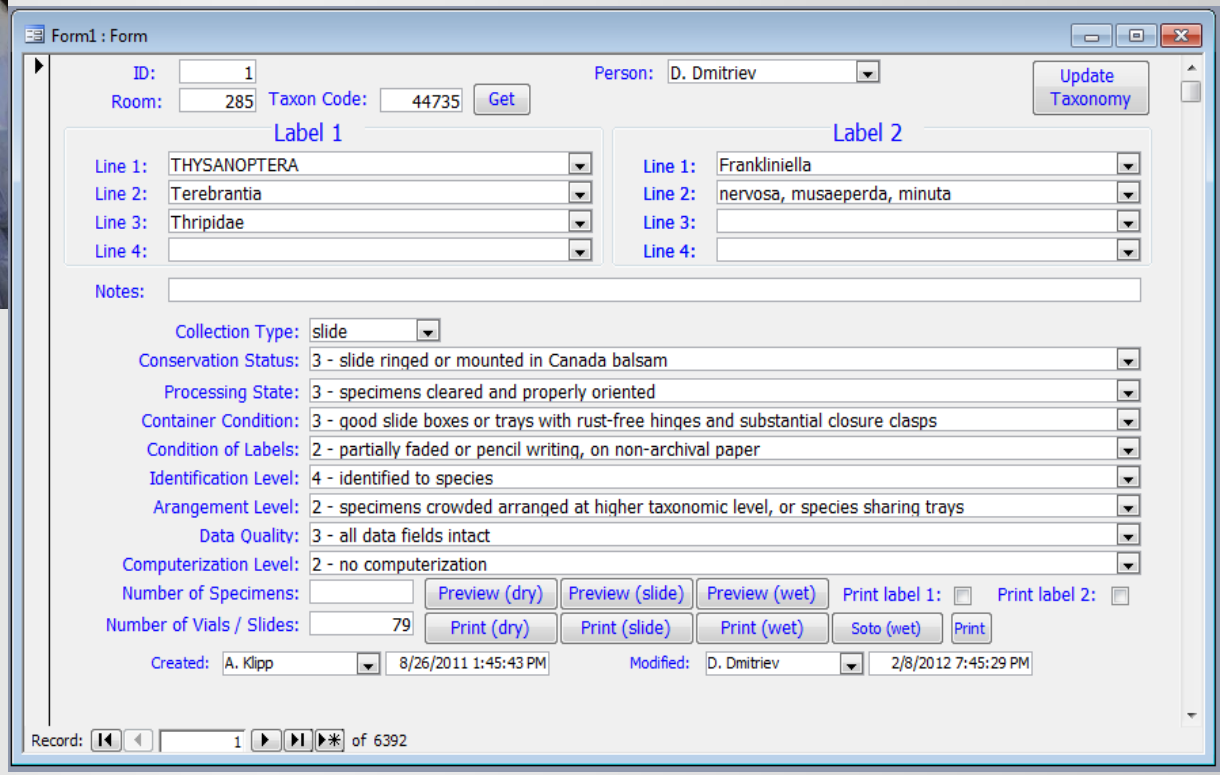
Collection profiling

Amy Bader

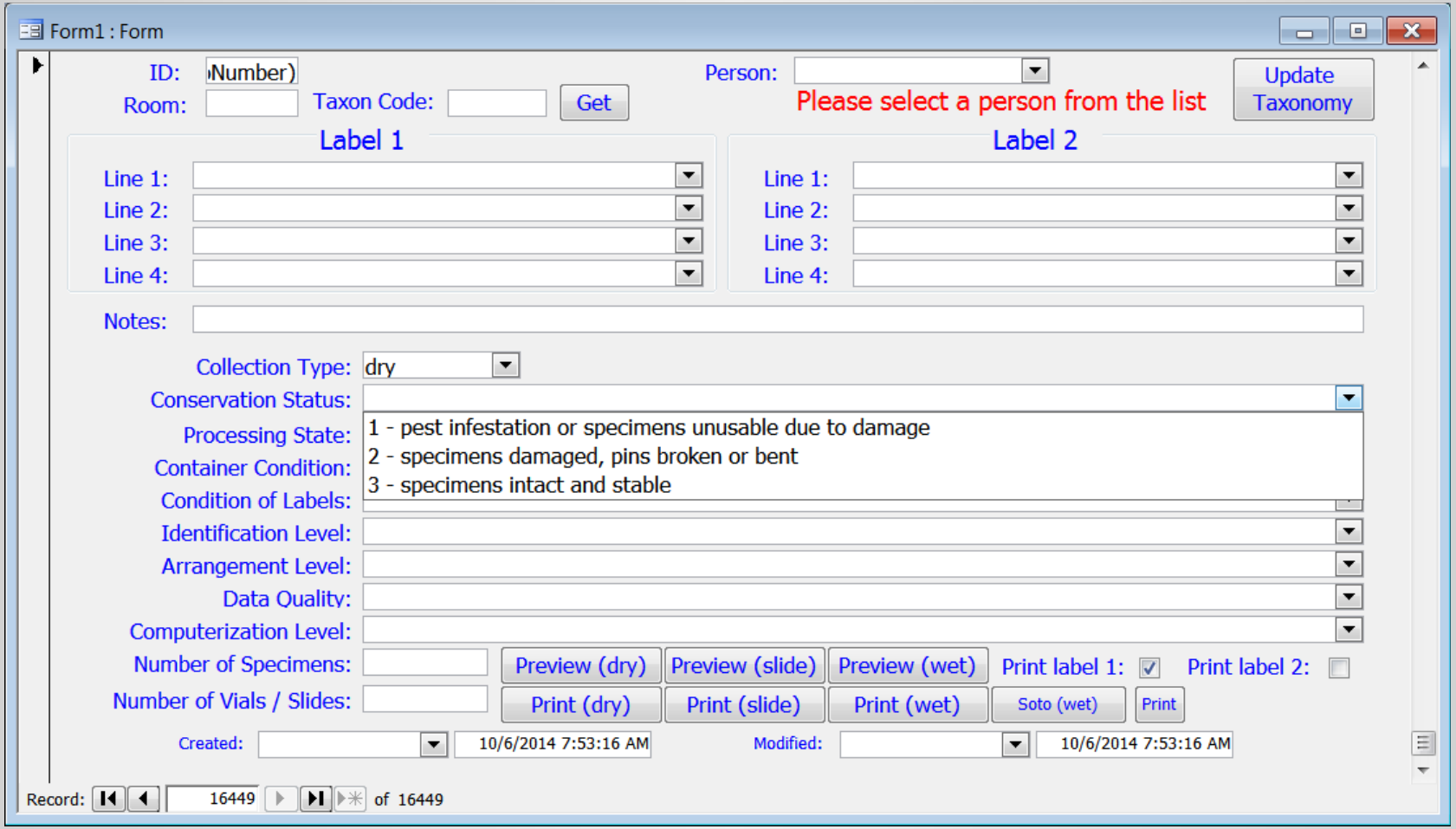
