How quickly can new accessions generate citations of a collection?

A collection manager’s perspective

Tommy McElrath, Illinois Natural History Survey, Champaign IL, 61820, USA

@INHSInsects @monotomidae

ORCID 0000-0003-0390-4227
Before I try to answer that question ...
INHS Insect Collection - our watershed

1,133,456 collection objects (digitized)
2,756,380 specimens (digitized)
INHS Insect Collection - the streams

Collection objects created by all project members
Big picture
Growing or dormant season burns: the effects of burn season on bee and plant communities

Brenna L. Decker & Alexandra N. Harmon-Threatt

Biodiversity and Conservation 28, 3621–3631 (2019) | Cite this article

1316 Accesses | 16 Citations | 14 Altmetric | Metrics
What tools do we have right now?
TaxonWorks: Tools - removing stream debris

Digitizing specimens faster (what collections people think about)
TaxonWorks: Tools - removing stream debris

Digitizing specimens faster (what collections people think about)
## TaxonWorks: Tools - removing stream debris

**Import data faster**

### Created imports

#### montelongo g...

**DwC-A Occurrences**  
Status: Ready

- Imported: 16
- Ready: 0
- NotReady: 0
- Errored: 0
- Failed: 0
- Unsupported: 0

#### morgan bees ...

**DwC-A Occurrences**  
Status: Ready

- Imported: 317
- Ready: 0
- NotReady: 0
- Errored: 0
- Failed: 0
- Unsupported: 0

### DwC-A Workbench

**Decker Burn Dataset**  
2184 records.

<table>
<thead>
<tr>
<th>Option</th>
<th>Status</th>
<th>occurrenceID</th>
<th>institutionID</th>
<th>institutionCode</th>
<th>collectionCode</th>
<th>catalogNumber</th>
<th>datasetName</th>
<th>basionyms</th>
<th>reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Imported</td>
<td>INHS Insect Collection 825911</td>
<td><a href="http://gbif.org/institution/Insect">http://gbif.org/institution/Insect</a>...</td>
<td>INHS</td>
<td>825911</td>
<td>DeckertBurnData</td>
<td>PreservedSpecimen</td>
<td>SCIN</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Imported</td>
<td>INHS Insect Collection 825912</td>
<td><a href="http://gbif.org/institution/Insect">http://gbif.org/institution/Insect</a>...</td>
<td>INHS</td>
<td>825912</td>
<td>DeckertBurnData</td>
<td>PreservedSpecimen</td>
<td>SCIN</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Imported</td>
<td>INHS Insect Collection 825913</td>
<td><a href="http://gbif.org/institution/Insect">http://gbif.org/institution/Insect</a>...</td>
<td>INHS</td>
<td>825913</td>
<td>DeckertBurnData</td>
<td>PreservedSpecimen</td>
<td>SCIN</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Imported</td>
<td>INHS Insect Collection 825914</td>
<td><a href="http://gbif.org/institution/Insect">http://gbif.org/institution/Insect</a>...</td>
<td>INHS</td>
<td>825914</td>
<td>DeckertBurnData</td>
<td>PreservedSpecimen</td>
<td>SCIN</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Imported</td>
<td>INHS Insect Collection 825915</td>
<td><a href="http://gbif.org/institution/Insect">http://gbif.org/institution/Insect</a>...</td>
<td>INHS</td>
<td>825915</td>
<td>DeckertBurnData</td>
<td>PreservedSpecimen</td>
<td>SCIN</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Imported</td>
<td>INHS Insect Collection 825916</td>
<td><a href="http://gbif.org/institution/Insect">http://gbif.org/institution/Insect</a>...</td>
<td>INHS</td>
<td>825916</td>
<td>DeckertBurnData</td>
<td>PreservedSpecimen</td>
<td>SCIN</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
TaxonWorks: Tools - increasing river size to the delta

