De-centralized but global: Redesigning biodiversity data aggregation for improved engagement and impact

Nico Franz, Ed Gilbert & Beckett Sterner

Biodiversity Knowledge Integration Center – Arizona State University

Keynote Session – June 10, 2019

3rd Annual Digital Data Conference – Yale University, New Haven, CO
0. Key message: Biodiversity informatics remains young and fresh; especially if we aim to incentivize experts/enthusiasts in publishing high-quality, "data-intelligent" biodiversity data products.
Structure of presentation

0. Key message: Biodiversity informatics remains young and fresh; especially if we aim to incentivize experts/enthusiasts in publishing high-quality, "data-intelligent" biodiversity data products.

1. "New at ASU" (and more generally): The NEON Biorepository Data Portal – a Darwin Core-based ecological monitoring and forecasting resource.
Structure of presentation

0. Key message: Biodiversity informatics remains young and fresh; especially if we aim to incentivize experts/enthusiasts in publishing high-quality, "data-intelligent" biodiversity data products.

1. "New at ASU" (and more generally): The NEON Biorepository Data Portal – a Darwin Core-based ecological monitoring and forecasting resource.

2. Rethinking centralized biodiversity data aggregation: Diagnosis and components of a de-centralized complement.
Structure of presentation

0. Key message: Biodiversity informatics remains young and fresh; especially if we aim to incentivize experts/enthusiasts in publishing high-quality, "data-intelligent" biodiversity data products.

1. "New at ASU" (and more generally): The NEON Biorepository Data Portal – a Darwin Core-based ecological monitoring and forecasting resource.

2. Rethinking centralized biodiversity data aggregation: Diagnosis and components of a de-centralized complement.
Introducing the **NEON Biorepository Data Portal**

https://biorepo.neonscience.org
Introducing the NEON Biorepository Data Portal

National Ecological Observatory Network (NEON)

- Continental scale tool for monitoring and forecasting ecological change
- Designed to run for **30 years** (~2018 – 2048), **81 sites** distributed throughout the US, including AL, HI, PR.
- Funded by NSF (~ $50-60 million annually), operated by Battelle with ASU as subcontractor.
Introducing the NEON Biorepository Data Portal

Climate monitoring tower & instrumentation

Example NEON site at the Santa Rita Experimental Range

Organismal sampling gear & field/domain team.
Introducing the NEON Biorepository Data Portal

Domain facilities process samples and send them to the NEON Biorepository at ASU.
Introducing the **NEON Biorepository Data Portal**

Where are we located?

**734 W Alameda Drive, Tempe, AZ 85282**
[https://www.asu.edu/map/interactive/?id=120&mrkld=66206](https://www.asu.edu/map/interactive/?id=120&mrkld=66206)

- Renovated and activated in the Fall of 2014.
- **Open to the public** (weekdays 9 am – 5 pm).
- **Free parking** available right at entrance.
Introducing the NEON Biorepository Data Portal
More information at Tuesday's poster session!

- Using the portal
- Access to samples
- New developments
- Informatics
- Research
- Prospective M.Sc./Ph.D. applicants and postdocs
- Early and broad
A short demonstration

https://biorepo.neonscience.org
[after a short BioRepo portal demo (Culicidae / SRER Plants)] *


Change in species richness per site.

• NEON organismal sampling produces a unique, taxon-/region-/time-constrained, change-focused data signal.

* Thanks to Dr. Kelsey Yule, NEON Biorepository @ ASU.
To increase trust, change the social design behind aggregated biodiversity data

Nico M Franz, Beckett W Sterner

Database, Volume 2018, Issue 100, https://doi.org/10.1093/database/bax100
Published: 04 January 2018

