

Nomenclature over 5 years on in TaxonWorks: Approach, implementation, limitations and outcomes

Presented by: Matt Yoder on behalf of the Species File Group and Collaborators

Matt's orcid: orcid.org/0000-0002-5640-5491

TDWG 2021

Who is the Species File Group (SFG)?



Deborah Paul



Dmitry Dmitriev



Yuri Ruskov



Dmitry Mozherin



Rich Flood



Geoff Ower



R. Edward DeWalt



José Luis Periera



Hernán Lucas Periera



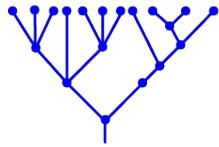
Matt Yoder

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



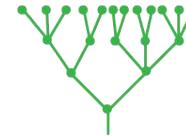
Backbones*

Phylogenetic



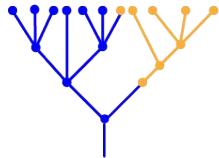
E.g. A quantitative phylogeny

Nomenclature



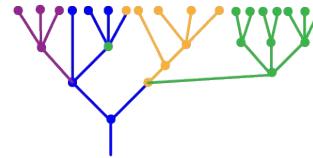
E.g. A TaxonWorks nomenclatural dataset

Natural/Biological



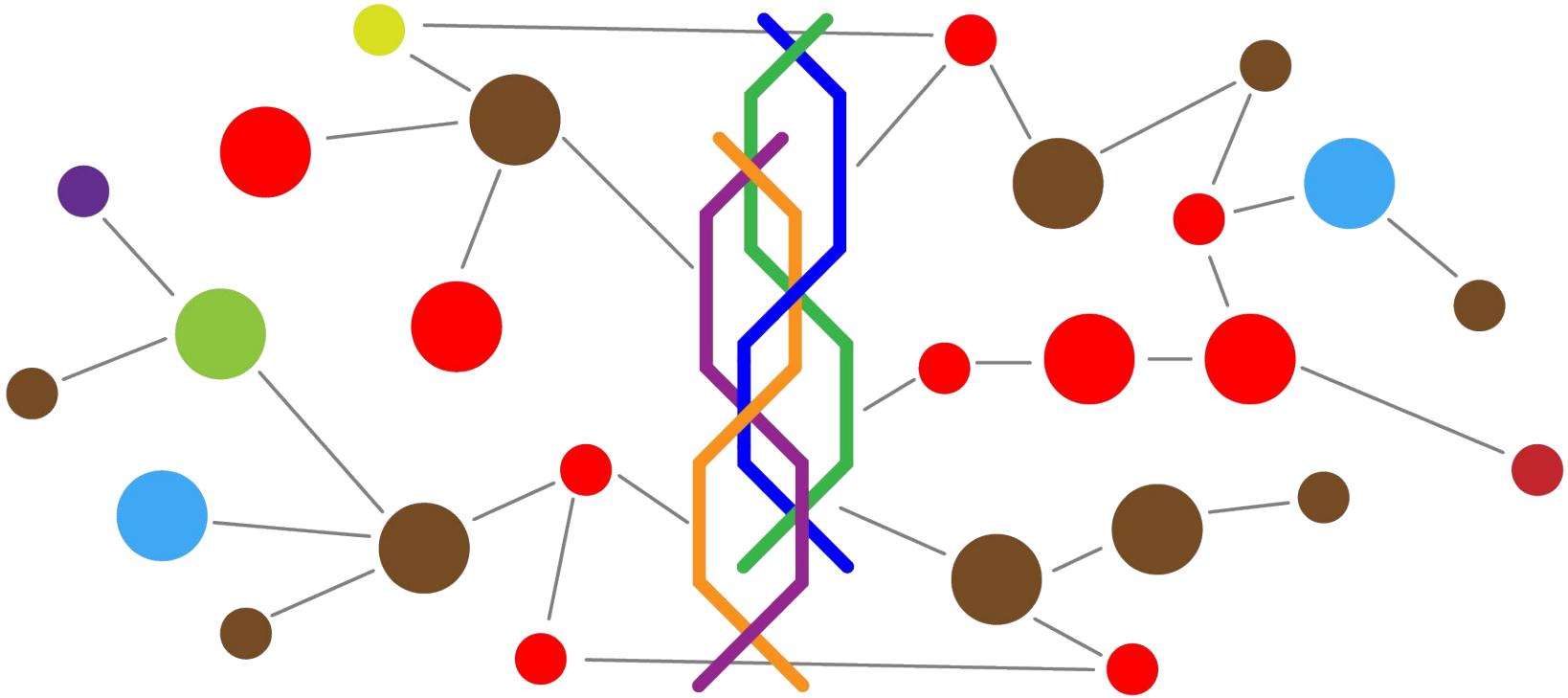
E.g. A taxonomist's classification of biological taxa/concepts.

