



Rod Page @rdmpage

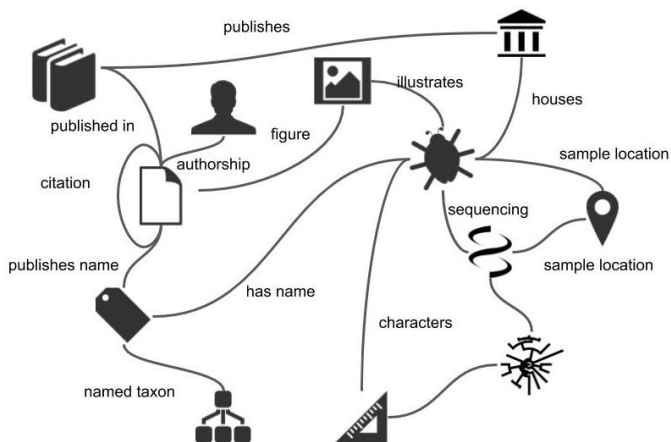
<http://iphylo.blogspot.com>

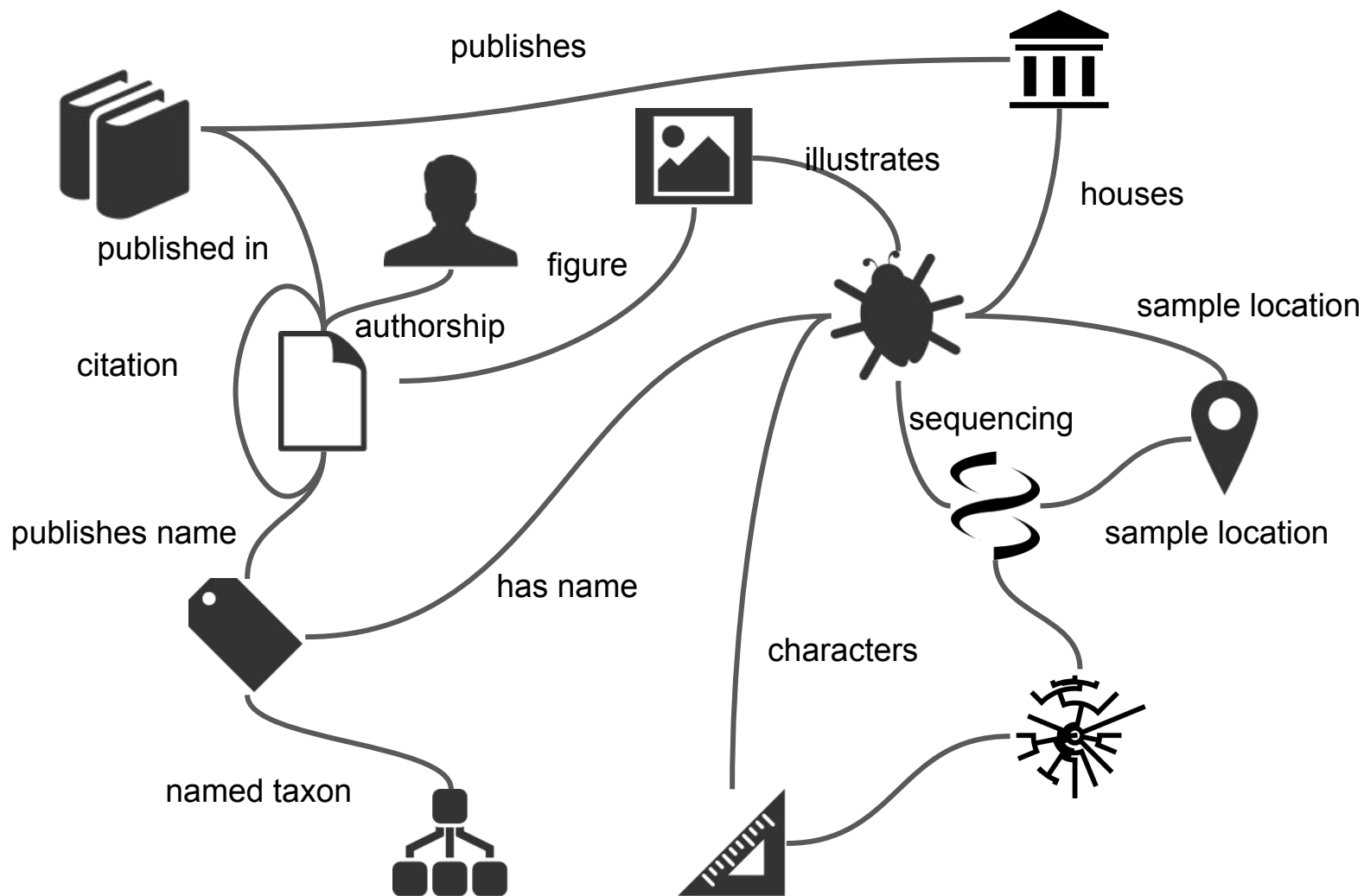
# Ozymandias: A biodiversity knowledge graph