Abstract
Growing concerns about the quality of aggregated biodiversity data are lowering trust in large-scale data networks. Aggregators frequently respond to quality concerns by recommending that biologists work with original data providers to correct errors ‘at the source.’ We show that this strategy falls systematically short of a full diagnosis of the underlying causes of distrust. In particular, trust in an aggregator is not just a feature of the data signal quality provided by the sources to the aggregator, but also a consequence of the social design of the aggregation process and the resulting power balance between individual data contributors and aggregators. The latter have created an accountability gap by downplaying the authorship and significance of the taxonomic hierarchies—frequently called ‘backbones’—they generate, and which are in effect novel classification theories that operate at the core of data-structuring processes. The Darwin Core standard for sharing occurrence records plays an under-appreciated role in maintaining the accountability gap, because this standard lacks the syntactic structure needed to preserve the taxonomic coherence of data packages submitted for aggregation, potentially leading to inferences that no individual source would support. Since high-quality data packages can mirror competing and conflicting classifications, i.e. unsettled systematic research, this plurality must be accommodated in the design of biodiversity data integration. Looking forward, a key directive is to develop new technical pathways and social incentives for experts to contribute directly to the validation of taxonomically coherent data packages as part of a greater, trustworthy aggregation process.

Issue Section: Perspective/Opinion

• https://doi.org/10.1093/database/bax100

• Critique of monolithic aggregation designs.
Rethinking centralized biodiversity data aggregation (2)

- https://doi.org/10.1093/database/bax100

- Critique of monolithic aggregation designs.

- Here we reformulate the issue in relation to the direction of data flow in hierarchical biodiversity data aggregation networks.
To increase trust, change the social design behind aggregated biodiversity data

Nico M Franz, Beckett W Sterner

Database, Volume 2018, 2018, bax100, https://doi.org/10.1093/database/bax100
Published: 04 January 2018  Article History

Abstract
Growing concerns about the quality of aggregated biodiversity data are lowering trust in large-scale data networks. Aggregators frequently respond to quality concerns by recommending that biologists work with original data providers to correct errors at the source. We show that this strategy falls systematically short of a full diagnosis of the underlying causes of distrust. In particular, trust in an aggregator is not just a feature of the data signal quality provided by the sources to the aggregator, but also a consequence of the social design of the aggregation process and the resulting power balance between individual data contributors and aggregators. The latter have created an accountability gap by downplaying the authorship and significance of the taxonomic hierarchies—frequently called ‘backbones’—they generate, and which are in effect novel classification theories that operate at the core of data-structuring processes. The Darwin Core standard for sharing occurrence records plays an under-appreciated role in maintaining the accountability gap, because this standard lacks the syntactic structure needed to preserve the taxonomic coherence of data packages submitted for aggregation, potentially leading to inferences that no individual source would support. Since high-quality data packages can mirror competing and conflicting classifications, i.e. unsettled systematic research, this plurality must be accommodated in the design of biodiversity data integration. Looking forward, a key directive is to develop new technical pathways and social incentives for experts to contribute directly to the validation of taxonomically coherent data packages as part of a greater, trustworthy aggregation process.

Issue Section: Perspective/Opinion

- https://doi.org/10.1093/database/bax100
- Critique of monolithic aggregation designs.
- Here we reformulate the issue in relation to the direction of data flow in hierarchical biodiversity data aggregation networks.
- Then we outline a complementary solution that is not centralized, yet which can potentially reach global coverage.
Rethinking centralized biodiversity data aggregation

- Critique of monolithic aggregation designs.
- Here we reformulate the issue in relation to the direction of data flow in hierarchical biodiversity data aggregation networks.

Then we outline a complementary solution that is not centralized, yet which can potentially reach global coverage.
Centralized biodiversity data aggregation

- Many collections remain offline.
Centralized biodiversity data aggregation

• Other, individual collections, may have a web presence.
Centralized biodiversity data aggregation

- Multi-collections software applications can bring institutional datasets on-line; yet this alone may not suffice to **publish DwC-Archive data** according to **FAIR standards**.
Centralized biodiversity data aggregation

• The Integrated Publishing Toolkit (IPT) allows individual or multi-collection datasets to **become discoverable** as DwC-Archive packages to higher-level aggregators.
Centralized biodiversity data aggregation

- There are also **web-only, mid-level portal applications** that support live collection management and can publish "up" through the (IPT).