Staging

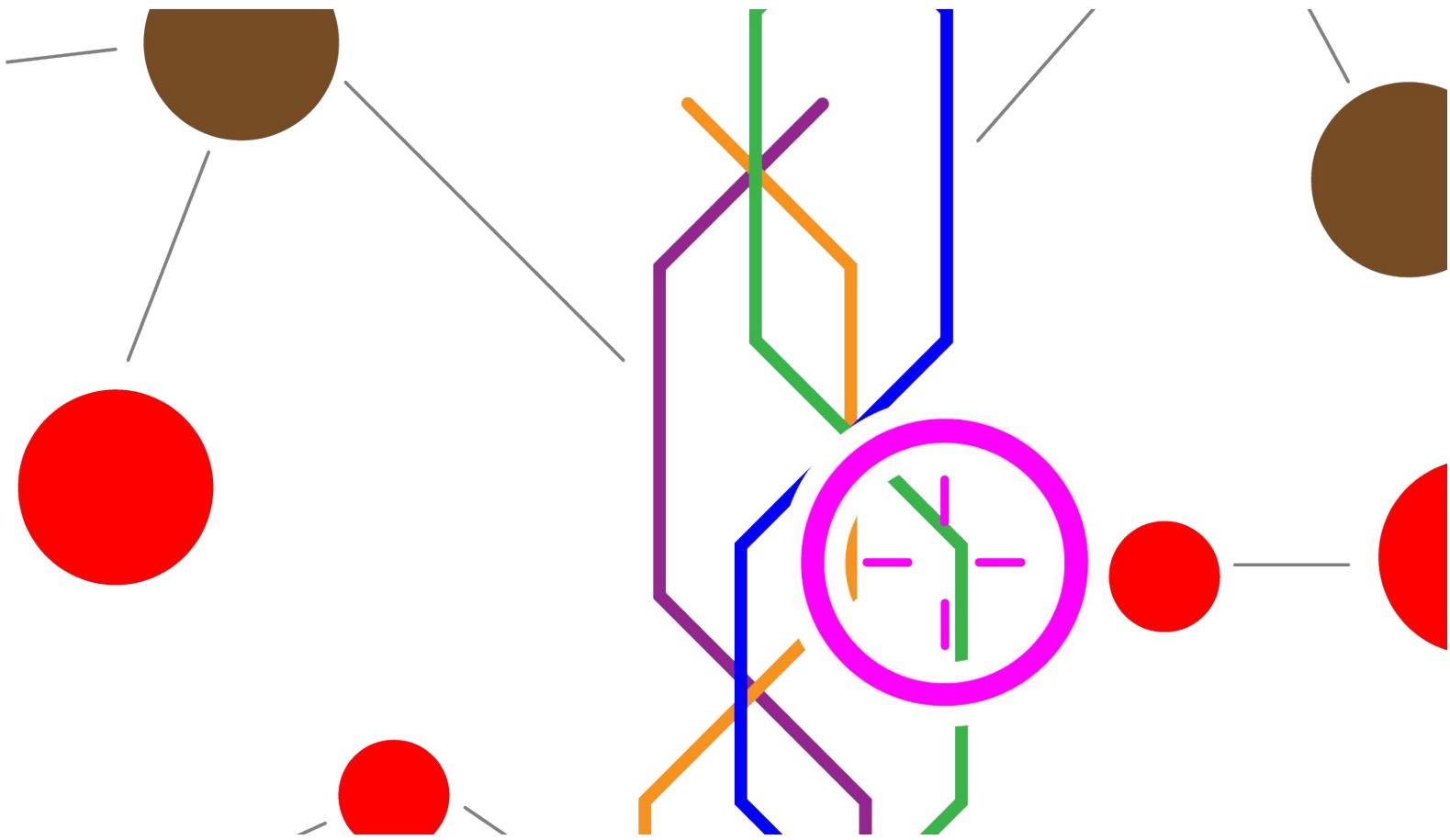


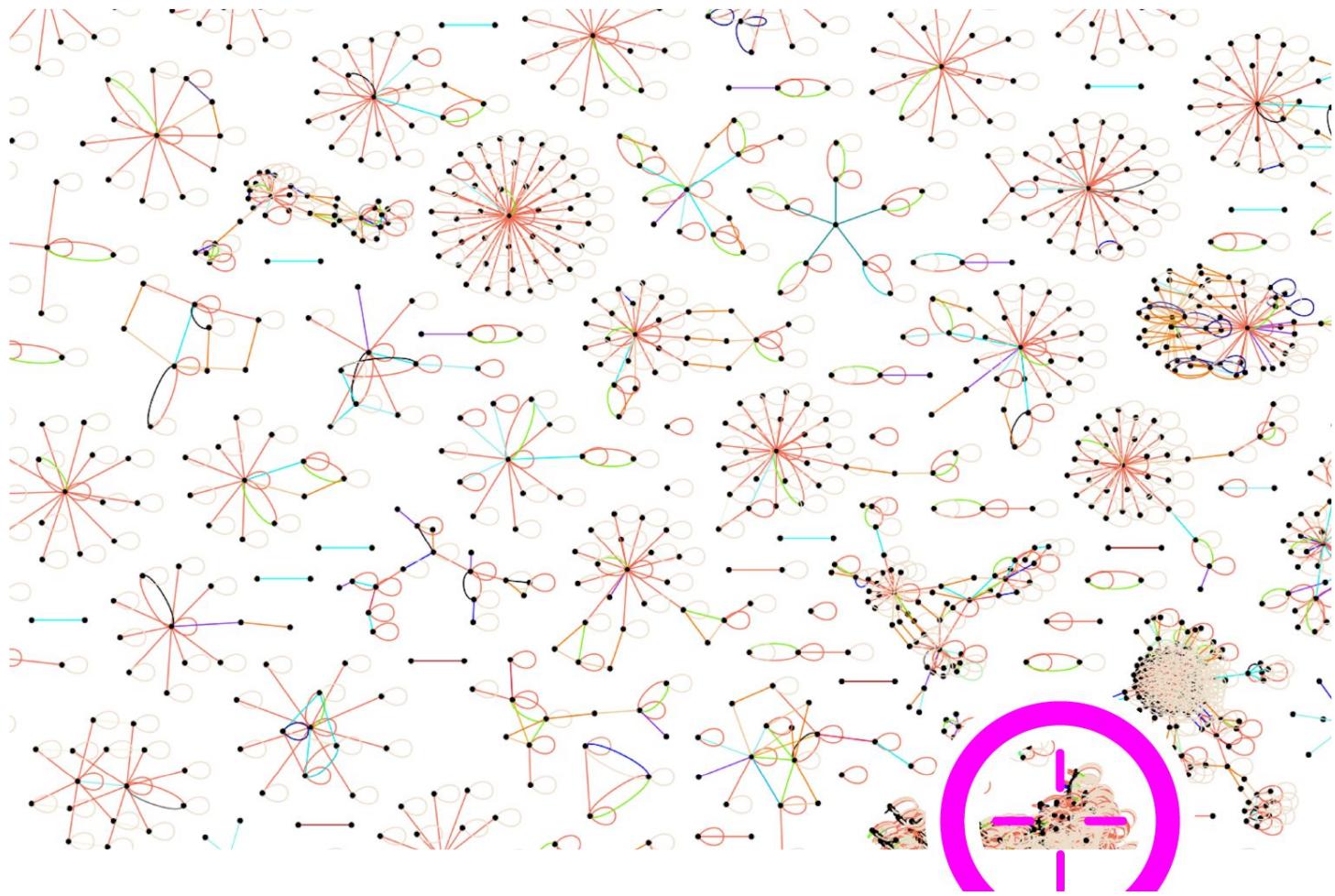
E.g. A classification supporting the (physical) arrangement of specimens in a Natural History Collection