processing loan return

INHS collection profiling

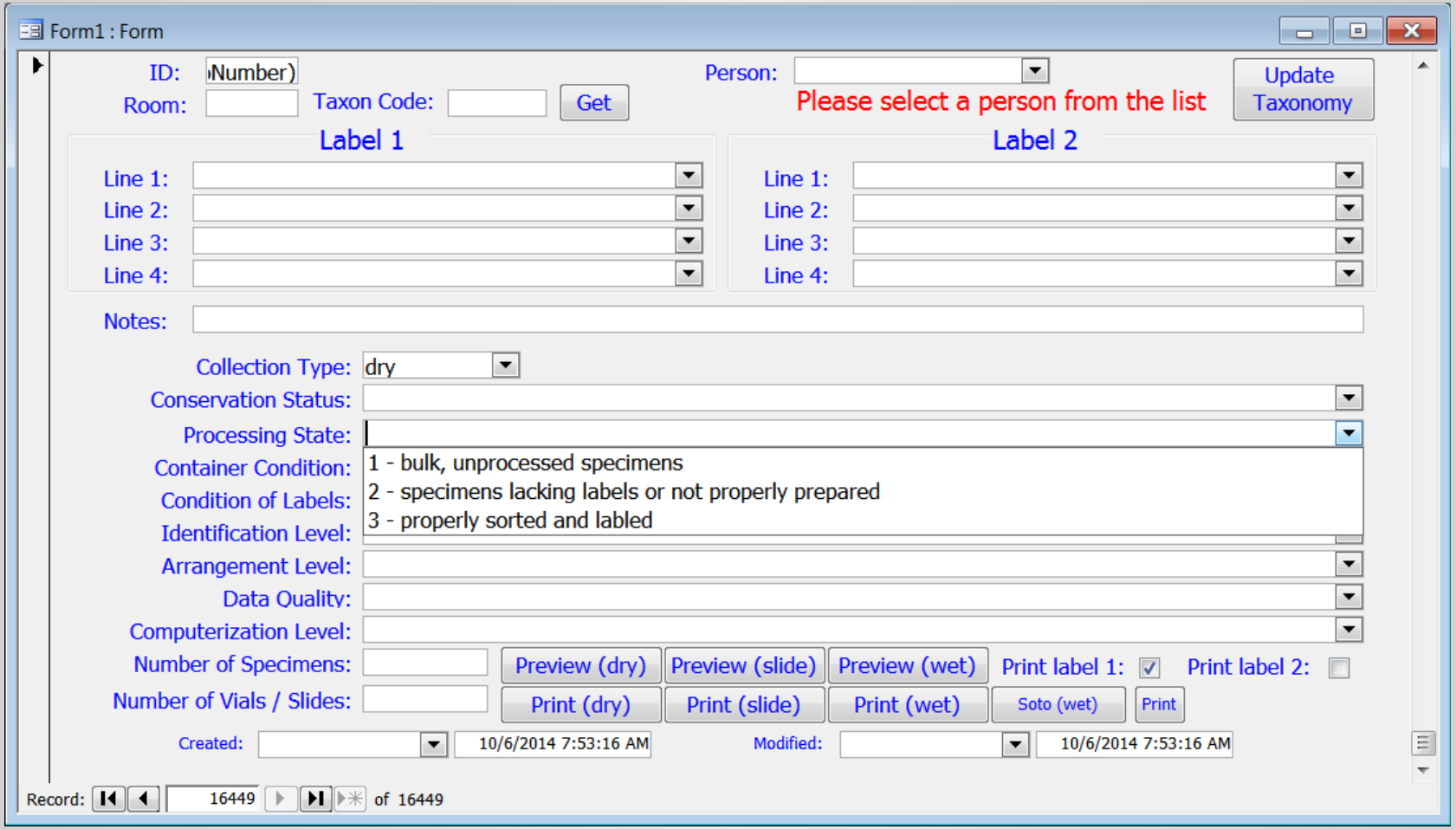
(Colin Favret et al. 2007)



