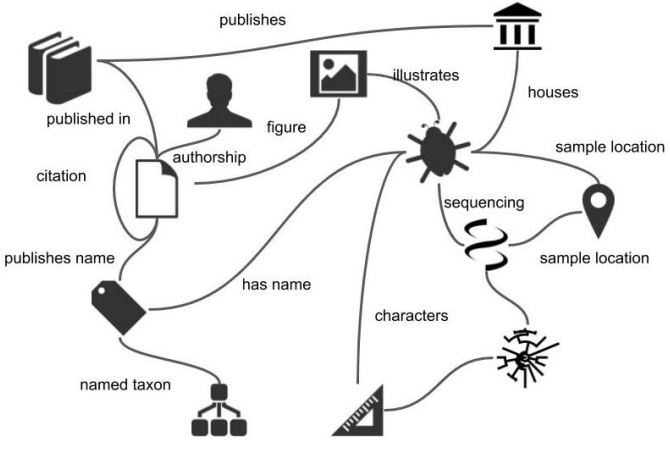
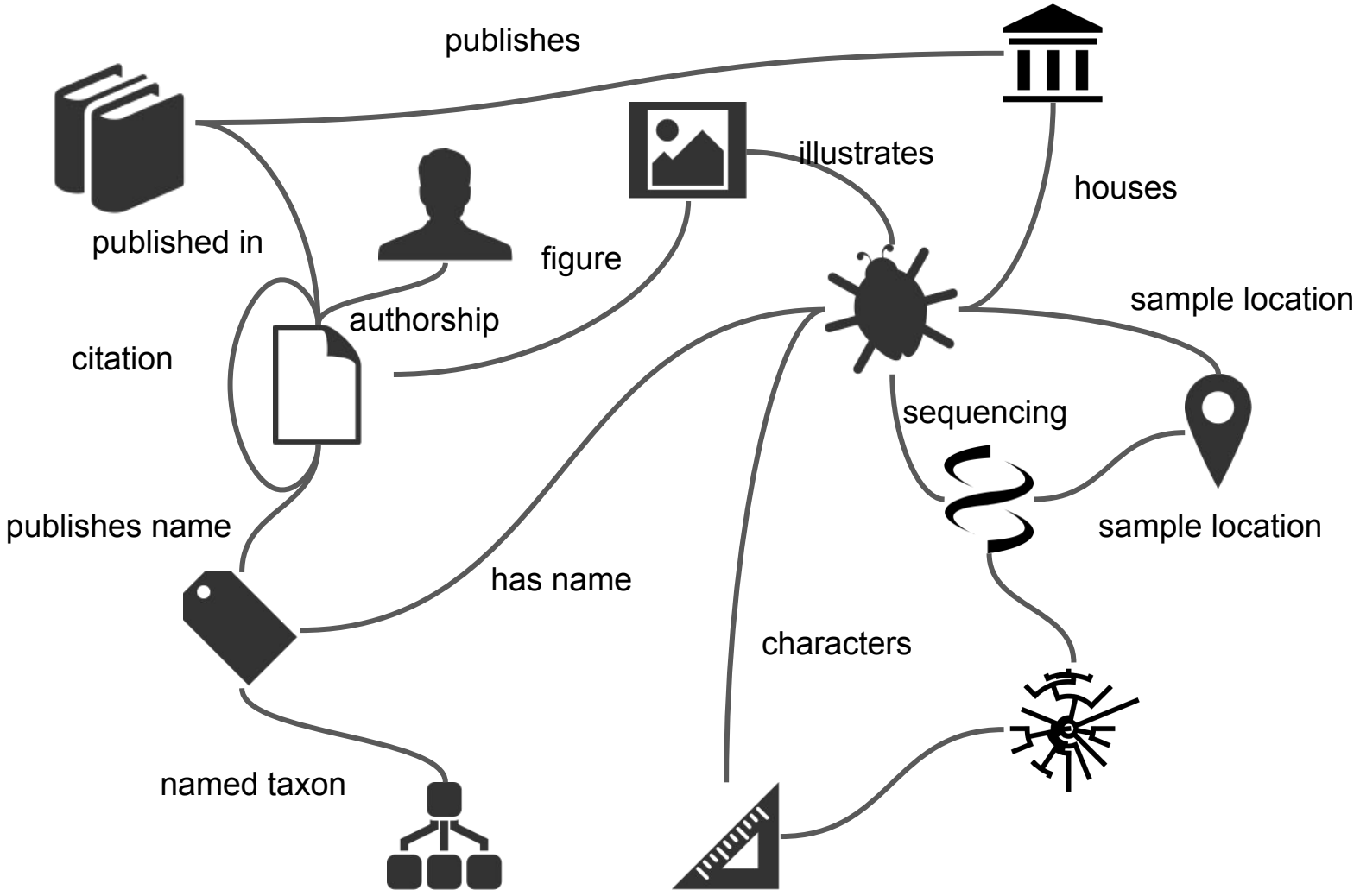
Rod Page @rdmpage

http://iphylo.blogspot.com

Ozymandias: A biodiversity knowledge graph

#knowledgegraph #semanticweb #linkeddata

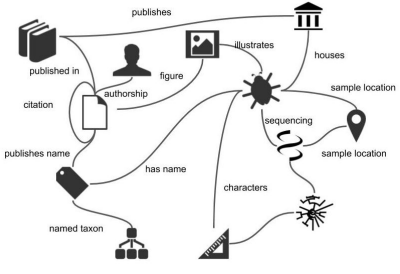




Biodiversity knowledge graph

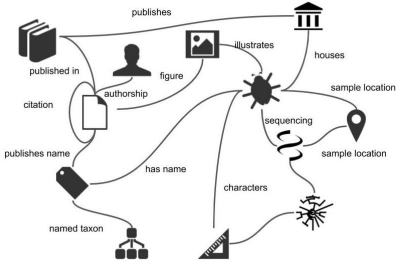