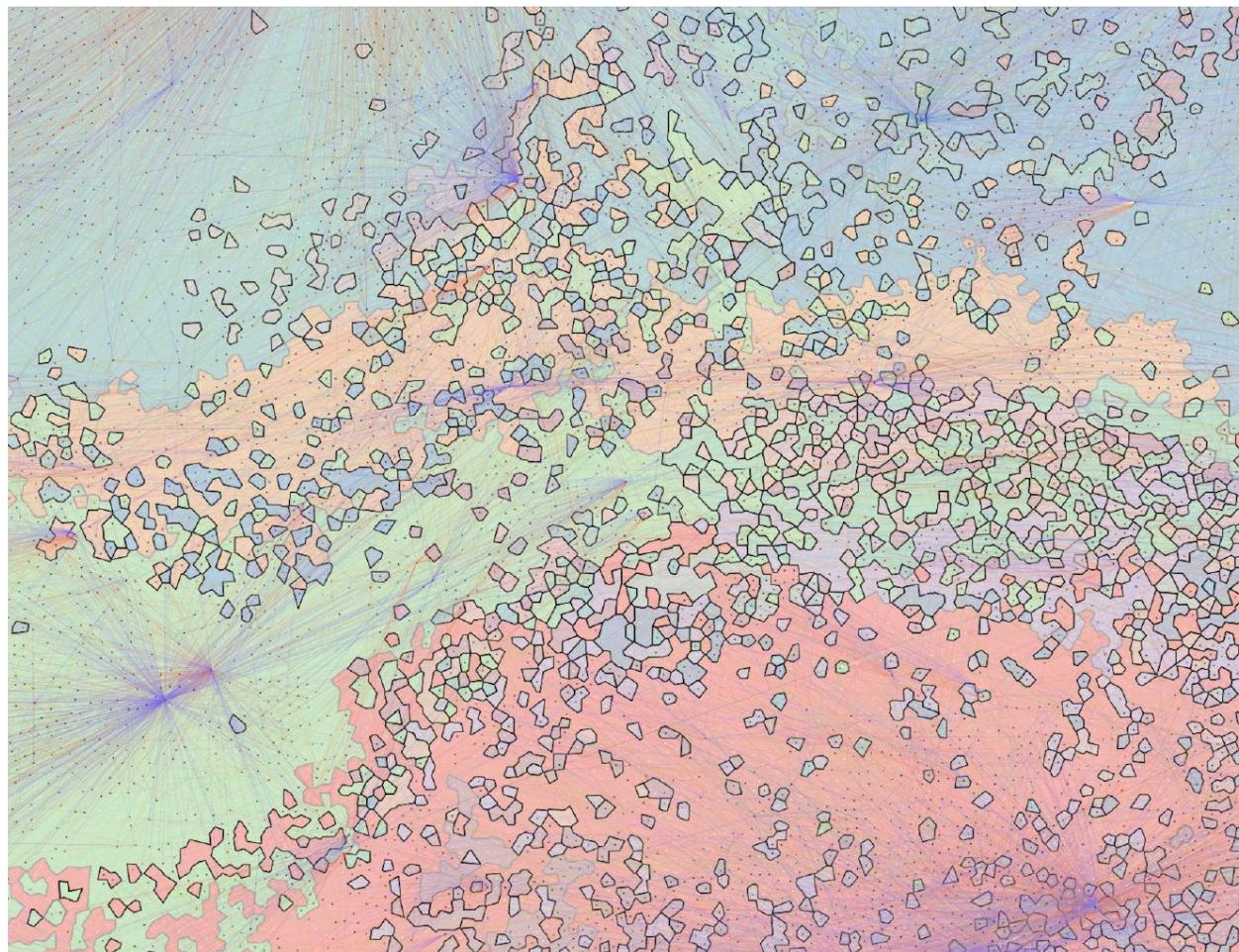
*There are others, and, are all backbones classifications/classifiers?