- **Web-Only, "Mid-Level" Portals**
  - Live-Managed Collection

- **Integrated Publishing Toolkit**
  - DwC-As Are Discoverable

- **Live Data Management**
  - No or Institutional Aggregation
  - IPT Needed for DwC-A Publishing (FAIR)
Centralized biodiversity data aggregation

- Symbiota portals also support "snapshot collections" – i.e., periodical, manually triggered batch re-/uploads of static versions that are live-managed elsewhere.
Centralized biodiversity data aggregation

• Symbiota portals also have a custom, fully built-in, IPT-analogous "Darwin Core Archive Publishing" module.
Example of Symbiota's DwC-A Publishing module (SCAN: ASUCOB)
Centralized biodiversity data aggregation

- Multiple, community-themed portals – each with unique live/snapshot collection profiles – can periodically receive **reciprocal snapshot updates**.
Highest-level aggregators typically only support collection snapshots!

- Continental to Global Aggregators
- Overlapping Sets of Snapshots

- Web-Only, "Mid-Level" Portals
  - Live-Managed Collection
  - Snapshot Collection (Version)

- Integrated Publishing Toolkit
  - DwC-As Are Discoverable

- Live Data Management
  - No or Institutional Aggregation
  - IPT Needed for DwC-A Publishing (FAIR)
This hierarchy sustains an **imbalance in directional data flow**: Annotations on global datasets are hard to pull downwards.
Moreover, by the time we reach the top, most experts/enthusiasts no longer feel at home (cf. Wenger 2000).
Wenger 2000: **Communities of practice and social learning systems**

<table>
<thead>
<tr>
<th>Table 2. Boundary Dimensions</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Coordination</strong></td>
</tr>
<tr>
<td><strong>Engagement</strong></td>
</tr>
<tr>
<td><strong>Imagination</strong></td>
</tr>
<tr>
<td><strong>Alignment</strong></td>
</tr>
</tbody>
</table>
Connecting data and expertise: a new alliance for biodiversity knowledge

Donald Hobern, Brigitte Baptiste, Kyle Copas, Robert Guralnick, Andrea Hahn, Edwin van Huis, Eun-Shik Kim, Melodie McGeoch, Isayvani Naicker, Laetitia Navarro, Daniel Noesgaard, Michelle Price, Andrew Rodrigues, Dmitry Schigel, Carolyn A. Sheffield, John Wieczorek

Abstract

There has been major progress over the last two decades in digitising historical knowledge of biodiversity and in making biodiversity data freely and openly accessible. Interlocking efforts bring together international partnerships and networks, national, regional and institutional projects and investments and countless individual contributors, spanning diverse biological and environmental research domains, government agencies and non-governmental organisations, citizen science and commercial enterprise. However, current efforts remain inefficient and inadequate to address the global need for accurate data on the world’s species and on changing patterns and trends in biodiversity. Significant challenges include imbalances in regional engagement in biodiversity informatics activity, uneven progress in data mobilisation and sharing, the lack of stable persistent identifiers for data records, redundant and incompatible processes for cleaning and interpreting data and the absence of functional mechanisms for knowledgeable experts to curate and improve data.
Designing for strong data communities
De-centralized, but global

- Independent, themed portal communities maintain **live-managed collections**.

[Diagram of portal communities with labels A, B, and C, and numbers 1 to 14 indicating live-managed collections.]

A: Consortium of Northeastern Herbaria
B: [Southwest Environment Information Network]
C: Smithsonian Tropical Research Institute

A–F = Portals  |  1–20 = Collections
"Live-Managed" Collection
De-centralized, but global

- **Partial, relevant collection snapshot subsets** are represented.
De-centralized, but global

- Even **partial, relevant portal snapshot subsets** are ingestible, with provenance.

A–F = Portals  |  1–20 = Collections  (Partial)
- "Live-Managed" Collection
- "Snapshot" Collection (Vs.)

A: Consortium of Northeastern Herbaria
B: [Southwest Environment Information Network]
C: Smithsonian Tropical Research Institute
D: National Ecological Observatory Network
De-centralized, but global

- Some research-themed portals may only include partial collection snapshots.
De-centralized, but global

- Highly configurable **portal-to-portal APIs** negotiate the flow of data **between live and snapshot collection instances**.
De-centralized, but global

- As API services are optimized, the distinction between live and snapshot collection management increasingly falls away.