Obstacles to building knowledge graphs

- Technical
- Social



Obstacles to building knowledge graphs

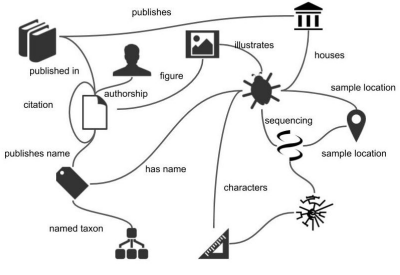
- Need globally unique, persistent identifiers (**how to label the nodes of the graph**)
- Need to create and agree on vocabularies (**how to label the edges of the graph**)
- Need to agree how to transmit the graph
- Who stores the global graph?
- “Killer apps”



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"
}
```

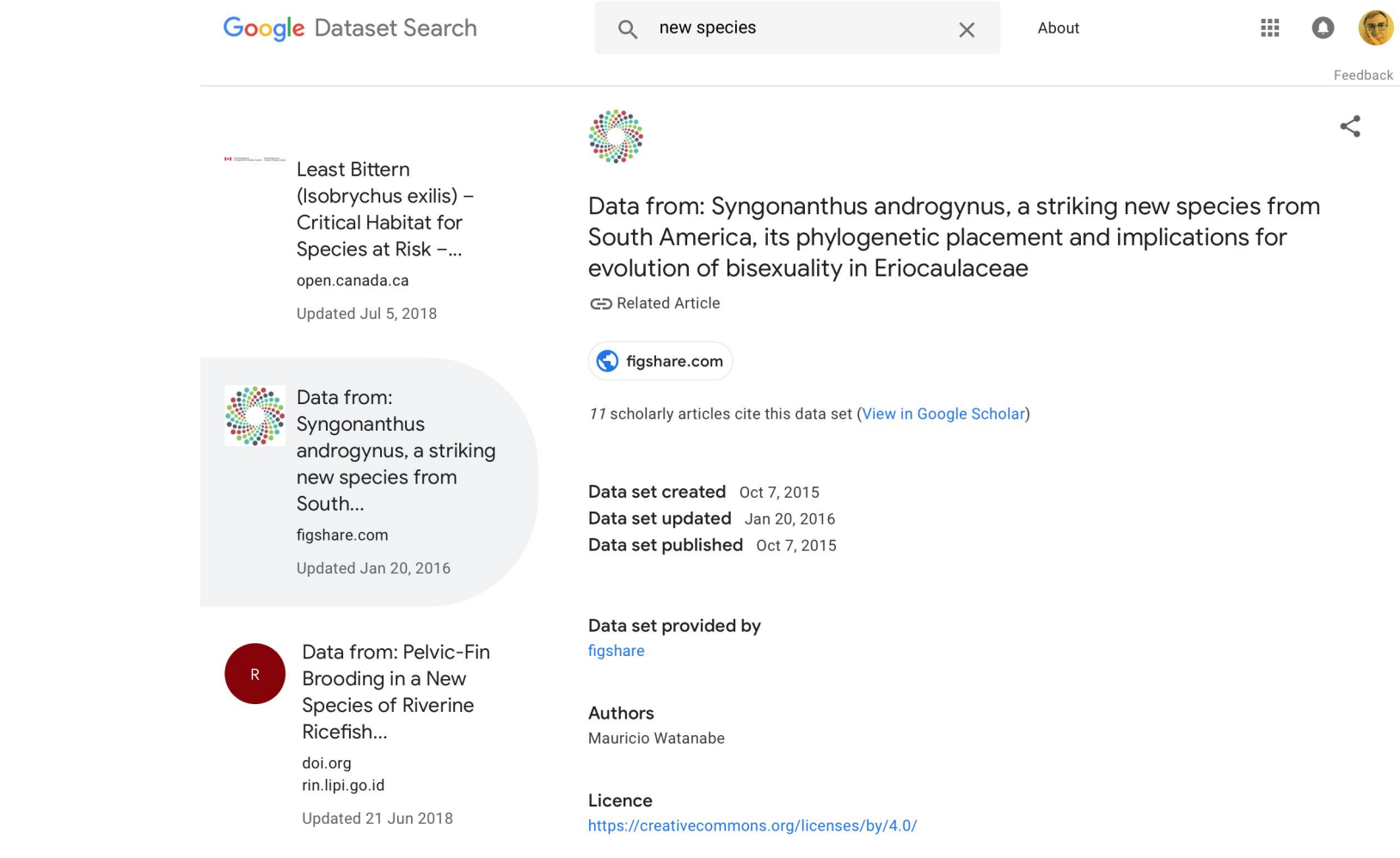
William Gibson @GreatDismal

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

WIKIDATA