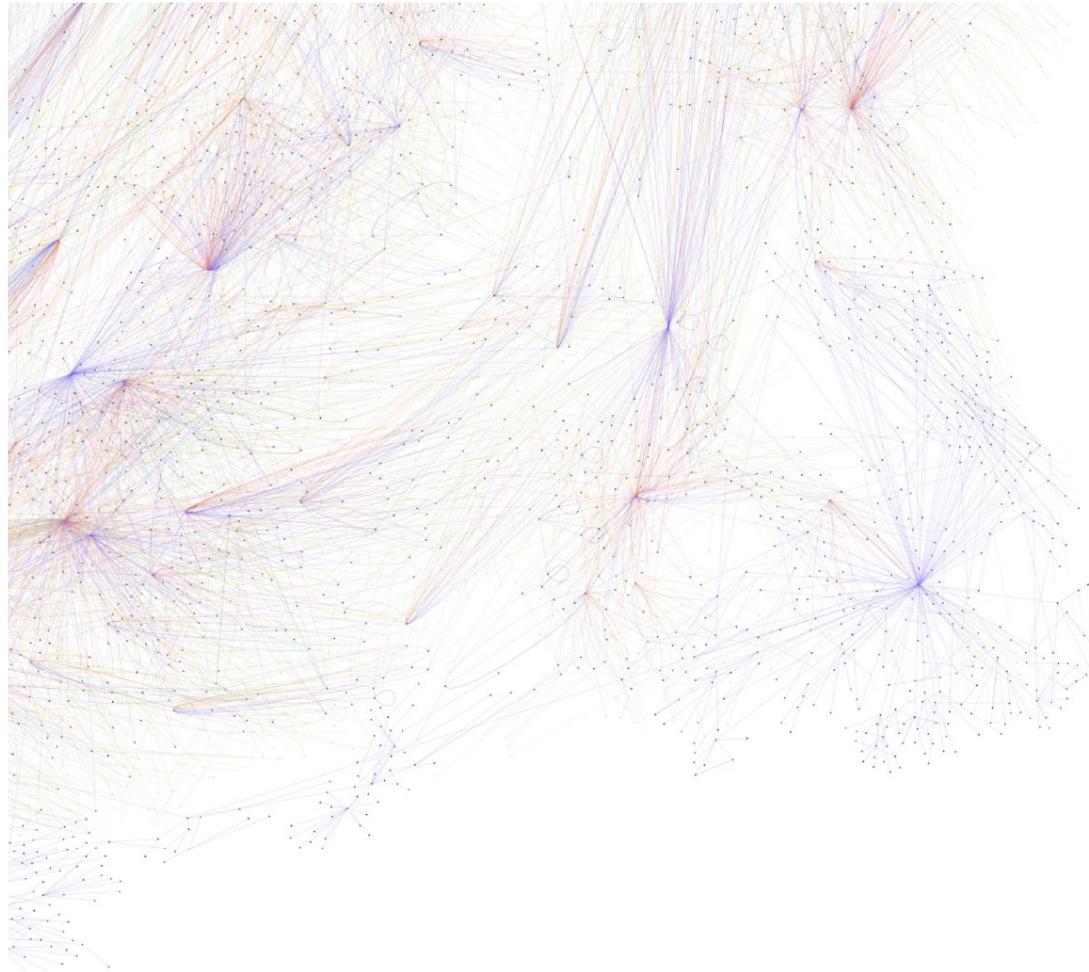


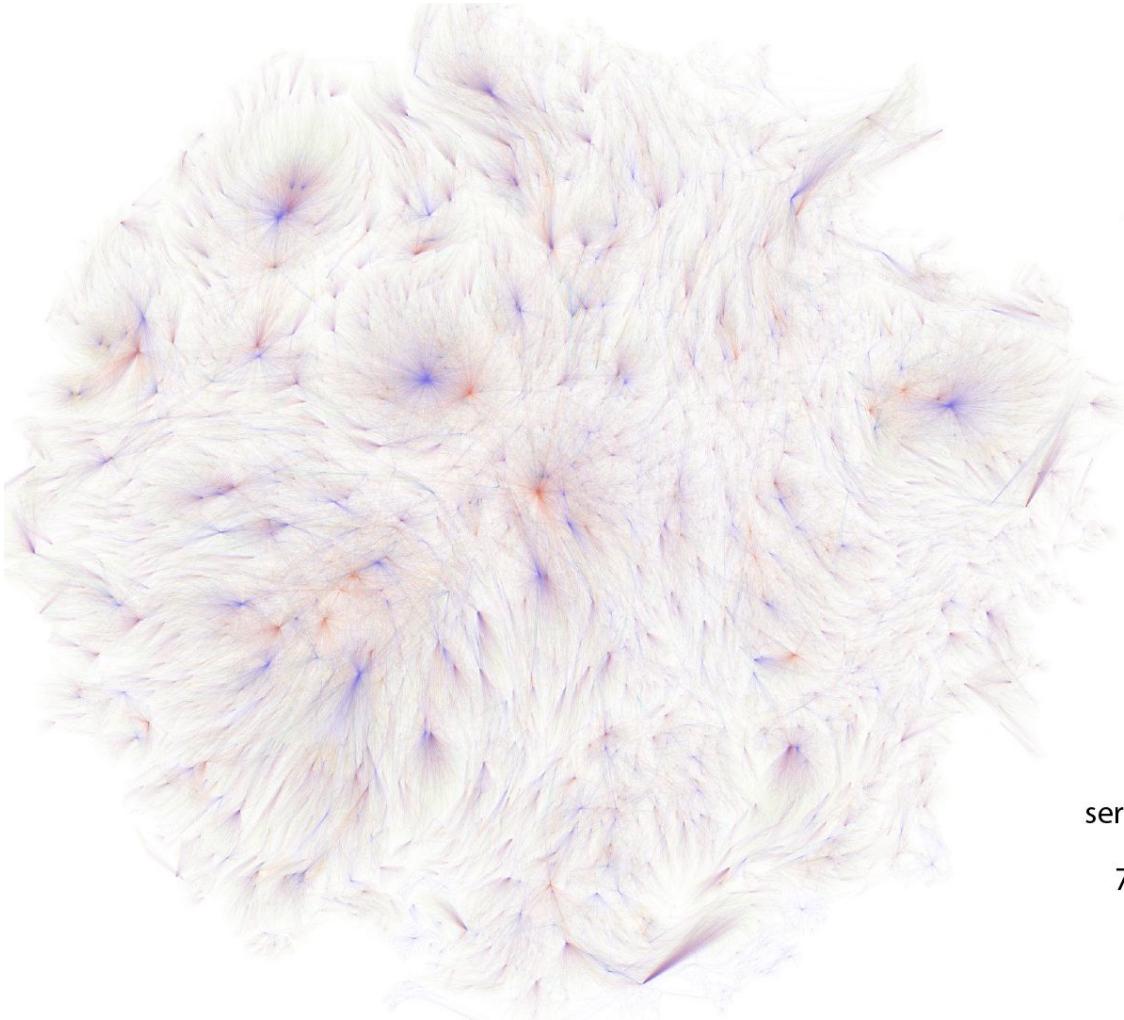
A backbone's strength grows as data are integrated across it,
or is a backbone the collective integration?











A pretty looking.... bone?

Dmitry Dmitriev et al., 3i - Auchenorrhyncha

~15 years of taxonomic research

nomenclature only

~1 million rows

~ 90 edge types

dot/graphviz

scripted off the API

serialized to RDF perhaps 100 - 150 million triples?

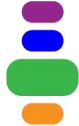