Export data faster
**TaxonWorks: Tools - trace elements**

Linkages —> DarwinCore

![DwC Dashboard](image)
TaxonWorks: Tools - trace elements

Linkages -> Global Identifiers
## TaxonWorks: Tools - trace elements

### Linkages —> Global Identifiers

<table>
<thead>
<tr>
<th>Term</th>
<th>Source (TaxonWorks)</th>
<th>Original (GBIF)</th>
<th>Interpreted (GBIF)</th>
<th>Remarks</th>
</tr>
</thead>
<tbody>
<tr>
<td>associatedMedia</td>
<td><a href="https://fig.tax%D0%BE%D0%BDworks.org/api/vl/images/7eda9f5762a28b0ed3a0f7ee24dd10f">https://fig.taxонworks.org/api/vl/images/7eda9f5762a28b0ed3a0f7ee24dd10f</a></td>
<td><a href="https://fig.tax%D0%BE%D0%BDworks.org/api/vl/images/7eda9f5762a28b0ed3a0f7ee24dd10f">https://fig.taxонworks.org/api/vl/images/7eda9f5762a28b0ed3a0f7ee24dd10f</a></td>
<td>Excluded</td>
<td></td>
</tr>
<tr>
<td>basisOfRecord</td>
<td>PreservedSpecimen</td>
<td>PreservedSpecimen</td>
<td>PRESERVED_SPECIMEN</td>
<td>Altered</td>
</tr>
<tr>
<td>catalogNumber</td>
<td>INHS Insect Collection 868734</td>
<td>INHS Insect Collection 868734</td>
<td></td>
<td></td>
</tr>
<tr>
<td>class</td>
<td>Arachnida</td>
<td>Arachnida</td>
<td></td>
<td></td>
</tr>
<tr>
<td>coordinateUncertaintyInMeters</td>
<td>10.0</td>
<td>10.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>country</td>
<td>United States</td>
<td>United States</td>
<td>United States of America</td>
<td>Altered</td>
</tr>
<tr>
<td>county</td>
<td>Platt</td>
<td>Platt</td>
<td></td>
<td></td>
</tr>
<tr>
<td>dateIdentified</td>
<td>2018</td>
<td>2018</td>
<td></td>
<td></td>
</tr>
<tr>
<td>day</td>
<td>11</td>
<td>11</td>
<td></td>
<td></td>
</tr>
<tr>
<td>decimalLatitude</td>
<td>39.998473</td>
<td>39.998473</td>
<td></td>
<td></td>
</tr>
<tr>
<td>decimalLongitude</td>
<td>-88.636857</td>
<td>-88.636857</td>
<td></td>
<td></td>
</tr>
<tr>
<td>dwcClass</td>
<td>Arachnida</td>
<td>Arachnida</td>
<td></td>
<td></td>
</tr>
<tr>
<td>endDateOfYear</td>
<td>131</td>
<td>131</td>
<td></td>
<td></td>
</tr>
<tr>
<td>family</td>
<td>Ixodida</td>
<td>Ixodida</td>
<td></td>
<td></td>
</tr>
<tr>
<td>footprintWKT</td>
<td>POINT Z (-88.636857 39.998473 0)</td>
<td>POINT Z (-88.636857 39.998473 0)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>genus</td>
<td>Ixodes</td>
<td>Ixodes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>georeferenceProtocol</td>
<td>A geospatial point translated from verbatim values recorded on human-readable media (e.g. paper specimen label, field notebook).</td>
<td>A geospatial point translated from verbatim values recorded on human-readable media (e.g. paper specimen label, field notebook).</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
TaxonWorks: Tools - making more clouds

Linkages —> GBIFFERENCE

<table>
<thead>
<tr>
<th>Original (GBIF)</th>
<th>Interpreted (GBIF)</th>
<th>Remarks</th>
</tr>
</thead>
<tbody>
<tr>
<td><a href="https://sfg.taxonworks.org/api/v1/images/7eda965762fa28b0ed3a0f7ec24dd10f">https://sfg.taxonworks.org/api/v1/images/7eda965762fa28b0ed3a0f7ec24dd10f</a></td>
<td>PRESERVED_SPECIMEN INHS Insect Collection 868734</td>
<td>Altered</td>
</tr>
<tr>
<td>PreservedSpecimen</td>
<td>INHS Insect Collection 868734</td>
<td></td>
</tr>
<tr>
<td>INHS Insect Collection 868734</td>
<td>Arachnida</td>
<td></td>
</tr>
<tr>
<td>Arachnida</td>
<td>10.0</td>
<td></td>
</tr>
<tr>
<td>10.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>United States</td>
<td>United States of America</td>
<td>Altered</td>
</tr>
<tr>
<td>United States</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Piatt</td>
<td>Platt</td>
<td></td>
</tr>
<tr>
<td>2018</td>
<td>2018-01-01T00:00:00</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>11</td>
<td></td>
</tr>
</tbody>
</table>
TaxonWorks: Tools - making more clouds

Linkages —> Global Identifiers

Identifiers

- Q26494089
- https://orcid.org/0000-0003-0390-4227

Data attributes

- Email: tcm@illinois.edu
- Address: Illinois Natural History Survey; Prairie Research Institute; 1816 South Oak Street, MC 652; Champaign, IL 61820
- Honorific: Dr.
- PeopleID: 1672
- Country: U.S.A.