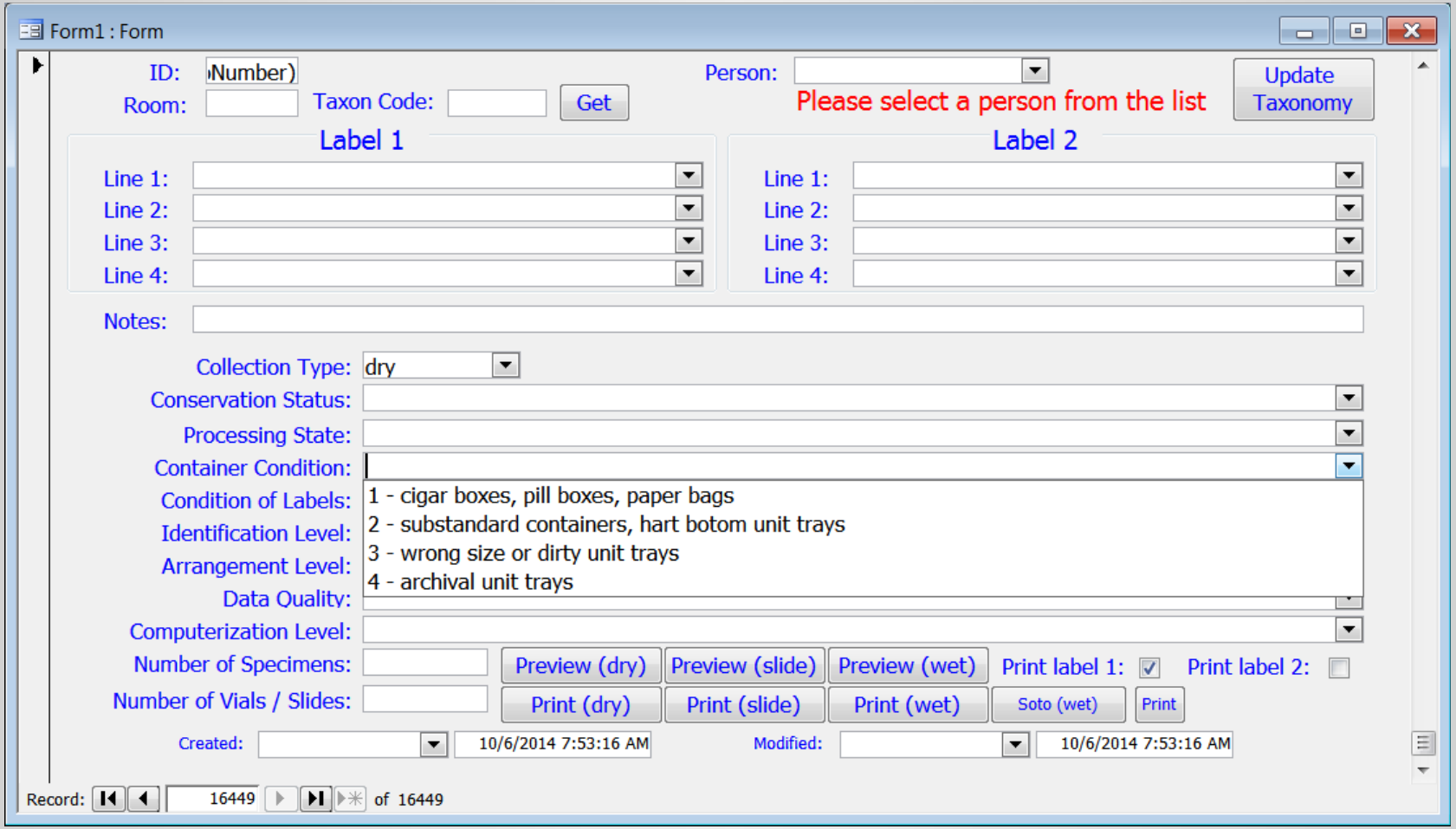
Collection profiling: Conservation status