#knowledgegraph

#semanticweb

#linkeddata



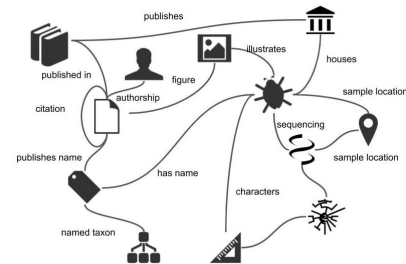


Biodiversity knowledge graph

# Obstacles to building knowledge graphs

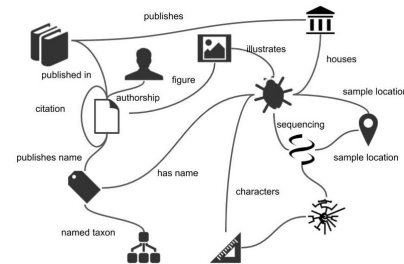
- Technical

- Social



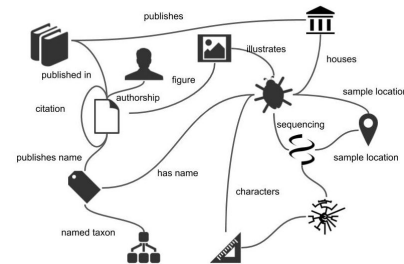
# Obstacles to building knowledge graphs

- Need globally unique, persistent identifiers  
**(how to label the nodes of the graph)**
- Need to create and agree on vocabularies  
**(how to label the edges of the graph)**
- Need to agree how to transmit the graph
- Who stores the global graph?
- “Killer apps”



# A new hope...

- The **identifier wars** are (nearly) over (DOIs FTW)
- Lots of domain-specific vocabularies, but **schema.org** is “good enough” for most things
- XML becoming a bedtime story to frighten the children  
**JSON is everywhere** (JSON-LD FTW).
- **Wikidata** as global knowledge graph
- apps?



```

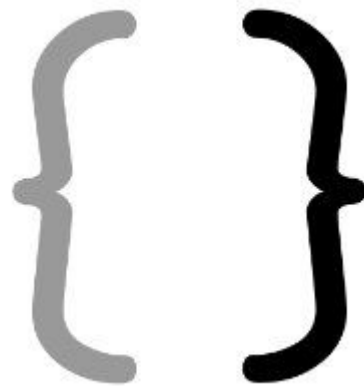
{
  "@context": "http://schema.org",
  "@id": "https://academic.oup.com/jmammal/article/95/5/943/984478/The-valid-generic-name-for-red-backed-voles",
  "@type": "ScholarlyArticle",
  "name": "The valid generic name for red-backed voles (Muroidea: Cricetidae: Arvicolinae): restatement of the case for
Myodes Pallas, 1811 ",
  "datePublished": "2014-10-31",
  "isPartOf": {
    "@id": "https://academic.oup.com/jmammal/issue/95/5",
    "@type": "PublicationIssue",
    "issueNumber": "5",
  },
  "url": "http://dx.doi.org/10.1644/14-MAMM-A-004",
  "publisher": "Oxford University Press",
  "sameAs": "https://academic.oup.com/jmammal/article/95/5/943/984478/The-valid-generic-name-for-red-backed-voles",
  "author": [{
    "name": "Carleton, Michael D.",
    "@type": "Person"
  }, {
    "name": "Gardner, Alfred L.",
    "@type": "Person"
  }, {
    "name": "Pavlinov, Igor Ya.",
    "@type": "Person"
  }, {
    "name": "Musser, Guy G.",
    "@type": "Person"
  }
  ],
  "description": "In view of contradictions in the recent literature, the valid genus-group name...",
  "pageStart": "943",
  "pageEnd": "959"
}

```




William Gibson @GreatDismal




The knowledge graph  
is **already** here  
(it's just not evenly distributed)



# Google Dataset Search (uses schema.org) <https://toolbox.google.com/datasetsearch>



About



Feedback



Least Bittern  
(Isobrychus exilis) –  
Critical Habitat for  
Species at Risk –...