Alternate values

- "T." alternate spelling of "Thomas Christian"
- "Thomas C." alternate spelling of "Thomas Christian"
- "Thomas" alternate spelling of "Thomas Christian"
- "TC" abbreviation of "Thomas Christian"
- "T.C." abbreviation of "Thomas Christian"
- "Tommy" alternate spelling of "Thomas Christian"
- "T.C." alternate spelling of "Thomas Christian"

Ixodes scapularis 1♂
det ED Struckhoff 2020
GUID: 2396f952-69ab-4bea-9a4b-22e96fecad4f
ex. pant leg; 11MAY2018 TC McElrath/DA Hennen

Ixodes scapularis ♂
det ED Struckhoff 2020

Ixodes scapularis Say, 1821

Black Legged Tick  In English  Collected in United States of America
Animalia > Arthropoda > Arachnida > Ixodida > Ixodidae > Ixodes

Collected By
Thomas McElrath  https://orcid.org/0000-0003-0390-4227
Derek Hennen  https://orcid.org/0000-0001-7005-1151

GBIF
As far as collections boat travels
Cleaning the river

Getting to the ocean

- Importers, importers, importers
- Integrate GUID management & creation into software
- More tools to link persistent identifiers with objects
  - 50-100 people w/GUIDS
  - 1000s without
- Make dataset -> aggregator workflow more seamless (1 button, not multiple)
- Data Validation & Cleaning tools
How to follow the rain?
INHS Insect Collection - the rain

Citations of Papers that cite the INHS Insect Collection (Google Scholar)

- Complete
- In progress
INHS Insect Collection - the rain

- Time from digitization to citation
  - 1 year, 2 months!
- Required physical loan
- Ignored local identifiers, GBIF DOI
- Specimen used previously?
- Can’t track individual specimens
- Limited by loan process/researcher time
- 9 such papers in 2022
  - Citations will increase, but indirectly
- Could be improved if researchers cite general dataset GBIF DOI directly
INHS Insect Collection - the rain

**iDigBio**
Query: dataset ID
1 Record (already found by GBIF)

**Symbiota**
Query: INHS https://biocoll.inhs.illinois.edu/portal
9 records, only 1 insect collection record

**SCAN**
Query: scan-bugs.org INHS
2 non-GBIF tracked insect collection records

Manual or Google Scholar tracked
INHS Insect Collection - the rain

![Graph showing the number of papers citing INHS Insect Collection specimens (GBIF) from 2000 to 2023. The graph indicates an increasing trend with peaks in 2021 and 2023.]

- **In progress**
- **Complete**

Papers that cite INHS Insect Collection specimens (GBIF)
### Specimens Used In

<table>
<thead>
<tr>
<th>Scientific Name</th>
<th>Collected By</th>
<th>Identified By</th>
<th>Date Collected</th>
<th>Date Identified</th>
<th>Institution</th>
<th>Catalog Number</th>
<th>Type Status</th>
<th>Basis Of Record</th>
</tr>
</thead>
</table>
Current and Future Habitat Suitability Models for Four Ticks of Medical Concern in Illinois, USA

by Heather L. Kopsco 1,*, Peg Gronemeyer 2, Nohra Mateus-Pinilla 1,2, and Rebecca L. Smith 1,3

1 Department of Pathobiology, College of Veterinary Medicine, University of Illinois Urbana-Champaign, Urbana, IL 61802, USA
2 Illinois Natural History Survey, Prairie Research Institute, University of Illinois Urbana-Champaign, Urbana, IL 61802, USA
3 Institute for Genomic Biology, University of Illinois Urbana-Champaign, Urbana, IL 61802, USA
* Author to whom correspondence should be addressed.