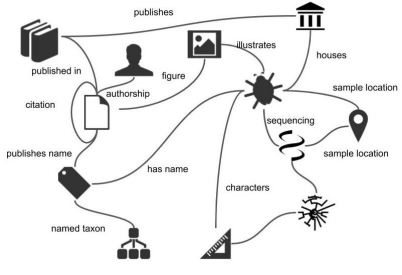
{ }

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



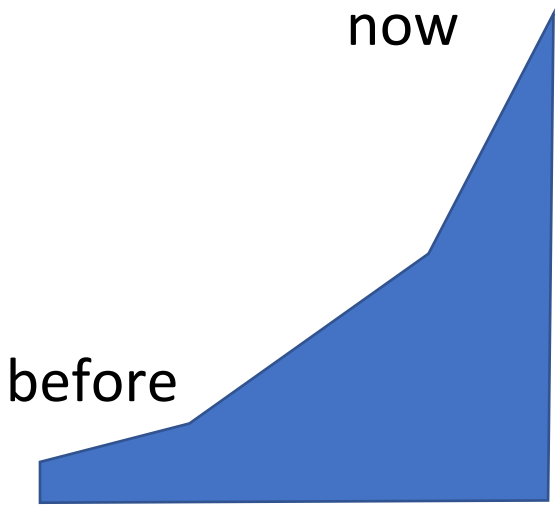
Obstacles to building knowledge graphs

- Technical
- ~~Social~~ Economic

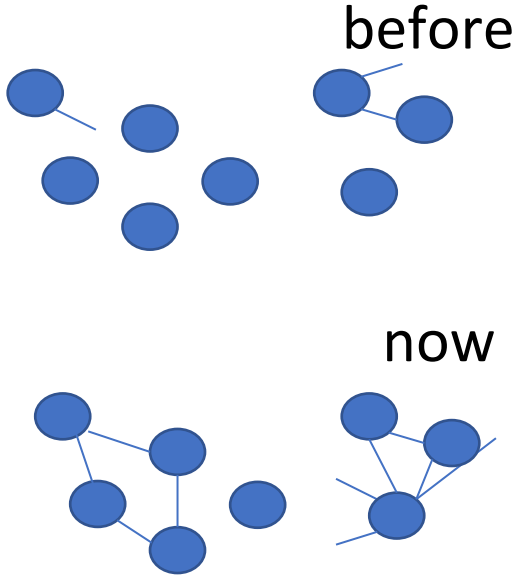


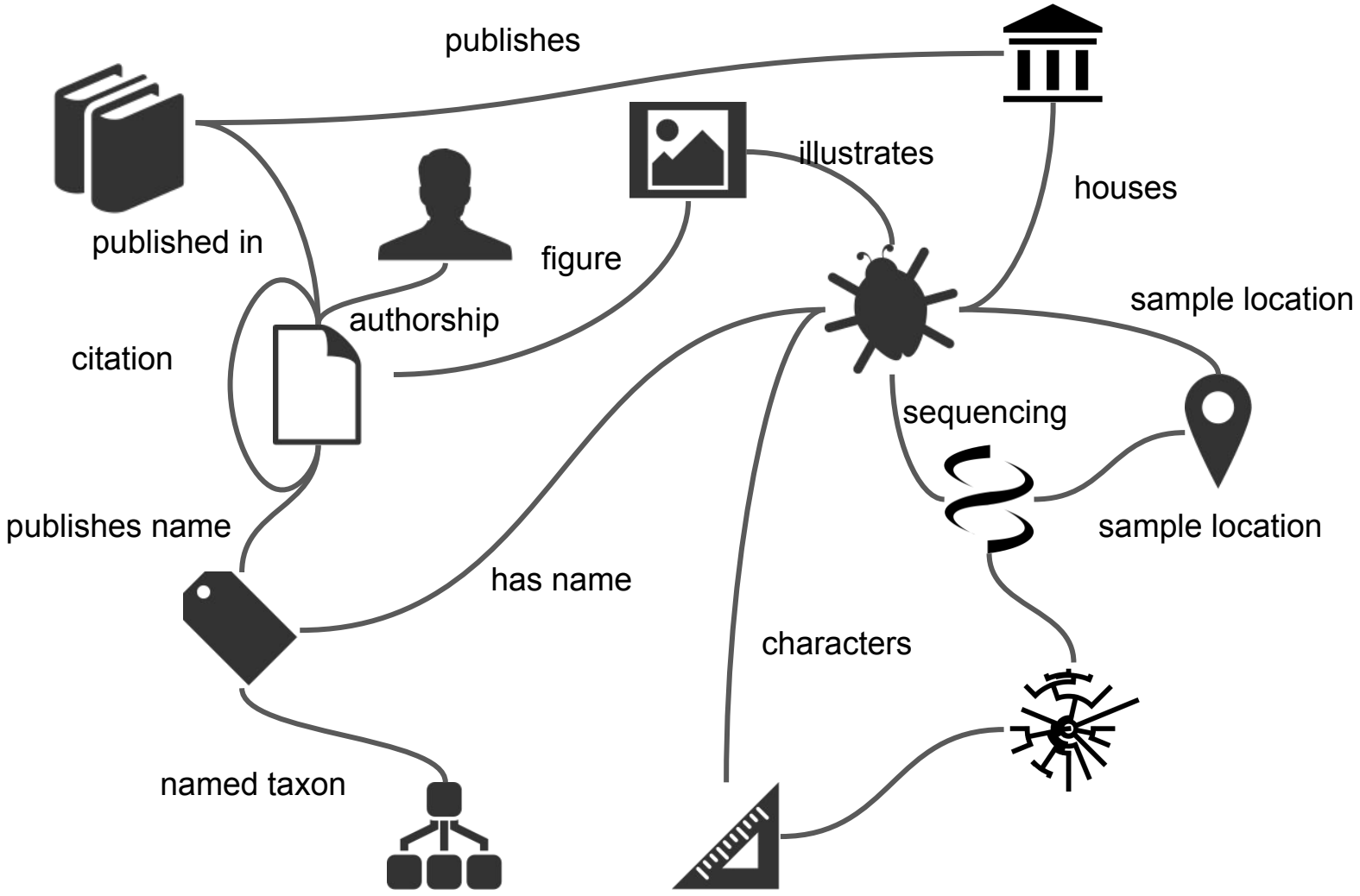
How do we measure progress?

