

# Distributed, but global in reach: Outline of a decentralized paradigm for biodiversity data intelligence

Nico Franz, Edward Gilbert & Beckett Sternier

**ASU Biocollections & BioKIC**



**Biodiversity Next 2019 – Main Conference**  
SI87 – Empowering the taxonomic community by  
linking information through names and taxonomy (2)

Aalmarktzaal, Stadsgehoorzaal, Leiden, The Netherlands  
Friday, October 25, 2019

# Rethinking centralized biodiversity data aggregation

The screenshot shows the homepage of the **DATABASE** journal, Volume 2018. The main article title is "To increase trust, change the social design behind aggregated biodiversity data" by Nico M Franz and Beckett W Sterner. The abstract discusses concerns about aggregated biodiversity data quality and the role of taxonomic 'backbones' in maintaining trust. A sidebar lists other articles in the volume.

**Article Contents**

- Abstract
- Introduction
- No simple diagnosis for data quality deficiencies
- Importance of trust for sustained use
- Matching accountability to responsibility
- Generation of novel systematic syntheses
- Role of the DwC standard
- Aggregation and authorship
- Consequences of disenfranchising taxonomic experts
- Backbone-based data signal distortion: an example
- An objection
- Conclusion
- Acknowledgements
- Funding

**Issue Section:** Perspective/Opinion

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

## Key message

Global aggregation via taxonomic 'backbones' is well suited when also betting that future systematic research will *not* play a great (or any) role in *restructuring* biodiversity knowledge.

# Rethinking centralized biodiversity data aggregation

The screenshot shows the homepage of the DATABASE journal. At the top, the journal title 'DATABASE' is displayed with the subtitle 'The Journal of Biological Databases and Curation'. Below the title, there are navigation links for 'Volumes', 'Submit', 'Alerts', and 'About'. A dropdown menu for 'All Database' is also present. On the left side, there is a sidebar with the journal logo and the text 'Volume 2018' and '2018'. The main content area features a large image of the journal cover for Volume 2018. The article abstract is titled 'To increase trust, change the social design behind aggregated biodiversity data' by Nico M Franz and Beckett W Sterner. It includes the DOI (<https://doi.org/10.1093/database/bax100>), publication date (04 January 2018), and article history. Below the abstract, there are links for 'PDF', 'Split View', 'Cite', and 'Permissions'. The 'Abstract' section discusses concerns about aggregated biodiversity data quality and the social design of aggregation processes. The 'Article Contents' sidebar lists various sections of the article, including 'Abstract', 'Introduction', 'No simple diagnosis for data quality deficiencies', 'Importance of trust for sustained use', 'Matching accountability to responsibility', 'Generation of novel systematic syntheses', 'Role of the DwC standard', 'Aggregation and authorship', 'Consequences of disenfranchising taxonomic experts', 'Backbone-based data signal distortion: an example', 'An objection', 'Conclusion', 'Acknowledgements', and 'Funding'.

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

## Key message(s)

Global aggregation via taxonomic 'backbones' is well suited when also betting that future systematic research will *not* play a great (or any) role in *restructuring* our biodiversity knowledge.

However, it is a viable alternative expectation that new systematics research will *continue* to matter greatly.

**Q:** How to *redesign* aggregation for the latter?

# Centralized biodiversity data aggregation

- Many collections remain offline.

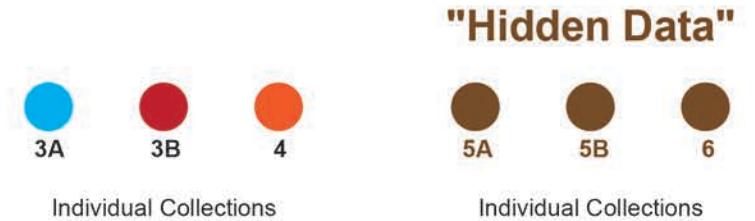
"Hidden Data"



Individual Collections

# Centralized biodiversity data aggregation

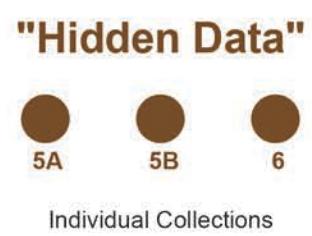
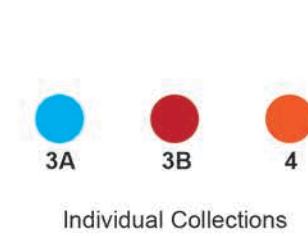
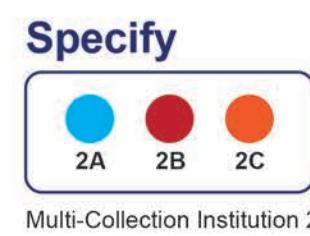
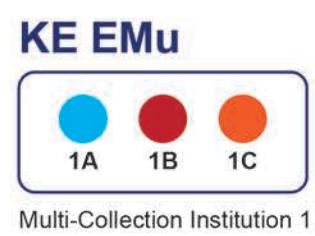
- Other, individual collections, may have a web presence.



# Centralized biodiversity data aggregation

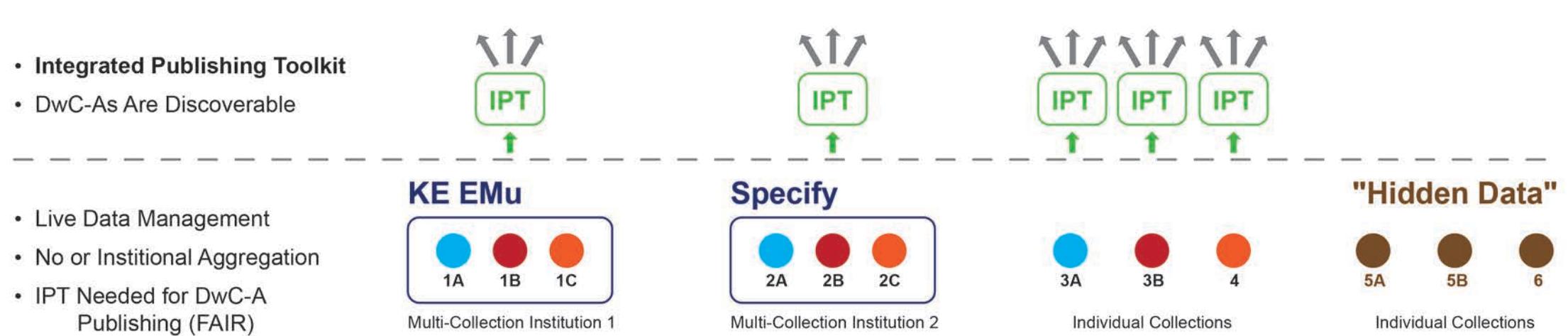
- Multi-collections software applications can bring institutional datasets on-line; yet this alone may not suffice to **publish DwC-Archive data** according to [FAIR standards](#).

- Live Data Management
- No or Institutional Aggregation
- IPT Needed for DwC-A Publishing (FAIR)



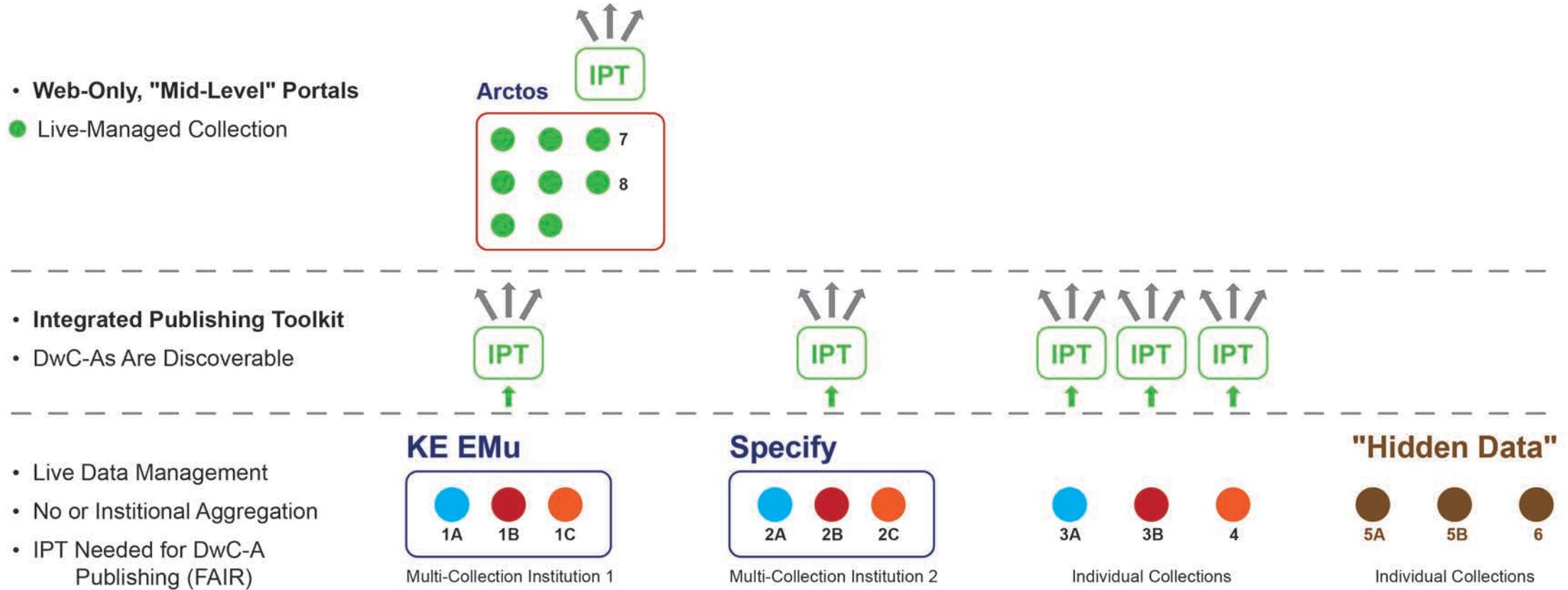
# Centralized biodiversity data aggregation