[open.canada.ca](https://open.canada.ca)

Updated Jul 5, 2018



Data from:  
Syngonanthus  
androgynus, a striking  
new species from  
South...

[figshare.com](https://figshare.com)

Updated Jan 20, 2016



Data from: Pelvic-Fin  
Brooding in a New  
Species of Riverine  
Ricefish...

[doi.org](https://doi.org)  
[rin.lipi.go.id](https://rin.lipi.go.id)

Updated 21 Jun 2018



Data from: Syngonanthus androgynus, a striking new species from South America, its phylogenetic placement and implications for evolution of bisexuality in Eriocaulaceae

 Related Article



11 scholarly articles cite this data set ([View in Google Scholar](#))

**Data set created** Oct 7, 2015

**Data set updated** Jan 20, 2016

**Data set published** Oct 7, 2015

**Data set provided by**  
[figshare](https://figshare.com)

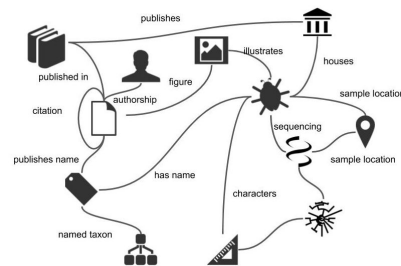
**Authors**  
Mauricio Watanabe

**Licence**  
<https://creativecommons.org/licenses/by/4.0/>



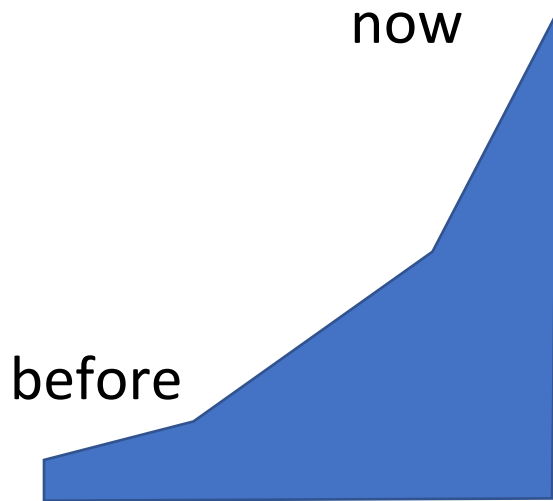
# Obstacles to building knowledge graphs

