

Describing the Earth's Species: Semantics, Interfaces and Communities form a Grand Challenge for FAIR Publishing

Matt Yoder,
Illinois Natural History Survey
orcid.org/0000-0002-5640-5491

Twitter: @mjyoder



Who is the Species File Group?



Deborah Paul



Dmitry Dmitriev



Yuri Ruskov



Dmitry Mozzherin



Rich Flood



Geoff Ower



R. Edward DeWalt



José Luis Pereira



Hernán Lucas Pereira



Matt Yoder

<https://speciesfilegroup.org/about.html>

Overview

Introduce biodiversity informatics from a taxonomist's perspective, and outline the scope of some of our challenges.

Introduce TaxonWorks, some of its design, goals, functions, and aspirations.

Discuss how the broader community might interact with our work, and reflect on what the FAIR digital framework is asking of us.



Taxonomist

Describes the Earth's biodiversity (e.g. species) and the evolutionary relationships among it.

Foundation

Physical specimens preserved in natural history collections.

A taxonomist's job

Find, collect, and (re)describe millions of species.

Digitize the past. There are millions of specimens in collections not digitized.

Describe new anatomy, or molecular data using natural language, figures, images, DNA, and all sorts of other data.

Reconcile their taxon concept against everything ever published, a 250 year citation window.

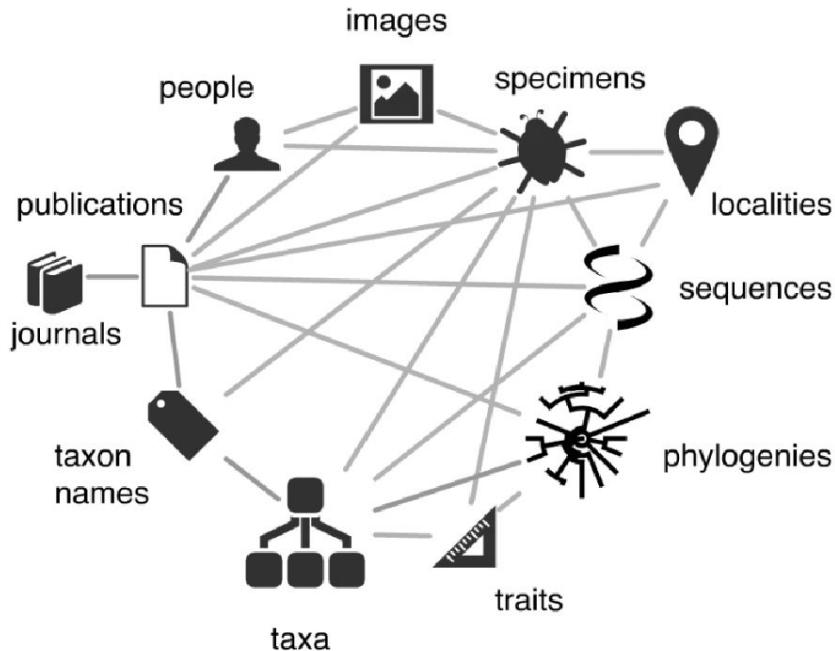
Mint, or re-use a name for their concept.

Publish.