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



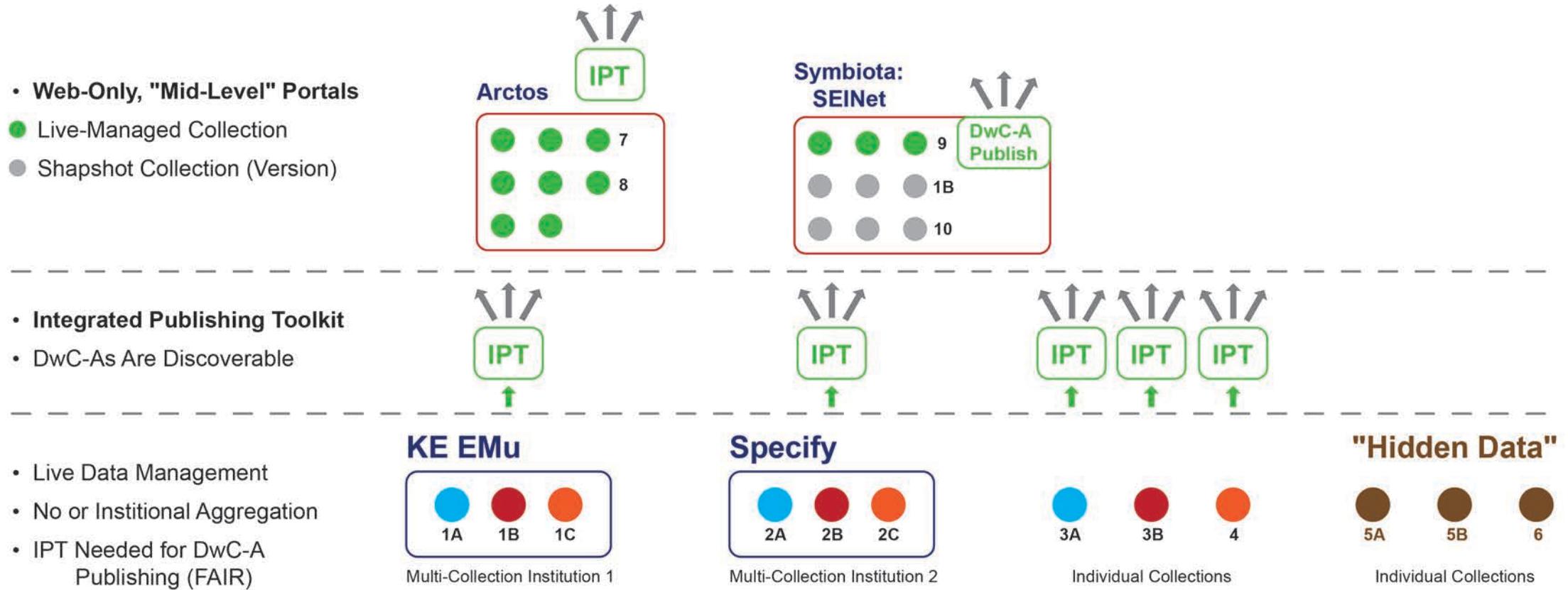
# Centralized biodiversity data aggregation

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



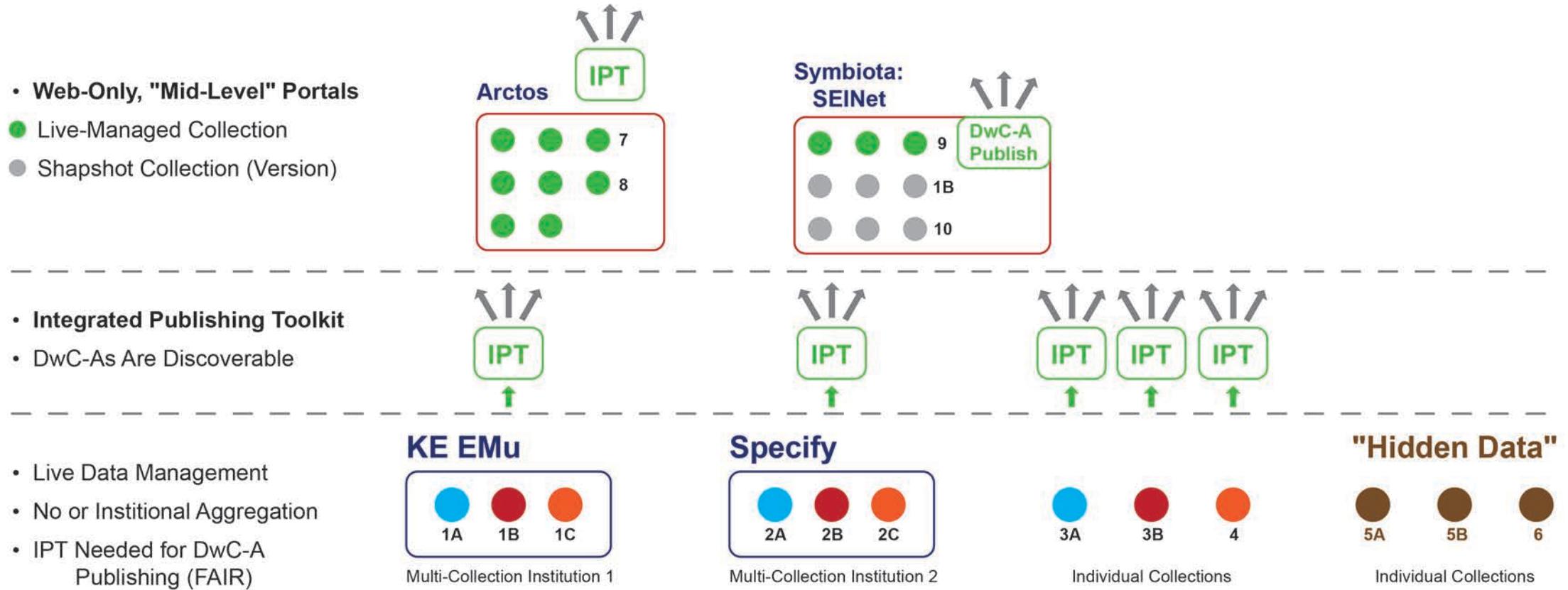
# Centralized biodiversity data aggregation

- 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: ASUHIC)

serv.biokic.asu.edu/ecdysis/collections/datasets/datapublisher.php

# ecdysis

A portal for live-data arthropod collections

Home | Search | Images | Inventories | Interactive Tools | Welcome! Noof. My Profile. Logout. Sitemap

Manage Collection Management in Darwin Core Archive Publisher

### Darwin Core Archive Publishing

Arizona State University Hasbrouck Insect Collection

Use the controls below to publish occurrence data from this collection as a Darwin Core Archive (DwC-A) file. DwC-A files are a single compressed ZIP file that contains one or several data files along with a meta.xml document that describes the content. The occurrence data file is required, but identifications (determinations) and image metadata are optional. Fields within the occurrences.csv file are defined by the Darwin Core exchange standard. We recommend that you also review instructions for Publishing Occurrence Data to GBIF and Publishing Occurrence Data to GBIF.

RSS Feed: <https://serv.biokic.asu.edu:443/ecdysis/webservices/dwc/rss.xml>

Creating DwC-A file ASU-ASUHIC\_DwC-A.zip  
Creating occurrence file (09:49:34 AM)...  
Done!! (09:49:40 AM)  
Creating meta.xml file (09:49:46 AM)...  
Done!! (09:49:47 AM)  
Creating image file (09:49:47 AM)...  
Done!! (09:49:48 AM)  
Creating occurrence file (09:49:48 AM)...  
Done!! (09:49:48 AM)  
Creating ent.xml (09:49:48 AM)...  
Done!! (09:49:48 AM)  
Mapping data to RSS feed...  
Done!!