Insects 2023, 14(3), 213; https://doi.org/10.3390/insects14030213

Received: 23 November 2022 / Revised: 11 February 2023 / Accepted: 13 February 2023 / Published: 21 February 2023

(This article belongs to the Special Issue Surveillance and Control of Vectors of Zoonotic Agents)

391 occurrences included in download

<table>
<thead>
<tr>
<th>INCLUDES RECORDS FROM 7 DATASETS</th>
<th>DOWNLOAD AS TXT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Harold W. Manter Laboratory of Parasitology Collection (HWML) Parasite Collection (Arctos)</td>
<td>4</td>
</tr>
<tr>
<td>New Mexico State Collection of Arthropods</td>
<td>2</td>
</tr>
<tr>
<td>Canadian Museum of Nature Parasite Collection</td>
<td>1</td>
</tr>
<tr>
<td>CHAS Entomology Collection (Arctos)</td>
<td>2</td>
</tr>
<tr>
<td>iNaturalist Research-grade Observations</td>
<td>282</td>
</tr>
<tr>
<td>INSDC Host Organism Sequences</td>
<td>7</td>
</tr>
<tr>
<td>Illinois Natural History Survey Insect Collection</td>
<td>113</td>
</tr>
</tbody>
</table>
Ixodes scapularis 1° det ED Struckhoff 2020

GBIF

Current and Future Habitat Suitability Models for Four Ticks of Medical Concern in Illinois, USA

by Heather L. Kopsco 1, 2, Peg Gronemeyer 2, 3, Nohra Mateus-Pinilla 2, 4, and Rebecca L. Smith 1

1 Department of Pathobiology, College of Veterinary Medicine, University of Illinois Urbana-Champaign, Urbana, IL 61802, USA
2 Illinois Natural History Survey, Prairie Research Institute, University of Illinois Urbana-Champaign, Urbana, IL 61802, USA
3 Institute for Genomic Biology, University of Illinois Urbana-Champaign, Urbana, IL 61802, USA
4 Author to whom correspondence should be addressed.

insects 2023, 14(3), 213; https://doi.org/10.3390/insects14030213

Received: 23 November 2022 / Revised: 11 February 2023 / Accepted: 13 February 2023 / Published: 21 February 2023

(This article belongs to the Special Issue Surveillance and Control of Vectors of Zoonotic Agents)

Download Browse Figures Review Reports Versions Notes

OCCURRENCE | 11 MAY 2018

Ixodes scapularis Say, 1821

Black Legged Tick In English Collected in United States of America Animalia > Arthropoda > Arachnida > Ixodida > Ixodidae > Ixodes

Collected By

Thomas McElrath id https://orcid.org/0000-0003-0390-4227
Derek Hennen id https://orcid.org/0000-0001-7005-1151

GUID: 2396f952-69ab-4bea-9a4b-22e96fecad4f

Ixodes scapularis ♂
det ED
Struckhoff 2020

Total time from collection to citation:
~4 years, 10 months

Discovered/linked via Bionomia/GBIF

Paper Published

Current and Future Habitat Suitability Models for Four Ticks of Medical Concern in Illinois, USA

~01MAR2023

21FEB2023
Ixodes scapularis 1♂
det ED Struckhoff 2020

Collected 11MAY2018
Determined/Imaged 13MAR2020

Bottleneck: Taxonomic impediment & Digitization Funds + 1 year, 10 months
Bottleneck: TaxonWorks Development of DWC Exporter
Fixed!
- 8 months
Bottleneck: none! hypothetically <1 day
- 1 year six months
Bottleneck: publishing
Not within purview of this talk
Collected: 11MAY2023
Determined/Imaged: 12MAY2023
Databased: 13MAY2023
Uploaded: 14MAY2023
Downloaded: 01DEC2023

Total hypothetical time from collection to citation: ~6.5 months

Phyconomus marinus, det TC McElrath, 12MAY2023

Discovered/linked via Bionomia/GBIF: ~14DEC2023

Paper Published: 01DEC2023
Total hypothetical time from collection to citation:

\[\sim 6.5 \text{ months}\]

And all it took was:
- solving the taxonomic impediment
- writing & publishing a paper in <6 months
- persistent identifiers with frictionless data
INHS Insect Collection - the rain

- Exponential growth of use of specimens
- Specimen used previously can be tracked
- Individual specimen can be tracked
- Not limited by collection manager time - only researcher
- 53(54) papers in 2023 so far
  - On track to exceed 2022 by 15 papers
- Collection Managers do NOTHING to get citations of specimens
Phycomonomus marinus
det TC McElrath 12MAY2023

GBIF

GBIF

Discovered/linked via Bionomia/GBIF

~14DEC2023

Paper Published

01DEC2023
Collected 11MAY2023

Determined/Imaged 12MAY2023

Uploaded 13MAY2023

Downloaded 14MAY2023

Phyconomus marinus
det TC McElrath
12MAY2023

No DOI to connect specimens to paper
Manual tracking
Time to discovery? 1 day - years
Back to old system