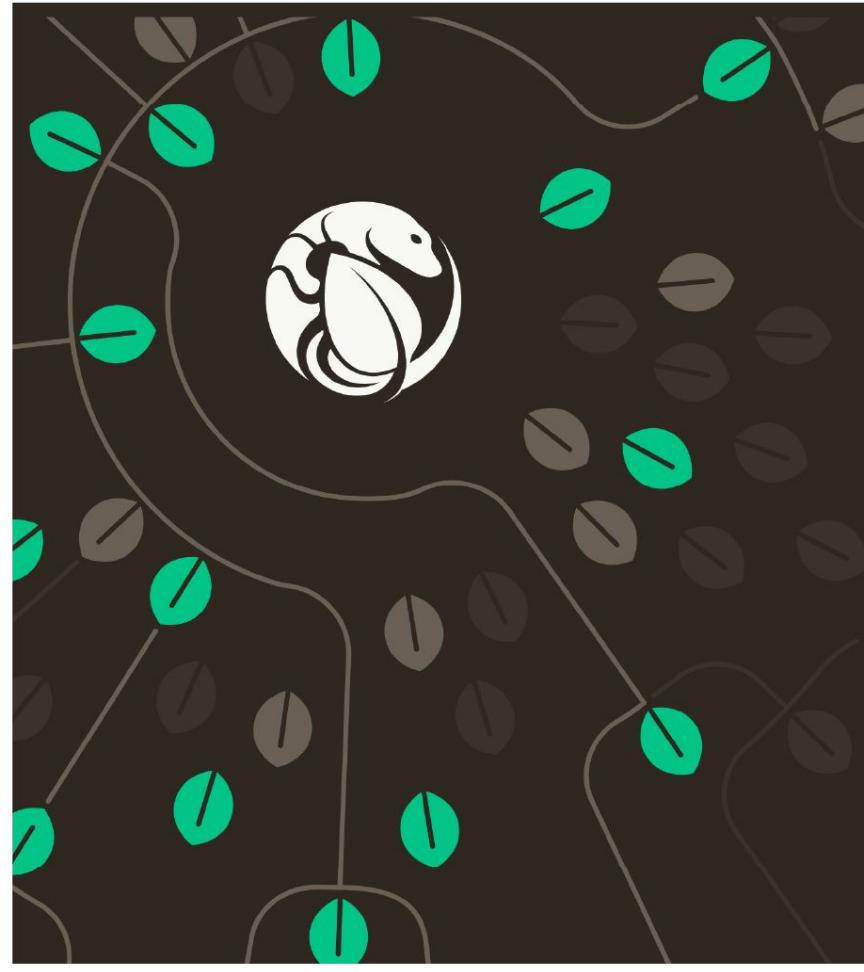
Example: DOI: 10.3897/zookeys.380.5755

... a graph of life

Rod Page, 2015



<https://iphyllo.blogspot.com/>



TaxonWorks

Describe life.

TaxonWorks is an integrated web-based workbench for taxonomists and biodiversity scientists.

It allows you to capture, organize, and enrich your data; share it with collaborators; and package it for analysis and publication.

Twitter: @TaxonWorks | <https://taxonworks.org>



TaxonWorks

Software

Open source
Rails
Vue.js, Javascript
Postgresql/PostGIS
Unit tests
CI builds
ORM - Ontology influenced, graph aware

Service

SFG - Ansible, Docker, Kubernetes

Features

Project-based workbench
Supports individuals, small teams
52 classes of biodiversity data
9 classes of annotations
“Tasked” based UI
Various batch imports (e.g. BibTeX)
Numerous “Domain” exports (e.g. DwC Archive)
JSON API - api.taxonworks.org





Filter sources

Source hub

 Show filter Show JSON Request

Filter

Reset

Select all | Download CSV | Download formatted | BibTeX

1

1 - 15 of 15 records. 500 records per page.