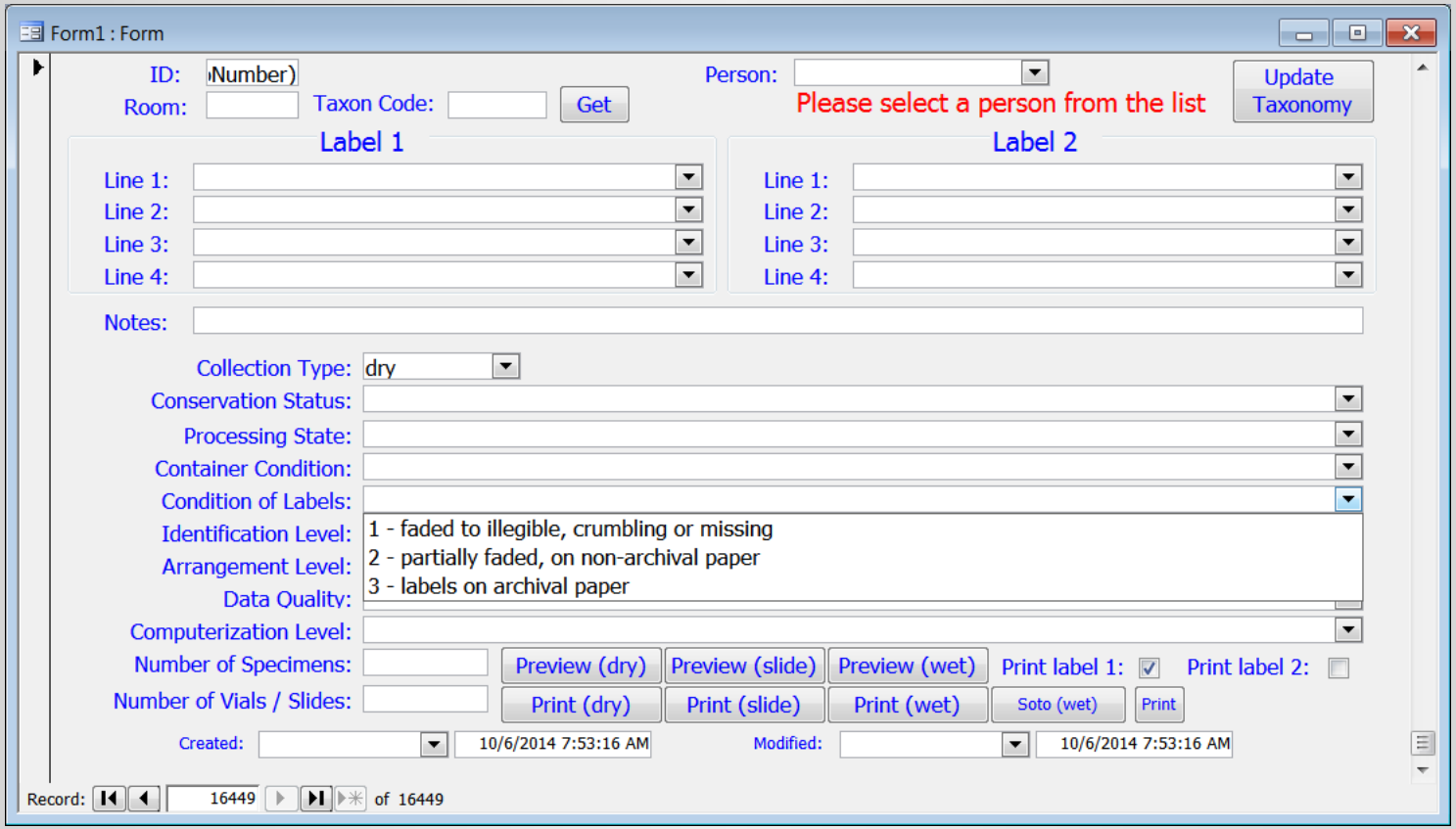
Collection profiling: Processing state



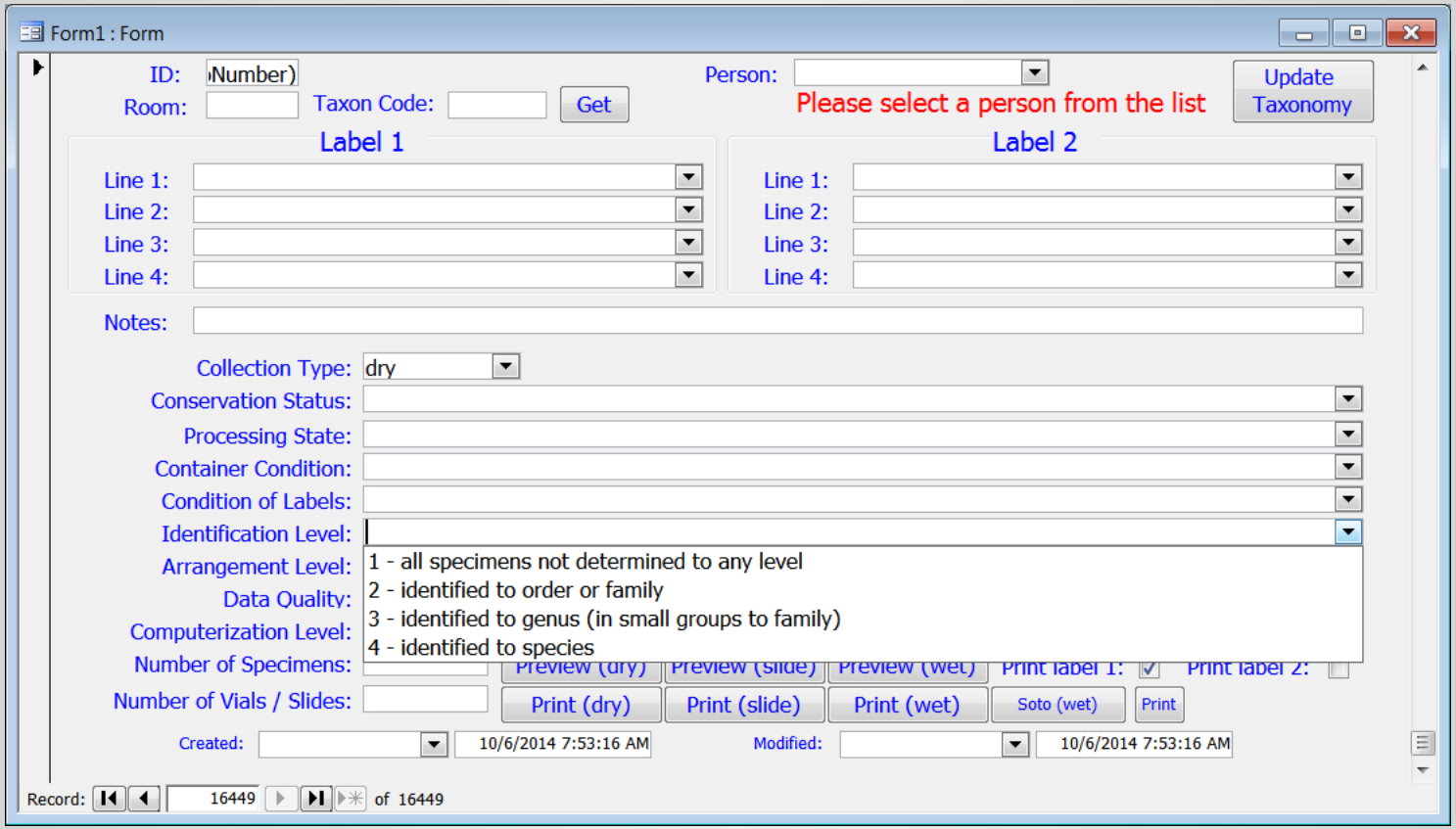
Collection profiling: Container condition