Paper Published 01DEC2023
Phyconomus marinus
det TC McElrath
12MAY2023

Discovered/linked via Bionomia/GBIF

~14DEC2023

Determined/Imaged
12MAY2023

Uploaded
13MAY2023

Downloaded
14MAY2023

Paper Published
~14DEC2023

01DEC2023
Collections/Datasets buried in:
- Materials/Methods (okay, but not great)
- Acknowledgments (worse, but still in main paper)
- Supplemental files (groan)
- Completely stripped (inexcusable)
<table>
<thead>
<tr>
<th>DOI Tracking?</th>
<th>Global vs. Local</th>
<th>Specimen level tracking?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yes</td>
<td>Global</td>
<td>Yes</td>
</tr>
<tr>
<td>Yes</td>
<td>Local</td>
<td>Yes</td>
</tr>
<tr>
<td>No</td>
<td>Global/Local</td>
<td>No</td>
</tr>
<tr>
<td>No</td>
<td>Local</td>
<td>No</td>
</tr>
</tbody>
</table>
Keep the cycle going

- CMs provide citation link alongside loaned specimens
- Publishers REQUIRE data management plan
- Grounds for Rejection/Revision
  - “data on request from the authors”
  - ONLY aggregator cited, not individual datasets
  - No verbatim data anywhere (not repeatable!)
  - Identifiers stripped from primary data everywhere

---

Data is not available upon request

Ian Hussey

Many journals now require data sharing and require articles to include a Data Availability Statement. However, several studies over the past two decades have shown that promissory notes about data sharing are rarely abided by, and that data is generally not available upon request. This has negative consequences for many essential aspects of scientific knowledge production, including independent verification of results, efficient secondary use of data, and knowledge synthesis. Here, I assessed the prevalence of data sharing upon request in articles employing the Implicit Relational Assessment Procedure published within the last 5 years. Of 52 articles, 42% contained a Data Availability Statement, most of which stated that data was available upon request. This rose from 0% in 2018 to 100% in 2022. Only 25% of articles’ authors actually shared data upon request. Among articles stating that data was available upon request, only 17% shared data upon request. The presence of Data Availability Statements was not associated with higher rates of data sharing ($p = .80$). Results replicate those found elsewhere: data is generally not available upon request, and promissory Data Availability Statements are typically not adhered to. Issues, causes, and implications are considered.

https://psyarxiv.com/jbu9r/
Replenish the water table

Keep the cycle going

**ECOLOGICAL MONOGRAPHS**

**Do Nearctic hover flies (Diptera: Syrphidae) engage in long-distance migration? An assessment of evidence and mechanisms**

C. Scott Clem, Keith A. Hobson, Alexandra N. Harmon-Threatt

First published: 19 July 2022 | [https://doi.org/10.1002/ecm.1542](https://doi.org/10.1002/ecm.1542) | Citations: 2

---

**A century of Illinois hover flies (Diptera: Syrphidae): museum and citizen science data reveal recent range expansions, contractions, and species of potential conservation significance**

C. Scott Clem, Lily V Hart, Thomas C McElrath

*Journal of Insect Science, Volume 23, Issue 4, July 2023, 13,*

[https://doi.org/10.1093/jisesa/iead051](https://doi.org/10.1093/jisesa/iead051)

Published: 03 August 2023 | Article history ▼
Replenish the water table

Caveats

• Building infrastructure costs $$
  • GBIF DOI registry
  • Server time/space costs $$
  • Data infrastructure development
• Data infrastructure not perfect
  • still not seeing my paper that cited three different GBIF DOIs on GBIF website
• More aggregators = more exposure
Acknowledgements

- Species File Group
  - Matt Yoder
  - Debbie Paul
  - Hernán Lucas Pereira
  - José Luis Pereira
  - Debbie Paul
- Funding:
  - Species File Group
  - Ed DeWalt
  - TaxonWorks community
    - Michelle Kohler
    - Lily Hart
    - Brian Fisher
    - Kojun Kanda
    - Gareth Powell
- Matt Gimmel
- Heidi Hopkins
- Chris Grinter
- Student workers
- Volunteers
- Symbiota Community
  - Katie Pearson
  - Jenn Yost
Provide & Cite DOIs
Let the water flow