Plants, herbivores, and parasitoids:

Tri-trophic digitization strategies

ADBC Thematic Collections Network

Kimberly Watson, Robert Naczi, Melissa Tulig
Randall Schuh, Katja Seltmann

Association of Southeastern Biologists
April 12, 2013

Tri-Trophic Example

Plants

Crop Plants

Produce fruits and tubers of significant agricultural and economic importance.

Poaceae: corn, wheat, rice
Fabaceae: soybean, hay
Solanaceae: tomato, potato

Herbivores

Hemiptera

Pierce plant stems, leaves. Reduce plant vigor, transmit disease, reduce harvest yield.

Aphids, scales, true bugs, cicadas, potato leafhoppers

Photo: www.alexanderwild.com

Parasitoids

Hymenoptera

Parasitoid wasps lay eggs inside aphid; larvae consume host from inside out, emerging from "mummy" as adults.

Agricultural pest control

Photo: www.alexanderwild.com

Species of Interest: North American Biota

Plant Hosts

Family	species
Apiaceae	250
Asteraceae	2,400
Chenopodiaceae	250
Cupressaceae	30
Cyperaceae	850
Fabaceae	850
Fagaceae	97
Grossulariaceae	53
Juglandaceae	17
Lamiaceae	240
Oleaceae	35
Pinaceae	66
Poaceae	1,400
Polygonaceae	440
Rhamnaceae	75
Rosaceae	360
Salicaceae	123
Scrophulariaceae	430
Solanaceae	85
Zygophyllaceae	15

Total species: 8,066

Herbivores

Hemiptera	species
Coccoidea (scale insects)	986
Aphidoidea (plant lice)	1,532
Psylloidea (jumping plant lice)	176
Auchenorrhyncha (cicadas, hoppers)	4,629
Heteroptera (true bugs)	3,827

Total species: 11,150

Parasitoids

Hymenoptera	species
Aphelinidae	212
Encyrtidae	490
Mymaridae	187
Signiphoridae	19
Trichogrammatidae	131

Total species: 1,039

Plants

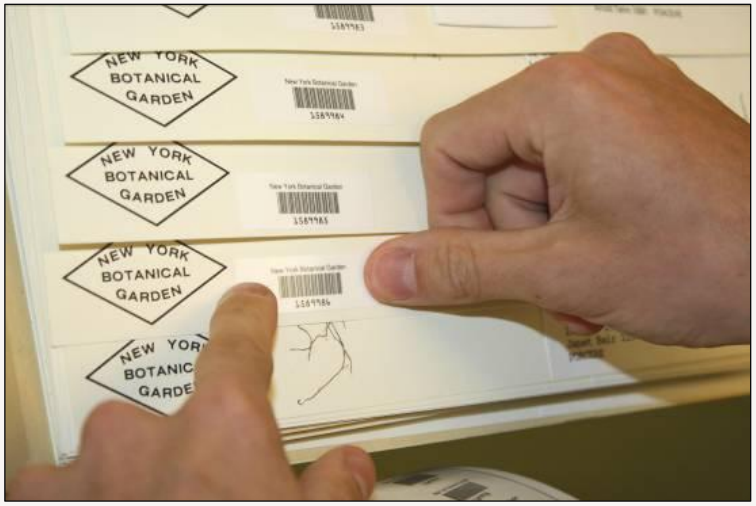
Participating Herbaria

Institutions (14)	Specimens already digitized	% Georeferenced	Specimens to be digitized
Eastern Michigan University	0	0	10,000
Illinois Natural History Survey	308,000	17	94,000
Iowa State University	46,000	0	102,000
Miami University	14,000	5	35,000
Missouri Botanical Garden	247,000	25	101,000
New York Botanical Garden	102,000	30	274,000
University of Colorado	51,000	0	67,000
University of Illinois	0	0	30,000
University of Kansas	129,000	65	97,000
University of Maine	100,000	0	34,000
University of Michigan	26,000	0	115,000
University of Minnesota	93,000	10	70,000
University of Texas	105,000	10	105,000
University of Wisconsin	120,000	50	90,000
Total	**1,341,000**		**1,224,000**
GRAND TOTAL			**2,565,000**

Additional data contributors: Consortium of California Herbaria, Consortium of Pacific Northwest Herbaria, Southwest Environmental Information Network

Rapid Data Entry

- **Generate “skeletal” database records for all specimens to be digitized**
  - Barcode
  - Scientific name (as filed)
  - Use Tropicos® authority files from the Missouri Botanical Garden
- **Average ±150-200 records per hour**
- **Send existing specimen data to NY**
  - Complete specimen records
  - Georeferenced (if available)
  - Darwin Core format

