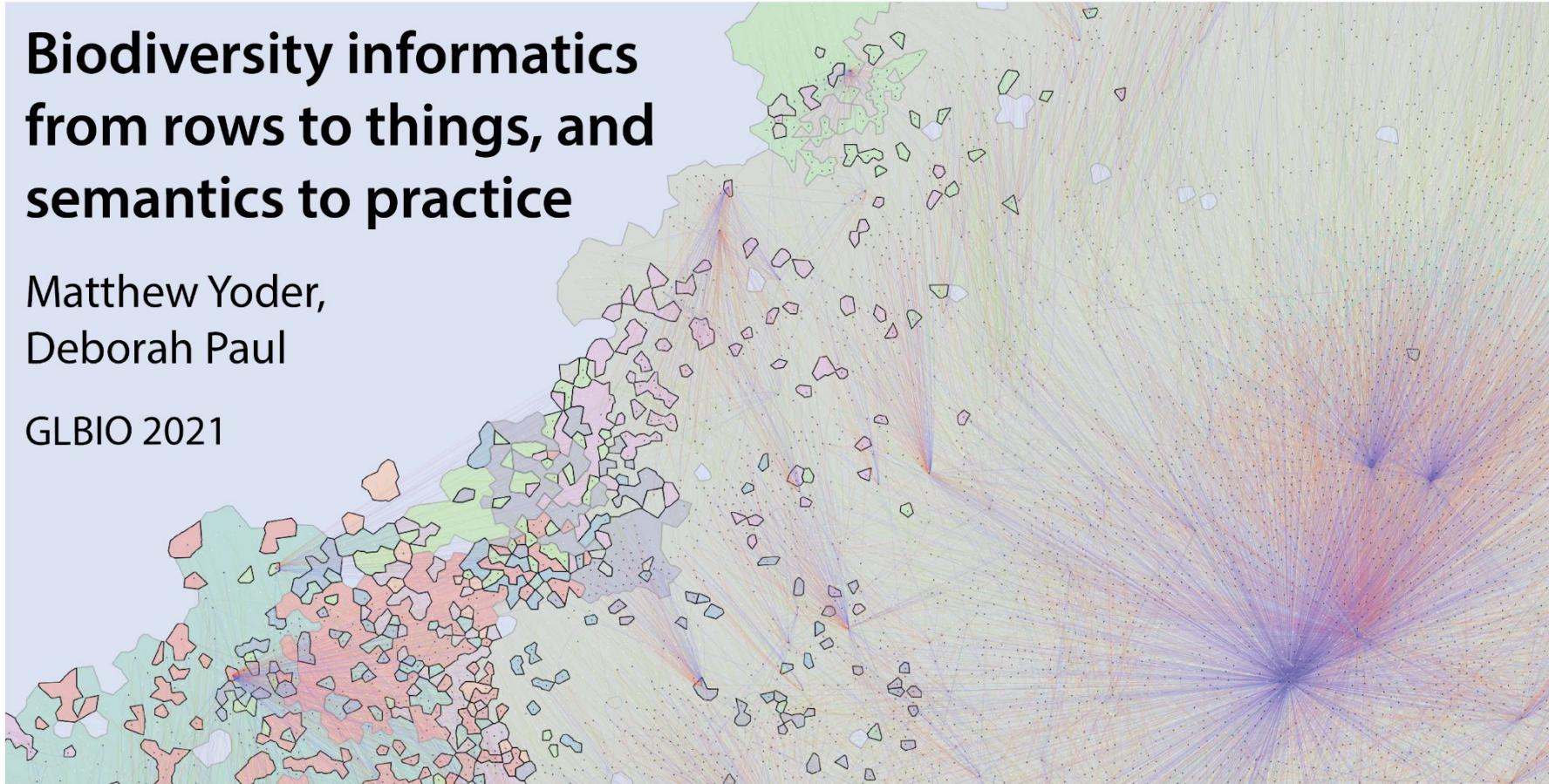


Biodiversity informatics from rows to things, and semantics to practice

Matthew Yoder,
Deborah Paul

GLBIO 2021



Overview

Why biodiversity (informatics)?