Linear growth (easy)



Connectivity (hard)





A biodiversity knowledge graph

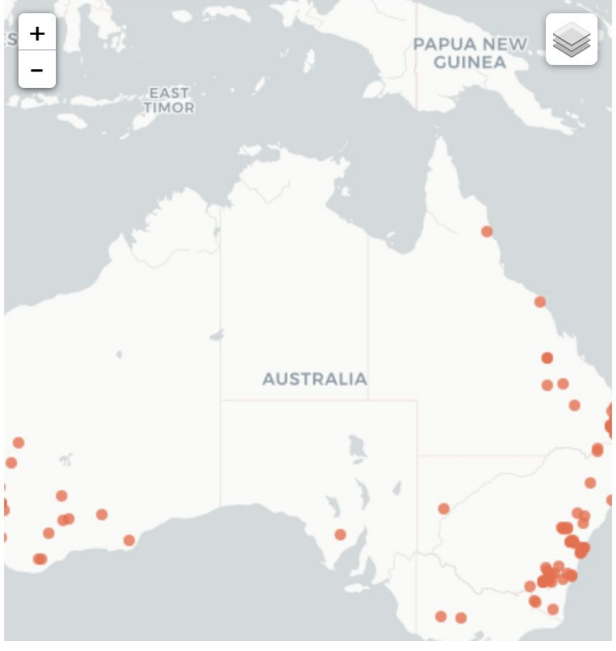
Acupalpa Kröber, 1912

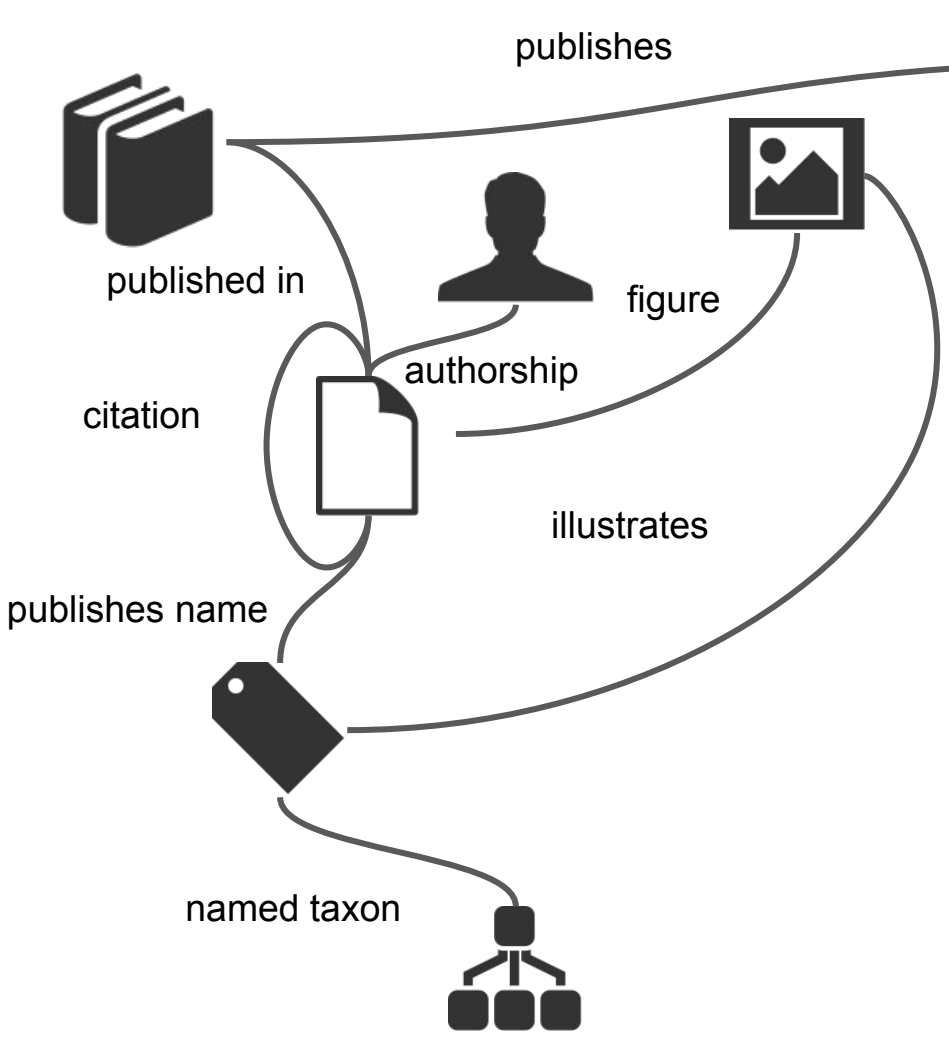
JSON

genus **Accepted** **Name authority: AFD**

Overview | Gallery | Names | Classification | Records | Literature | Sequences | Data Partners

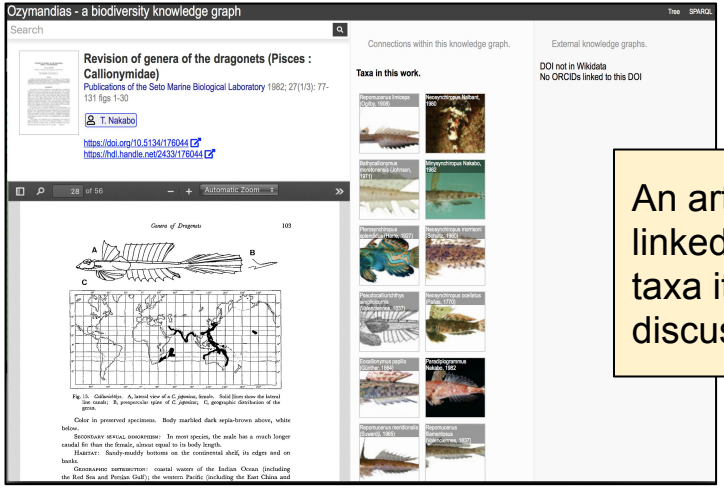
Occurrence records map (333 records)



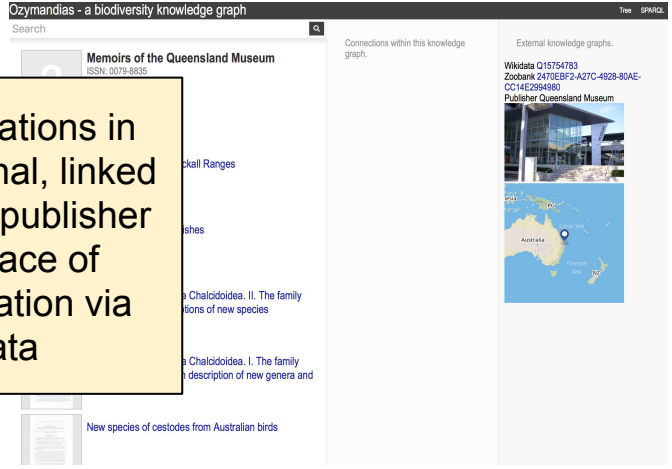


