# The evolution of databasing at the Illinois Natural History Survey Insect Collection (USA: Illinois): lessons learned from migrating three decades of digital data into TaxonWorks **Thomas C. McElrath<sup>1</sup>**, Dmitry A. Dmitriev<sup>1,2</sup>, Matthew J. Yoder<sup>1,2</sup> 1 = Insect Collection; 2 = Species File Group; Illinois Natural History Survey, Prairie Research Institute, University of Illinois, Champaign, IL, USA

### Abstract

Parasite

The Illinois Natural History Survey Insect Collection (INHS-IC) recently migrated from several legacy collection management systems and into TaxonWorks, a biodiversity data management system developed by the SpeciesFile group and its collaborators at INHS. The migration process had several immediate outcomes. First, it unified semantics over five separate databases that were throttling collection management. Second, new interfaces illuminated past issues arising from inconsistent use, conflicting goals, and poorly constrained standards. Third, the migration itself introduced issues originating from the complex mapping of old to new data models. Once the migration was in place, we had the opportunity to quickly iterate on improvements in collections management workflows in conjunction with developers of TaxonWorks. Adoption of the new system allowed us to fully absorb digital assets, including several hundred thousand images and collection objects in collection management " "limbo". We discuss how to manage such a move, what we could have done better, and how valuable the relationship between museum professionals and behind-the-scenes coders can be. As we move closer to recording our millionth digital accession, we are excited to imagine the possibilities that a flexible and evolving biodiversity workbench means for innovation within collections digitization and downstream uses of digital data.

#### Methods

The INHS-IC is one of the first projects to adopt TaxonWorks for day-to-day collection management, e.g. specimen digitization, loan management, and label printing. Our pre-migration data was complex, originating over several decades and persisting in three different relational databases (FileMaker Pro for collection objects, collecting events, people, and loans; Microsoft Access for container profiles; and web tool with a SQL back end for data entry). These data were migrated to TaxonWorks via a custom script written by DAD and MJY. The migration was tested and refined on and off over a one year period in synergistic parallel with the development of required features in TaxonWorks.

# Results

Nearly 1 million collection objects were migrated along with their associated collecting event, people, loan object, georeferences, locality, and container tables. Once in Taxon-Works, we were able to run queries against a unified dataset for the first time (Fig. 3, 4).

Outstanding issues with the migration are currently tracked on GitLab. There we record various "to-do's", fixes, or uploads, or script-based curation that need to occur. Problems encountered following the migration include: 1) duplicate or improperly delimited namespaces, 2) a subset of collection objects that were not assigned collecting events (Fig. 4), 3) and various fields were not initially migrated.

In the six months since migration, 42,641 data objects have been "depicted" in our data-

### Discussion

**THE GOOD:** The migration to TaxonWorks was a marked advance in our digital infrastructure. At the onset it allowed us to use data model and code underlying TaxonWorks in a script-framework to model and ingest our diverse legacy data, creating new, semantically rich relationships in the process. Our new ability to natively manage images has led to increased data validation and subsequent correction of longstanding issues with transcription quality (Fig. 8). Furthermore, image handling has lead to the integration of decades worth of images (Figs. 1, 5-7) generated from a variety of previously funded National Science Foundation grants, daily curation, and the INHS-IC portion of the Terrestrial Parasite Tracker Thematic Collections Network (TPT-TCN) grant.

As TaxonWorks is web-based we have greatly increased flexibility to work anywhere. Students and hourlies were able to capture images of specimens for later transcription, which has served us well during the global COVID-19 pandemic. Over 9,000 specimens for the Terrestrial Parasite Tracker project have been uploaded and are being transcribed remotely.

TaxonWorks as a platform has also seen immediate benefits from the migration process. Issue (bug, feature requests) reporting increased sharply following the migration, with 102 more issues being reported in the six months following migration than before (Fig. 4). As more daily users encountered bugs, workflow constraints, or errors in migration, the corresponding number of issues reported and then closed also increased. Through a series of brainstorming a design sessions, followed by iterative development, the authors imagined, created, and improved several TaxonWorks custom workflows or "tasks" (Figs. 6-9). These workflows have since been used by multiple other collections using TaxonWorks to ingest

**THE BAD:** Once all the data was unified into one location, many issues with data standardization became readily apparent. Because of the disjunct nature of digitization, different projects had a variety of standard operating procedures, with some generating only images, others only verbatim data, or others only parsed data. With the robust data architecture of TaxonWorks, these data are all handled in parity, and value can be added as needed or required at later points in time. For example, the numerous specimens generated with only verbatim label data can now be bulk parsed in an in-development text-string annotator.