Tag	Id	Cached	Year	Type	Documents	In project
<input type="checkbox"/>	49315	Abraham, R. & Joswig, W. (1985) Parasitization of fly puparia from dead snails by <i>Spalangia erythromera</i> (Hymenoptera: Pteromalidae) and <i>Basasyls semele</i> (Hymenoptera: Diapriidae). <i>Spixiana, München</i> 8, 285–287.	1985	Source::Bibtex		
<input type="checkbox"/>	49558	Aguilar-Menezes, E.L., Menezes, E.B. & Loiacono, M.S. (2003) First record of <i>Coptera haywardi</i> Loiacono (Hymenoptera: Diapriidae) as a parasitoid of fruit-infesting Tephritidae (Diptera) in Brazil. <i>Neotropical Entomology</i> 32(2), 355–358.	2003	Source::Bibtex		
<input type="checkbox"/>	50190	Amini, A., Sadeghi, H., Lofalizadeh, H. & Notton, D. (2014) Parasitoids (Hymenoptera: Pteromalidae, Diapriidae) of <i>Carponyx vesuviana</i> Costa (Diptera: Tephritidae) in South Khorasan province of Iran. <i>Biharean Biologist</i> 8(2), 122–123.	2014	Source::Bibtex		
<input type="checkbox"/>	54284	Bin, F. (1972b) Ritrovamento dell'ospite e descrizione del maschio di <i>Basalsyls helicola</i> (Kieff.) (Hymenoptera, Proctotrupoidea, Diapriidae). <i>Entomologica, Bari</i> 8, 35–44. [1972]	1972	Source::Bibtex		
<input type="checkbox"/>	113740	Coon, B.R., Harms, N.E., Cuda, J.P. & Grodowitz, M.J. (2014) Laboratory biology and field population dynamics of <i>Trichopria columbiana</i> (Hymenoptera: Diapriidae), an acquired parasitoid of two hydilla biological control agents. <i>Biocontrol Science and Technology</i> 24, 123–1264. https://doi.org/10.1080/13652427.2014.933311	2014	Source::Bibtex		
<input type="checkbox"/>	65502	Guillén, L., Aluja, M., Equihua, M. & Sivinski, J. (2002) Performance of two fruit fly (Diptera: Tephritidae) pupal parasitoids (<i>Coptera haywardi</i> [Hymenoptera: Diapriidae] and <i>Pachycerepeodes vindimiae</i> [Hymenoptera: Pteromalidae]) under different environmental soil conditions. <i>Biological Control</i> 23(3), 219–227.	2002	Source::Bibtex		
<input type="checkbox"/>	74349	Kazimírová, M. & Vallo, V. (1992) Influence of larval density of Mediterranean fruit fly (<i>Ceratitis capitata</i> , Diptera, Tephritidae) on parasitism by a pupal parasitoid, <i>Coptera occidentalis</i> (Hymenoptera, Proctotropoidea, Diapriidae). <i>Acta Entomologica Bohemoslovaca</i> 89, 179–185.	1992	Source::Bibtex		
<input type="checkbox"/>	76041	Kühlhorn, F. (1962) Über parasitische Hautflügler in Viehställen (Ichneumonidae, Braconidae, Diapriidae, Proctotropidae, Eulophidae und Pteromalidae). <i>Zeitschrift für Angewandte Zoologie</i> 49, 525–538.	1962	Source::Bibtex		
<input type="checkbox"/>	82253	Monteiro, M.R. & Prado, A.P.dO (2000) <i>Trichopria</i> sp. (Hymenoptera: Diapriidae) attacking the pupae of <i>Chrysomya putoria</i> (Wiedemann) (Diptera: Calliphoridae) in a poultry facility. <i>Anais da Sociedade Entomológica do Brasil</i> 29(1), 159–167.	2000	Source::Bibtex		
<input type="checkbox"/>	85243	Ogloblin, A.A. (1957b) Los insectos de las Islas Juan Fernandez. 35. Myrmidae, Ceraphronidae, Diapriidae y Sclerionidae (Hymenoptera). <i>Revista Chilena de Entomología</i> 5, 413–444.	1957	Source::Bibtex		
<input type="checkbox"/>	114239	Shimbori, E.M., Costa, V.A. & Zucchi, R.A. (2020) Annotated checklist and illustrated key to parasitoids (Hymenoptera: Diapriidae, Eulophidae and Pteromalidae) of fruit flies (Diptera, Tephritidae) in Brazil. <i>Zootaxa</i> 4858, 53–70. https://doi.org/10.11646/zootaxa.4858.1.3	2020	Source::Bibtex		

Rich bibliography management



Manage controlled vocabulary

Keyword	Topic	Predicate	ConfidenceLevel	BiocurationGroup	BiocurationClass
---------	-------	-----------	-----------------	------------------	------------------

A keyword is the value used in tag

Name	<input type="text"/>
Definition	<p>Definition (minimum length 20 characters)</p> <input type="text"/>

Name	Definition	Uses	Show	Edit	Pin	Destroy
Beatty Spider Project	This is for specimens photographed during the Beatty Spider Project.		Show			
body part (not genitalia)	Dissected body part (not genitalia).		Show			
Check transcription	This tag is for specimens who have doubtful or confusing verbatim label data as transcribed, that cannot lead to a verified location or georeference. The specimens should have the label data re-checked by someone with more knowledge.		Show			
Collecting Event to be filled out	This is for Collecting Events that have identifiers, but no associated data. Generally, the data is being sought out.		Show			
Collection Objects to Review	This is for Collection Objects to review later		Show			
Duplicate	This is used for specimens that are potentially true duplicates. These should be examined and matched to determine if one is possibly better than the other specimens and determining which is the true specimen, and if true duplicates, should be reassigned catalog numbers.		Show			
Emily	Emily's Salticidae Project		Show			
Flipped images	This is for specimens that were accidentally uploaded upside down into the SQED image processing filter.		Show			



An Interactive Key to Species of *Erasmoneura* and *Rossmoneura*

[View](#) | [Observation matrix hub](#) | [Edit](#) | [TNT](#) | [nexus](#)
[Download](#) | [TNT](#) | [nexus](#) | [NeXML](#) | [OTU contents](#)