The Ozymandias graph takes data from the Atlas of Living Australia (ALA), the Australian Faunal Directory (AFD), CrossRef, the Biodiversity Heritage Library (BHL), BioStor, ORCID, Wikispecies and Wikidata and crosslinks it using shared identifiers. The resulting set of data and links is added to a triple 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.

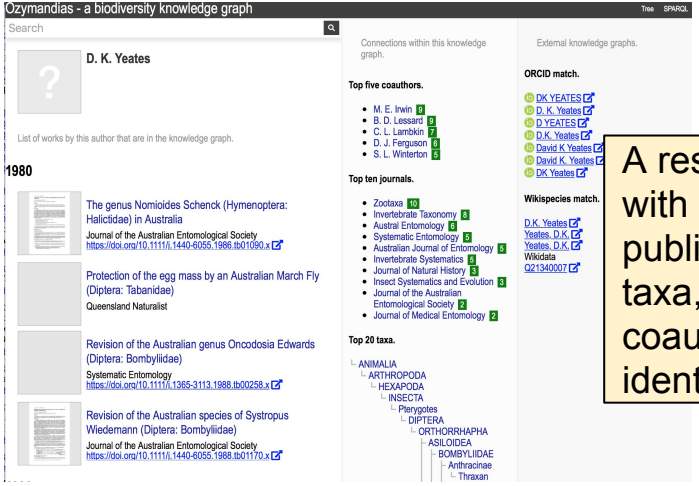
https://ozymandias-demo.herokuapp.com