A–F = Portals | 1–20 = Collections (Partial)
- "Live-Managed" Collection = Direct Updates
- "Snapshot" Collection (Vs.) = API Updates

A: Consortium of Northeastern Herbaria
B: [Southwest Environment Information Network]
C: Smithsonian Tropical Research Institute
D: National Ecological Observatory Network
E: Cooperative Taxonomic Resource for American Myrtaceae

Portal-to-Portal API Services
De-centralized, but global

- API service configurations include **filtered**, source-/sink-approval contingent data pushes and pulls.

A–F = Portals  |  1–20 = Collections  (Partial)
- "Live-Managed" Collection  = Direct Updates
- "Snapshot" Collection (Vs.)  = API Updates
De-centralized, but global

- API service configurations allow **filtered** \{collection, taxon, region, etc.\}, source-/sink-approval contingent **data pushes and pulls**.

---

A: Consortium of Northeastern Herbaria  
B: [Southwest Environment Information Network]  
C: Smithsonian Tropical Research Institute  
D: National Ecological Observatory Network  
E: Cooperative Taxonomic Resource for American Myrtaceae
De-centralized, but global

- Portal-to-portal **API configurations** become the "substrate" upon which the communities realize their "modes of belonging".
De-centralized, but global

- The de-centralized network is **broadly extensible** between closely (high data flow) or remotely (low data flow) related communities.

A–F = **Portals** | 1–20 = **Collections** (Partial)
- "Live-Managed" Collection = Direct Updates
- "Snapshot" Collection (Vs.) = API Updates

Higher-/Lower-volume data flow

---

**A:** Consortium of Northeastern Herbaria
**B:** [Southwest Environment Information Network]
**C:** Smithsonian Tropical Research Institute
**D:** National Ecological Observatory Network
**E:** Cooperative Taxonomic Resource for American Myrtaceae

---

**C. STRI**

- 13
- 14
- 2P
- 4P
- 9P

**DwC-A Publish**;

- **13P**
- **8P**
- **17P**
- **18P**

**E. CoTRAM**

- **3P**
- **8P**
- **17P**
- **18P**

**Automated & Filtered Push / Pull Data Updates**

**Portal-to-Portal API Services**

**Collection-Region-Taxon-Configurable**

**D. NEON**

- 15
- 16
- 17
- AP
- BP

---

**DwC-A Publish**

- **8P**
- **9P**
- **10P**
- **12P**

**B. SEINet**

- 11
- 12
- 1P
- 3P
- 5P
- 6P
- 7P

---

**A:** CNH

- 1
- 2
- 3
- 4
- 5
- 6
- 7
- 8P
- 10P
- 12P

---

**E. CoTRAM**

- **3P**
- **13P**
- **8P**
- **17P**
- **18P**

---

**C. STRI**

- 13
- 14
- 2P
- 4P
- 9P

---

**D. NEON**

- 15
- 16
- 17
- AP
- BP

---

**B. SEINet**

- 11
- 12
- 1P
- 3P
- 5P
- 6P
- 7P

---

**A:** CNH

- 1
- 2
- 3
- 4
- 5
- 6
- 7
- 8P
- 10P
- 12P

---

**E. CoTRAM**

- **3P**
- **13P**
- **8P**
- **17P**
- **18P**

---

**C. STRI**

- 13
- 14
- 2P
- 4P
- 9P

---

**D. NEON**

- 15
- 16
- 17
- AP
- BP

---

**B. SEINet**

- 11
- 12
- 1P
- 3P
- 5P
- 6P
- 7P

---

**A:** CNH

- 1
- 2
- 3
- 4
- 5
- 6
- 7
- 8P
- 10P
- 12P
De-centralized, but global

- On the basis of a shared API service culture, a de-centralized data portal network can potentially grow to attain **global coverage**.
Designing for expert/enthusiast access
BioCache: Global access through custom research portal instances

- Researchers create and **register "via single handshake" a new portal instance.**

BioCache = Project-Driven Portal Instance(s)
"Handshake" with one portal + "DNS Propagation"

A–F = Portals  |  1–20 = Collections  (Partial)
- "Live-Managed" Collection = Direct Updates
- "Snapshot" Collection (Vs.) = API Updates

Higher-/Lower-volume data flow

---

**Diagrams:**
- **C. STRI**
- **D. NEON**
- **E. CoTRAM**
- **A: CNH**
- **B: SEINet**

---

A: Consortium of Northeastern Herbaria
B: [Southwest Environment Information Network]
C: Smithsonian Tropical Research Institute
D: National Ecological Observatory Network
E: Cooperative Taxonomic Resource for American Myrtaceae
F: Red de Herbarios del Noroeste de México
BioCache: Global access through custom research portal instances

- Research instances enable repeatable, global data queries and re-ingestion.