Code	OTU	Ground color of dorsum	Coloration of vertex	Coloration of vertex midline	Anteclypeus coloration	Pronotum color pattern	Color pattern of mesonotum	Coloration of thoracic venter	Presence of crossbands on forewings	Presence of oblique pattern on forewing	Presence of a brown spot on forewing inner apical cell	Presence of a brown spot in forewing apical cell II	Clavus color pattern	Crown fore margin	Length of forewing outer apical cell	Length of 2 nd abdominal appendages	Width of aedeagus	Width of aedeagus shaft in lateral view	Position and length of aedeagus distal processes	Position and length of aedeagus proximal processes	Shape of pygofer dorsal appendage
Code	OTU	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1	<i>Erasmoneura atra</i> (Johnson, 1935)	3	6	0	0	3	2	1	0	0	1	3	0	0	0	1	2	0	8	2	0
2	<i>Erasmoneura bipentagona</i> (Beamer, 1927)	0	0.3	0	0	3	1	1	0	2	0	1	3	0	1	1	3	1	0	1	0
3	<i>Erasmoneura caerulea</i> (Beamer, 1937)	3	6	1	0	3	2	1	0	4	0	0	3	0	0	0	1	2	0	8	2
4	<i>Erasmoneura emelianovi</i> Dmitriev & Dietrich, 2007	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	2	0	8	1
5	<i>Erasmoneura fulmina</i> (McAtee, 1920)	0.2	6	0	1	3	12	2	0.1	0	0	1	3	0	1	0	1	3	1	0	4
6	<i>Erasmoneura latiloba</i> Dmitriev, 2008	1	0	0	1	3	2	2	0	4	0	0	3	0	1	1	1	3	1	8	0
7	<i>Erasmoneura margaritae</i> Dmitriev & Dzubanov, 2007	1	6	0	0	3	5	1	0	4	0	0	3	0	1	1	1	0	8	0	0
8	<i>Erasmoneura mixta</i> (Beamer, 1932)	0.1	0	0	0.1	0.3	0.2	0.2	0.7	0.4	0	0	0.3	0	1	0	1	2	0	8	0
9	<i>Erasmoneura nigerrima</i> (McAtee, 1920)	3	6	0	0	3	2	1	0	4	0	0	3	0	1	0	0	3	1	8	2
10	<i>Erasmoneura nigra</i> (Gillette, 1898)	3	6	0	0	3	2	1	0	0	0	1	3	0	1	0	1	2	1	8	2
11	<i>Erasmoneura rubricata</i> (Van Duzee, 1909)	0	0	0	0	0	0	0	0	0.4	0	0	0.3	0	1	0	1	3	1	8	0
12	<i>Erasmoneura tricuspidata</i> Dmitriev & Zahniser, 2017	0	6	0	0	3	2	1	0	2	0	1	3	0	1	1	1	3	1	4	0
13	<i>Erasmoneura variabilis</i> (Beamer, 1929)	1	6	0	1	3	2	1	0	0	0	1	3	0	1	0	1	3	1	0	4
14	<i>Erasmoneura vulnerata</i> (Fitch, 1851)	1	6	0	1	3	2	1	0	0	0	1	3	0	1	0	1	3	1	0	4
15	<i>Rossmoneura calva</i> (Beamer, 1946)	0	3	0	1	2	1	1	0	2	0	0	1	1	1	0	2	2	1	6	1
16	<i>Rossmoneura carbonata</i> (McAtee, 1920)	0	13	0	1	2	1	1	0	0.2	0	1	13	1	1	1	2	2	1	6	4
17	<i>Rossmoneura tecta</i> (McAtee, 1920)	0	13	0	1	2	1	1	0	0.2	1	1	1	1	1	0	2	2	1	4	1

Build observation matrices

Search 9813a8a6-45dd-448b-a9a6-fbaa700006093 det. *Psychoda phalaenoides* Linnaeus, 1758 ✓ [Spe]

« Id < Identifier > Id > Identifier > Recent Save Save and new Reset

Object details

Total Biocurations

LifeStage: Immature, Pupae, Exuvia, Egg(s), Adult

Sex: Unknowable Sex, Gynandromorph, Male, Female

Add to container

Digitize specimens

Accession metadata

Determinations

OTU: Quick, Recent, Pinboard, Search

Select an OTU Echinorrhynchida ✓

Collecting Event

Selector: Pinboard, Search

Terrestrial Parasite Tracker
Photographed by: Noah Seo
Date: 10/11/2021

Pinboard, Search

Verbatim

Label: Search

Clone from spec

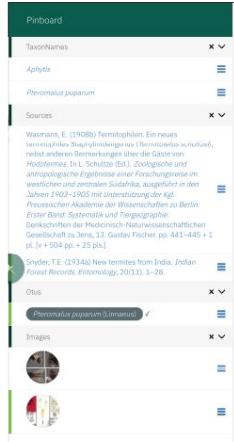
Locality: Provide verbatim latitude/longitude to preview location on map.