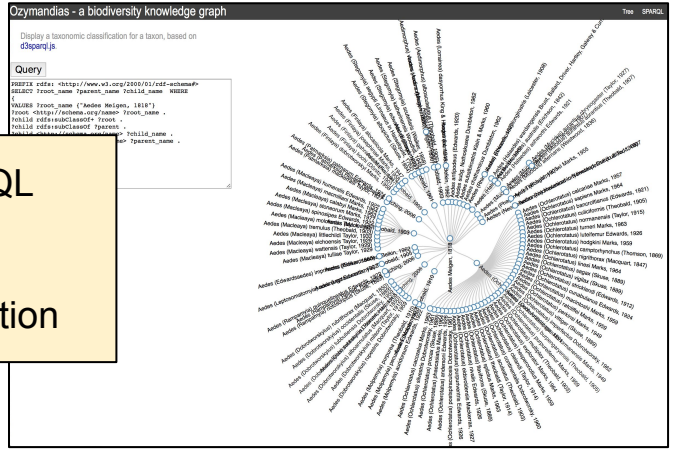
An article linked to the taxa it discusses



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



A researcher with lists of publications, taxa, coauthors, and identifiers



A SPARQL query to display a classification

Demo

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

Technical Details

Data sources

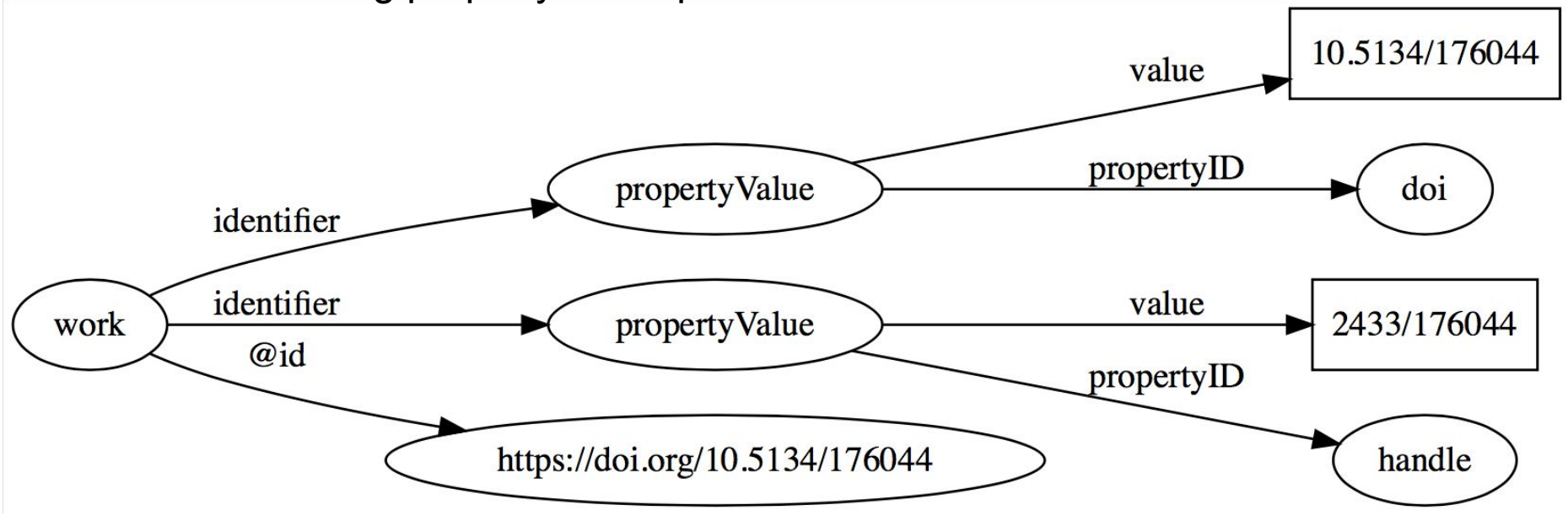
- CSV to **MySQL** to N-triples (AFD)