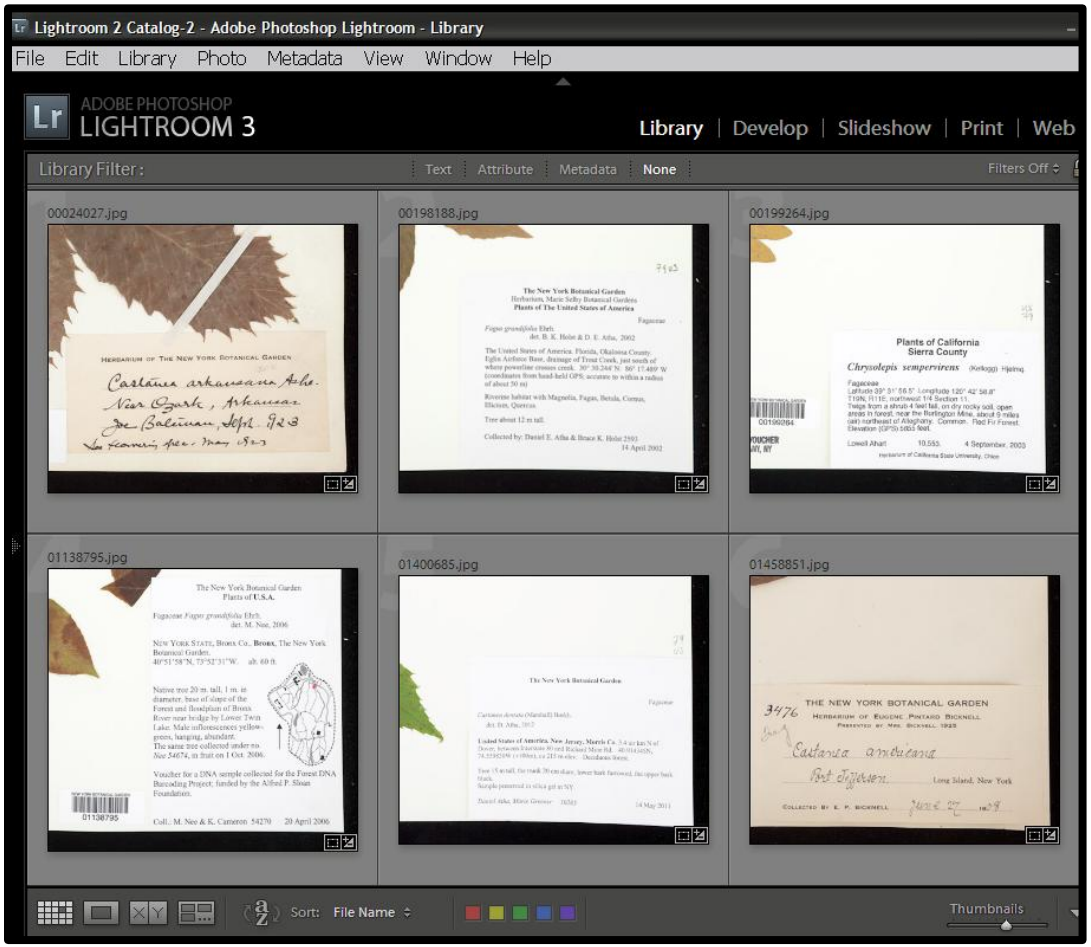


Rapid Image Capture

- **Imaging equipment**
  - 21 Megapixel DSLR camera
  - Macro Autofocus lens, 55 mm
  - Photo-eBox, with even illumination
  - Copy stand
- **Photograph every specimen**
  - Image file name = Barcode
- **Average ±80-120 exposures per hour**
- **Send JPGs and skeletal data to NY**

Batch Image Post-Processing

- **Quality control**
- Convert to grayscale
- Autolevels
- Crop to lower half or lower right corner of sheet
- Crop to label
- Flag specimens with >1 barcode per sheet
- Export JPGs of labels only