Species File Group

Research and Development opportunities

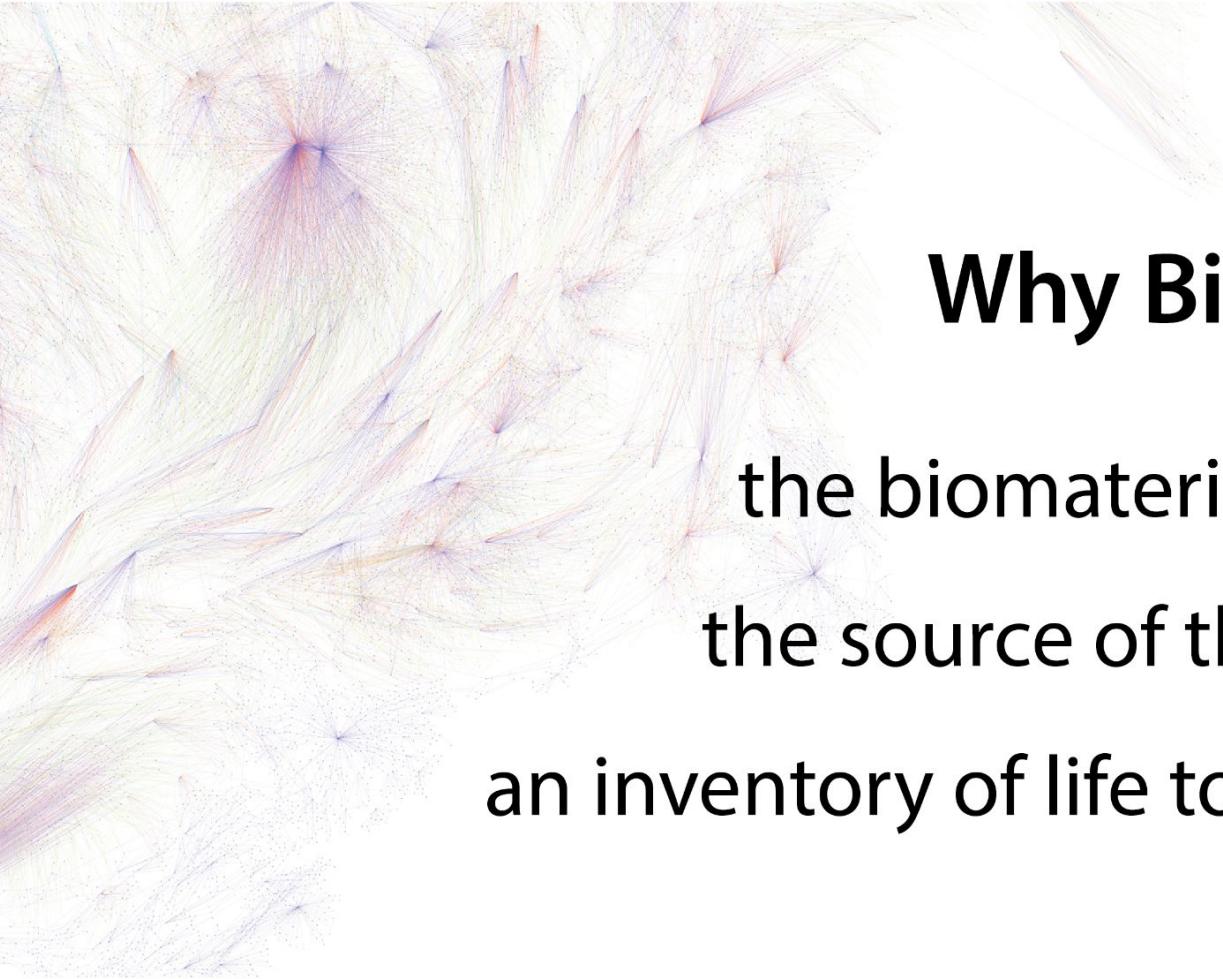
Taxon names and nomenclature

Interlude - Identifiers

Informatics use-cases

Interlude - Rows to things

Driven by questions



Why Biodiversity?

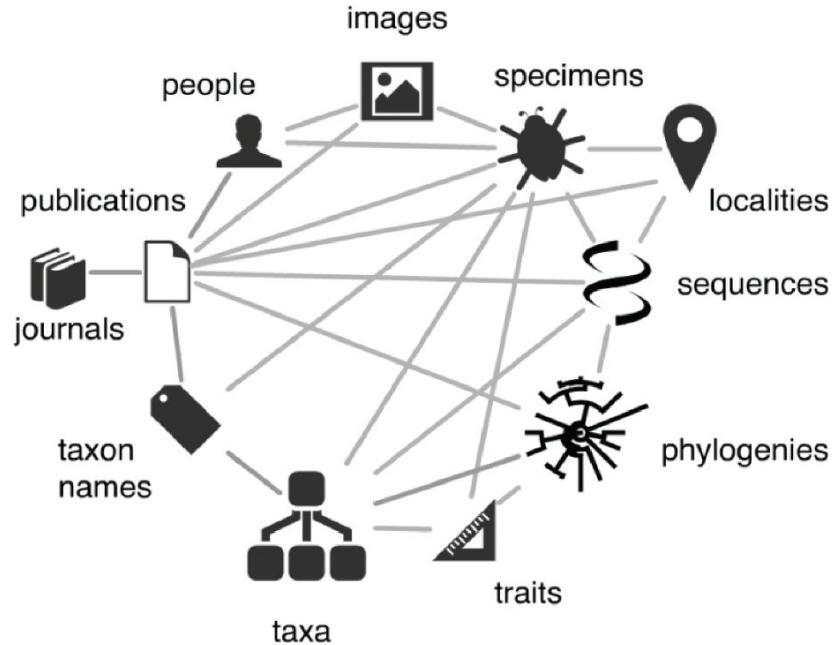
the biomaterial to save a life

the source of the coronavirus

an inventory of life to bring to Mars

...a graph of life

Rod Page, 2015



Species File Group

Research and development opportunities in informatics

Interfaces

Collaborations

Algorithms

Ontologies

Outreach



Global Names Architecture



HOME

NEWS

APPS

WORKSHOPS

PAPERS

DOCS

SOURCE

Global Names Architecture

Scientific Names Services

The Global Names Architecture (GNA) is a system of web-services which helps people to register, find, index, check and organize biological scientific names and interconnect on-line information about species.

Questions?

Drop a note at Global Names' Gitter chatroom.

[Chat on Gitter →](#)

Work with Names

[Find →](#)

[Verify →](#)

[Parse →](#)

Register Names

If you are a zoologist -- you can officially register species electronically.