BioCache = Project-Driven Portal Instance(s)
"Handshake" with one portal + "DNS Propagation"
Supports network-wide, repeated data ingestion

A–F = Portals | 1–20 = Collections (Partial)
- "Live-Managed" Collection = Direct Updates
- "Snapshot" Collection (Vs.) = API Updates

Higher-/Lower-volume data flow

A: Consortium of Northeastern Herbaria
B: [Southwest Environment Information Network]
C: Smithsonian Tropical Research Institute
D: National Ecological Observatory Network
E: Cooperative Taxonomic Resource for American Myrtaceae
F: Red de Herbarios del Noroeste de México
**BioCache: Global access through custom research portal instances**

- Valued-added data can return to all (live) source collections via filtered pulls.

BioCache = Project-Driven Portal Instance(s)

"Handshake" with one portal + "DNS Propagation"
Supports network-wide, repeated data ingestion

Annotations: Filtered pull back into live sources

A–F = Portals | 1–20 = Collections (Partial)

- "Live-Managed" Collection = Direct Updates
- "Snapshot" Collection (Vs.) = API Updates

Higher-/Lower-volume data flow

---

A: Consortium of Northeastern Herbaria
B: [Southwest Environment Information Network]
C: Smithsonian Tropical Research Institute
D: National Ecological Observatory Network
E: Cooperative Taxonomic Resource for American Myrtaceae
F: Red de Herbarios del Noroeste de México
Stay tuned, it's underway
Taxonomic data intelligence for Darwin Core occurrences

- https://doi.org/10.3897/rio.2.e10610

- Engaging expert/enthusiast communities ⇔ need for pluralism and democracy for and among taxonomic perspectives in biodiversity data aggregation designs.
Alignment by Alan Weakley (http://herbarium.unc.edu/flora.htm)

\[
\begin{array}{c|c|c|c}
\{AZ, NM\} & \{CO, WY, MT, AB, eBC\} & \{OR, WA, wBC\} \\
Abies lasiocarpa var. arizonica & Abies lasiocarpa var. lasiocarpa "A" & \\
Abies bifolia "B" & Abies lasiocarpa & \\
B\text{'}a & A*B & A\text{'}b \\
"B, not a" & "A and B" & "A, not b" \\
\end{array}
\]

Distribution

  => USDA Plants Database
  => Flora of North America

Split-concept labels
Taxonomic data intelligence for Darwin Core occurrences

• https://doi.org/10.3897/rio.2.e10610

• Engaging expert/enthusiast communities ⇔ need for pluralism and democracy for and among taxonomic perspectives in biodiversity data aggregation designs.

• Spatial reasoning tools (RCC–5) can help attain consistent and comprehensive taxonomic meaning mappings ⇔ intelligence for data integration across evolving or conflicting views.
Controlling the taxonomic variable: Taxonomic concept resolution for a southeastern United States herbarium portal

Nico Franz, Edward Gilbert, Bertram Ludäscher, Alan Weakley

Executive summary -

Overview. Taxonomic names are imperfect identifiers of specific and sometimes conflicting taxonomic perspectives in aggregated biodiversity data environments. The inherent ambiguities of names can be mitigated using syntactic and semantic conventions developed under the taxonomic concept approach. These include: (1) representation of taxonomic concept labels (TCLs: name sec. source) to precisely identify name usages and meanings, (2) use of parent/child relationships to assemble separate taxonomic perspectives, and (3) expert provision of Region Connection Calculus articulations (RCC–5: congruence, [inverse] inclusion, overlap, exclusion) that specify how data identified to different-sourced TCLs can be integrated. Application of these conventions greatly increases trust in biodiversity data networks, most of which promote unitary taxonomic ‘synteses’ that obscure the actual diversity of expert-held views. Better design solutions allow users to control the taxonomic variable and thereby assess the robustness of their biological inferences under different perspectives. A unique constellation of prior efforts — including the powerful Symbiota collections software platform, the Euler/X multi-taxonomy alignment toolkit, and the "Weakley Flora" which entails 7,000 concepts and more than 75,000 RCC-5 articulations — provides the opportunity to build a first full-scale concept resolution service for SERNEC, the Southeast Regional Network of Expertise and Collections, currently with 60 member herbaria and 2 million occurrence records.