The use of collaborative issue-tracking (e.g. GitHub/GitLab) has greatly simplified the tracking of issues with such a large migration, as well as improvements to data presentation, use, and workflows. Over 50% of the issues created on our GitLab have been solved, with numerous others on their way to being fixed. Explicit problem narration has allowed for better communication between collections personnel and developers.

**THE UGLY?:** Looking back, even more communication between collection personnel and developers could have helped avoid some of the issues that have been revealed since the migration. Because some of the collection personnel arrived following many of the initial dis-Cussions about how data is modeled in TaxonWorks, they were not able to check that they migrated correctly until differences in understanding showed themselves far down the line. While these problems are not uncorrectable, time and energy could have been saved if all steps were discussed in more detail. Migration of the INHS-IC data and initial use of the database for collection management highlighted the needs for specialized interfaces for col-Iection management, which were designed and tested for the last two years in collaboration with Species File Group developers.

**THE FUTURE:** Moving forward, the digital infrastructure of INHS will continue to benefit from Collaborative issue-tracking, real-time development and open responses from developers, and the continuing refinement of digital data workflows. Depictions of all kinds of data will Iead to increased validation and communication with scientists using that data. Real-time queries about data quality will allow for targeted curation projects and lead to higher-quality data being shared with aggregators. Cross-collection meetings with others using Taxon-Works is synergistically identifying and prioritizing key improvements that once met will ben-

# Acknowledgments

The authors thank José Luis Pereira and Hernán Lucas Pereira (Argentina) for their coding prowess and bug-squashing. The authors also wish to thank the organizing committee of SPNHC for allowing the authors to change format to a poster. We also thank the Species File Group (https:// speciesfilegroup.org/) for funding the development of TaxonWorks.





![](_page_0_Figure_26.jpeg)