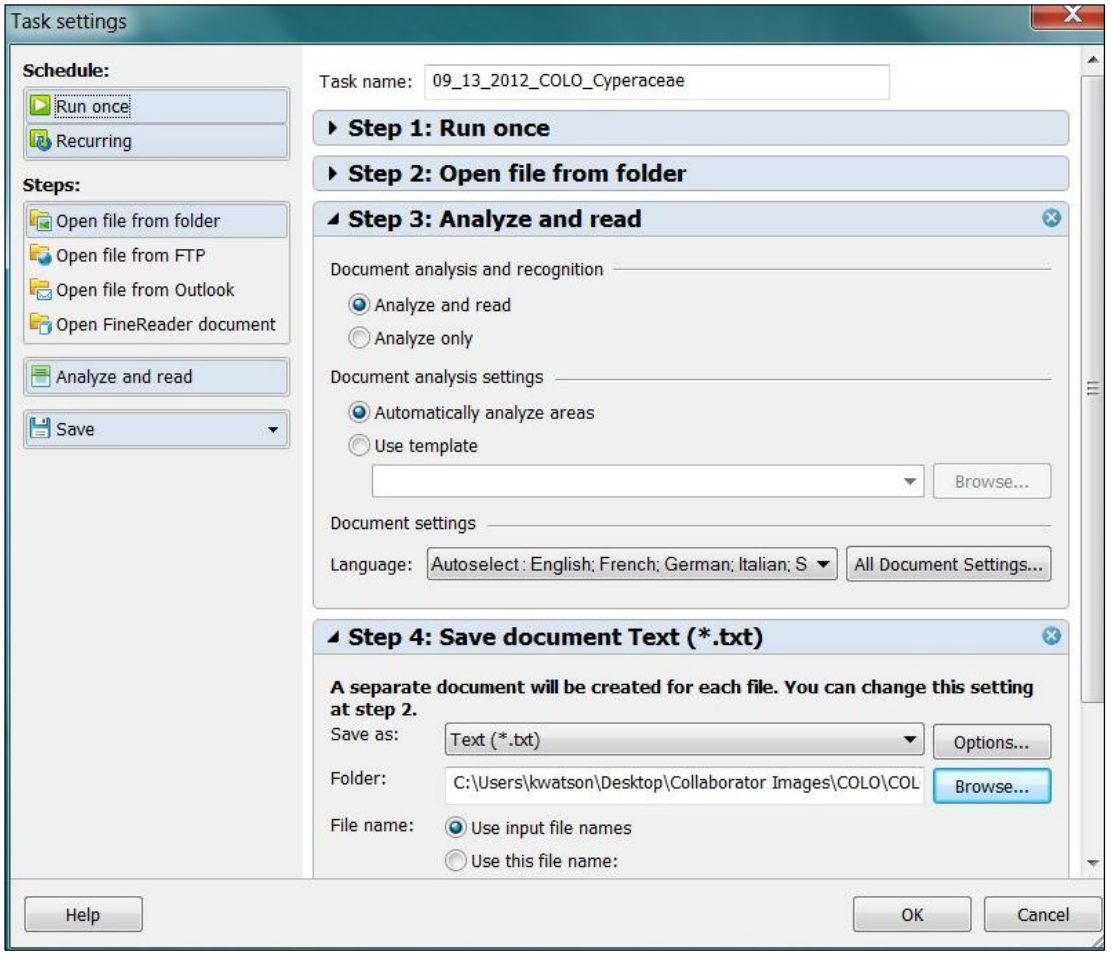


Batch OCR

ABBYY FineReader 11 Corporate Edition

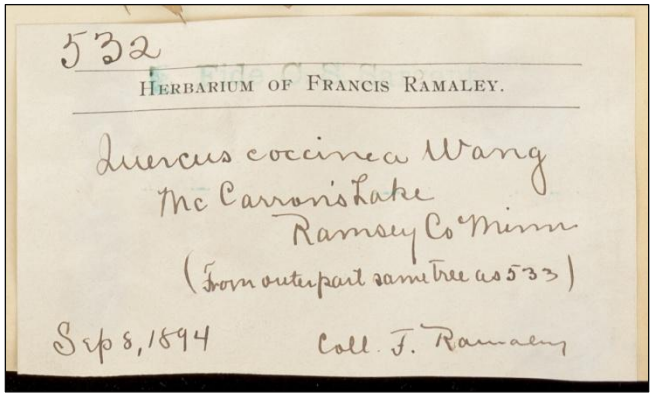
ABBYY Hot Folder

- Run once or recurring on a given directory
- Automatically analyze and read
- Autoselect Language
- Save as individual text files
  - Barcode.txt