[Zoobank →](#)



The content of this website is in public domain

Powered by Jekyll

Hosted on [GitHub](#)

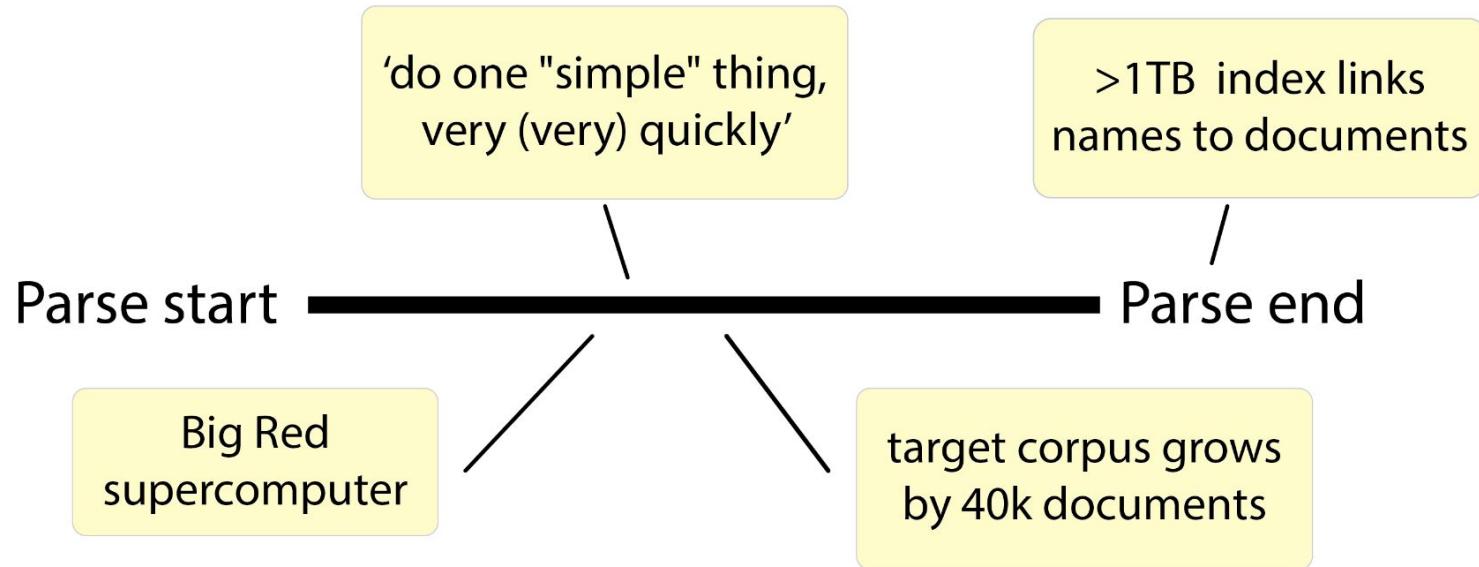
<https://globalnames.org/>





Global Names + HathiTrust

Find names in ~10% of everything ever published



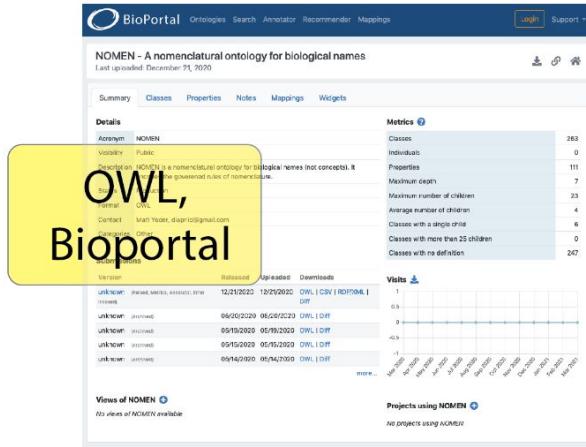
see also BHL

<https://wiki.htrc.illinois.edu/display/COM/GlobalNames+and+the+HathiTrust>

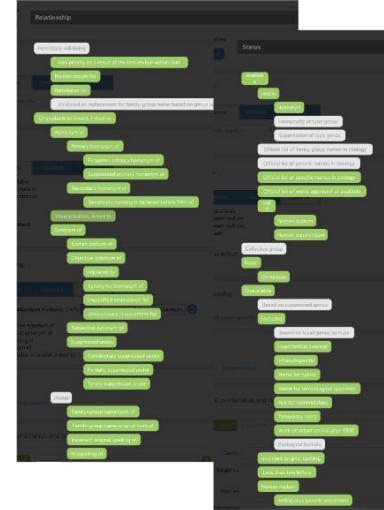


Names + meaning: NOMEN

ontology encoding
rules of nomencalture



Facilitates
UI and data
persistence



TaxonWorks

opportunities: reason, refine, integrate

<https://github.com/SpeciesFileGroup/nomen>

Catalogue of Life



Yuri Ruskov
Geoff Ower

community support



TaxonWorks