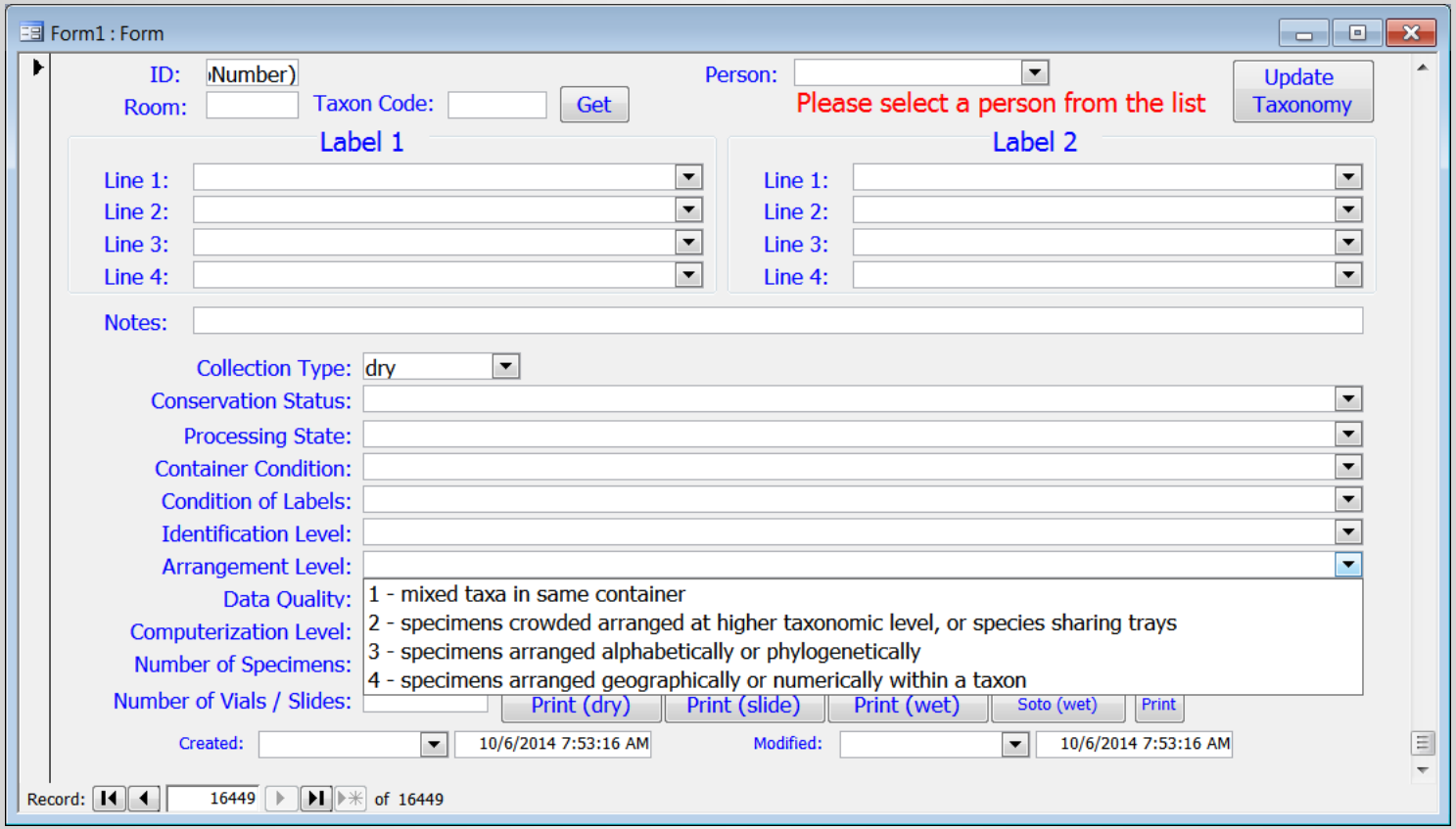
Collection profiling: Condition of labels



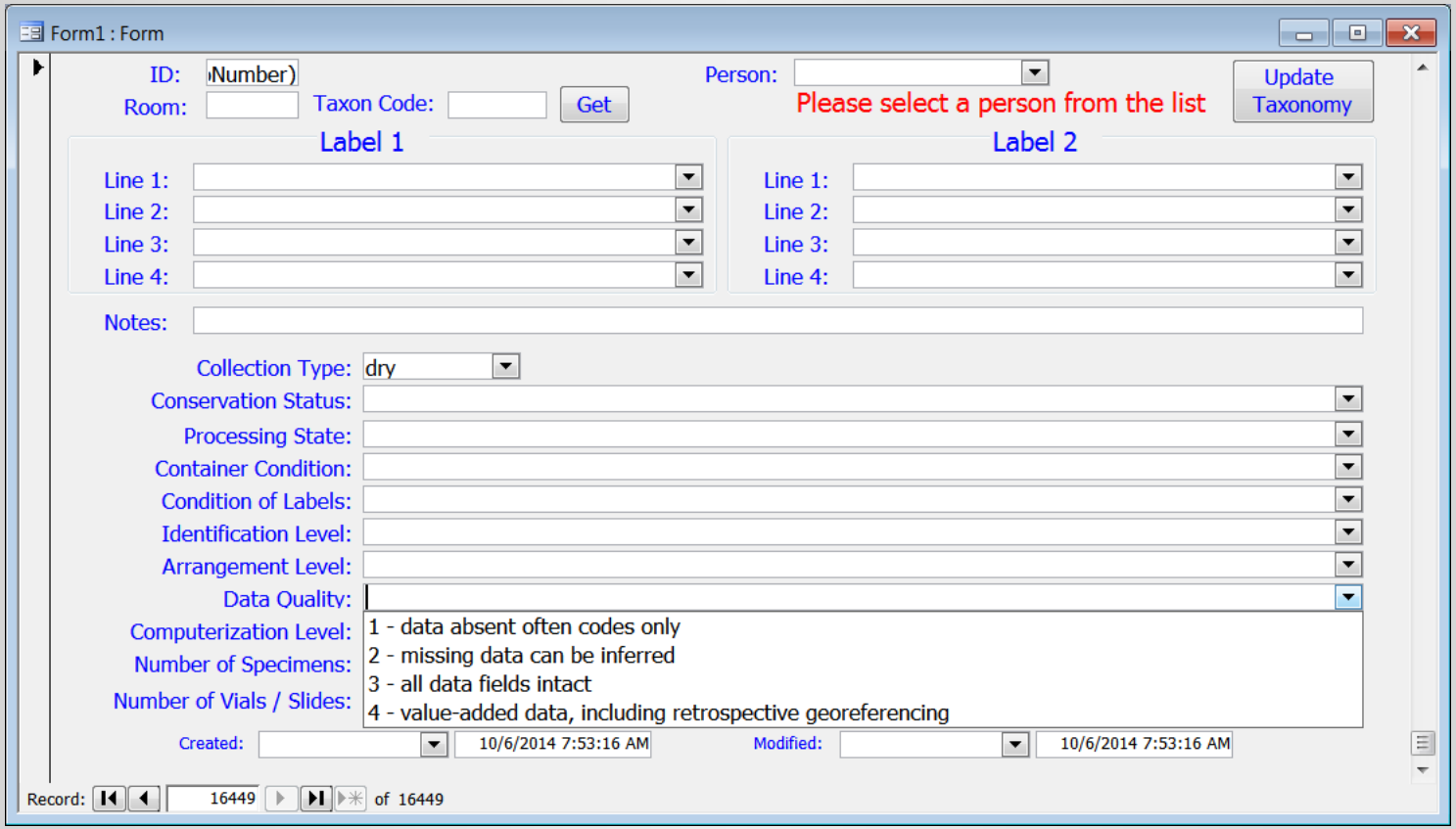
Collection profiling: Identification level