- Technical
- ~~Social~~ Economic

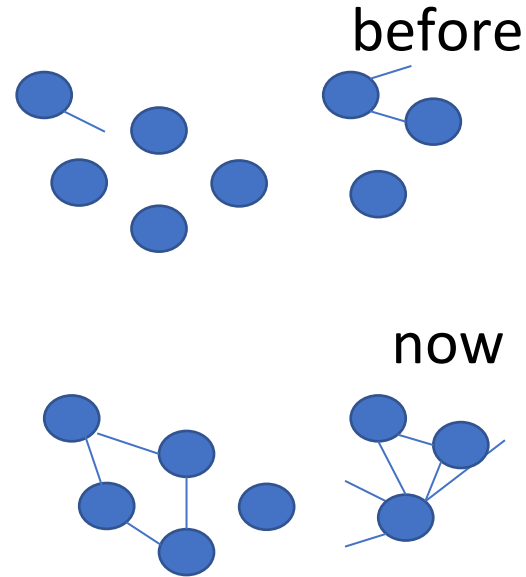


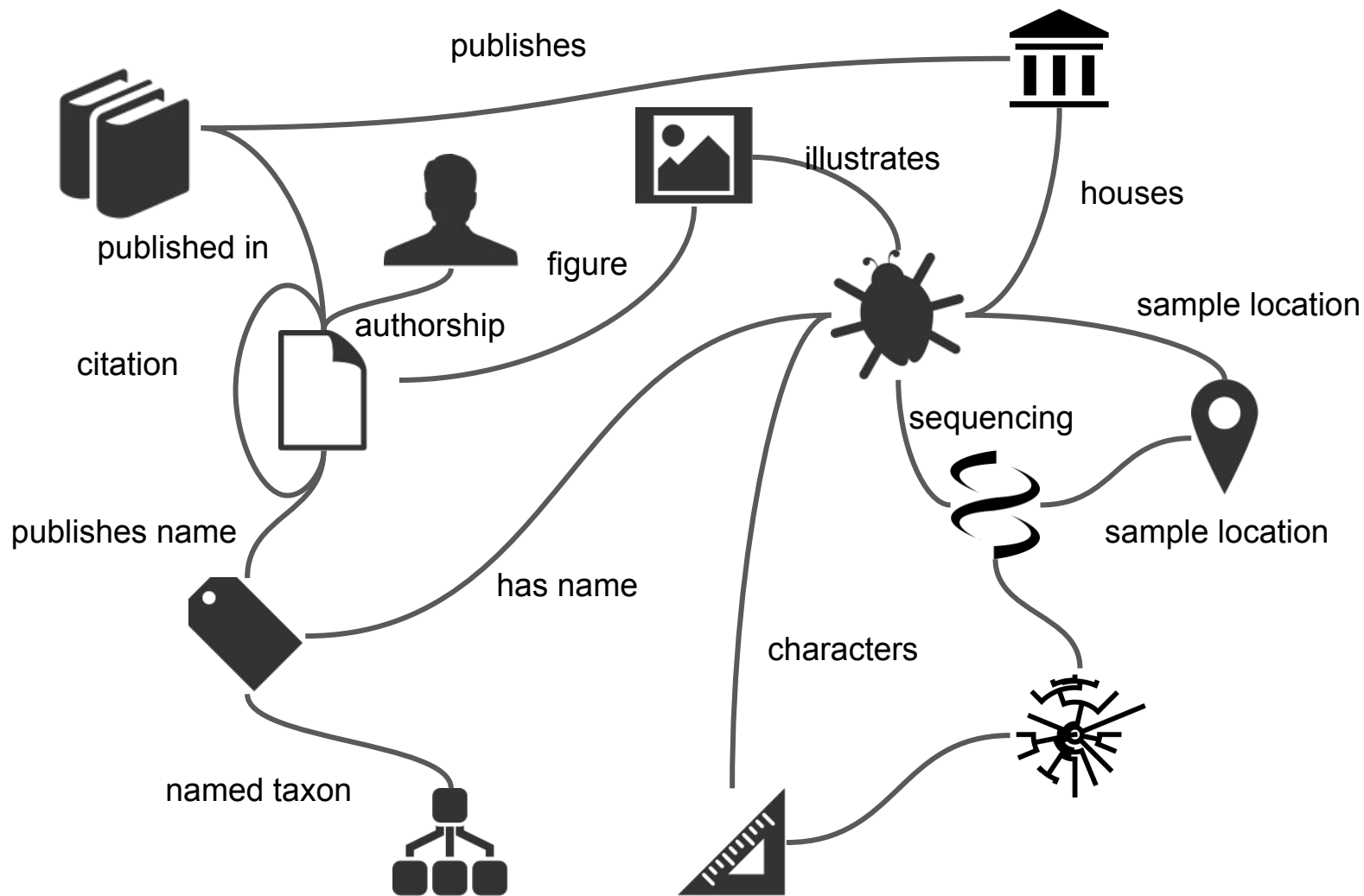
# How do we measure progress?

Linear growth (easy)



Connectivity (hard)