Intellectual merit. We have developed a multi-dimensional, step-wise plan to transition SERNEC’s data culture from name- to concept-based practices. (1) We will engage SERNEC experts through annual, regional workshops and follow-up interactions that will foster buy-in and ultimately the completion of 12 community-identified use cases. (2) We will leverage RCC-5 data from the Weakley Flora and further development of the Euler/X logic reasoning toolkit to provide comprehensive genus- to variety-level concept alignments for at least 10 major flora treatments with highest relevance to SERNEC. The visualizations and estimated >1 billion inferred concept-to-concept relations will effectively drive specimen data integration in the transformed portal. (3) We will expand Symbiota’s taxonomy and occurrence schemas and related user interfaces to support the new concept data, including novel batch and map-based specimen determination modules, with easy output options in Darwin Core Archive format. (4) Through combinations of the new

- https://doi.org/10.3897/rio.2.e10610
- Engaging expert/enthusiast communities ⇔ need for pluralism and democracy for and among taxonomic perspectives in biodiversity data aggregation designs.
- Spatial reasoning tools (RCC–5) can help attain consistent and comprehensive taxonomic meaning mappings ⇔ intelligence for data integration across evolving or conflicting views.
- Biological inferences become robust
"Controlling the taxonomic variable"

"Please select your preference (A - D); we can perform all translations."

Impact:

The (formerly) federal standard.

The [current] regional flora.

The best, latest consensus.

The bible.
Project remains pending, but look!
"Taxonomically intelligent data integration for a new Flora of Alaska"

Blog 1: Project roadmap

Posted by Cem on 2018-08-17

Hi! Welcome to this development blog, this website, and this project. Steffi and I were so pleased to get this NSF funding — many, many thanks to the ABI panelists, the four reviewers, and to NSF and US taxpayers.

In this blog, I'll lay out the main steps in the development of the informatics infrastructure. (For a general overview of the project, see here.) Here is an overview diagram:

The software and data elements to be constructed and integrated are:

1. A core database: a standard MySQL (MariaDB) database, running on a web-hosting platform (Dreamhost). Over the years I've played with many database paradigms, including XML databases (ExistDB) and graph databases (e.g., Astore, Allegrograph).

- NSF DBI 1759964
- PIs Ickert-Bond & Webb
- Reconciling Hulten, FNA & Pan-Arctic Flora
- See http://alaskaflora.org/
iNaturalist is mostly there already

Taxon Framework Relationships

<table>
<thead>
<tr>
<th>Relationship</th>
<th>Taxon Framework Relationship 332324</th>
</tr>
</thead>
<tbody>
<tr>
<td>Covered by a taxon framework for Class Magnoliopsida Sourced to Plants of the World Online</td>
<td></td>
</tr>
<tr>
<td>Relationship: Match</td>
<td></td>
</tr>
<tr>
<td>Plants of the World Online</td>
<td>iNaturalist.org</td>
</tr>
<tr>
<td>Species Trapa hyncana (parent: Genus Trapa)</td>
<td>Match</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Relationship</th>
<th>Taxon Framework Relationship 332323</th>
</tr>
</thead>
<tbody>
<tr>
<td>Covered by a taxon framework for Class Lycopsida Sourced to Plants of the World Online</td>
<td></td>
</tr>
<tr>
<td>Relationship: Deviation</td>
<td></td>
</tr>
<tr>
<td>Plants of the World Online</td>
<td>iNaturalist.org</td>
</tr>
<tr>
<td>Species Lycopodium obtusifolium (parent: )</td>
<td>One-to-one</td>
</tr>
</tbody>
</table>

We follow PPG I in recognizing Phlegmariaurus as a segregate of Hupezia.