79% of citations from sources older than 2000



Approach*

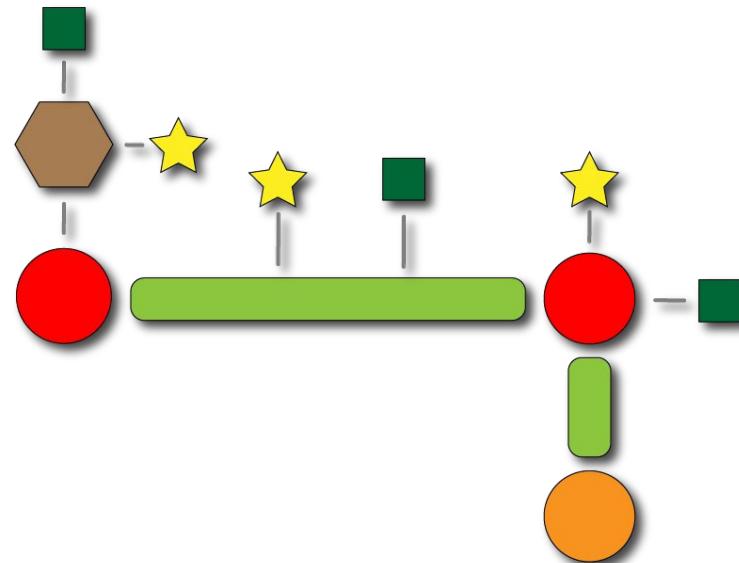
1. Consider rich “bespoke” knowledge-bases as background,
i.e. listen to the curators, taxonomists, and systematists
2. Don’t confuse biology and nomenclature
3. Model Things, not rows, or strings, or fields
4. Model all Things (almost) equally
5. Speak the standard, don’t be constrained by it

* To building a next-generation biodiversity informatics workbench.