Title: ASU-ASUHIC DwC-Archive X  
DwC-Archive: Darwin Core Archive: Arizona State University Hasbrouck Insect Collection  
EML: <https://serv.biokic.asu.edu:443/ecdysis/collections/datasets/itemhandler.php?collid=1>  
DwC-Archive File: [https://serv.biokic.asu.edu:443/ecdysis/content/dwca/ASU-ASUHIC\\_DwC-A.zip](https://serv.biokic.asu.edu:443/ecdysis/content/dwca/ASU-ASUHIC_DwC-A.zip)  
Publication Date: Thu, 24 Oct 2019 09:49:50

Publishing Information

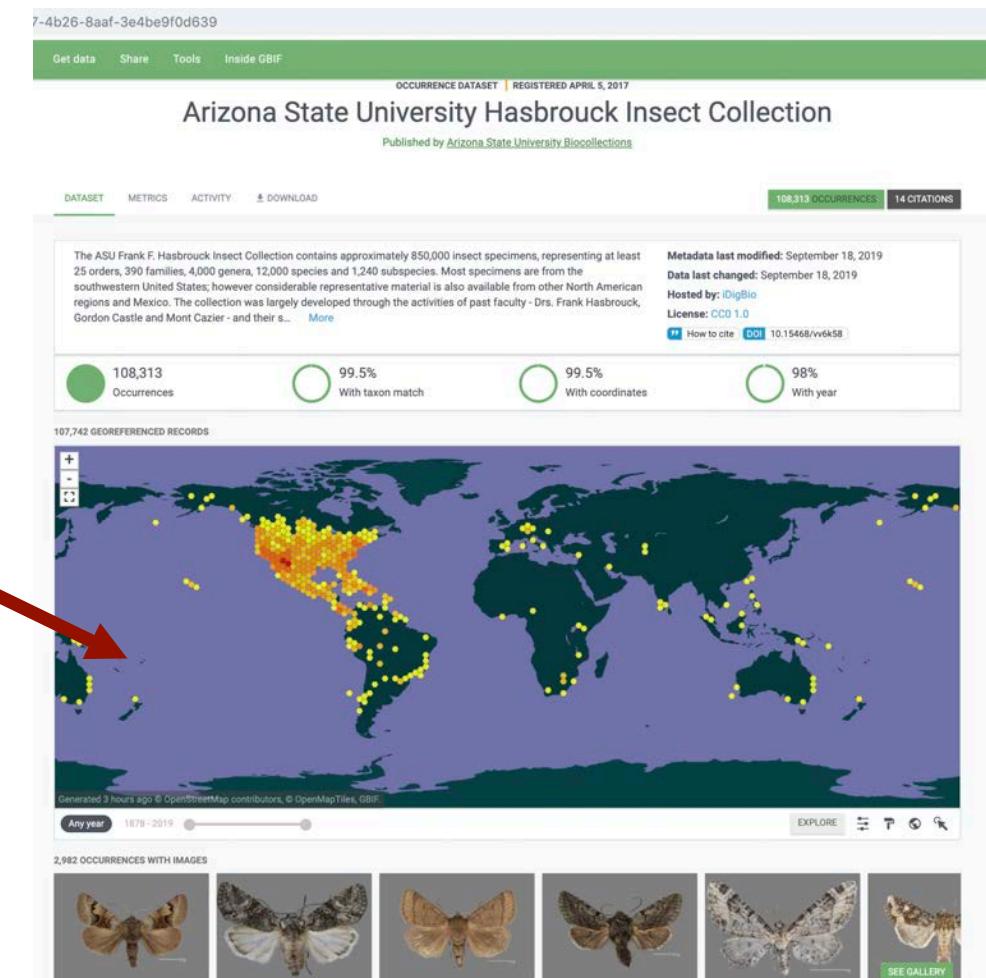
GUID source: symbiotaUUID

Public/Refresh DwC-A File

include Determination History  
 include Image URLs  
 Redact Sensitive Localities

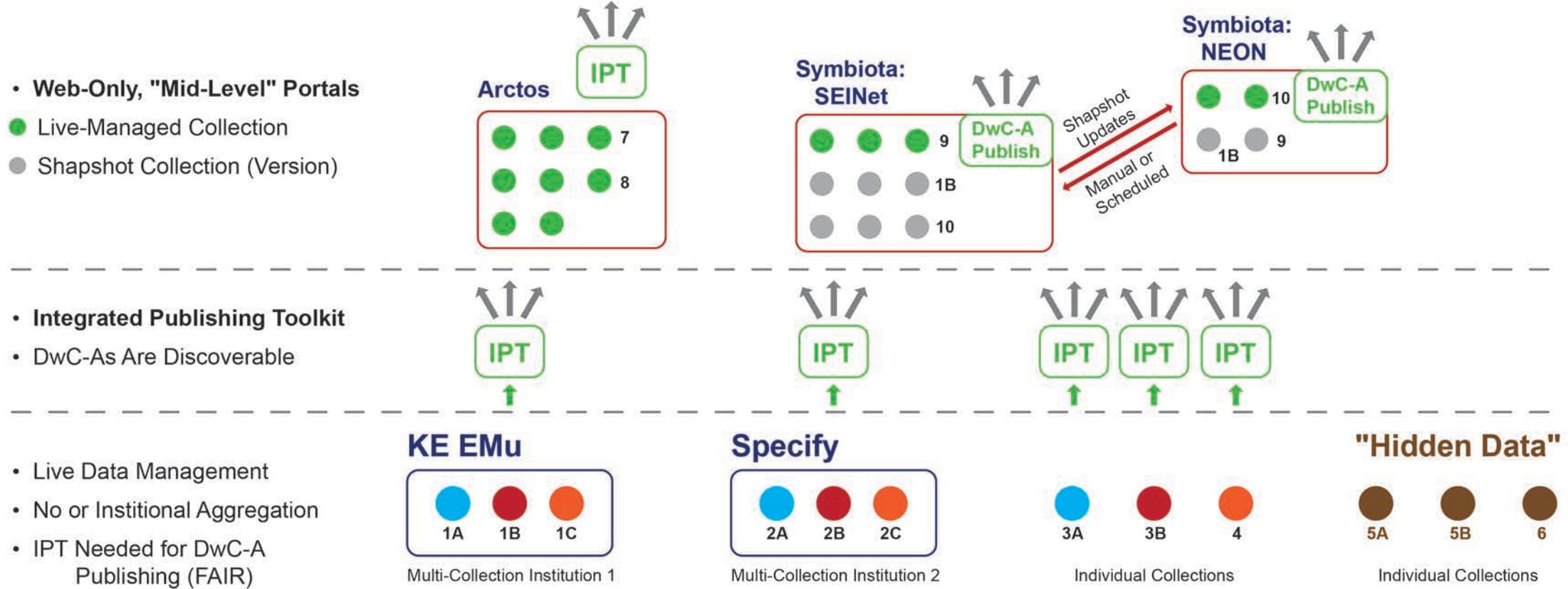
Create/Refresh Darwin Core Archive

ASUHIC  
October 24, 2019



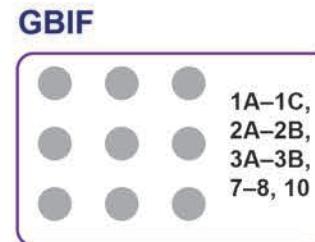
# Centralized biodiversity data aggregation

- Multiple, community-themed portals – each with unique live/snapshot collection profiles – can periodically receive **reciprocal snapshot updates**.

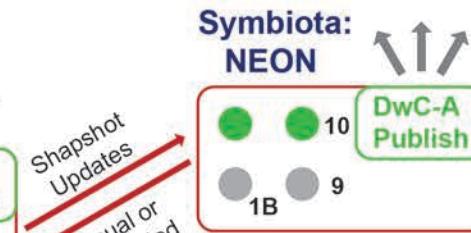
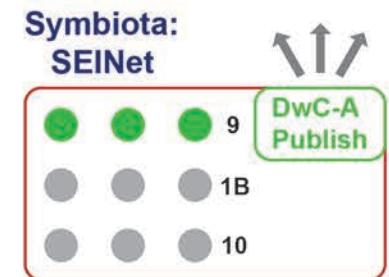
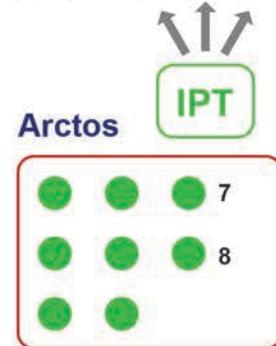


# Highest-level aggregators typically only support collection **snapshots**!

- Continental to Global Aggregators
- Overlapping Sets of Snapshots



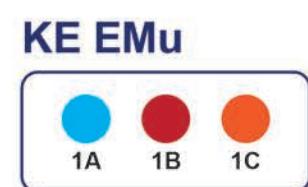
- Web-Only, "Mid-Level" Portals
- Live-Managed Collection
- Snapshot Collection (Version)



- Integrated Publishing Toolkit
- DwC-As Are Discoverable

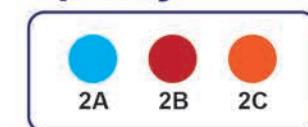


- Live Data Management
- No or Institutional Aggregation
- IPT Needed for DwC-A Publishing (FAIR)