Year, Month, Day: Now

Add Determination

Interfaces

There's a lot of semantics, how can we help?



Pinboard

Smart selectors

A screenshot of a software interface. At the top are tabs: 'Common' (selected), 'Advanced', and 'Show all'. Below them is a list of taxon types: 'Unavailable' (radio button), 'Nomen nudum' (radio button), 'Nomen dubium' (radio button), and 'Fossil' (radio button). A red arrow points from the 'Soft Validation' dialog box to the 'Nomen nudum' radio button. The 'Soft Validation' dialog box contains the following text:

Soft Validation

Taxon name [Fix all](#)

- ⚠ Etymology is missing
- ⚠ Missing relationship: Original genus is not selected
- ⚠ Missing relationship: Type species is not selected
- ⚠ Gender is not specified (possible gender is masculine)
- ⚠ Taxon name author role is not selected
- ⚠ Original publication is not selected

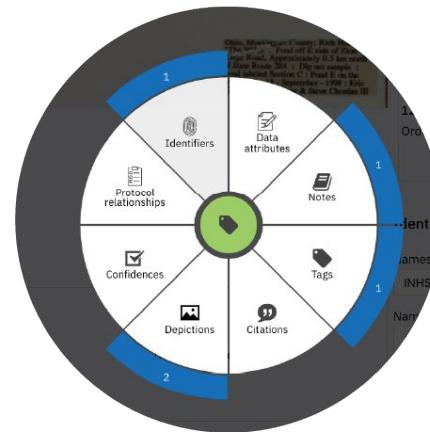
[Fix](#) Cached values should be updated

Soft validation

Cloning, recent, smart navigation, hot-key bindings, field locks...

UI/UX in the context of research: <https://bioip.github.io/>

Radials



<https://github.com/SpeciesFileGroup/svg-detailer>



Uniquify people

Filter

Search

Update match people

Referenced in project

Any
 Current

Person

Name
 Wildcard

Last name
 Wildcard

First name
 Wildcard

Active year

Before After

Born year

Before After

Died year

Before After

Levenshtein cutoff

Select person

25 people found

Search a person...					
	Cached	Lived	Active	Used	Role
<input checked="" type="checkbox"/> Selected	DeGeer, Carl	1720 - 1778	1741 - 1783		?

Match people

24 matches found

Search a person...				
	Cached	Lived	Active	Used
<input type="checkbox"/>	Alam, Mohammad Zahangeer	[? - ?]	[1964 - 2015]	[?]
<input type="checkbox"/>	Alam, Mohammad Zahangeer	[? - ?]	[? - ?]	[?]
<input type="checkbox"/>	de Geer	[? - ?]	[? - ?]	[?]
<input checked="" type="checkbox"/>	de Geer	[? - ?]	[? - ?]	[?]
<input type="checkbox"/>	de Geer	[? - ?]	[? - ?]	[?]
<input type="checkbox"/>	de Geer	[? - ?]	[? - ?]	[?]

Selected person

This person will remain.

Person to merge

 This person(s) will be deleted.

DeGeer, Carl		De Geer
Id	88372	125922
Type	Person::Vetted	Person::Unvetted
Last name	DeGeer	De Geer
First name	Carl	
Suffix		
Prefix		
Cached	DeGeer, Carl	De Geer
Year born	1720	
Year died	1778	
Year active start	1741	
Year active end	1783	
Created by id	278	281
Updated by id		281
Created at	2020-06-11T16:29:15.336Z	2021-03-04T17:17:17.208Z