NOMEN

-  TaxonName
 -  TaxonName subclass
 -  TaxonNameRelationship
 -  TaxonNameClassification
 -  Citation
 -  Annotation





Implementation

Project management

- Source being processed
- Source processed - Biology
- Source processed - Complete
- Source processed - Distribution
- Source Processed Except for Figures
- Source processed - Nomenclature
- Source unprocessed

Chronology

Browse nomenclature

Autocomplete

Toggle to biological taxa

Metadata views

Quick forms

Pin items

Navigation

The screenshot shows the TaxonWorks interface with the following sections:

- Browse nomenclature:** A search bar with placeholder "Select a taxon name" and a dropdown for "Redirect to valid name". Below it is a "Navigate" section with "Previous", "Random", and "Next" buttons, and tabs for "Valid", "Invalid", and "Both". A sidebar lists sources: Root, Lepidoptera, Aluctoidea, Aluctidae, Orneodes, Aluctinae, Lepidoptera, Aluctidae, Leach, 1815, and Orneodes Latreille, 1796.
- Metadata views:** A central area showing the history of the genus *Orneodes* Latreille, 1796. It includes a "Summary" section with validation status: "This name is not valid/accepted. The valid name is *Alucta* Linnaeus, 1758. The record is for reference. There are no records in this entry." Below this is a "Soft Validation" section with various error messages (e.g., etymology missing, gender not specified) and a "Relationships" section with a note about the original publication.
- Quick forms:** A horizontal bar at the top with buttons for "Project", "Administration", "Account", "Sign out", and "more".
- Pin items:** A feature allowing users to pin specific items to the interface.
- Navigation:** A circular navigation menu with various icons for different functions like search, browse, and analysis.

Depictions

Depictions



Infinite identifiers

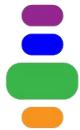


Annotate anywhere
(e.g. cite)

Validation (data quality)
Intelligent fixes

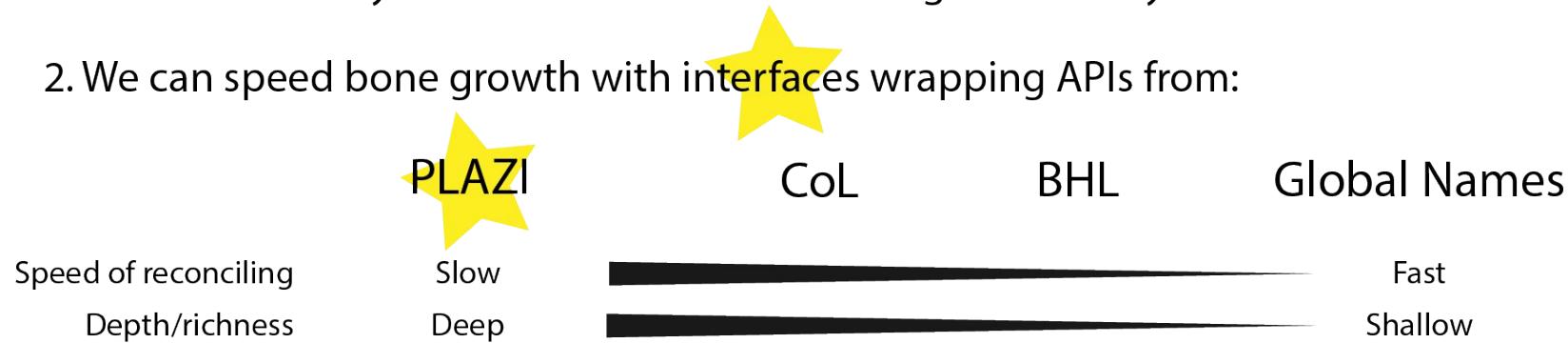
Custom attributes

- Data attributes**
- Taxonomic rank: 50.00
 - Original Genus: ORIGINAL GENUS UNDETERMINED
 - Availability: 100% objective synonym
 - Valid name: Alucta
 - BurnettNo: 20548.00
 - Genus_code: 08-09-00-v02.50
 - Petro_G: PetroG-09-00-G00
 - Frontal_image: l0c0500_3PQ
 - Back_image: l0c0500_3PQ
 - GENUS_NOM: Any & Fischer (1991) stated - A junior synonym of the genus of *Alucta* Linnaeus, 1758.
 - GENUS_REF: 3260
 - TS_REF: 40229
 - TS_GENUS: Phalaenidae
 - TS_FAMILY: Lepidoptera
 - TS_AUTHOR: Linnaeus
 - TS_YEAR: 1758
 - TS_PAGE_COMMENT: (but included by Latreille as *Phalaenae hexadactyla* Falterlin, an incorrect authority) by subsequent monotypy