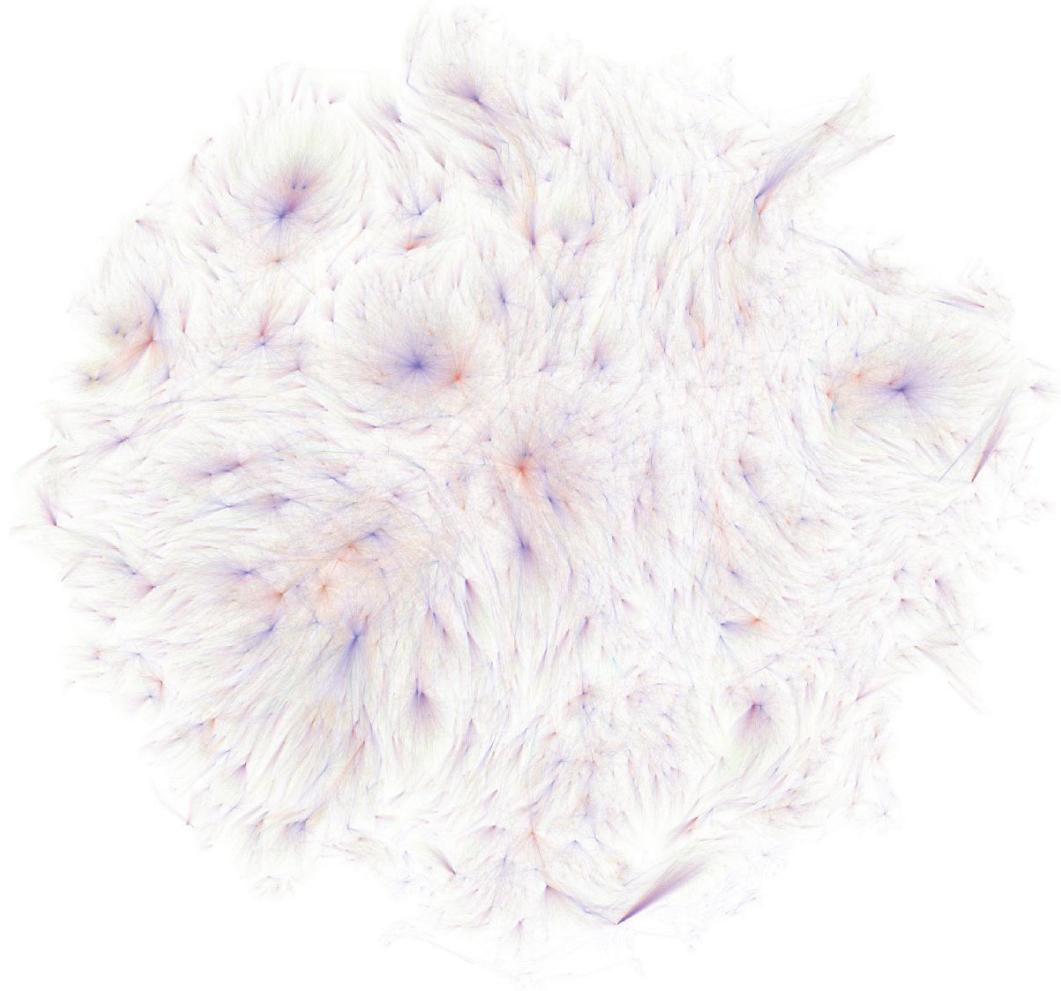
The screenshot shows a detailed view of a species record in the TaxonWorks Species File. At the top, there's a navigation bar with links like 'Species File', 'Search', 'New', 'Recent', 'Type', 'Key', 'Help', and 'Wiki'. Below the navigation is a search bar with the placeholder 'Search' and a magnifying glass icon. The main content area displays a species card for 'ILLIESIA'. The card includes a thumbnail image of a specimen, the genus name 'ILLIESIA', and a brief description: 'Some 15 species have been described from the Americas, mostly from Mexico and Central America. The type species is I. longistylis.' There are also sections for 'Assessments' and 'References', each with a small thumbnail image.