A biodiversity knowledge graph

# *Acupalpa* Kröber, 1912

JSON

genus Accepted Name authority: AFD

Overview

Gallery

Names

Classification

Records

Literature

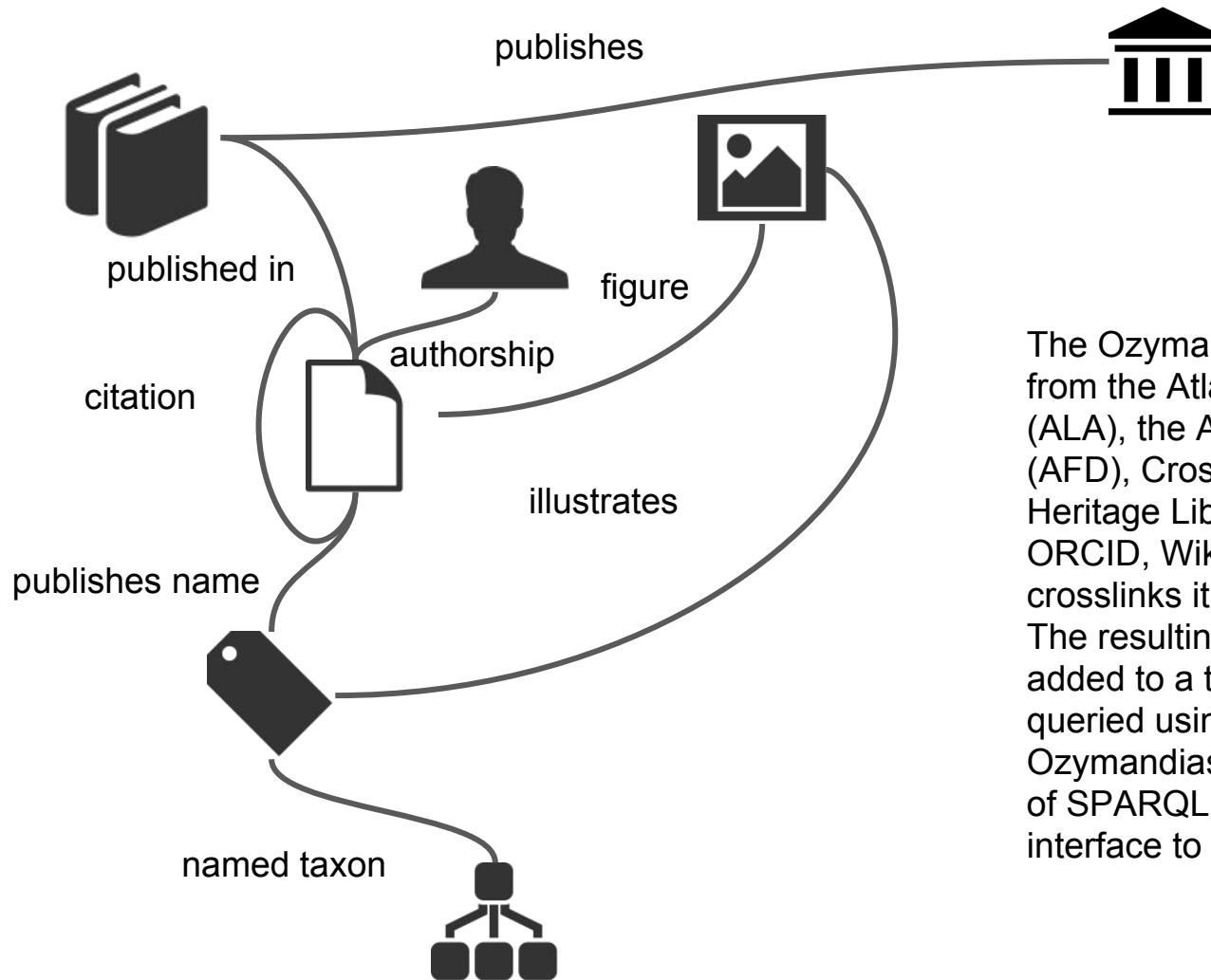
Sequences

Data Partners



## Occurrence records map (333 records)





The Ozymandias graph takes data from the Atlas of Living Australia (ALA), the Australian Faunal Directory (AFD), CrossRef, the Biodiversity Heritage Library (BHL), BioStor, ORCID, Wikispecies and Wikidata and crosslinks it using shared identifiers. The resulting set of data and links is added to a triple store which can be queried using SPARQL. The Ozymandias web site uses a small set of SPARQL queries to create a simple interface to the knowledge graph.

[illegible]

Publications in  
a journal, linked  
to the publisher  
and place of  
publication via  
Wikidata

Ozymandias - a biodiversity knowledge graph	Yes	SPARQL
<p>Search</p> <p><b>Memoirs of the Queensland Museum</b> ISSN: 0079-8835</p> <p><b>Publications in print, linked to digital content by the publisher Place of publication via metadata</b></p>	<p>Connections within this knowledge graph.</p>	<p>External knowledge graphs.</p> <p>Wikidata Q15754783 ZooBank 2470E8F2-A27C-4628-80AE-CC14E2984980 Publisher Queensland Museum</p>
<p>skull Ranges</p> <p>fishes</p> <p>Chalcidoidea. II. The family descriptions of new species</p> <p>Chalcidoidea. I. The family description of new genera and</p> <p>New species of cestodes from Australian birds</p>		

A SPARQL query to display a classification

**Ozymandias - a biodiversity knowledge graph**

Search

**D. K. Yeates**

List of works by this author that are in the knowledge graph.