Multi-Collection Institution 1

**Specify**



Multi-Collection Institution 2

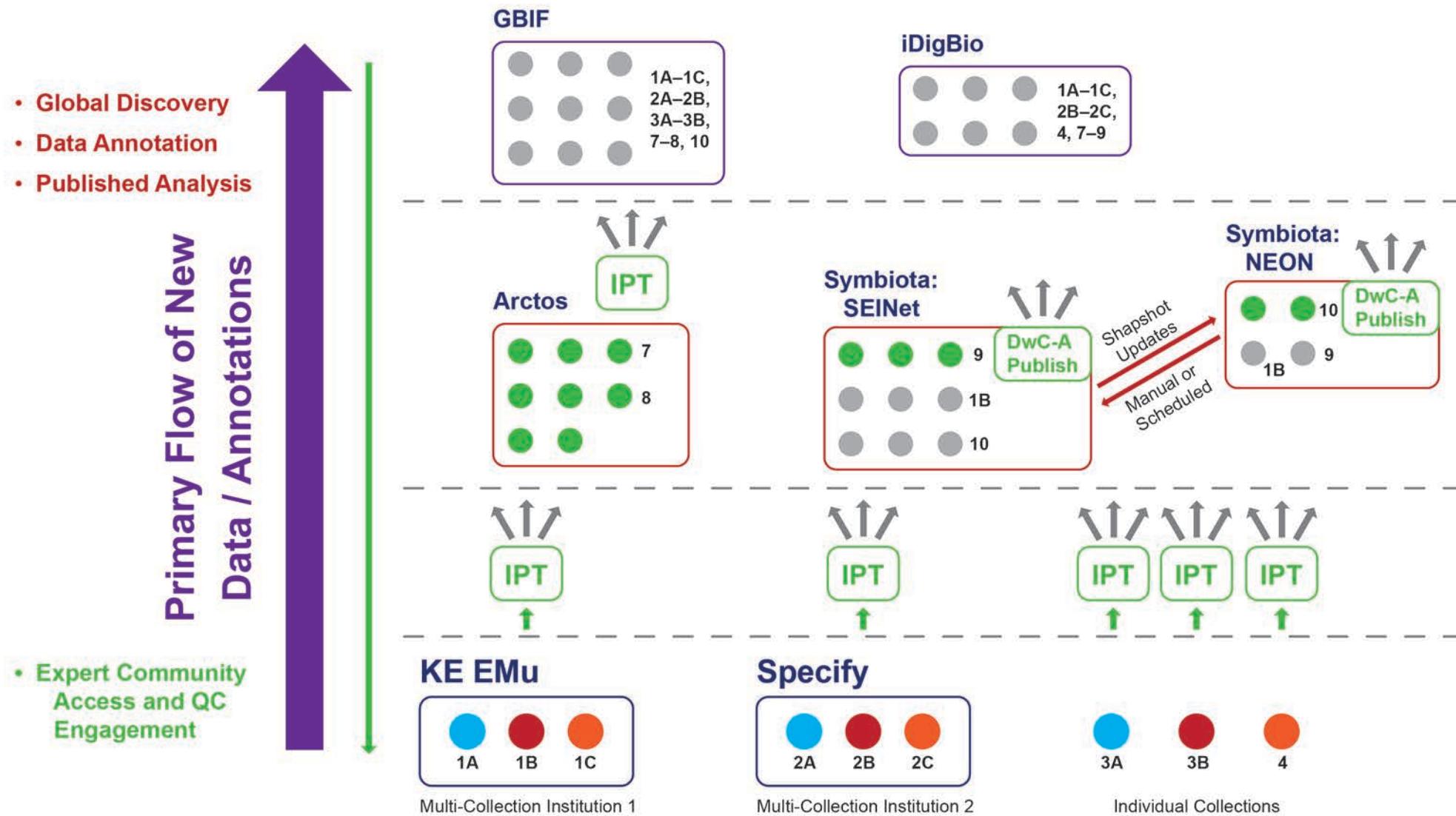
**"Hidden Data"**



Individual Collections

Individual Collections

This hierarchy sustains an **imbalance in directional data flow:**  
**Annotations on global datasets are hard to pull downwards.**

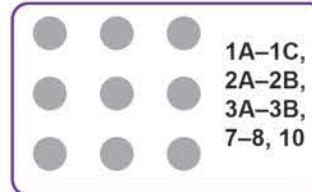


Moreover, by the time we reach the top, most **experts/enthusiasts no longer feel at home** (cf. [Wenger 2000](#)).

- All-Inclusive Data Product of the "Global Community"

Communities  
of Practice

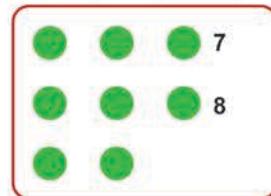
GBIF



iDigBio

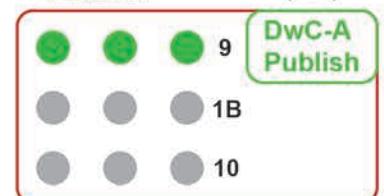


Arctos



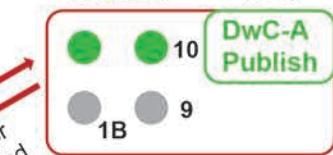
IPT

Symbiota:  
SEINet

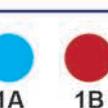


Symbiota:  
NEON

Snapshot  
Updates  
Manual or  
Scheduled

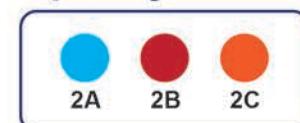


IPT



Multi-Collection Institution 1

Specify



Multi-Collection Institution 2



Individual Collections

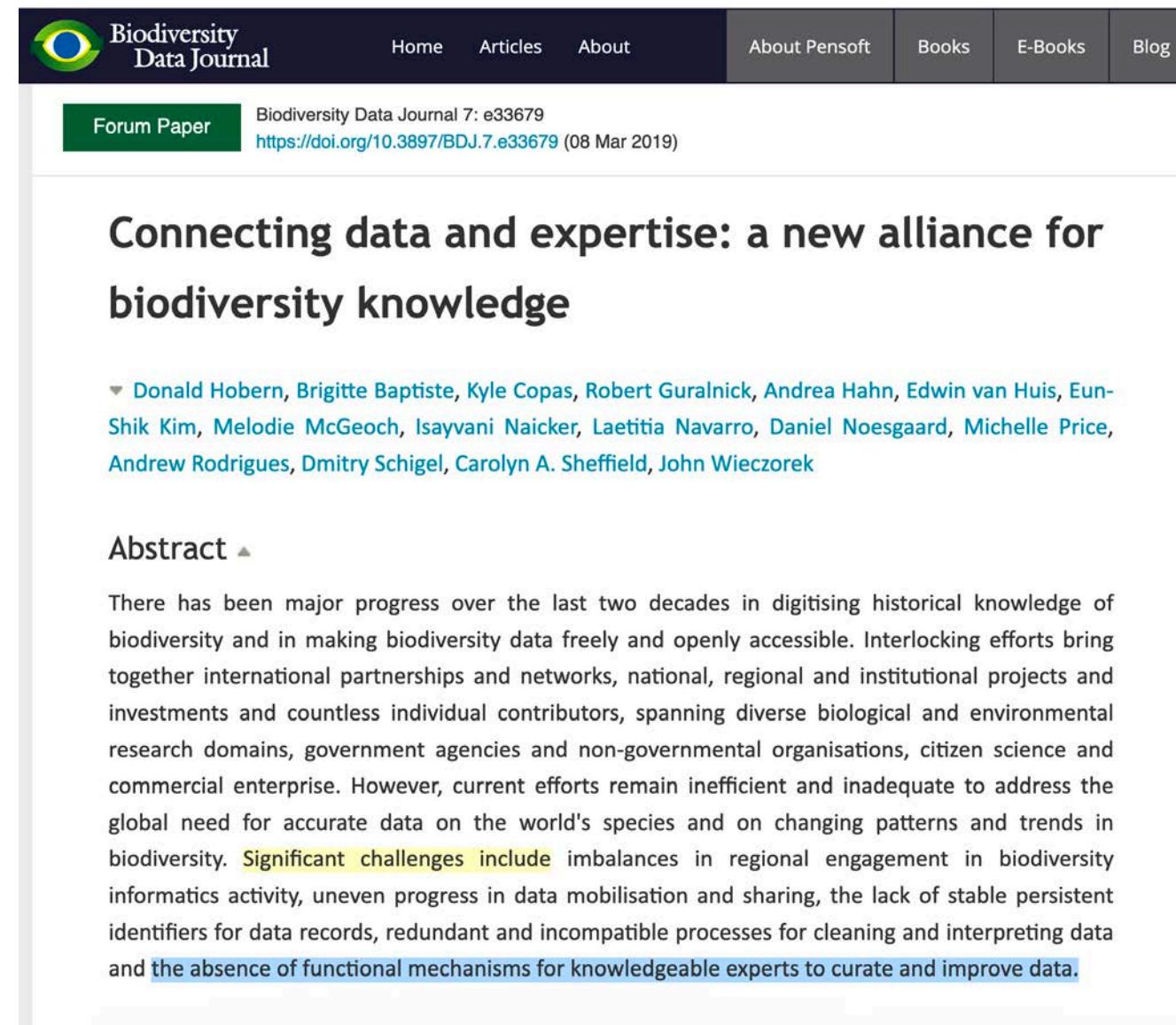
- Data Products Driven by Individuals or Small Groups (Collections, Institutions)

# Wenger 2000: Communities of practice and social learning systems

**Table 2.** Boundary Dimensions

	Coordination	Transparency	Negotiability
Engagement	What opportunities exist for joint activities, problem-solving, and discussions to both surface and resolve differences through action?	Do people provide explanations, coaching, and demonstrations in the context of joint activities to open windows on to each other's practices?	Are joint activities structured in such a way that multiple perspectives can meet and participants can come to appreciate each other's competences?
Imagination	Do people have enough understanding of their respective perspectives to present issues effectively and anticipate misunderstandings?	What stories, documents, and models are available to build a picture of another practice? What experience will allow people to walk in the other's shoes? Do they listen deeply enough?	Can both sides see themselves as members of an overarching community in which they have common interests and needs?
Alignment	Are instructions, goals, and methods interpretable into action across boundaries?	Are intentions, commitments, norms, and traditions made clear enough to reveal common ground and differences in perspectives and expectations?	Who has a say in negotiating contracts and devising compromises?