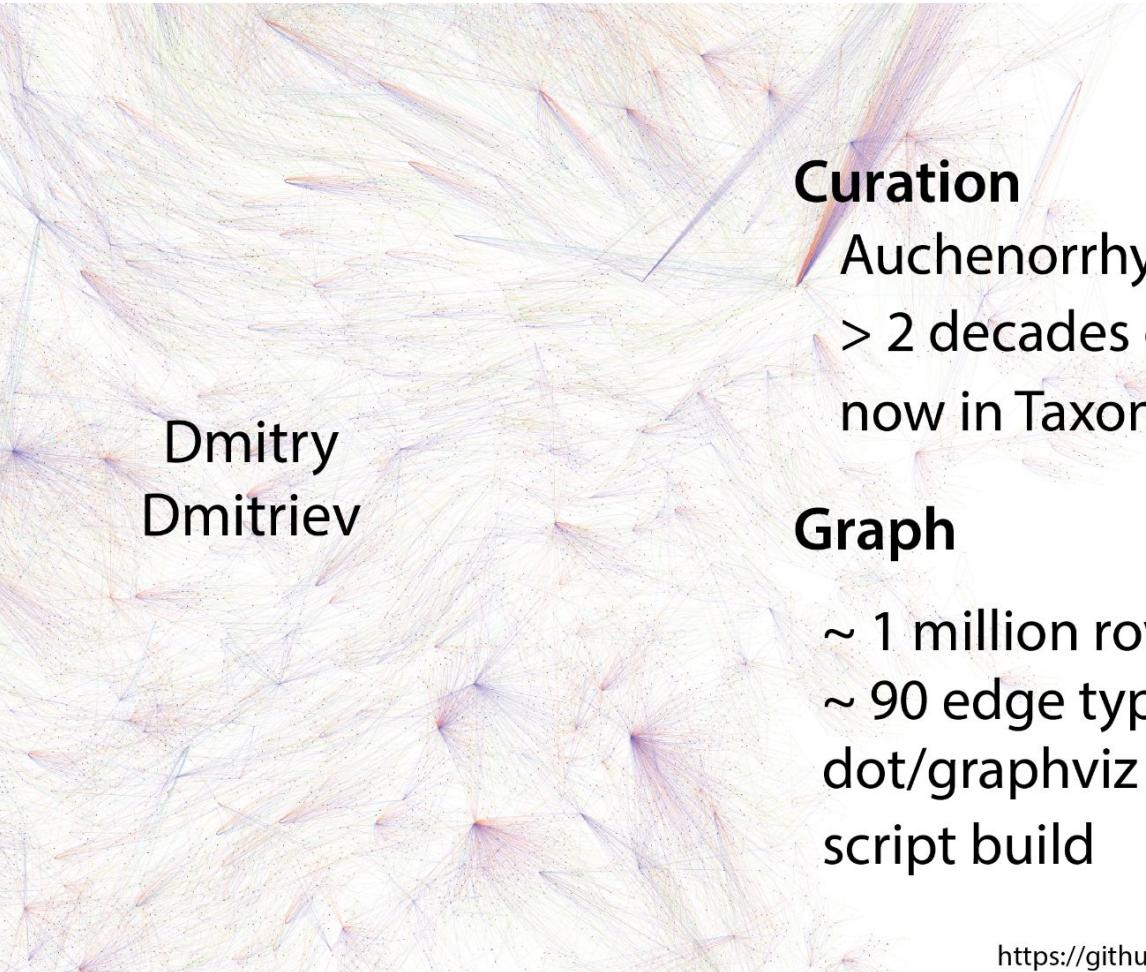
Species Files

16 contributing projects

export
CoL-DP format

Nomenclature as a graph





Curation

Auchenorrhyncha

> 2 decades of data curation

now in TaxonWorks with NOMEN

Graph

~ 1 million rows

~ 90 edge types

dot/graphviz

script build

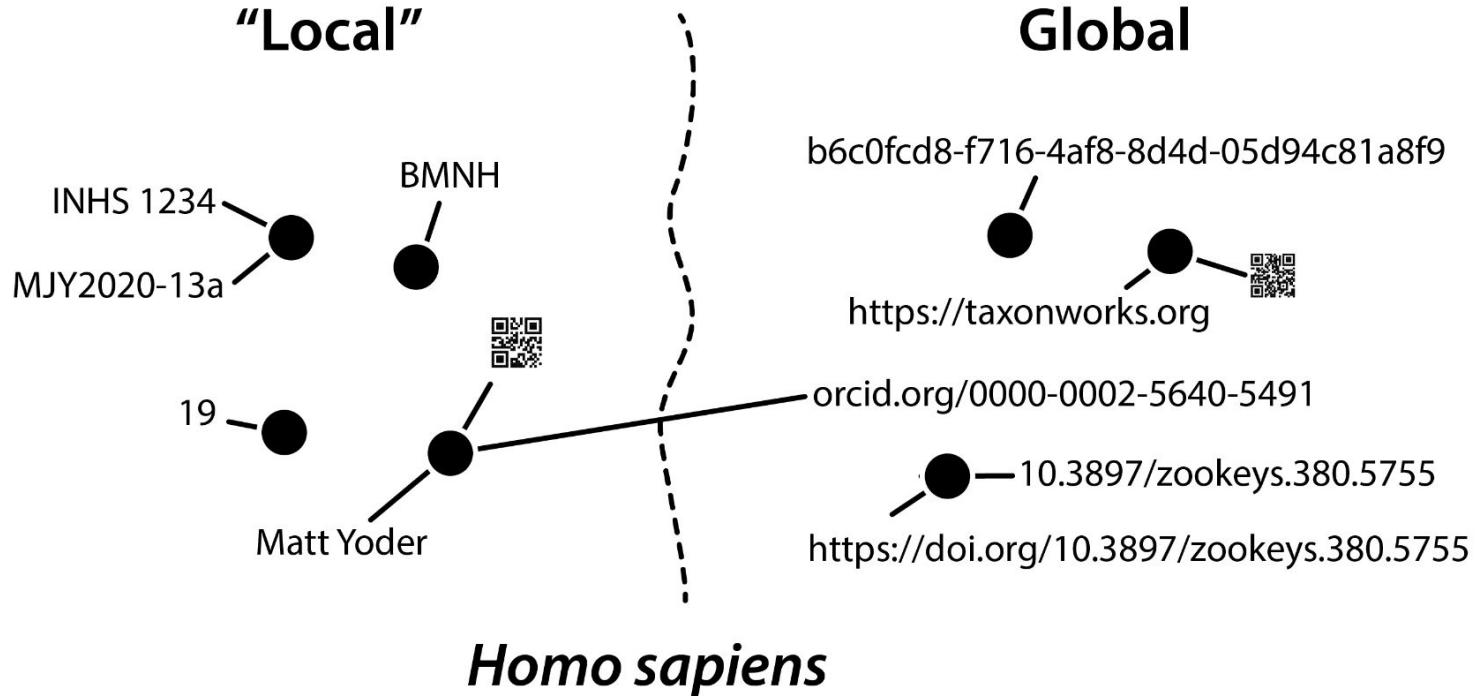
extensible to
~ 20my rows



Opportunities
subgraph analysis
network diffs
visualization
query optimization
RDF representation

Interlude - Identifiers

... so what are taxon names?