+	Assign	Overview metadata	Review	
Insect Collection 508843	Obiect metadata	Summarv		
Really .	Added to all collection objects accordingly	Create	Create and Create a	nd
		Create	next new	Nuke
Insect Collection 508844	☐ Is metadata depiction	Catalogu     Preparati	e number will be added.	
INH Bankar Colo SOB B	O none	Reposito	ry type will be added.	
	Odown -> across			
Insect Collection 508845	→ across -> down			
Macon coll-	Quick Recent Pinboard	Search		
Insoct Collection 508846	Search	Q		
Insect Collection 508846	INHS Insect Collection			
TINHS Sob, B18	Identifier End			
nsect Collection 508847	<b>508834</b> 508850			
Proma to C	Repository			
	Search a repository	Q		
nsect Collection 508848	Illinois Natural History Survey 🕥			
Numeric Colores				
Biterion Biterion Non	Preparation			
nsect Collection 508849 6/3	O Bulk dry O Pin			
ty, 6245	O Pill box O Vial			
		Legend	Stage images	2
		Legend Triangle indicates data pres	Stage images Sence Purple borders indicate record is in progress	
		Legend <i>Triangle indicates data pres</i> Identifier(s)	Stage images Sence Purple borders indicate record is in progress	
O map Original image Recalcu	ulate Buffered breakdown	Legend Triangle indicates data press Identifier(s) Buffered collecting end Ender the second seco	Stage images Sence Purple borders indicate record is in progress event	
O map Original image Recalcu	ulate Buffered breakdown	Legend Triangle indicates data press Identifier(s) Buffered collecting es Buffered determination Triangle indicates data press Identifier(s) Buffered collecting es Estimates data press Identifier(s) Identif	Stage images Purple borders indicate record is in progress event ion	
O map Original image Recalcu Collecting event labels Use [a] Illinois, Jackson Co., Little Grau	ulate Buffered breakdown Determination labels Use [s] nd Canyon, Det. R. E. DeWalt, 2016	Legend Triangle indicates data press identifier(s) identifier(s) Buffered collecting et i Buffered determination Buffered other labels i	sence Purple borders indicate record is in progress	
O map Original image Recalcu Collecting event labels Use [a] Illinois, Jackson Co., Little Grav about 5.8 mi SW Murphysboro,	ulate Buffered breakdown Determination labels Use [s] nd Canyon, , [37.68028,	Legend Triangle indicates data press Identifier(s) Buffered collecting e E Buffered determination Buffered other labels Collecting event	sence Purple borders indicate record is in progress	
Imap       Original image       Recalculation         Collecting event labels       Use [a]         Illinois, Jackson Co., Little Granabout 5.8 mi SW Murphysboro, -89 39472], 15 September 197         Richardson	Ulate Buffered breakdown Determination labels Use [s] nd Canyon, , [37.68028, 73, M. W.	Legend Triangle indicates data press Identifier(s) Buffered collecting e E Buffered determination Collecting event Collecting event	sence Purple borders indicate record is in progress went ion s (5)	
Omap Original image Recalcu Collecting event labels Use [a] Illinois, Jackson Co., Little Gran about 5.8 mi SW Murphysboro, -89 39472], 15 September 19 Richardson	Ulate Buffered breakdown Determination labels Use [s] nd Canyon, , [37.68028, 73, M. W.	Legend Triangle indicates data press Identifier(s) Udentifier(s) Buffered collecting er U Buffered determination Collecting event U Taxon determination	sence Purple borders indicate record is in progress went ion (s)	
O map Original image Recalcu Collecting event labels Use [a] Illinois, Jackson Co., Little Gran about 5.8 mi SW Murphysboro, -89 39472], 15 September 19' Richardson Identifier Use [g] Specimen	Ulate Buffered breakdown Determination labels Use [s] Det. R. E. DeWalt, 2016 (37.68028, 73, M. W.	Legend Triangle indicates data press Identifier(s) Udentifier(s) Buffered collecting end Collecting event Collecting event Collecting event Collecting event Collecting event Collecting event Collecting event	Stage images sence Purple borders indicate record is in progress went ion (s)	<image/>
O map Original image Recalcu Collecting event labels Use [a] Illinois, Jackson Co., Little Grad about 5.8 mi SW Murphysboro, -89 39472], 15 September 19 Richardson Identifier Use [g] Specimen vt	Ulate Buffered breakdown Determination labels Use [s] Det. R. E. DeWalt, 2016 Use [h]	Legend Triangle indicates data press Identifier(s) Uffered collecting et Euffered determination Collecting event Collecting event Taxon determination	sence Purple borders indicate record is in progress went ion (s)	<image/>
map       Original image       Recalculation         Collecting event labels       Use [a]         Illinois, Jackson Co., Little Gravatoria       about 5.8 mi SW Murphysboro, -89 39472], 15 September 19°         Richardson       Identifier       Use [g]         Specimen       vt	Ulate Buffered breakdown Determination labels Use [s] Det. R. E. DeWalt, 2016 Use [h]	Legend Triangle indicates data press Legend Triangle indicates data press Ledentifier(s) Duffered collecting e Duffered determination Collecting event Collecting event Collecting event Collecting event	sence Purple borders indicate record is in progress went ion (s)	