# Community Access ⇔ Engagement ⇔ Quality ⇔ Trust ⇔ Use & Impact



The screenshot shows the homepage of the Biodiversity Data Journal. At the top, there is a navigation bar with links for Home, Articles, About, About Pensoft, Books, E-Books, and Blog. Below the navigation bar, a green button labeled "Forum Paper" is visible. To its right, the journal's title "Biodiversity Data Journal 7: e33679" and the DOI "https://doi.org/10.3897/BDJ.7.e33679 (08 Mar 2019)" are displayed. The main content area features a large, bold title: "Connecting data and expertise: a new alliance for biodiversity knowledge". Below the title, a list of authors is provided: 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. An "Abstract" section follows, containing a detailed paragraph about the progress in digitising biodiversity knowledge and the challenges faced. The abstract highlights imbalances in regional engagement, uneven progress in data mobilisation, lack of stable persistent identifiers, redundant processes, and the absence of functional mechanisms for curating data.

Biodiversity Data Journal 7: e33679  
<https://doi.org/10.3897/BDJ.7.e33679> (08 Mar 2019)

## 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**.

A–F = Portals | 1–20 = Collections

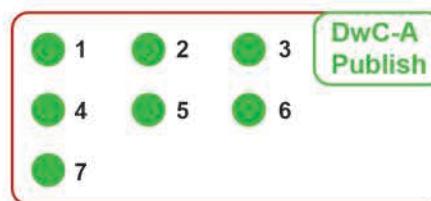
● "Live-Managed" Collection



A: Consortium of Northeastern Herbaria

B: [Southwest Environment Information Network]

C: Smithsonian Tropical Research Institute



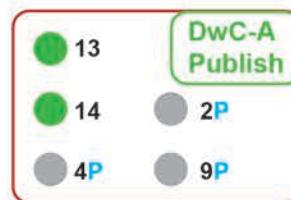
# De-centralized, but global

- **Partial, relevant collection snapshot subsets** are represented.

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

- "Live-Managed" Collection
- "Snapshot" Collection (Vs.)

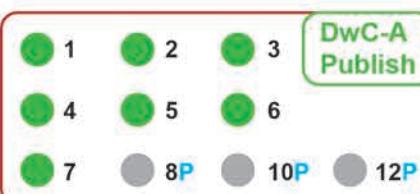
C. STRI



A: Consortium of Northeastern Herbaria

B: [Southwest Environment Information Network]

C: Smithsonian Tropical Research Institute



A: CNH



B. SEINet

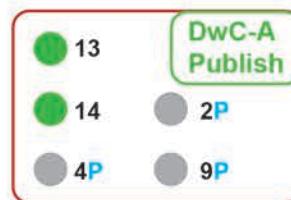
# De-centralized, but global

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

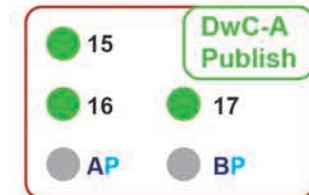
A–F = Portals | 1–20 = Collections (Partial)

- "Live-Managed" Collection
- "Snapshot" Collection (Vs.)

C. STRI



D. NEON



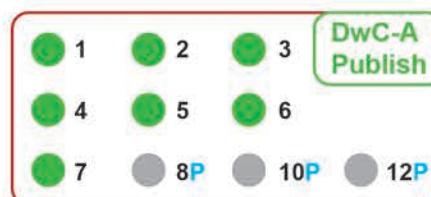
A: Consortium of Northeastern Herbaria

B: [Southwest Environment Information Network]

C: Smithsonian Tropical Research Institute

D: National Ecological Observatory Network

A: CNH



B. SEINet



# De-centralized, but global

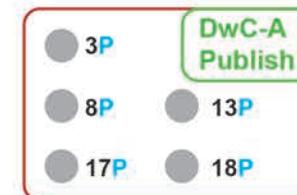
- Some research-themed portals may only include partial collection snapshots.

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

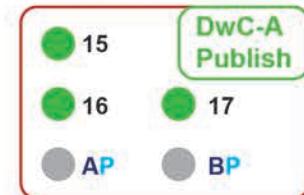
- "Live-Managed" Collection
- "Snapshot" Collection (Vs.)



E. CoTRAM



D. NEON



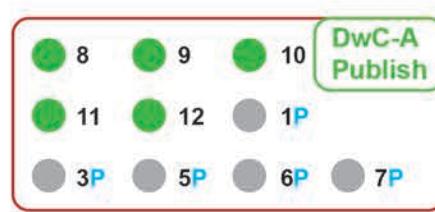
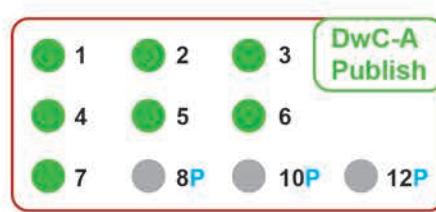
A: Consortium of Northeastern Herbaria

B: [Southwest Environment Information Network]

C: Smithsonian Tropical Research Institute

D: National Ecological Observatory Network

E: Cooperative Taxonomic Resource  
for American Myrtaceae

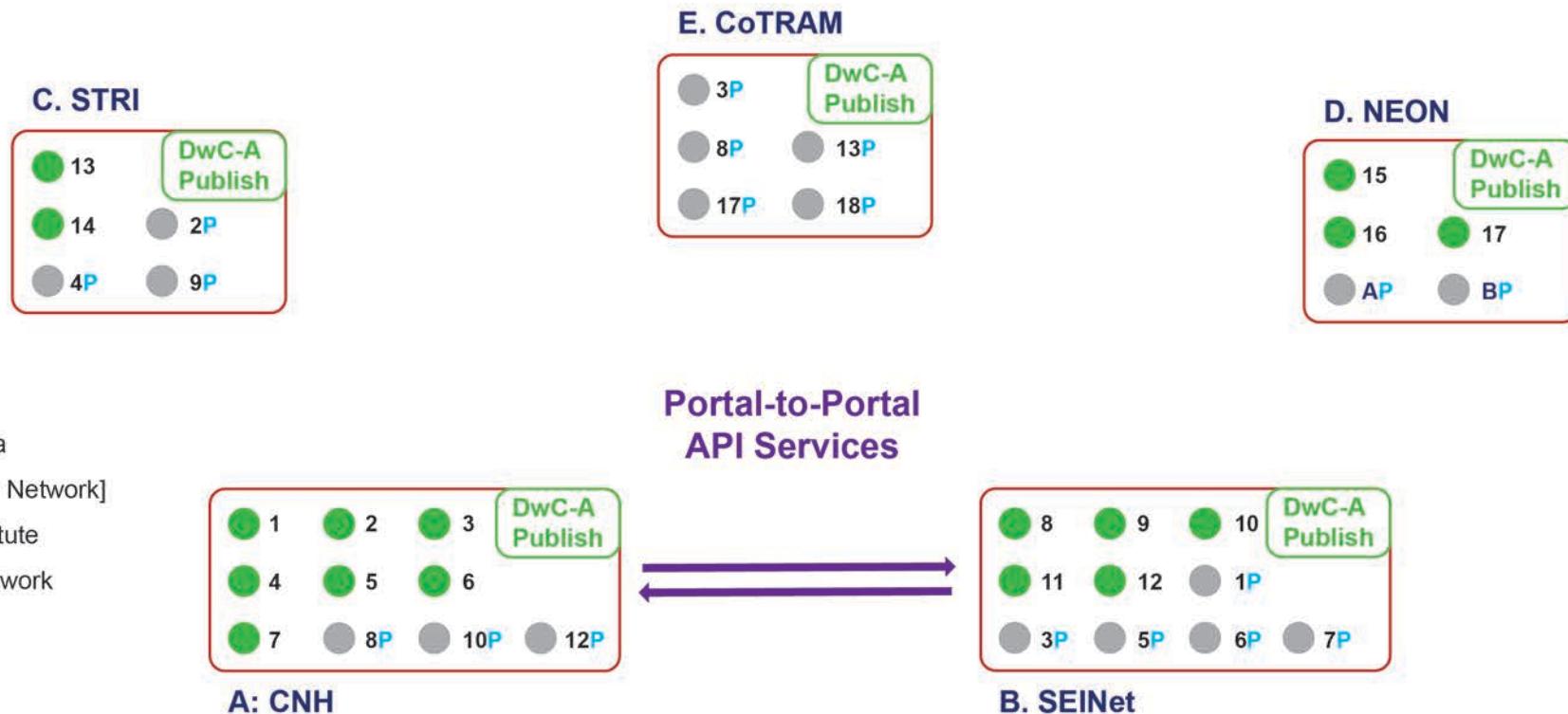


# De-centralized, but global

- Highly configurable **portal-to-portal APIs** negotiate the flow of data **between live and snapshot collection instances**.