Limitations

1. We need to unify/centralize a wealth of existing community-created documentation
2. We can speed bone growth with interfaces wrapping APIs from:



3. We can work on standards, DwC and CoLDp are limited in their capacity to use NOMEN, impacting export and import quality
4. “Full” support for Botanical nomenclature
5. Output, output, output



Outcomes

Global

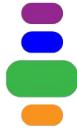
5.8% of the Catalogue of Life arrives through TaxonWorks
That's nomenclature for **117,547** (biological) species.

People

14308 hours of (human) curation
~40 curators/week strengthening backbones

Sharing

5339 People are shared over > 1 project, **101185** added
43573 Sources are shared over > 1 project, **121010** added



Output

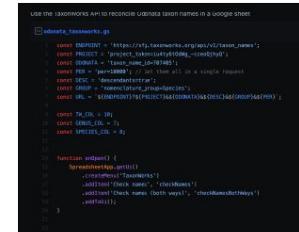
Exports

DwC Archive,
NeXML,
Nexus,
CoLDP,
CSV,
BiBTeX
SQL

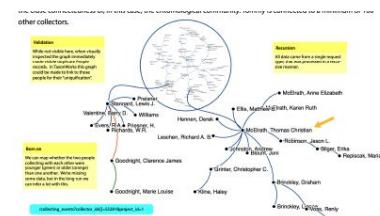
JSON API Code



Autocomplete



Genomic pipelines



Data analysis

1 step publishing.

Websites

for biologists *and* developers...

The image is a collage of screenshots from several biodiversity informatics platforms, each showing a different aspect of data management and analysis:

- TaxonWorks**: Shows the "Filter sources" interface with sections for "In Project" and "Search". It includes filters for "Text in full citation", "Type (Text, Biotics, Website, Person)", "Authors", "Date (Start year, End year)", and "Serials".
- DwC-A Workbench**: Displays a table of 10000 v2-9999 records with columns for Options, Status, catalogNumber, identifiedBy, dateIdentified, TW:CollectionObjectLifeStageSet, preparations, loanDestination, typeStatus, occurrenceRemarks, TW:CollectionObjectListNotes, eventID, and recordedBy. A progress bar at the top indicates the status of imports.
- Repository**: Features a central circular navigation hub with links to Identifiers, Data attributes, Notes, Tags, Confidences, Depictions, and Citations. It also includes sections for Determination, CollectingEvent, and Verbatim, along with a map view.
- Filter images**: A grid of small images of insects, likely beetles, used for identification or comparison.



TaxonWorks Together

December 6th-11th

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

Weekly meetings

Species File Group Community Liaison - Debbie Paul - dlpaul@illinois.edu



Citations

TaxonWorks is open source, follow links from <https://taxonworks.org>

The autocomplete by José Luis Pereira: https://github.com/SpeciesFileGroup/taxonworks_autocomplete

The 3i based graph was first published at GLBIO 2021. See links from <https://github.com/mjy/presentations>

“The complete taxonomy of world harvestmen” is by Adriano Kury and collaborators at <https://wcolite.com/>

“Odonata Central” is at <https://www.odonatacentral.org>

The genomics pipeline Google app integration of Odonate names by Raef Fraenze is at
<https://gist.github.com/rafelafrance/d1e85fe07e0697614b9271d3798dfa35>

The NOMEN ontology used in TaxonWorks is at <https://github.com/SpeciesFileGroup/nomen>,
and see https://biss.pensoft.net/articles.php?journal_name=biss&id=20284

The SFG explores the creation of public taxon pages, join us here: <https://github.com/SpeciesFileGroup/page>

More on backbone types: https://github.com/SpeciesFileGroup/taxonworks_doc/tree/master/concepts

Stats updated in real time for known instances of TaxonWorks: <https://stats.taxonworks.org>

Kojun Kanda shared his beetle images.