map       Original image       Recalculation         Collecting event labels       Use [a]         Illinois, Jackson Co., Little Gravatoria       about 5.8 mi SW Murphysboro, -89 39472], 15 September 19°         Richardson       Identifier       Use [g]         Specimen       vt	Ulate       Buffered breakdown         Ind Canyon,       Determination labels Use [s]         Ind Canyon,       Det. R. E. DeWalt, 2016         (J37.68028,       Det. R. E. DeWalt, 2016         Use [h]       Use [h]	Legend Triangle indicates data pres Identifier(s) Uffered collecting e Uffered determination Uffered other labels Uffered other labels Uffered other labels Uffered other labels Taxon determination	sence Purple borders indicate record is in progress went ion s (s)	
map       Original image       Recalculation         Collecting event labels       Use [a]         Illinois, Jackson Co., Little Gradabout 5.8 mi SW Murphysboro, -89 39472], 15 September 19°         Richardson         Identifier       Use [g]         Specimen         vt	Ulate Buffered breakdown           Determination labels         Use [s]           Det. R. E. DeWalt, 2016         Determination labels           Use [h]         Use [h]	Legend Triangle indicates data pres Legend Identifier(s) Udentifier(s) Buffered collecting e Collecting event Collecting event Collecting event Collecting event	sence Purple borders indicate record is in progress went ion s (s)	<image/>
map       Original image       Recalculation         Collecting event labels       Use [a]         Illinois, Jackson Co., Little Gran         about 5.8 mi SW Murphysboro,         -89 39472], 15 September 19°         Richardson         Identifier       Use [g]         Specimen         vt         Identifiers	Utate Buffered breakdown Determination labels Use [s] Det. R. E. DeWalt, 2016 Use [h] Use [h]	Legend Triangle indicates data pres Legend Identifier(s) Duffered collecting e E Buffered determination Collecting event Collecting event Collectin	sence Purple borders indicate record is in progress went ion (s)	<image/>
map       Original image       Recalculation         Collecting event labels       Use [a]         Illinois, Jackson Co., Little Grad       about 5.8 mi SW Murphysboro,         about 5.8 mi SW Murphysboro,       -89 39472], 15 September 19°         Richardson       Identifier         Vt	Late Buffered breakdown Determination labels Use [s] Det. R. E. DeWalt, 2016 Use [h] Use [h] Identifier	Legend Triangle indicates data pres Identifier(s) Uffered collecting e Uffered determination Uffered other labels Uffered othe	Stage images Purple borders indicate record is in progress went ion (s)	<image/>
map       Original image       Recalculation         Collecting event labels       Use [a]         Illinois, Jackson Co., Little Grau       about 5.8 mi SW Murphysboro,         about 5.8 mi SW Murphysboro,       -89 39472], 15 September 19°         Richardson       Identifier         Use [g]       Specimen         vt       Identifiers         Identifiers       Snmespace         INHS Insect Collection       Interview	Jate       Buffered breakdown         Determination labels       Use [s]         Det. R. E. DeWalt, 2016       Determination         Use [h]       Use [h]         Identifier       Identifier         Q       891323	Legend Triangle indicates data pres i dentifier(s) i dentifier(s) ii d	sence Purple borders indicate record is in progress went ion s (s)	<image/>
map Original image Recalculation   Collecting event labels Use [a]   Illinois, Jackson Co., Little Grau about 5.8 mi SW Murphysboro, -89 39472], 15 September 19°   Richardson   Identifier Use [g]   Specimen   vt   Identifiers   Namespace   INHS Insect Collection   Namespace	Late Buffered breakdown          Image: Determination labels       Use [s]         Determination labels       Use [s]         Det. R. E. DeWalt, 2016         Use [h]         Use [h]         Identifier         Q       891323         Identifier	Legend Triangle indicates data pres i dentifier(s) i dentifier(s) Buffered collecting e i Buffered other labels i Collecting event i Taxon determination	sence Purple borders indicate record is in progress went ion s	<image/>