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

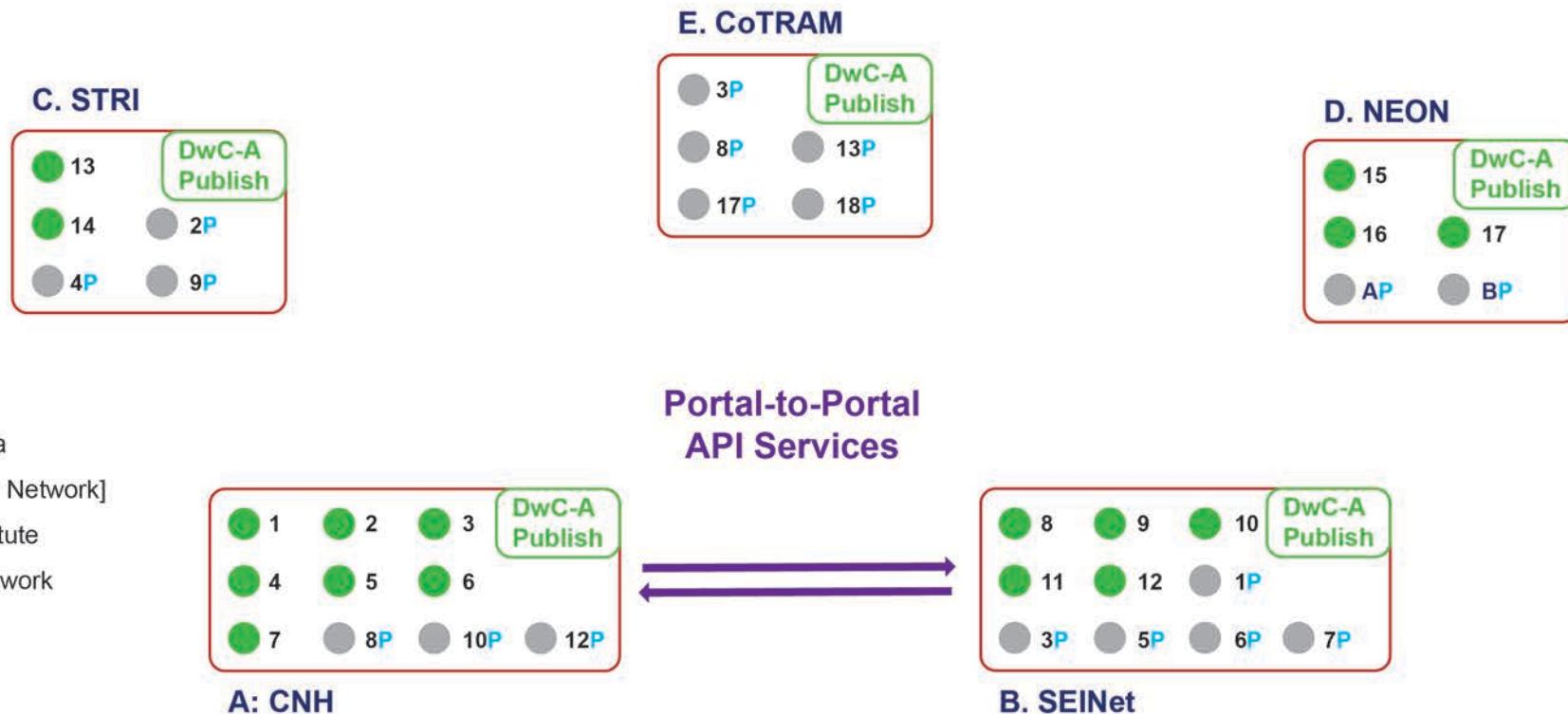
E: Cooperative Taxonomic Resource  
for American Myrtaceae

# 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

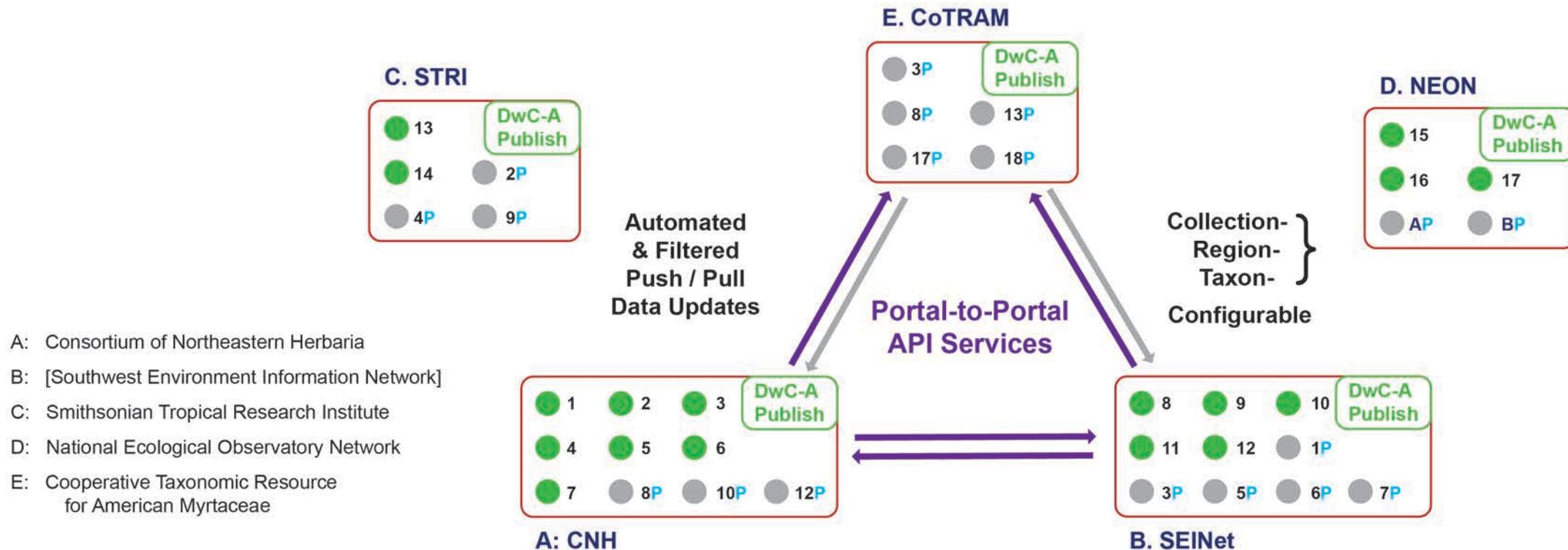
E: Cooperative Taxonomic Resource  
for American Myrtaceae

# De-centralized, but global

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

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

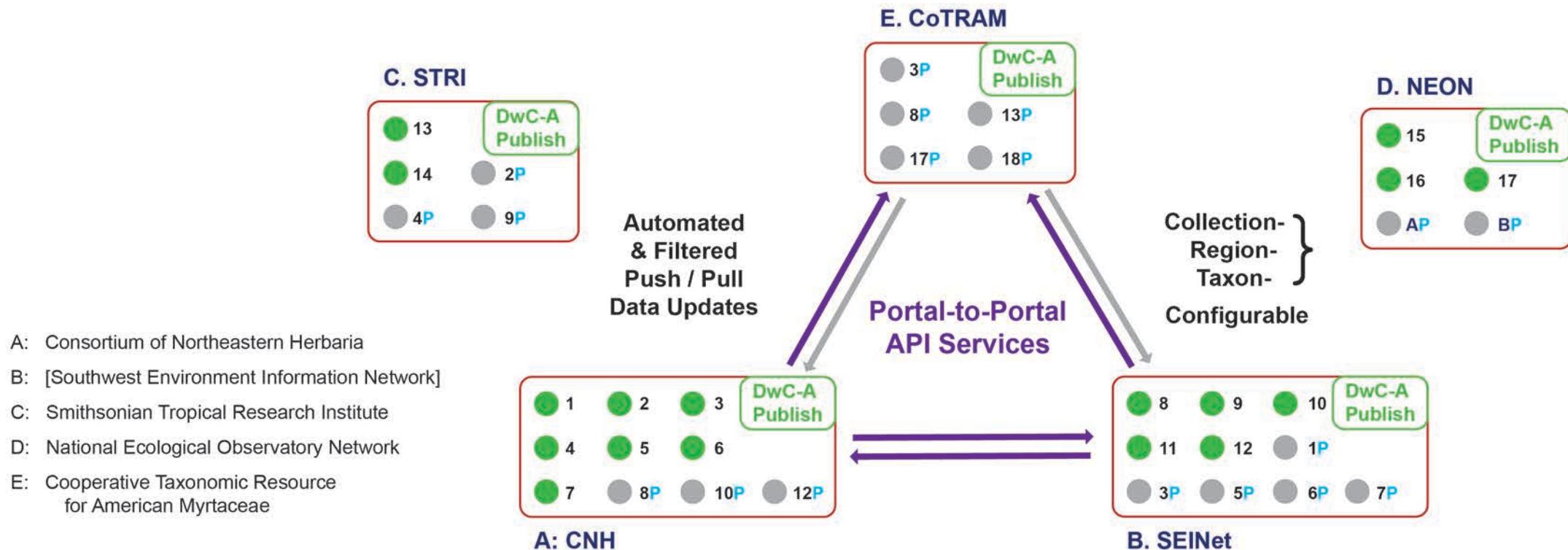
- "Live-Managed" Collection = Direct Updates
- "Snapshot" Collection (Vs.) = API Updates