**1980**

The genus Nomioides Schenck (Hymenoptera: Halictidae) in Australia  
Journal of the Australian Entomological Society  
<https://doi.org/10.1111/j.1440-6055.1986.tb01090.x>

Protection of the egg mass by an Australian March Fly (Diptera: Tabanidae)  
Queensland Naturalist

Revision of the Australian genus Oncodossia Edwards (Diptera: Bombyliidae)  
Systematic Entomology  
<https://doi.org/10.1111/j.1365-3113.1986.tb00256.x>

Revision of the Australian species of Systropus Wiedemann (Diptera: Bombyliidae)  
Journal of the Australian Entomological Society  
<https://doi.org/10.1111/j.1440-6055.1988.tb01170.x>

Connections within this knowledge graph.

**Top five coauthors.**

- M. E. Irwin
- B. D. Lessard
- C. L. Lambkin
- D. J. Ferguson
- S. L. Winterton

**Top ten journals.**

- Zootaxa
- Invertebrate Taxonomy
- Australian Entomology
- Systematic Entomology
- Austral Journal of Entomology
- InsectBiology Systematics
- Journal of Natural History
- Isis
- Isis Systematics and Evolution
- Journal of the Australian Entomological Society
- Journal of Medical Entomology

**Top 20 taxa.**

- ANIMALIA
  - ARTHROPODA
    - HEXAPODA
      - INSECTA
        - Pterygotes
          - DIPTERA
            - ORTHOPTERRAPHA
              - ASILOIDEA
                - BOMBYLIIDAE
                  - Anthraxinae
                    - Trogonini

External knowledge graphs.

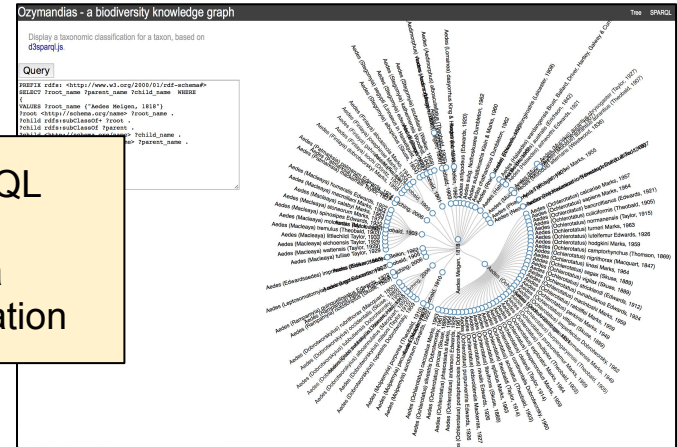
**ORCID iD.**

  - DK YEATES
  - D. K. Yeates
  - D. YEATES
  - D. K. Yeates
  - David K Yeates
  - David K. Yeates
  - DK Yeates

**Wikispecies match.**

  - D.K. Yeates
  - Yeates, D.K.
  - Yeates, D.K.
  - Wikidata
  - Q2134400?

**A re with pub tax coa ider**



# Demo

<https://ozymandias-demo.herokuapp.com/>

# Technical Details

## Data sources

- CSV to **MySQL** to N-triples (AFD)
- JSON to **CouchDB** (document store) to N-triples (ALA)
- JSON-LD to N-triples (**Zenodo**)