- JSON to **CouchDB** (document store) to N-triples (ALA)
- JSON-LD to N-triples (**Zenodo**)

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

PHP and Javascript web app (SPARQL queries)

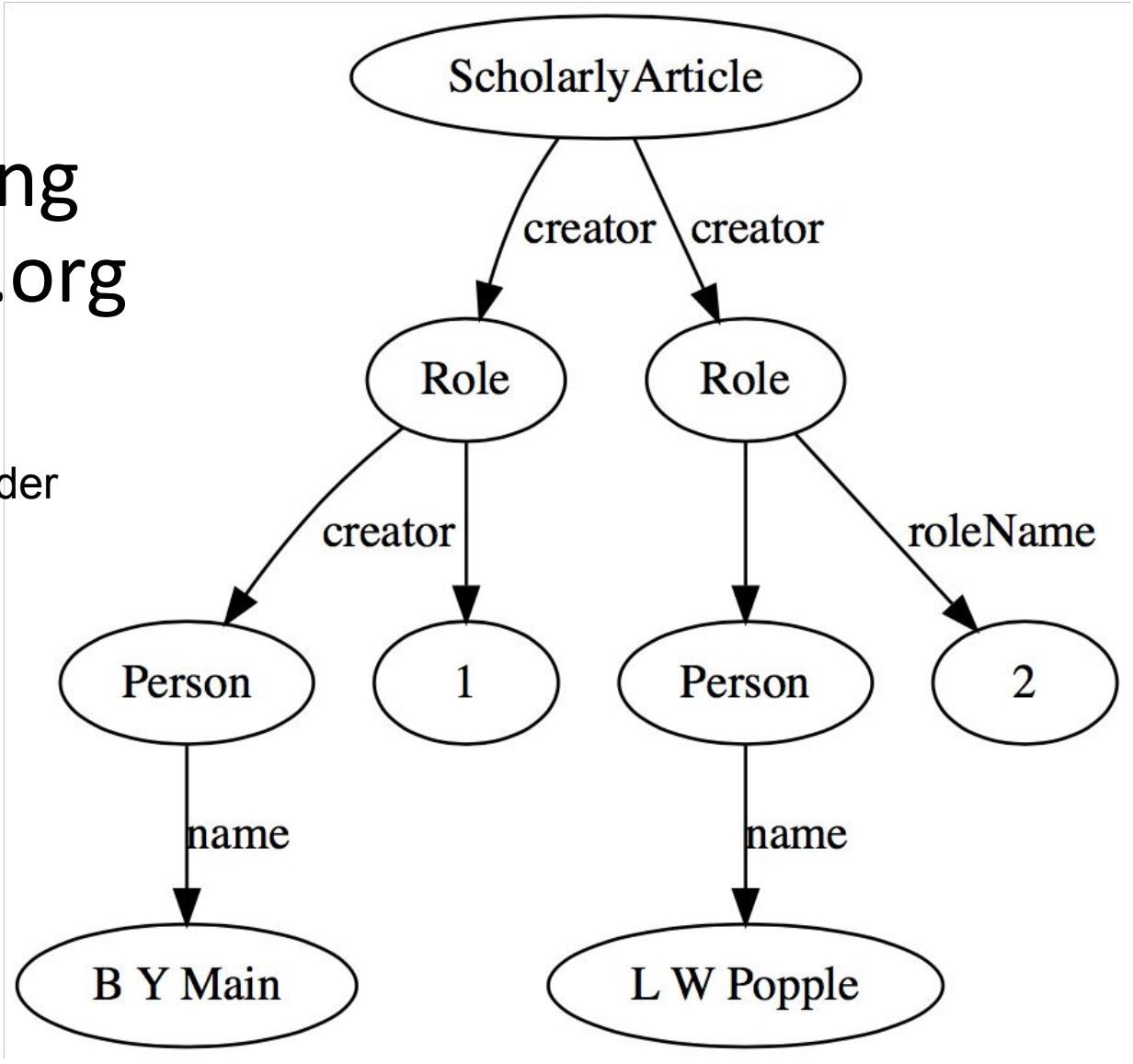
Multiple identifiers for same thing

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

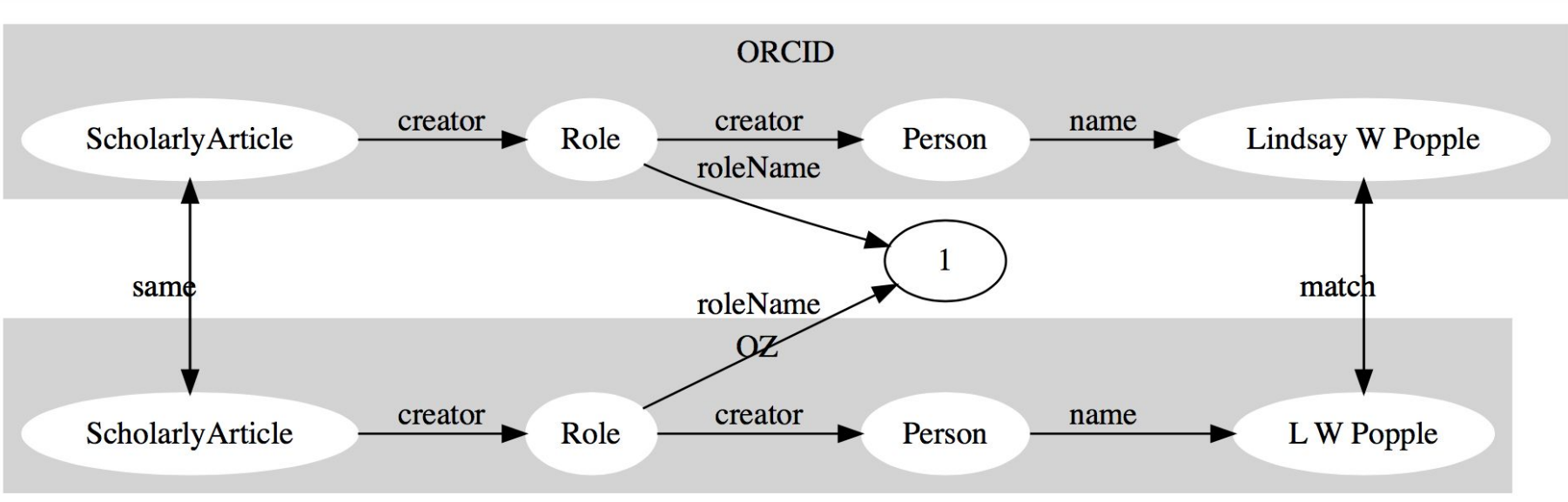


Author ordering using schema.org “Role”

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**