# De-centralized, but global

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

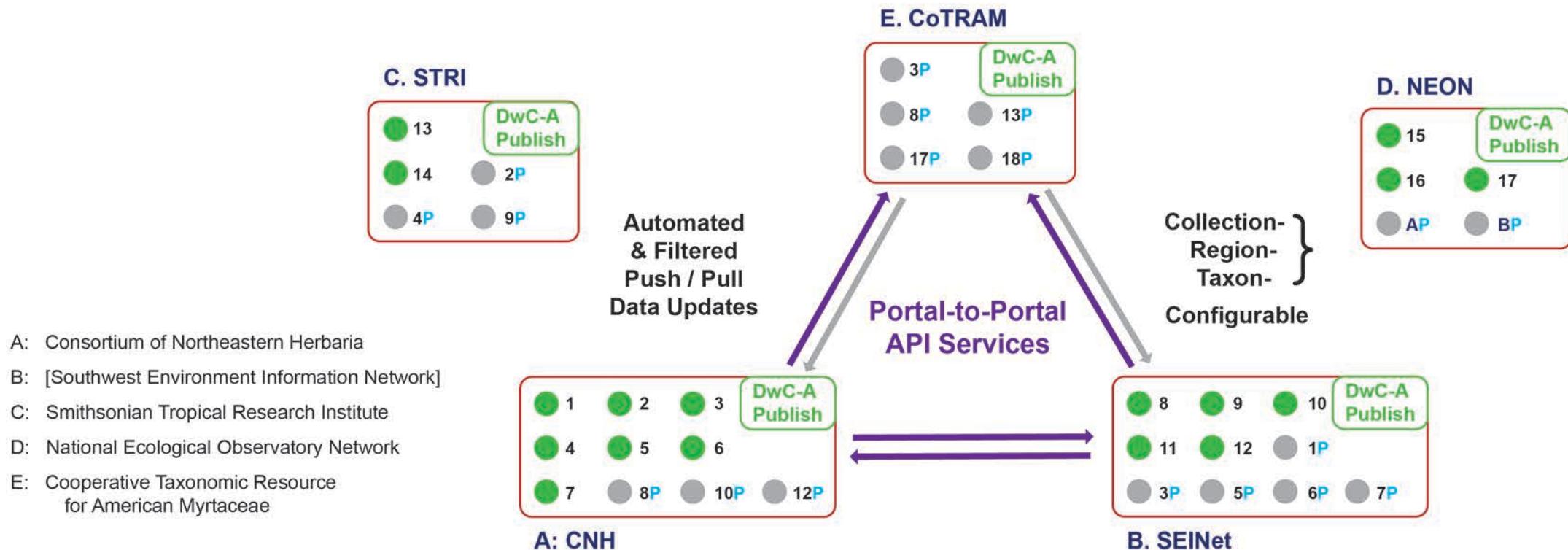
A–F = Portals | 1–20 = Collections (Partial)  
● "Live-Managed" Collection = Direct Updates  
● "Snapshot" Collection (Vs.) = API Updates



# De-centralized, but global

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

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

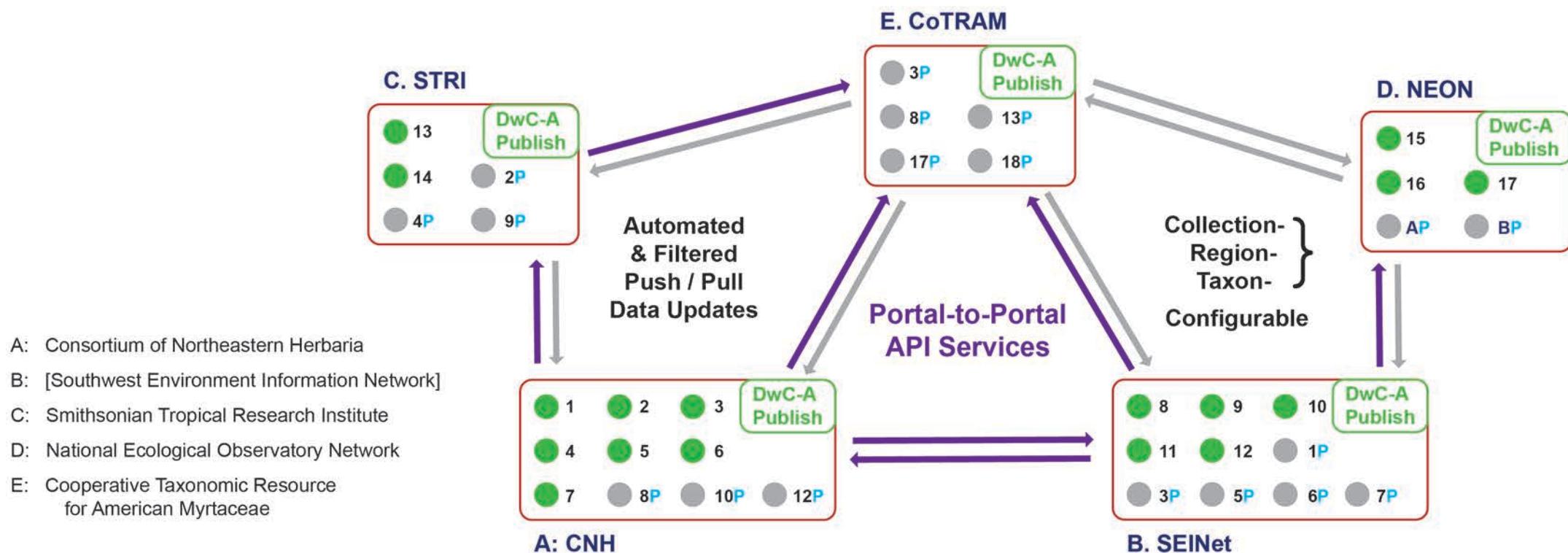


- A: Consortium of Northeastern Herbaria
- B: [Southwest Environment Information Network]
- C: Smithsonian Tropical Research Institute
- D: National Ecological Observatory Network
- E: Cooperative Taxonomic Resource for American Myrtaceae

# 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

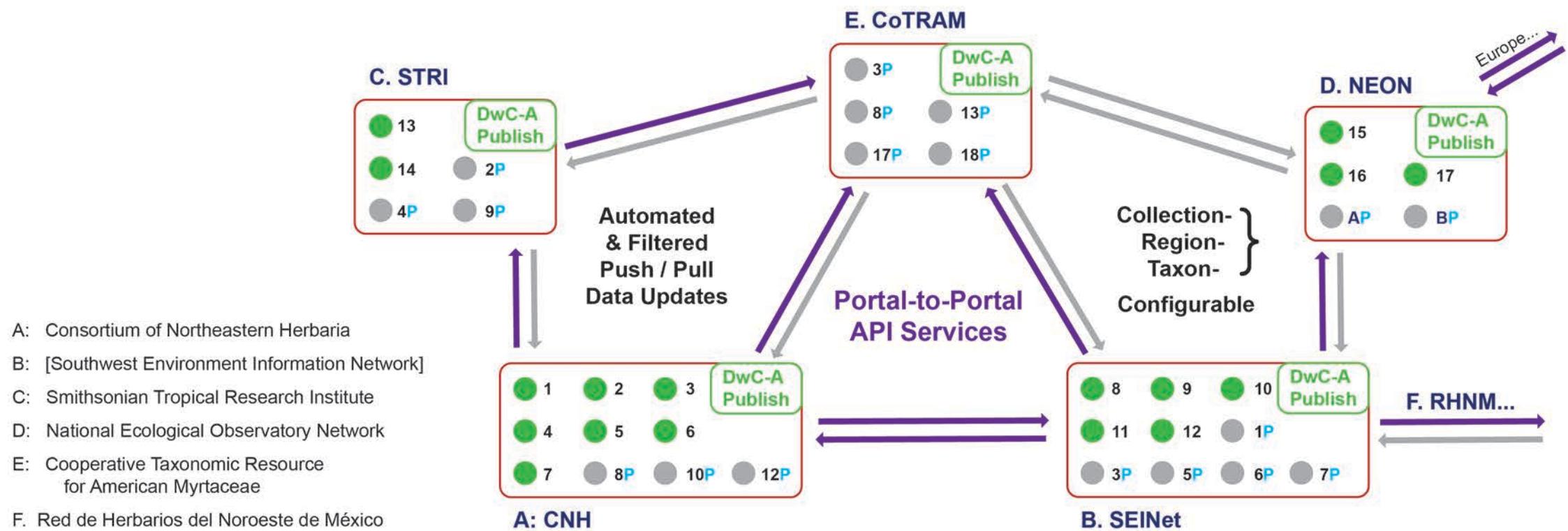


# De-centralized, but global

- On the basis of a shared API service culture, a de-centralized data portal network can potentially grow to attain **global coverage**.