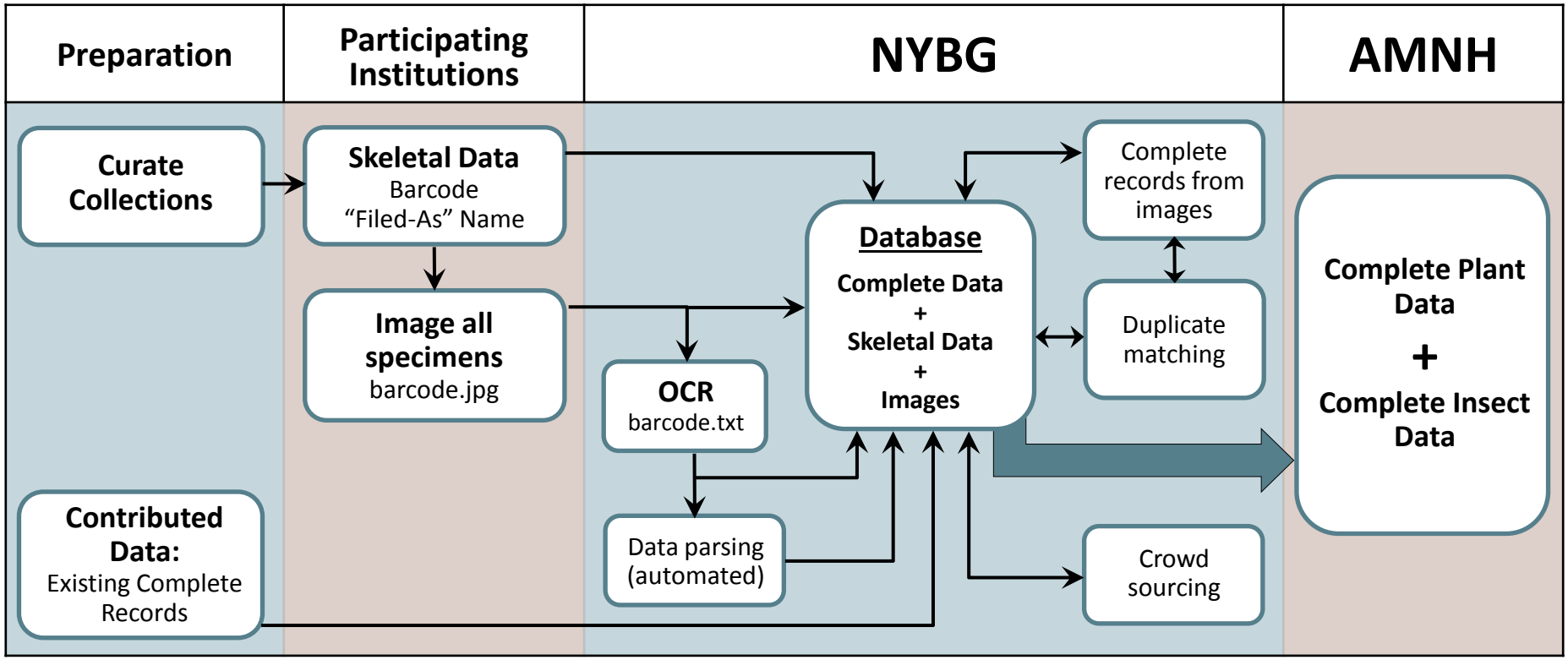
Populate Skeletal Records

- Merge individual text files into one table using Powershell script
- Skeletal record = Barcode, Scientific name, JPG image, editable OCR text
- Search, group, enter data for several collections at once



Microsoft Excel - COLO_09_20_12_Cyperaceae_OCR.xlsx	
01048842.txt	PLANTS OF BRITISH COLUMBIA CANADA MANNING PARK Carex nigricans C. A. Meyer Blackwall Peak north of Ranger Station on Hope-Princeton Highway. Common in moist,, grassy area by creek below cliffs on north slope; alt. about 6000 No. 11591 - J. A. Calder August 9/53 D,, B. O. Savile Botany Division,Science Service Department of Agriculture, Ottawa, Canada,
01044247.txt	PLANTS OF BRITISH COLUMBIA, CAN. Carex media R,, Br,, Four miles on road to Jesmond from crossroads near Kelly Lake; approx. 51°04'N 121' Occasional in wet, mossy area near creek; alt. 4400' . No. 18719 Coll. J.A. Calder July 12, 1956 J.A. Parmelee R.L. Taylor Botany Division, Science Service Department Agriculture, Ottawa, Canada
01048859.txt	PLANTS OF BRITISH COLUMBIA, CAN. Carex nigricans C.A. Meyer Itcha Mountains 26 miles NE of Anahim Lake; approx. 52°43,N 124°54,W Common in almost all moist habitats in the Itcha1 s from tree line to mountain sumr No. 20249 Coll: J. A. Calder Aug. 16-19, J.A.Parmelee 1956 R. L. Taylor Botany Division
01050301.txt	PLANTS OF BRITISH COLUMBIA, CANADA QUEEN CHARLOTTE ISLANDS Carex macloviana ssp. pachystachya (Cham.) Hult, Common in swale in wet maadowland back from sea beach. Outskirts of Sandspit, Moresby Island. No. 21821 J.A. Calder June 27, 1957 D.B.O. Savile, R.L. Taylor DEPARTMENT OF AGRICULTURE, OTTAWA, CANADA
01044270.txt	PLANTS OF BRITISH COLUMBIA, CANADA Carex media R. Br. 3l/2 miles ESE of Barker ville. Occasional in moist mossy area at edge of coniferous woods; alt. 4000 feet. No. 14259 Coll:J.A.Calder August 8, 1954. D. B. O. Savile, J.M. Ferguson. botany division, science service DEPARTMENT OF AGRICULTURE, OTTAWA. CANADA I86«i—ism—334
01050699.txt	PLANTS OF BRITISH COLUMBIA, CANADA Carex parrvana Dewey