map       Original image       Recalcular         Collecting event labels       Use [a]         Illinois, Jackson Co., Little Grad       about 5.8 mi SW Murphysboro,         -89 39472], 15 September 19°       Richardson         Identifier       Use [g]       Specimen         vt       Identifiers       Interview         Identifiers       Namespace       INHS Insect Collection         Namespace       Select a namespace       Select a namespace	Llate       Buffered breakdown         Determination labels       Use [s]         Det. R. E. DeWalt, 2016       Det. R. E. DeWalt, 2016         Use [h]       Identifier         Identifier       891323         Identifier       Q         Identifier       Q	Legend Triangle indicates data pres i dentifier(s) i dentifier(s) iiiiiiiiiiiiiiiiiiiiiiiiiiiiiiiiiiii	sence Purple borders indicate record is in progress went ion s	<image/>
map       Original image       Recalculation         Collecting event labels       Use [a]         Illinois, Jackson Co., Little Grad about 5.8 mi SW Murphysboro, -89 39472], 15 September 19°         Richardson         Identifier       Use [g]         Specimen         vt         Identifiers         Namespace         INHS Insect Collection         Namespace         Select a namespace         Muffered collecting event	Jate       Buffered breakdown         Determination labels       Use [s]         Ind Canyon,       Det. R. E. DeWalt, 2016         [37.68028,       Table of the second seco	Legend Triangle indicates data pres Identifier(s) Buffered other labels Collecting event Collecting event Coll	sence Purple borders indicate record is in progress went ion s	
map       Original image       Recalculation         Collecting event labels       Use [a]         Illinois, Jackson Co., Little Gran         about 5.8 mi SW Murphysboro,         -89 39472], 15 September 197         Richardson         Identifier       Use [g]         Specimen         vt         Identifiers         Namespace         INHS Insect Collection         Namespace         Select a namespace         Murphysboro,         Linois, Jackson Co., Litt         Grand Canyon, about 5.8 mi         SW Murphysboro,         SW Murphysboro,	Jate       Buffered breakdown         Determination labels       Use [s]         Det. R. E. DeWalt, 2016       Det. R. E. DeWalt, 2016         Use [h]       Identifier         Use [h]       Identifier         Buffered determinations       Rgrypnia straminae, F         Det. R. E. DeWalt, 2016       Det. R. E. DeWalt, 2016	Legend Triangle indicates data pres Udentifier(s) Buffered collecting e Collecting event Collecting event Collecti	sence Purple borders indicate record is in progress went ion s (6)	
map       Original image       Recalculation         Collecting event labels       Use [a]         Illinois, Jackson Co., Little Grand about 5.8 mi SW Murphysboro, -89 39472], 15 September 19°         Richardson         Identifier       Use [g]       Specimen         vt         Identifiers         Namespace         INHS Insect Collection         Namespace         Select a namespace         Select a namespace         Illinois, Jackson Co., Litt         Grand Canyon, about 5.8 mi         SW Murphysboro, [37.68028, -89.39472], 15 September 1         M. W. Richardson	ulate       Buffered breakdown         Ind Canyon,       Determination labels Use [s]         Identifier       Det. R. E. DeWalt, 2016         Use [h]       Identifier         Use [h]       Identifier         Identifier       891323         Identifier       Identifier         Id	Legend Triangle indicates data pres Udentifier(s) Udentifier(s) Buffered collecting e Collecting event Collecting	sence Purple borders indicate record is in progress went ion s (s)	
map       Original image       Recalculation         Illinois, Jackson Co., Little Grad about 5.8 mi SW Murphysboro, -89 39472], 15 September 19' Richardson       Identifier Use [g]       Specimen         Identifier       Use [g]       Specimen       Vt         Identifiers       Namespace       INHS Insect Collection         Namespace       Select a namespace       Select a namespace         V.       Buffered collecting event       [111inois, Jackson Co., Litt         Grand Canyon, about 5.8 mi       Sw Murphysboro, [37.68028, -89.39472], 15 September 1         M. W. Richardson       M. W. Richardson	ulate       Buffered breakdown         Determination labels       Use [s]         nd Canyon,       Det. R. E. DeWalt, 2016         [J7.68028,       Det. R. E. DeWalt, 2016         Use [h]       Identifier         Use [h]       Identifier         Buffered determinations       891323         Identifier       Setle         Buffered determinations       Agrypnia straminae, F         Det. R. E. DeWalt, 2016       Det. R. E. DeWalt, 2016	Legend Triangle indicates data pres Legend Tiangle indicates data pres Legend Identifier(s) Buffered collecting e Collecting event Collecting event	sence Purple borders indicate record is in progress went ion s (c)	