Interlude - Identifiers

... identifiers identifying identifiers?

TAXON IDENTIFIERS

Encyclopedia of Life ID : [344](#) • GND ID : [4027110-9](#) • BNCF Thesaurus ID : [634](#) • National Diet Library ID : [00566605](#) • NCBI taxonomy ID : [50557](#) • ITIS TSN : [99208](#) • Freebase ID : [/m/03vt0](#) • WoRMS-ID for taxa : [1307](#) • PSH ID : [909](#) • Fauna Europaea ID : [4](#) • Dyntaxa ID : [4000074](#) • Fossilworks taxon ID : [56637](#) • GBIF taxon ID : [216](#) • ZooBank ID for name or act : [9584517B-E62D-4456-B9A0-6B65B8273B6C](#) • TED topic ID : [insects](#) • Art & Architecture Thesaurus ID : [300310470](#) • New Zealand Organisms Register ID : [dd35a426-1309-4fef-b664-68c32e849434](#) • EPPO Code : [1INSEC](#) • Quora topic ID : [Insects](#) • BugGuide ID : [52](#) • Nederlands Soortenregister ID : [145014](#) • JSTOR topic ID : [insects](#) • UNESCO Thesaurus ID : [concept2461](#) • Bibliothèque nationale de France ID : [11932119n](#) • ADW taxon ID : [Insecta](#) • Encyclopædia Britannica Online ID : [animal/insect](#) • U.S. National Archives Identifier : [10640332](#) • BabelNet ID : [00046870n](#) • Encyclopædia Universalis ID : [insectes](#) • Great Russian Encyclopedia Online ID : [2250439](#) • Iconclass notation : [25F711](#) • Fauna Europaea New ID : [e1ed93e8-a324-41b1-bfe0-48ae4c2e0d58](#) • Open Food Facts food category ID : [insects](#) • IRMNG ID : [1096](#) • iNaturalist taxon ID : [47158](#) • Canadian Encyclopedia article ID : [insect](#) • Plant Parasites of Europe ID : [parasites/animalia/arthropoda/insecta](#) • PACTOLS thesaurus ID : [pcrtLeJDEv8jY1](#) • Klexikon article ID : [Insekten](#) • Library of Congress authority ID : [sh85066670](#) • MeSH descriptor ID : [D007313](#) • MeSH tree code : [B01.050.500.131.617](#) • ASC Leiden Thesaurus ID : [294917306](#) • Belgian Species List ID : [59113](#) • DR topic ID : [insektor](#) • OmegaWiki Defined Meaning : [1959](#) • Australian Educational Vocabulary ID : [scot/1354](#) • Great Aragonese Encyclopedia ID : [20877](#) • Visual Novel Database ID : [g2878](#) • Microsoft Academic ID : [2776290811](#) • French Vikidia ID : [Insecte](#) • English Vikidia ID : [Insect](#) • Spanish Vikidia ID : [Insecto](#) • Fossiilid.info ID : [260](#) • EuroVoc ID : [5676](#) • Encyclopedia of the Great Plains ID : [pe.034](#) • CALS Encyclopedia of Arkansas ID : [insects-5150](#) • e-WV: The West Virginia Encyclopedia ID : [853](#) • Thesaurus for Graphic Materials ID : [tgm005414](#) • UK Parliament thesaurus ID : [91679](#) • Iranica ID : [insects](#) • YSO ID : [1983](#) • NicoNicoPedia ID : [昆虫](#) • Open Library subject ID : [insects](#) • KBpedia ID : [Insect](#) • Colon Classification : [K86](#) • CAB ID : [11990](#) • Google News topics ID : [CAAqJQgKlh9DQkFTRVFvSUwyMHZNRE4yZERBU0JXVnVMVWRDS0FBUE](#) • BOLD Systems taxon ID : [82](#) • Encyclopedia of Melbourne ID : [EM00757b](#) • NBIC taxon ID : [89](#) • Plazi ID : [C8C4569D-5212-9366-6EFE-B1B6EDEFE016](#)

source: [wikidata: Q1390](#)



Opportunities

HCI

gold standards

AI training

vizualization

Facilitated

open source

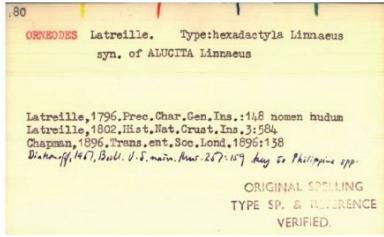
containerized

unit tested

APIs

Images and AI

Example 1



TaxonWorks v0.1.5 Beta

Browse nomenclature

Search a taxon name: Orneodes Advanced search Recent search

Navigate

Product Relation Next

Orneodes Latreille, 1796

History (9)