Plant Specimen Digitization Workflow



- Automated data parsing (e.g. SALIX 2, LABELx)
- Duplicate matching from complete records
- Manual data entry from image; crowd sourcing

Insects

Participating Entomological Institutions

Institutions (18)	Specimens already digitized	% Georeferenced	Specimens to be digitized
American Museum of Natural History	30,000	100	333,000
B. P. Bishop Museum, Honolulu	0	0	70,000
California Academy of Sciences	4,000	100	40,000
California Dept. Food & Agriculture	1,000	100	75,000
Carnegie Museum, Pittsburgh	0	0	15,000
Colorado State University	0	0	15,000
Cornell University	0	0	30,000
Illinois Natural History Survey	36,000	100	73,000
Mississippi State University	0	0	50,000
North Carolina State University	1,000	100	75,000
Oregon State University	1,000	100	40,000
Texas A&M University	15,000	100	150,000
University of California, Berkeley, Essig Museum	12,000	92	45,000
University of California, Riverside	14,000	100	75,000
University of Delaware	2,000	0	20,000
University of Kansas	0	0	50,000
University of Kentucky	0	0	35,000
University of Massachussetts, Amherst	10,000	0	15,000
Total	**126,000**		**1,206,000**
Grand Total			**1,332,000**

Additional data contributors: Canadian National Collection, Ottawa; University of California, Davis; Kansas State University