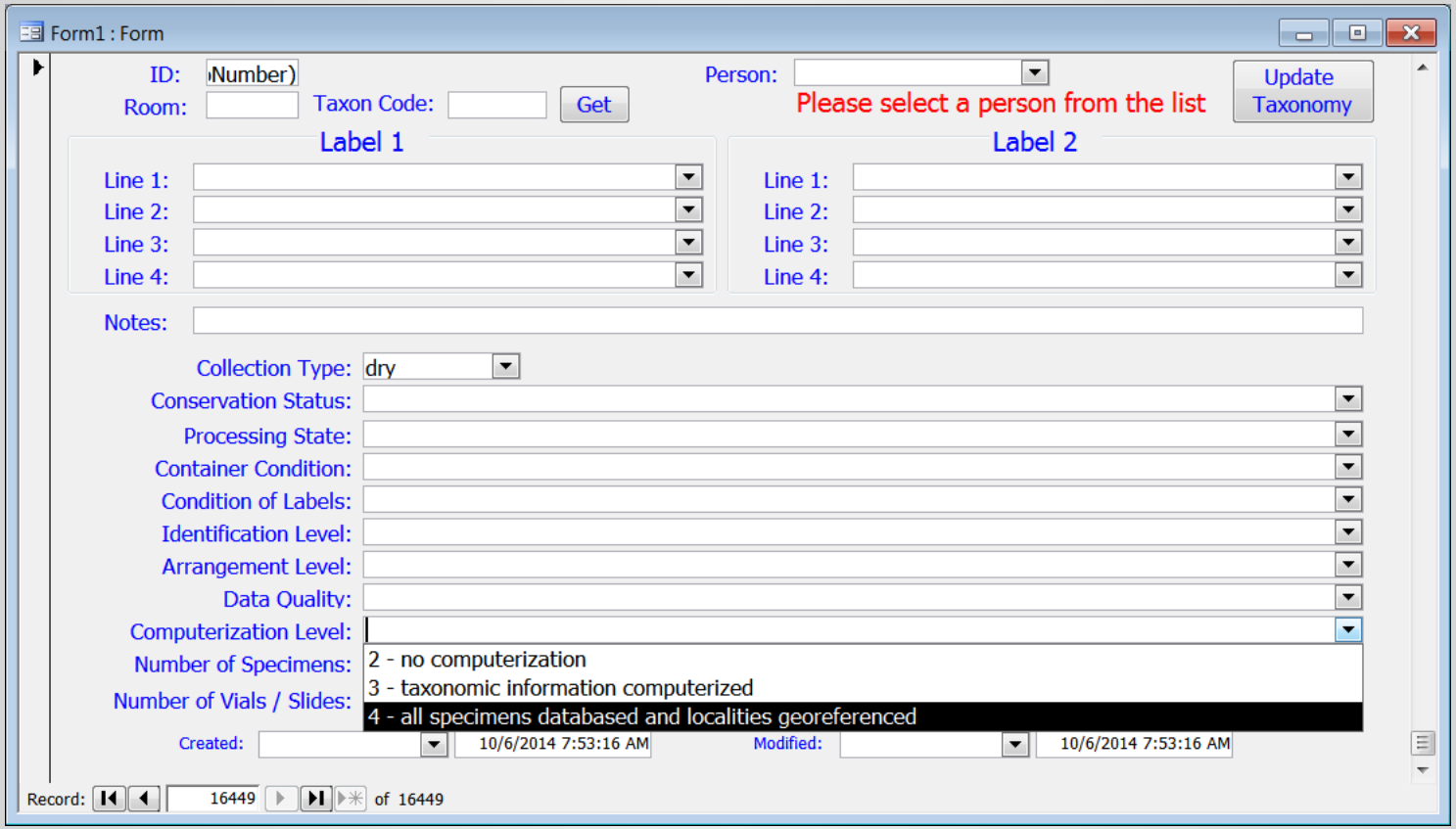
Collection profiling: Arrangement level



Collection profiling: Data quality

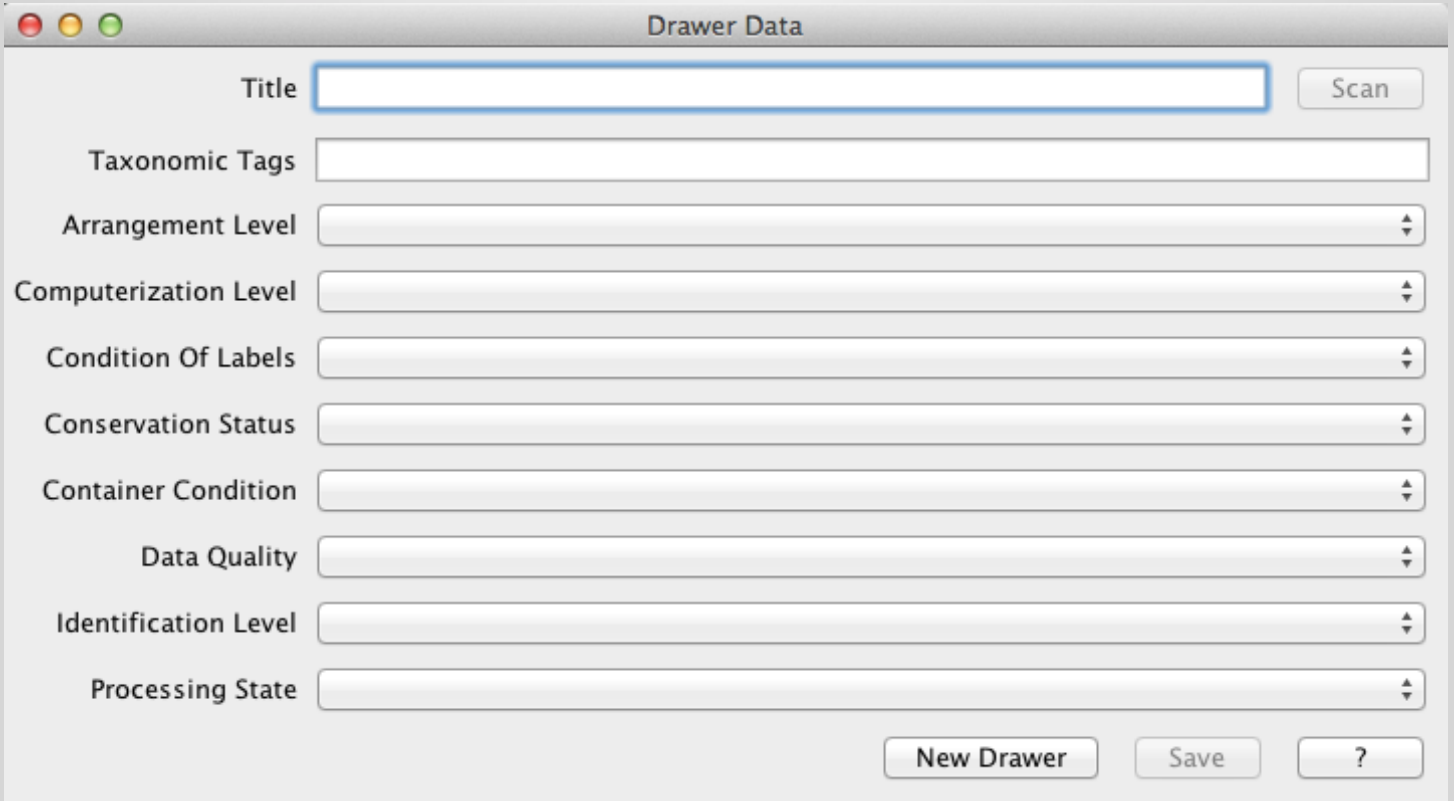


Collection profiling: Computerization level



Scanning - Data Entry

- Title must be unique
- After you have entered a title, press Scan
- Tags should include any relevant taxa
- Fill in as many quality menus as you can
- Don't forget to save!
- Once scanning is done, press New Drawer