- ✓ [Orneodes Linnaeus, 1758: 542](#) Type species by subsequent designation [Aulus hexadactylus Linnaeus, 1758](#)
- [Orneodes Linnaeus, 1758: 542](#) Type species by subsequent designation [Aulus hexadactylus Linnaeus, 1758](#)
- [Orneodes Latreille, 1796 \(objective synonym of \[Aulus Linnaeus, 1758\]\(#\)\)](#)
- [Latreille, 1802: 146](#) Type species by original designation [Aulus hexadactylus Linnaeus, 1758](#)
- [Latreille, 1802: 146](#) Objective synonym of [Aulus Linnaeus, 1758](#)
- [Aulus \(Aulus\) Huygeni, 1851 \(objective synonym of \[Aulus Linnaeus, 1758\]\(#\)\)](#)
- [Aulus \(Aulus\) Huygeni, 1851 \(objective synonym of \[Aulus Linnaeus, 1758\]\(#\)\)](#)
- [Orneodes \(Aulus\) Latreille, 1796](#) (nominotypical of [Orneodes Latreille, 1796](#))
- [Orneodes \(Aulus\) Latreille, 1796 \(nominotypical of \[Orneodes Latreille, 1796\]\(#\)\)](#)

Summary

This name is not valid according to the valid names in [SpeciesLink](#). There are 9 names registered in this entity. Records with a date range 31-dec-1958 to 31-dec-1958. The following taxon need to be renamed:

Soft validation

Taxon name

- ⚠ **Homonymy** is missing
- ⚠ **Gender** is not specified (possible gender is masculine)
- ⚠ **Thru name authorship** is not related
- ⚠ **Original publication** is not related, it is derived from the source
- ⚠ **Original authorship request** is not defined for the source and status name author roles

Relationships

- ⚠ **The original publication is not related**

Annotations

Citation

- Latreille, 1796:146

Data attributes

- Taxon: 93,000
- Original source: ORIGINAL GENUS UNDERDEPONED
- Author: Latreille, 1796
- Work_id: id_47
- Status: valid
- Date: 2019-06-29 00:00:00
- Front image: 00050.jpg
- Back image: 00050.jpg
- DOI: [https://doi.org/10.5872/taxonworks/1091](#) (DOI of the original description of [Orneodes Linnaeus, 1758](#))

```

JSON
{
  "id": 10,
  "names": "Orneodes",
  "parent_id": 8,
  "cached": "Orneodes",
  "cached_id": "<!--Orneodes-->",
  "vernacular_name": null,
  "vernacular_id": null,
  "native_name": null,
  "native_id": null,
  "cached_author_year": null,
  "etymology": null,
  "year_of_publication": 1796,
  "vernacular_authors": "Latreille",
  "rank": "species",
  "real_string": "Orneodes",
  "types": "Prototypic",
  "created_by_id": 26,
  "updated_by_id": 24,
  "project_id": 5,
  "citations": "132",
  "cached_original_combination": "Orneodes",
  "cached_original_combination_id": "<!--Orneodes-->",
  "cached_secondary_homonym": null,
  "cached_primary_homonym": "Orneodes",
  "created_at": "2020-05-09T00:00:00+00:00Z",
  "updated_at": "2021-07-16T00:00:00+00:00Z",
  "ancestors": "Orneodes Linnaeus, 1758",
  "verb_ids": null,
  "object_ids": "Orneodes",
  "object_label": "Orneodes Latreille, 1796",
  "global_id": "uid://taxon-works/Prototypic/10",
  "base_class": "TaxonName",
  "uri": "https://taxon-works.org/taxon-names/10.json",
  "object_id": "taxon-works/10",
  "origin_citation": {
    "id": 2,
    "pages": "146",
    "object_label": "Prototypic (<!--Orneodes-->) in Latreille, 1796:146",
    "name_label": null,
    "global_id": "uid://taxon-works/Citation/2",
    "base_class": "Citation",
    "uri_fers": "https://sif.taxonworks.org/taxon_names/2.json",
    "object_uris": "Citations/2",
    "data_attributes": {
      "original_combination": "Orneodes Latreille, 1796",
      "original_combination_id": "<!--Orneodes-->/Orneodes Linnaeus, 1758"
    },
    "status": "[ ]",
    "ancestor_ids": "[ ]",
    "parent": "[ ]",
    "typo_taxon_name_relationship": "[ ]"
  }
}

```

> 500k digitized cards

curated over decades

accessible metadata



Images and AI

Example 2



segmented
images

Rich Flood

TaxonWorks v0.17.0 - Issues

Task: Transcribe depiction

Spec depiction: 56679 Next without data Next TODO max Original image Recalculate Buffered transcription

Collecting event labels: Use [a] Ohio, Muskingum County: Rich Hill TWP p32 *Pond off E side of Zion Ridge Road, A; approximately 0.5 km north of State Route 284: Do not sample - Pond labeled Section C: v18 E Grade wilds map: 18 - September - 1998 - Eric Stata, Joe Beinlich & Steve Cheshire III Notes: Ohio Biological Survey Order: Coleoptera

Determination labels: Use [b] ***Dip SicnC55 Faunon. d Drekymat Det. Eric G. Chapman, 1999

Other labels: Use [c]

Identifiers: INHS Insect Collection Identifier: 892140

NameSpace: Name: [Tropisternus niger Orth.]