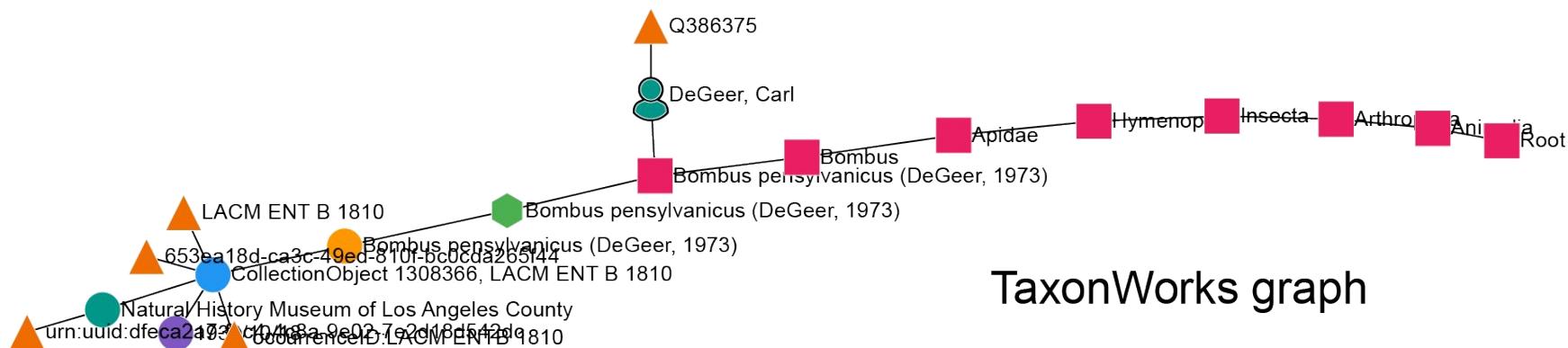
Use case: Getting to FAIR - from rows to Things