Streamlined Workflow for Rapid Data Entry

Curate and stage specimens

- Scientific name (determined by specialist)
- Collection event
- Sex

Pin barcode to each specimen

Capture complete label data, including host and sex

Streamlined Web Interface

Arthropod Easy Capture

Taxon

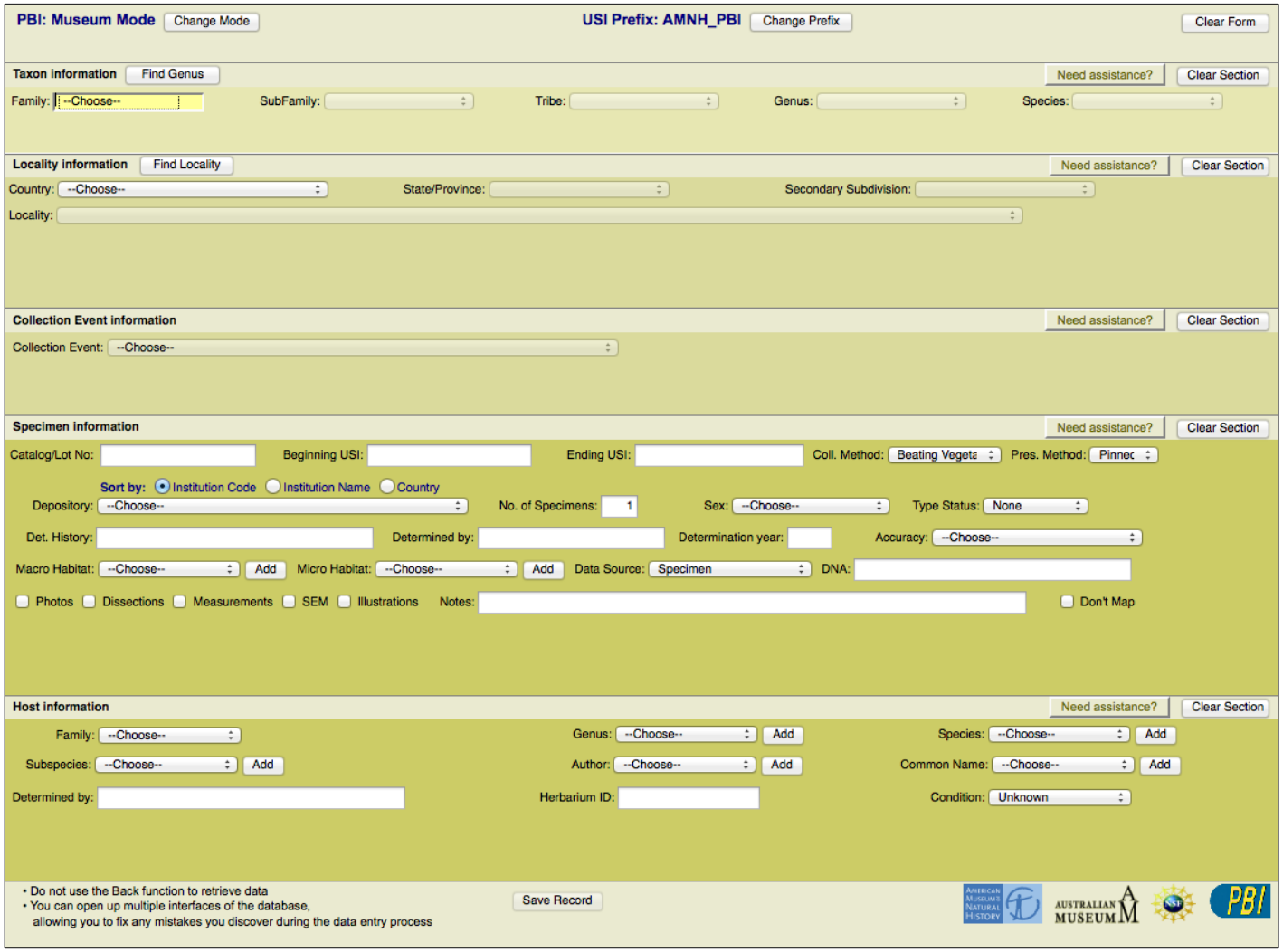
Locality

Collection event

Specimen data

Host data

- Plant
- Herbivore



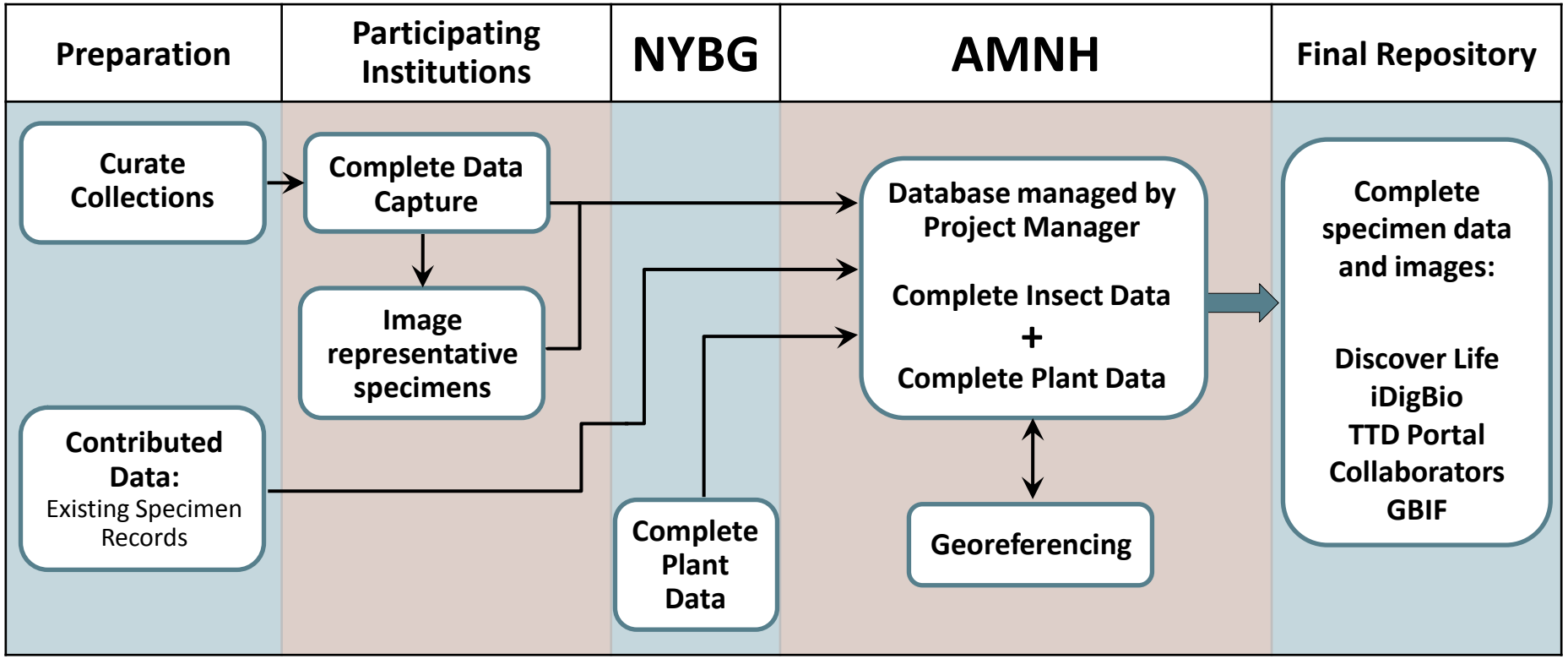
Schuh RT (2012) Integrating specimen databases and revisionary systematics. ZooKeys 209: 255–267, doi: 10.3897/zookeys.209.3288