<table>
<thead>
<tr>
<th>Relationship</th>
<th>Taxon Framework Relationship 332322</th>
</tr>
</thead>
<tbody>
<tr>
<td>Covered by a taxon framework for Class Lycopsida Sourced to Plants of the World Online</td>
<td></td>
</tr>
<tr>
<td>Relationship: Deviation</td>
<td></td>
</tr>
<tr>
<td>Plants of the World Online</td>
<td>iNaturalist.org</td>
</tr>
</tbody>
</table>
iNaturalist is mostly there already

Taxon Framework Relationship 71572

Covered by a taxon framework for Family Nymphalidae Sourced to Markku Savela’s Lepidoptera and Some Other Life Forms

<table>
<thead>
<tr>
<th>Relationship: Deviation</th>
<th>Markku Savela’s Lepidoptera and Some Other Life Forms</th>
<th>iNaturalist.org</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subtribe Eubagina (parent: Tribe Biblidini)</td>
<td>Many-to-many</td>
<td>Tribe Ageroniini (parent: Subfamily Biblidinae)</td>
</tr>
<tr>
<td>Subtribe Epicallina (parent: Tribe Biblidini)</td>
<td></td>
<td>Tribe Epicalini (parent: Subfamily Biblidinae)</td>
</tr>
<tr>
<td>Subtribe Callicorina (parent: Tribe Biblidini)</td>
<td></td>
<td>Tribe Callicorini (parent: Subfamily Biblidinae)</td>
</tr>
<tr>
<td>Subtribe Ageroniina (parent: Tribe Biblidini)</td>
<td></td>
<td>Tribe Biblidini (parent: Subfamily Biblidinae)</td>
</tr>
<tr>
<td>Subtribe Epiphilina (parent: Tribe Biblidini)</td>
<td></td>
<td>Tribe Catonephelini (parent: Subfamily Biblidinae)</td>
</tr>
<tr>
<td>Tribe Biblidini (parent: Subfamily Biblidinae)</td>
<td></td>
<td>Tribe Eubagini (parent: Subfamily Biblidinae)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Tribe Epiphilini (parent: Subfamily Biblidinae)</td>
</tr>
</tbody>
</table>

iNat splits Catonephelini/ina from Epicaliini/ina and has Eubagini, Epicalini, Callicorini, Ageroniini. Epiphilini as tribes sibling to Biblidini rather than as subtribes within Biblidini.

Downstream deviations for Epicallini 1

https://www.inaturalist.org/taxon_framework_relationships
Recent addition: Concept alignment for phylogenomic trees

Reliable theories of multi-tree node congruence **require expert judgment**.

Franz, N.M. et al. 2019. Verbalizing phylogenomic conflict: [...] . [https://doi.org/10.1371/journal.pcbi.1006493](https://doi.org/10.1371/journal.pcbi.1006493)
Hopeful conclusions

Key message: Biodiversity informatics remains young and fresh; especially if we aim
to incentivize experts/enthusiasts in publishing high-quality, "data-intelligent" biodiversity data products.
Hopeful conclusions

Key message: Biodiversity informatics remains young and fresh; especially if we aim to incentivize experts/enthusiasts in publishing high-quality, "data-intelligent" biodiversity data products.

There are many grassroots or federally supported projects in this domain that express enthusiasm for, and confidence in, a strong future for new, high-volume, perpetually data-restructuring systematic research; driving our evolving views of relations between DwC datasets and biological knowledge.
Hopeful conclusions

Key message: Biodiversity informatics remains young and fresh; especially if we aim to **incentivize experts/enthusiasts** in publishing high-quality, "data-intelligent" biodiversity data products.

There are many grassroots or federally supported projects in this domain that express **enthusiasm for, and confidence in, a strong future for** new, high-volume, perpetually **data-restructuring systematic research**; driving our evolving views of relations between DwC datasets and biological knowledge.

If you have the passion and stomach for that future, **join us now!**
Acknowledgments

• Gil Nelson, Alnycea Blackwell, Jillian Goodwin, and all other iDigBio and Yale Peabody organizers and sponsors of #Digidata 2019.

• ASU Biocollections & NEON Biorepository team.

• Euler/X team: Bertram Ludäscher, Shizhuo Yu, Jessica Cheng.

• If you wish to read one paper on aligning taxonomic concepts: https://doi.org/10.1093/sysbio/syw023