TDWG standard

Darwin Core Archive

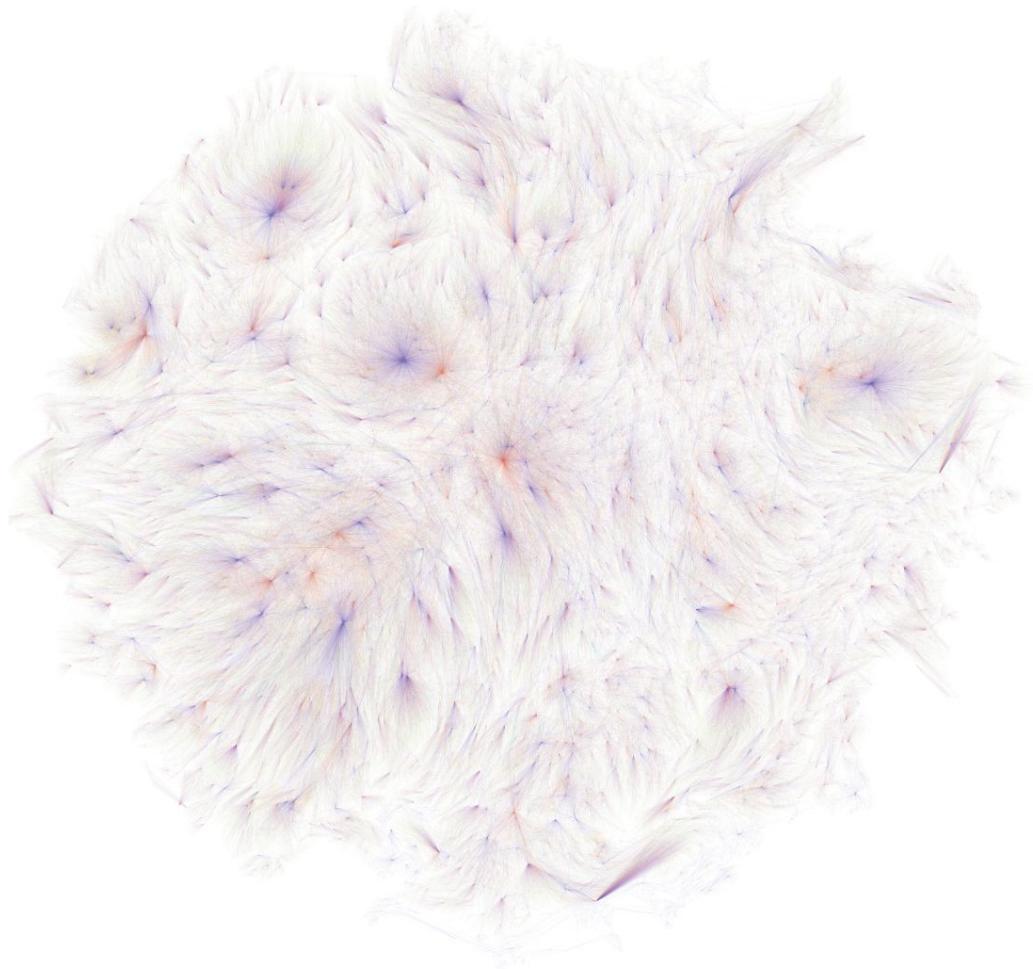
(

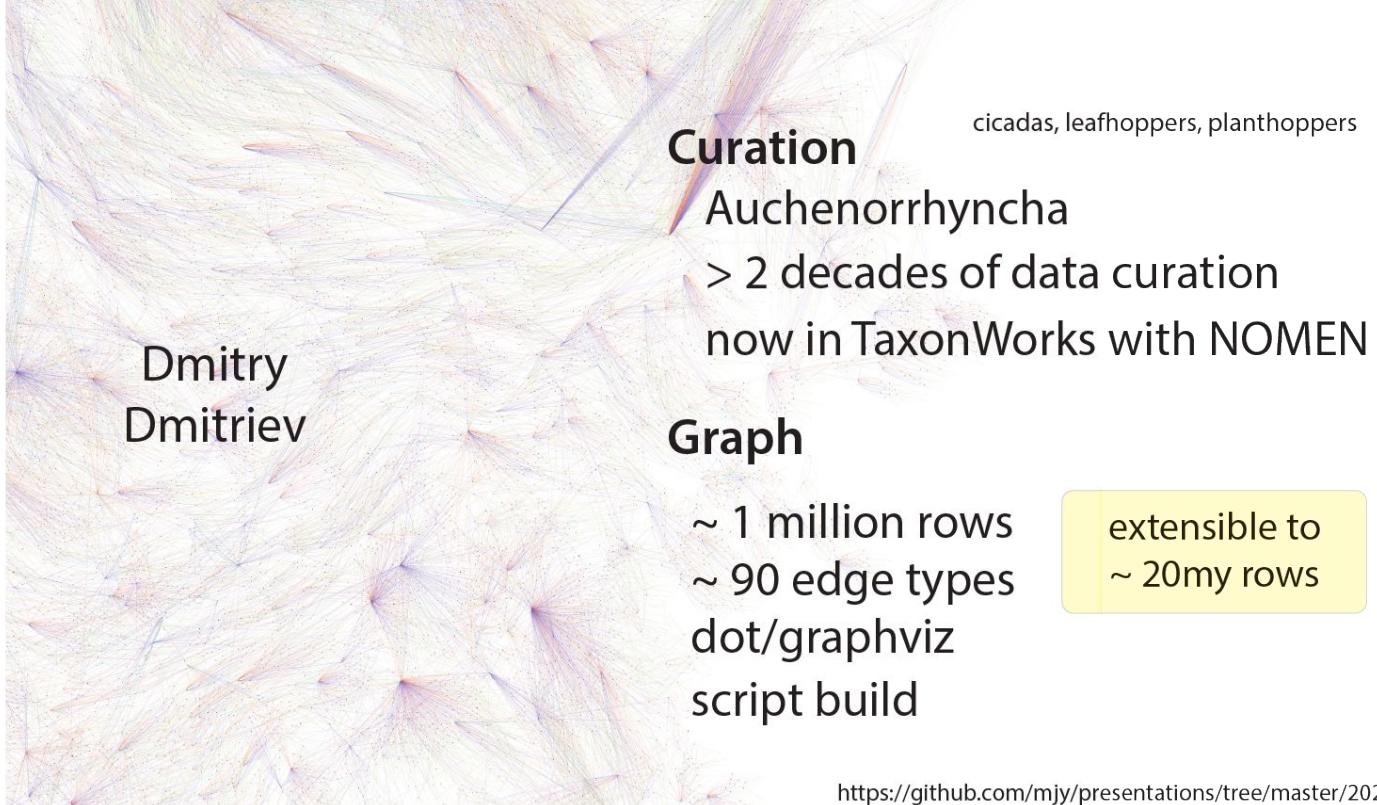
```
occurrences.csv + (~/Downloads/UCRC-ENT_DwC-A_Yoder_Sample) - VIM
1 id,institutionCode,collectionCode,ownerInstitutionCode,collectionID,basisOfRecord,occurrenceID,catalogNumber,otherCatalogNumbers,kingdom,
1 38999529,LACM,ENT,,3bf6519b-f105-4cab-bc9a-81f7cacca37,PreservedSpecimen,"LACM ENTB 1810","LACM ENTB 1810",,Animalia,Arthropoda,Insecta,,
2
```