Identifiers: INHS Insect Collection Identifier: 892140

NameSpace: Name: Select a ctsu

Buffered collecting event: Ohio, Muskingum County: Rich Hill TWP Ridge Road, Pond off E side of Zion Ridge Road, Approximately 0.5 km north of State Route 284: Do not sample - Pond labeled Section C: v18 E Grade wilds map: 18 - September - 1998 - Eric Stata, Joe Beinlich & Steve Cheshire III Notes: Ohio Biological Survey Order: Coleoptera

Buffered determinations: Tropisternus niger Orth. Drekymat Det. Eric G. Chapman, 1999

Buffered other labels: Total: 11 Note: Note: Tag with: Ohio Biological Survey

Save and next without data [d] Save and next [e] Save changes [f]

transcribed
data

Explore data at: <https://api.taxonworks.org>

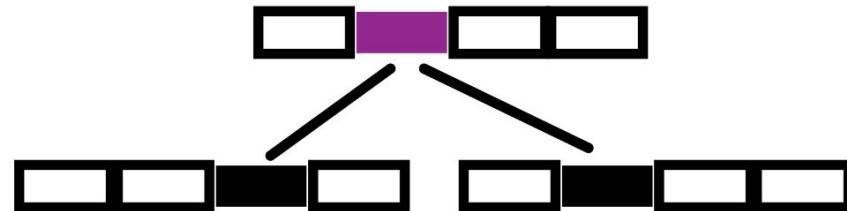
Interlude - Rows to things

What is the same?

Identity



Ontology



Algorithm



Human

Interlude - Rows to things

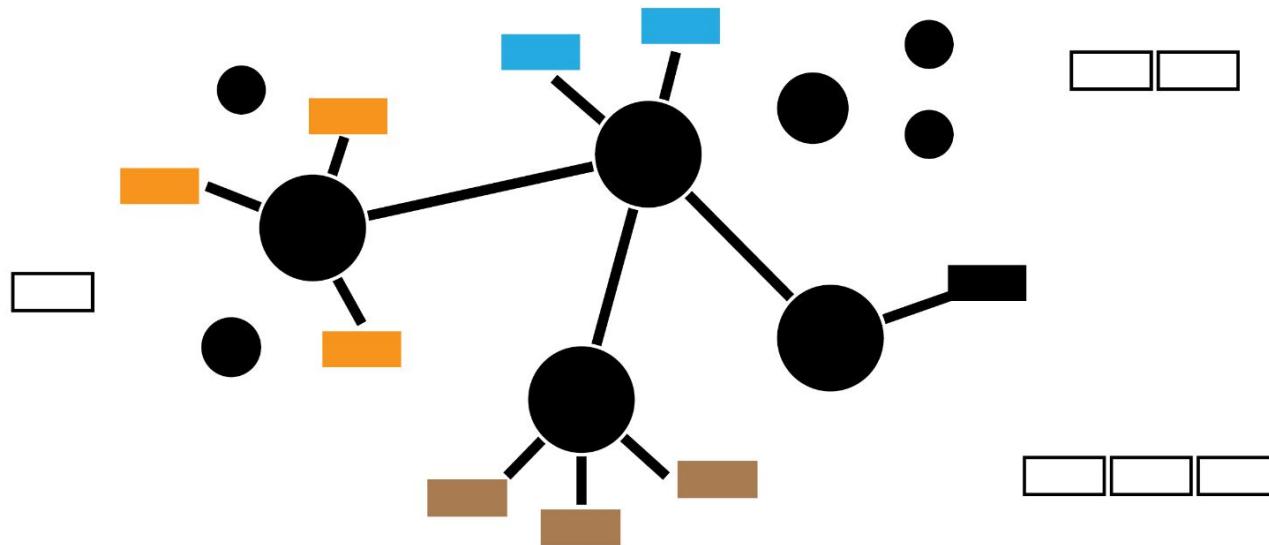
A row of data

e.g. DwC-A



Interlude - Rows to things

One row (e.g. DwC-A) maps to many things





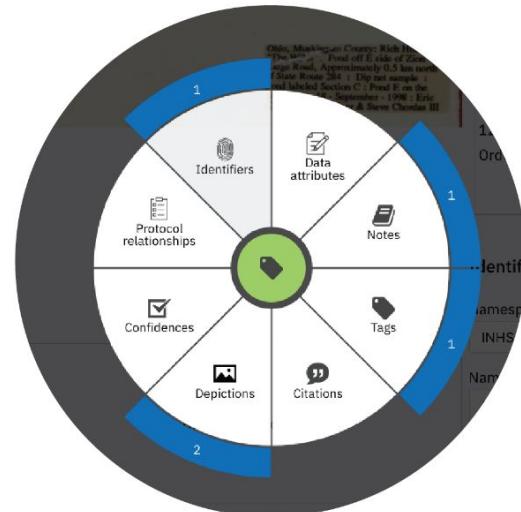
Exploring things

DwC-A rows to TaxonWorks things

Herán Lucas & José Luis Pereira

import
DwC-A

curate the
graph in TaxonWorks



Driven by questions

Why biodiversity informatics?

What is the source of the coronavirus?

How do I build an inventory of life
to bring to Mars?

Where is the biomaterial to save a life?



A collaboration

With the Species File Group and collaborators

Via open source

<https://github.com/SpeciesFileGroup>

<https://github.com/gnames/>

Via project collaborators

<https://stats.taxonworks.org>



A collaboration

... with you?

Community Liason: Deborah Paul, @dlpaul@illinois.edu

Chat: <https://gitter.im/SpeciesFileGroup/taxonworks>

Twitter: @TaxonWorks, @mjyoder

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

Events: <https://speciesfilegroup.org/events.html>

APIs: See TaxonWorks and Global Names homepages

Thank you!