Image Capture

- Use existing imaging stations at partner institutions
- Photograph representative male and female specimens for each species
- Generate high-resolution composite images
- Gather exemplars → Capture images → Composite and save RAW images → Crop images → Recursively add scale bars to images → Upload to online server
- Expect to produce about 20,000 new images

Tuxedo drakei Schuh

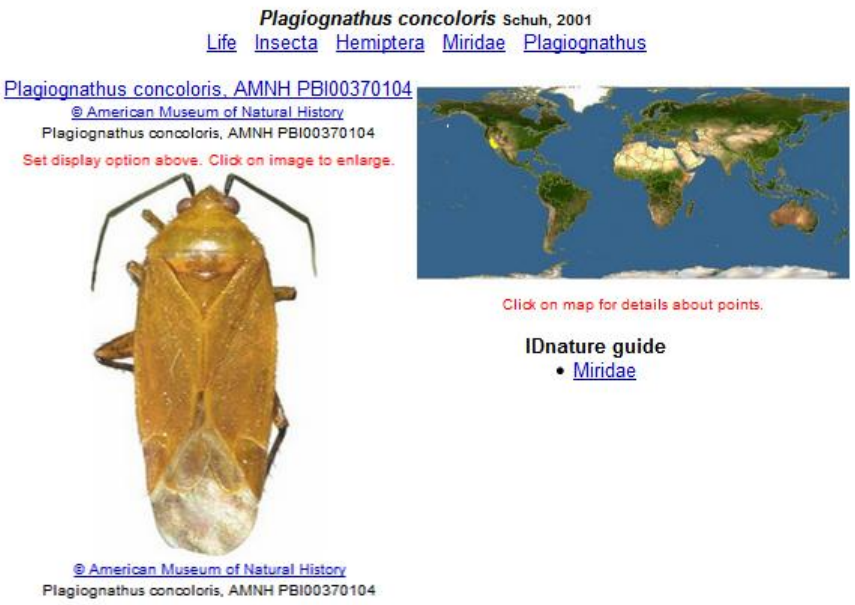
Insect Specimen Digitization Workflow

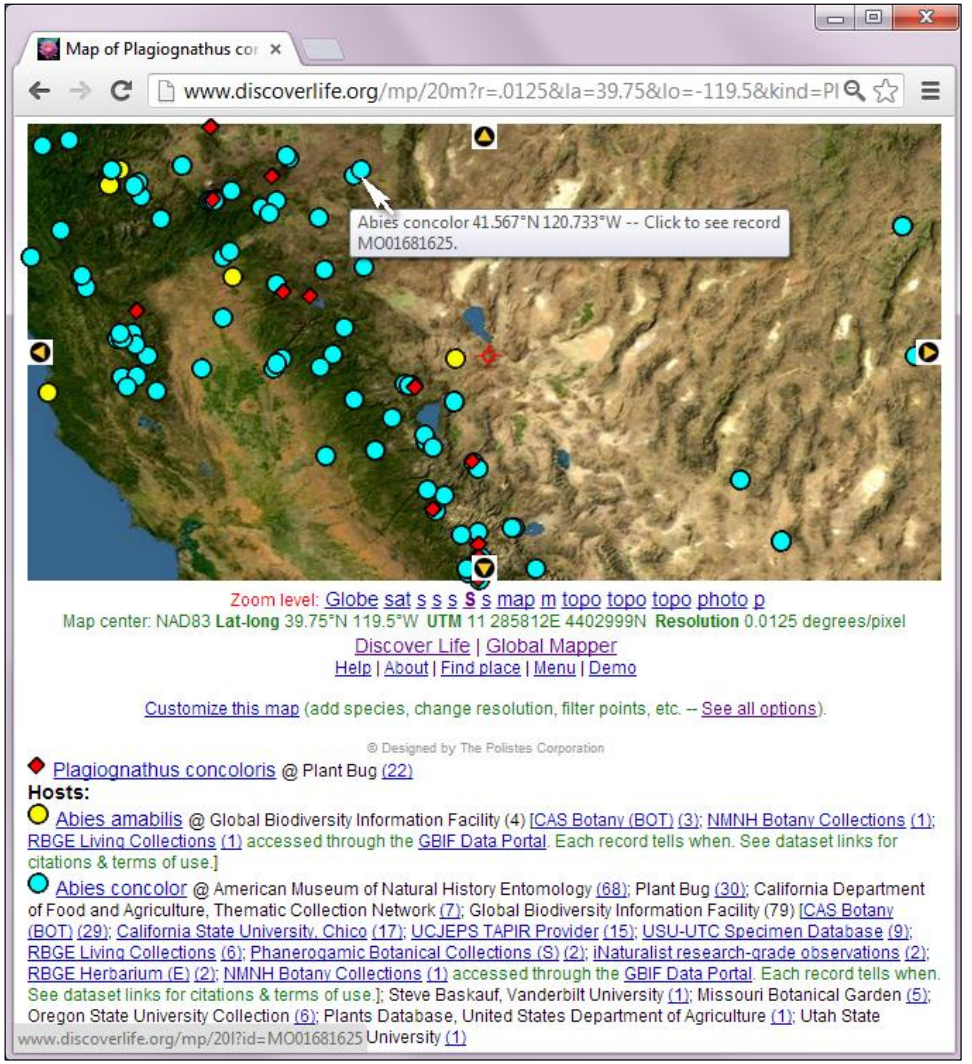


- All data aggregated at AMNH, managed by Project Manager, Katja Seltmann
- Georeference combined dataset, use automated tools (e.g. GEOLocate)
- Completed plant and insect data submitted to iDigBio, Discover Life, GBIF, etc.
- Completed plant specimen data repatriated to participating herbaria

Data Integration: Discover Life

- Generates species pages with specimen data, maps, and images
- Creates a linkage between host/herbivore/parasitoid data
- PBI database is already a data provider
- Data updated every ±24 hours





www.discoverlife.org/tttcn/

Digitization Challenges

- Insure accuracy of specimen identifications
- Assimilate and implement authority files for the groups where needed
- Expand existing insect database to include authority files for parasitoids, and plants
- Standardize and integrate data across databases
- Duplicate specimens with differing names; how to report discrepancies to partners
- Train botanical collaborators to manage data and images, use imaging equipment
    - Digitization and data management experience
    - Technical support
- Maintain data over the long-term
- Long-term archival image storage for all institutions, 36+ TB of RAW files
- Communicate efficiently and effectively among participants from >30 institutions

Tri-Trophic TCN Partners

BOTANY

- **Robert Naczi, New York Botanical Garden**
- **Robert Magill, Missouri Botanical Garden**
- **Richard Rabeler, University of Michigan**
- **Melissa Tulig, New York Botanical Garden**
- Barbara Thiers, New York Botanical Garden