Omap       Original image       Recalculation         Collecting event labels       Use [4]         Illinois, Jackson Co., Little Grau         about 5.8 mi SW Murphysboro,         -89 39472], 15 September 19"         Richardson         Identifier       Use [9]         Specimen         vt         Identifiers         Namespace         INHS Insect Collection         Namespace         Select a namespace         Numerphysboro, [37.68028, -89.39472], 15 September 19"         W.         Buffered collecting event         [111nois, Jackson Co., Litt         Grand Canyon, about 5.8 mi         SW Murphysboro, [37.68028, -89.39472], 15 September 19"         M. W. Richardson	Jate       Buffered breakdown         Ind Canyon,       Determination labels Use [s]         Det. R. E. DeWalt, 2016       Det. R. E. DeWalt, 2016         Use [h]       Identifier         Use [h]       Identifier         Buffered determinations       Agrypnia straminae, F         Det. R. E. DeWalt, 2016       Det. R. E. DeWalt, 2016	Legend Triangle indicates data pres ↓ Identifier(s) ↓ Buffered collecting e ↓ Buffered other labels ↓ Taxon determination ↓ ↓ ↓ ↓ ↓ ↓ ↓ ↓ ↓ ↓ ↓ ↓ ↓	sence Purple borders indicate record is in progress went ion s (s)	
Imap       Original image       Recalculation         Collecting event labels       Use [a]         Illinois, Jackson Co., Little Granta         about 5.8 mi SW Murphysboro,         -89 39472], 15 September 197         Richardson         Identifier         Vt         Identifiers         Namespace         INHS Insect Collection         Namespace         Select a namespace         V.         Buffered collecting event         fillinois, Jackson Co., Litt         Grand Canyon, about 5.8 mi         Sw Murphysboro, [37.68028, -89.39472], 15 September 1         M. W. Richardson         Total       1	Jate       Buffered breakdown         Determination labels       Use [s]         Ind Canyon, (137.68028, 73, M. W.       Det. R. E. DeWalt, 2016         Use [h]       Identifier         Use [h]       Identifier         Identifier       891323         Identifier       Identifier	Legend   Triangle indicates data press   Identifier(s)   Identifier(s)   Image: State of the state of th	sence Purple borders indicate record is in progress went ion s	
Imap       Original image       Recalculation         Collecting event labels       Use [a]         Illinois, Jackson Co., Little Grantabout 5.8 mi SW Murphysboro, -89 39472], 15 September 19'         Richardson         Identifier       Use [g]       Specimenter         vt         Identifiers       Specimenter         vt       Identifiers         Namespace       INHS Insect Collection         Namespace       Select a namespace         Select a namespace       Select a namespace         Murphysboro, [37.68028, -89.39472], 15 September 19         M. W. Richardson	ulate Buffered breakdown  Determination labels Use [s] Det. R. E. DeWalt, 2016  Use [h]  Use [h]  Use [h]  Identifier  Selfered determinations  Agrypnia straminae, F Det. R. E. DeWalt, 2016  Tag with Ohio Biological Survey  Save and next w/out data [n]	Legend   Triangle indicates data press   Identifier(s)   Identifier(s) <	stage images Purple borders indicate record is in progress went ion s (s) (s) (s) (s) (s) (s) (s) (s) (s) (	
map       Original image       Recalculation         Collecting event labels       Use [a]         Illinois, Jackson Co., Little Grad         about 5.8 mi SW Murphysboro,         -89 39472], 15 September 19'         Richardson         Identifier       Use [g]         Specimen         vt         Identifiers         Namespace         INHS Insect Collection         Namespace         Select a namespace         Murphysboro, [37.68028, -89.39472], 15 September 1         M. W. Richardson         Total 1         Note         (3)       (3)	Jate       Buffered breakdown         Ind Canyon,       Determination labels Use [s]         Det. R. E. DeWalt, 2016       Det. R. E. DeWalt, 2016         Use [h]       Identifier         Use [h]       Identifier         Buffered determinations       891323         Identifier       Identifier         Suffered determinations       Agrypnia straminae, F         Det. R. E. DeWalt, 2016       Tag with Ohio Biological Survey         Save and next w/out data [n]	Legend Triangle indicates data pres Latertifier(s) Buffered collecting e Collecting event Collecting event Taxon determination Taxon determination X Save and next Save	sence Purple borders indicate record is in progress went ion (6)	

Fig. 6. "Transcribe depiction" task developed to parse verbatim label values from staged specimens. Color hexagons indicate "completeness"

![](_page_0_Picture_31.jpeg)

![](_page_0_Picture_32.jpeg)

## @INHSIIInois @monotomidae

@INHSInsects

![](_page_0_Figure_35.jpeg)

Fig. 9. "Comprehensive digitization" task with movable image overlay for easy transcription and validation. Specimen generated for the Terrestrial Parasite Tracker Thematic Collections Network grant.

H S ollection 382

Oct. 27, 1936

Frison-Mohr

Philichihyophada

Double-Crested

OCT. 27, 1936

FRISON - Mohr.

Cormorant

toxocerum

(Nity)