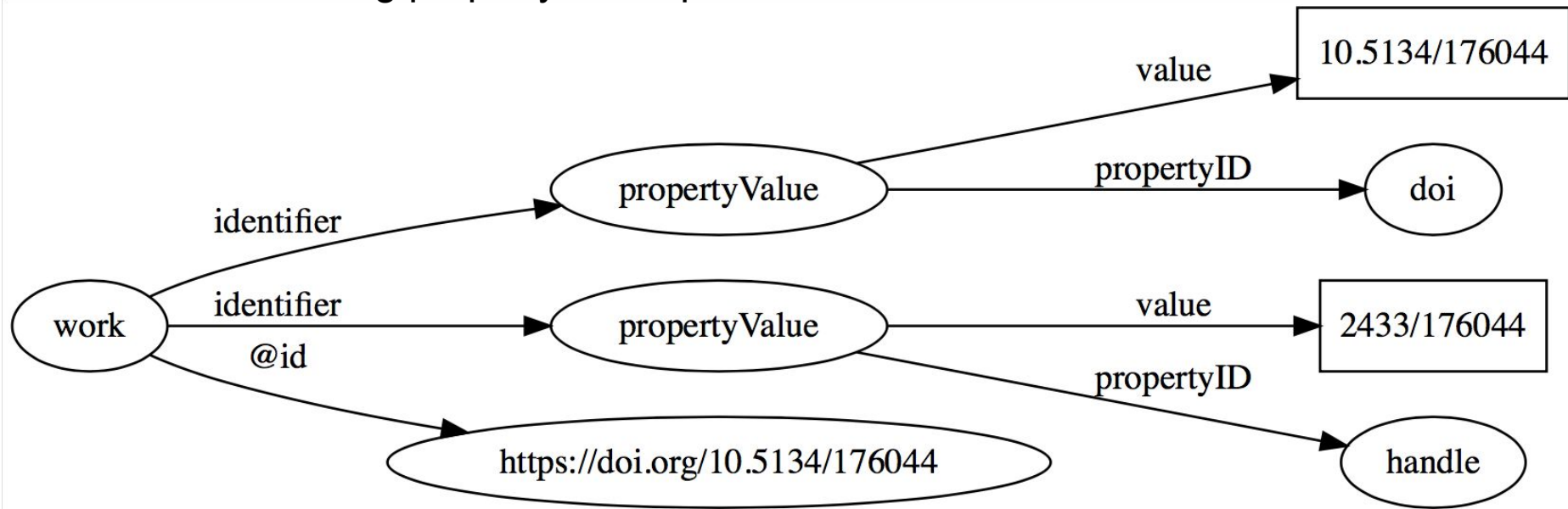
Lots of **cleaning** and **linking**, then import into **Blazegraph** (triple store)

PHP and Javascript web app (SPARQL queries)