- Kimberly Watson, New York Botanical Garden
- Margaret Koopman, Eastern Michigan University
- Loy Phillippe, Illinois Natural History Survey
- Deborah Lewis, Iowa State University
- Michael Vincent, Miami University
- Timothy Hogan, University of Colorado
- Mary Ann Feist, University of Illinois
- Craig Freeman, University of Kansas
- Christopher Cambell, University of Maine
- Anita Cholewa, University of Minnesota
- Beryl Simpson, University of Texas
- Kenneth Cameron, University of Wisconsin

Data Contributors

- Consortium of Pacific Northwest Herbaria
- Consortium of California Herbaria
- Southwest Biodiversity Consortium

ENTOMOLOGY

- **Randall Schuh, American Museum of Natural History**
- **Christiane Weirauch, University of California, Riverside**
- **John Heraty, University of California, Riverside**
- **Charles Bartlett, University of Delaware**
- **Benjamin Normark, University of Massachusetts, Amherst**
- **Katja Seltmann, American Museum of Natural History**
- Christine Johnson, American Museum of Natural History
- Neal Evenhuis, BP Bishop Museum, Honolulu
- David Kavanaugh, California Academy of Sciences
- Stephen D. Gaimari, California Dept. Food and Agriculture
- Chen Young, Carnegie Museum, Pittsburg
- Boris C. Kondratieff, Colorado State University
- James K. Liebherr, Cornell University
- Dmitry Dmitriev, Illinois Natural History Survey
- Richard Brown, Mississippi State University
- Andy Deans, North Carolina State University
- David Maddison, Oregon State University
- Christopher Marshall, Oregon State University
- John Oswald, Texas A&M University
- Kipling Will, University of California, Berkeley
- Caroline Chaboo, University of Kansas
- Michael Sharkey, University of Kentucky
- John Pickering, University of Georgia

Data Contributors

- Canadian National Collection, Ottawa
- University of California, Davis
- Kansas State University

American Museum of Natural History

The New York Botanical Garden

NSF

NSF Award#1115104

Thanks!

Tri-Trophic TCN: https://sites.google.com/site/ttdtcn/

Discover Life: www.discoverlife.org/tttcn/