TaxonWorks graph

Nomenclature as a graph





Dmitry
Dmitriev

Curation

cicadas, leafhoppers, planthoppers

Auchenorrhyncha

> 2 decades of data curation

now in TaxonWorks with NOMEN

Graph

~ 1 million rows

~ 90 edge types

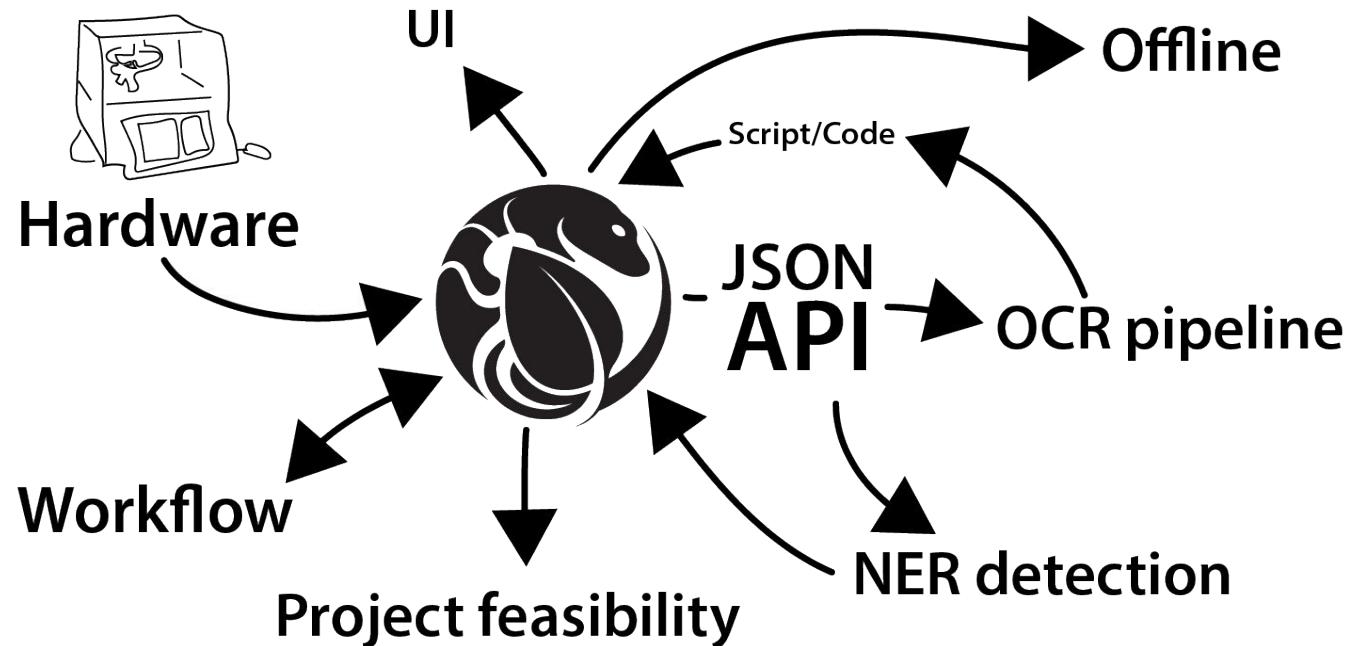
dot/graphviz

script build

extensible to
~ 20my rows

<https://github.com/mjy/presentations/tree/master/2021>

Taxon Works as a hub



Collaborate

Tippy-top	Meta-academics, e.g. studies “in-situ” within TaxonWorks
A little higher	Semantic cross-mapping (instances, or classes) AI training Gold standards Visualization
Low hanging fruit (but still up there)	Open-source improvements Wikidata integration API improvements API wrappers Documentation Jupyter notebook or similar integration

FAIRly burnt out

How findable is “F”?

How accessible is “A”?

How interoperable is “I”?

How reusable is “R”?

*Providing achievable milestones is critical.
Prototyping in a “real” tool is rewarding.*



TaxonWorks Together

December 6th-11th

speciesfilegroup.org/events.html

Weekly meetings

Species File Group Community Liaison

Debbie Paul - dlpaul@illinois.edu