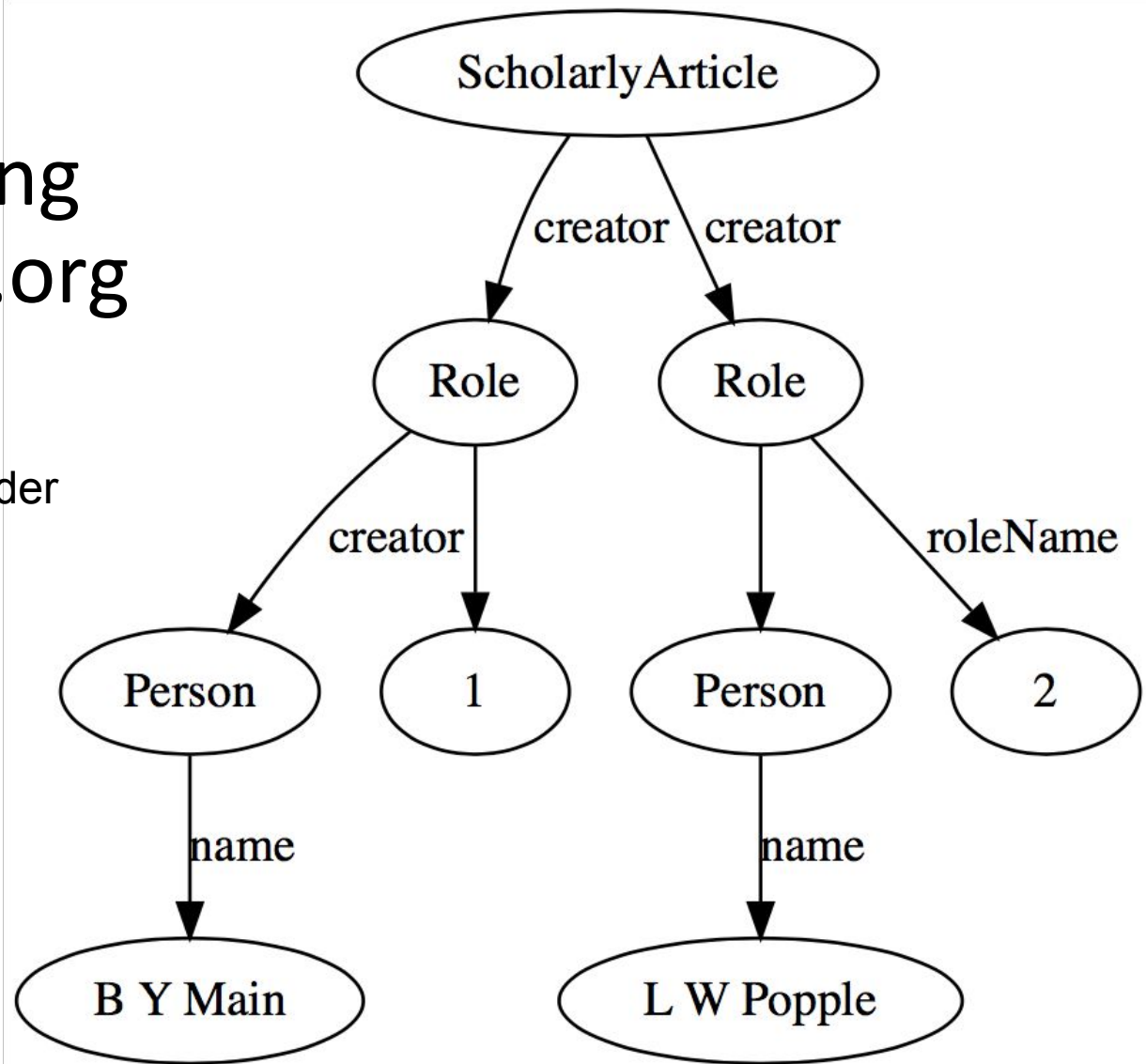
# Multiple identifiers for same thing

Use schema.org property-value pairs to include all identifiers

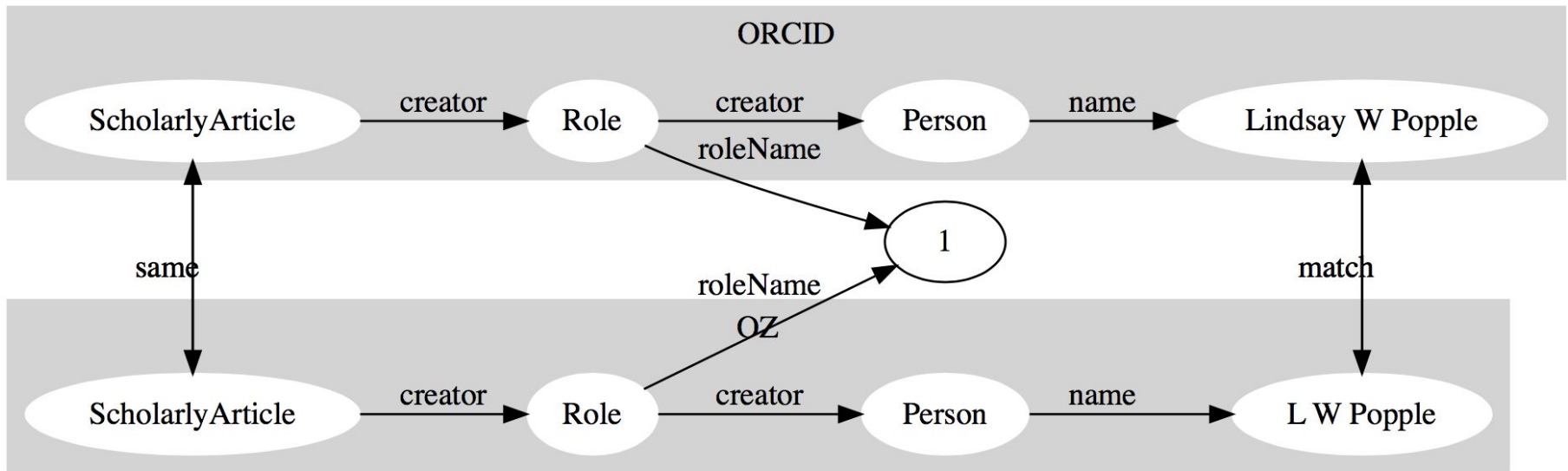


# Author ordering using schema.org “Role”

roleName property is order  
in list of authors



# Matching authors to identifiers



Same work in ORCID and Ozymandias (e.g., same DOI), author names in same position (“roleName”) are similar, so we can assign ORCID id for “Lindsay W Popple” to Ozymandias author “L. W. Popple”.

# What's next?

- Find identifiers for more entities (**strings to things**)
- Match more authors to **ORCID**, **Wikispecies**, and **Wikidata**
- Expand knowledge graph, e.g. **specimens** and **DNA sequences**
- SPARQL queries that **answer questions**