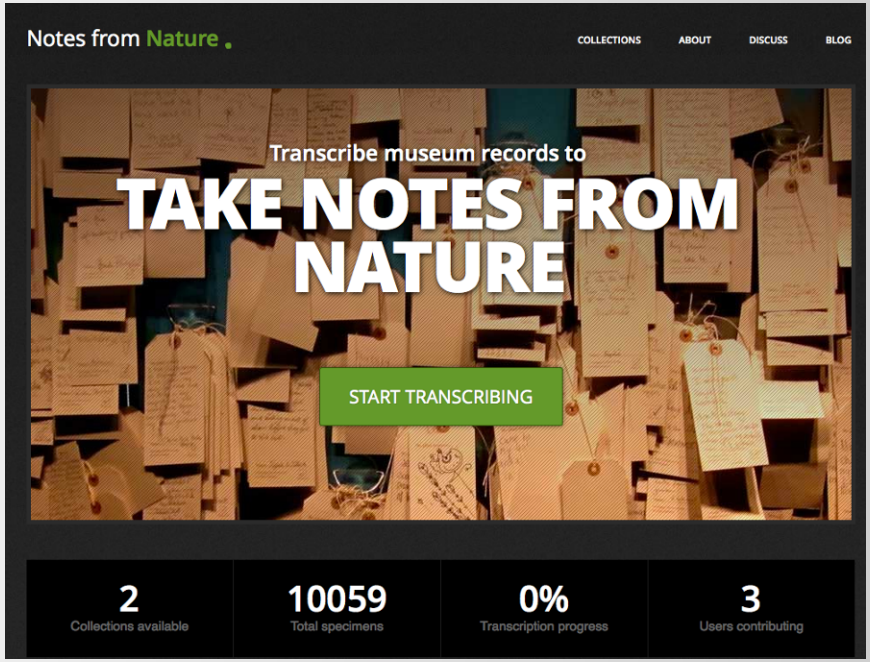


Collection profiling

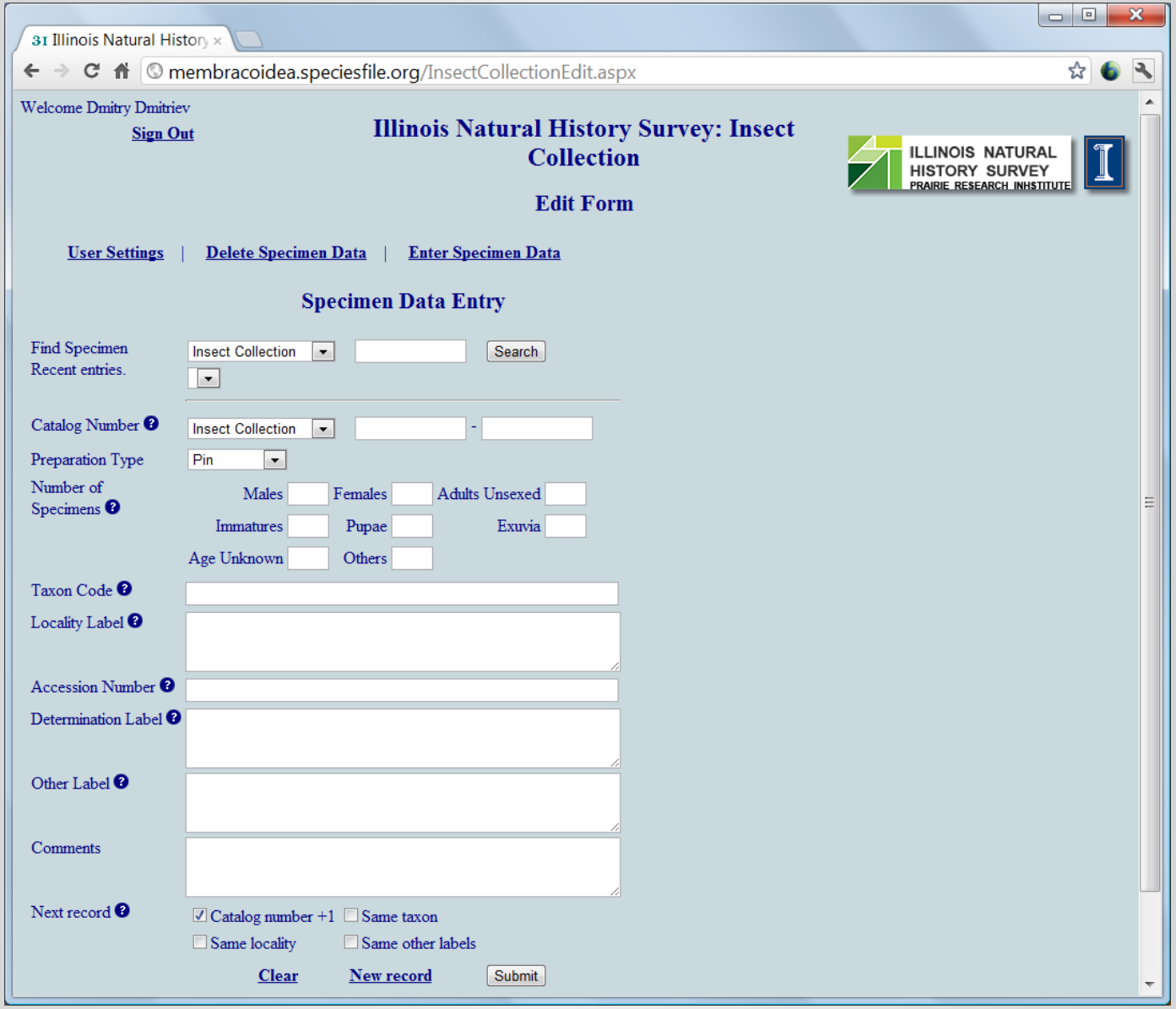
Collection profiling

Label data capture

- OCR doesn't work
  - labels partly obscured by specimens
  - many hand written
  - high error rates—time fixing errors > time required to manually enter data
- Manual entry of verbatim label more accurate
  - still expensive, time consuming, error prone
- Crowd-sourcing is most viable option
  - rapid
  - low cost
  - built-in redundancy to reduce errors
  - applications already available (e.g., Notes from Nature)



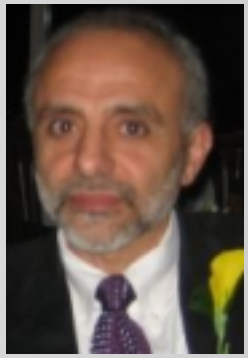
Online data entry form

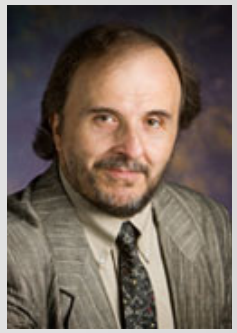


Parsing verbatim labels

- Matching with existing parsed records (~40% match)
- Finding date (~95% records)
- Finding coordinates (~85% records)
- Finding collector (~70% records)
- Finding identifier (~30% records)
- Student parsing
- Crowd-sourcing (have not been tested yet).

InvertNet UIUC Team





- Chris Dietrich – Director
  - Systematic Entomologist
- John Hart – CoPI
  - Computer Science - Graphics
- Nahil Sobh – CoPI
  - Computational Multiscale Nanosystems
- Umberto Ravaioli – CoPI
  - Computational Multiscale Nanosystems
- David Raila – Senior Collaborator
  - Computer Science – Sr. Research Programmer
- Others
  - Programmers, research assistants, hourlies

InvertNet Collaborating Curators

- A. Cognato, MSU
- G. Courtney, J. VanDyk, ISU
- J. Holland, Purdue
- R. Holzenthal, P. Tinerella, Minnesota
- P. Johnson, SDSU
- H. Klompen, M. Daly, OSU
- J. Rawlins, R. Davidson, J. Fetzner, Carnegie Museum
- D. Rider, G. Fauske, NDSU
- A. Short, Kansas
- R. Sites, Missouri
- D. Young, Wisconsin-Madison
- J. Zaspel, Wisconsin-Oshkosh
- G. Zolnerowich, KSU
- D. Rubinoff, U Hawaii
- T. Roberts, U Iowa