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

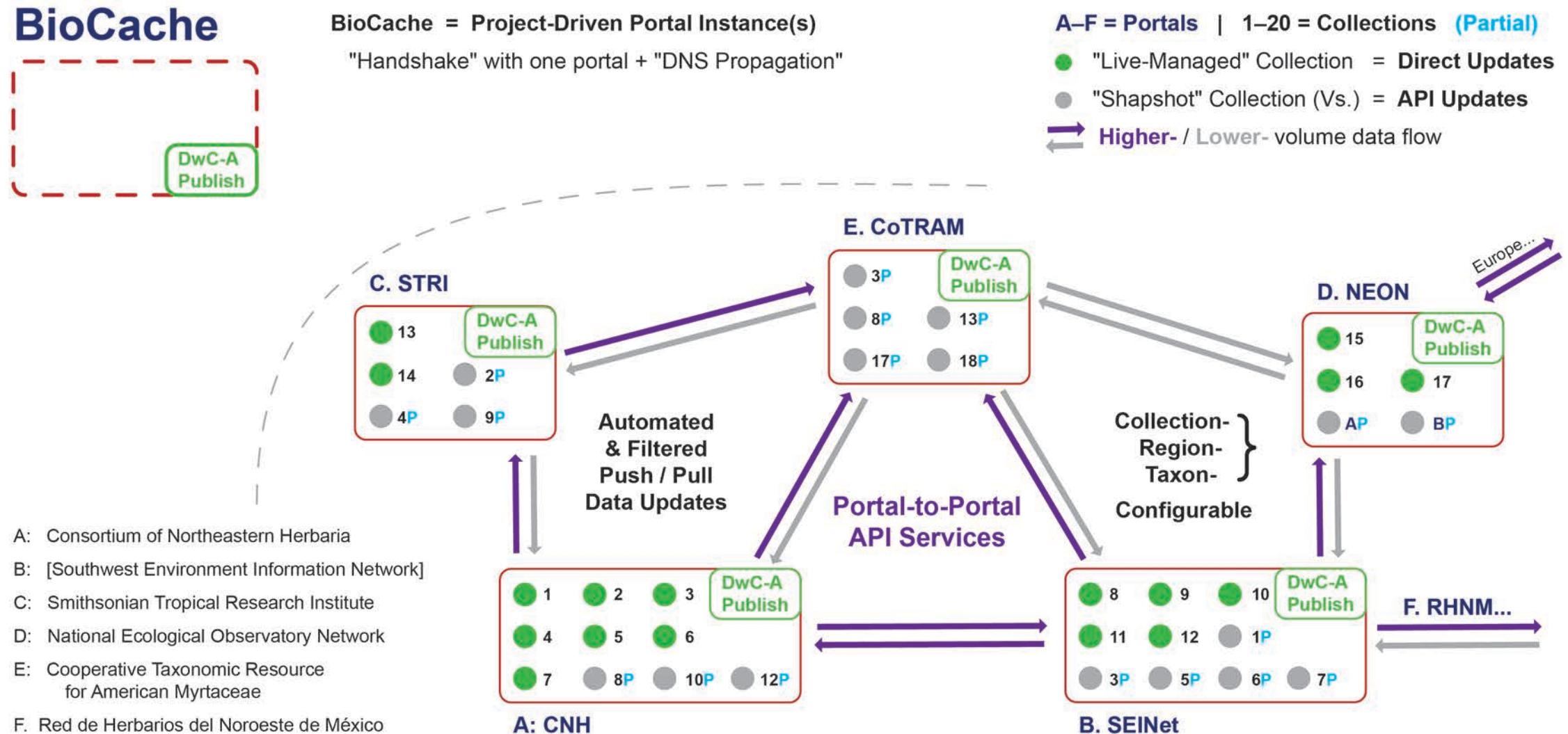
- "Live-Managed" Collection = Direct Updates
- "Snapshot" Collection (Vs.) = API Updates
- ↔ Higher- / Lower- volume data flow



Designing for expert/enthusiast access

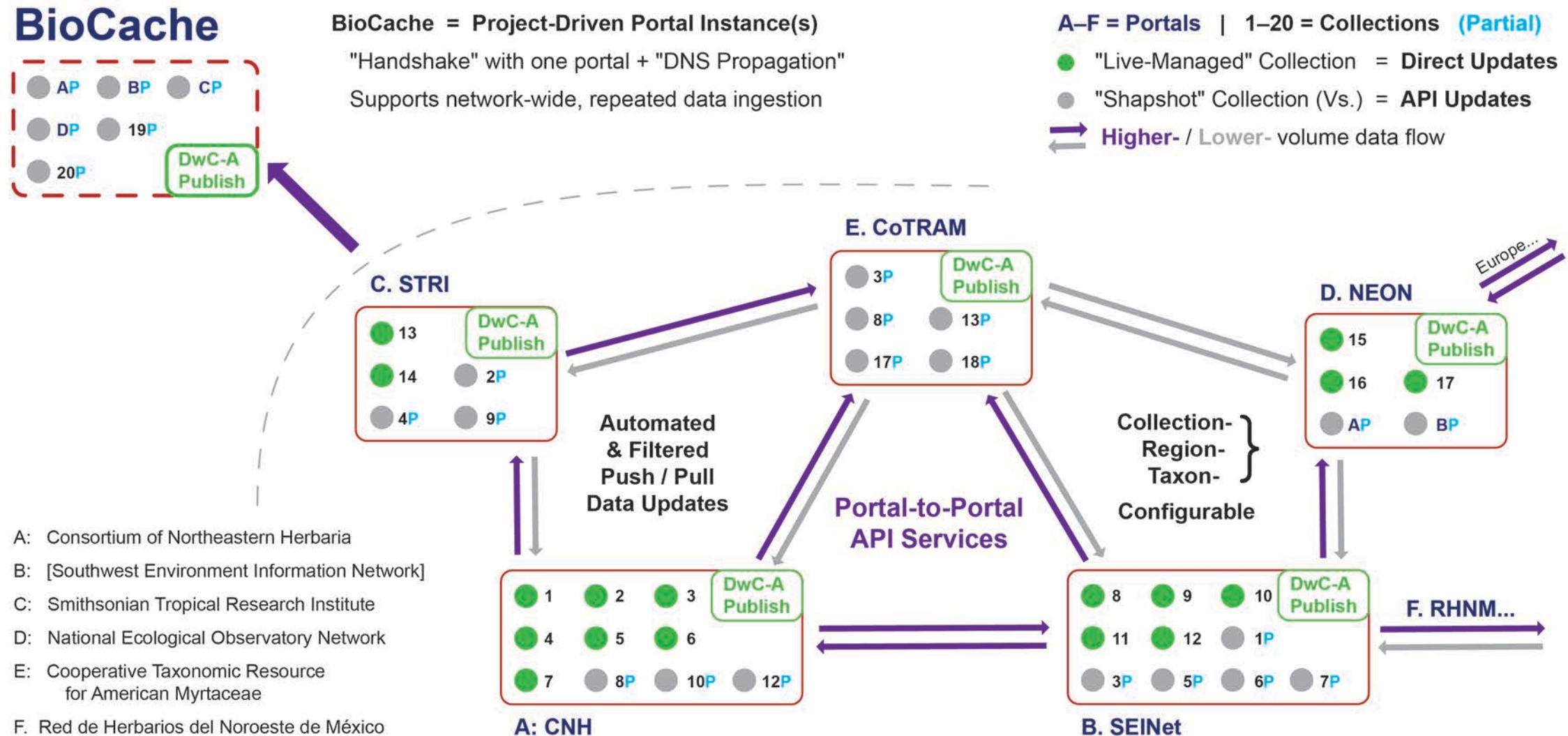
# BioCache: Global access through **custom research portal instances**

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



# BioCache: Global access through **custom research portal instances**

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



# Southwest U.S. Coleoptera portal as an early adopter

• Online

• 2 (part)

• 1 filter

• All sn

• **Taxo**

- Online Symbiota portal
  - 2 (partial) collections
  - 1 filtered portal extract
  - All snapshot versions
  - **Taxonomy is live-managed!**

**Enter a taxon**

Taxon: Amphidorini LeConte, 1862

Display authors      **DISPLAY TAXON TREE**  
 Display full tree below  
 family  
 Display species with subgenera

Animalia Linnaeus, 1758

Arthropoda Latreille, 1829

Hexapoda Latreille, 1825

Insecta Linnaeus, 1758

Pterygota

Neoptera

Coleoptera Linnaeus, 1758

Polypogon Emery, 1886

Cucujiformia Lameere, 1938

Tenebrionoidea Latreille, 1802

Tenebrionidae Latreille, 1802

Tenebrioninae

**Amphidorini** LeConte, 1862

*Eleodes* Eschscholtz, 1829

*Eleodes (Amphidora)* Eschscholtz, 1829

*Eleodes (Blapylis)* Horn, 1870

*Eleodes (Blapylis) bishopensis* Somerby and Doyen, 1976

*Eleodes (Blapylis) blanchardii* Blaisdell

*Eleodes (Blapylis) brunneipes* Casey, 1890

*Eleodes (Blapylis) caseyi* Blaisdell, 1909

*Eleodes (Blapylis) clavicornis* Eschscholtz, 1829

*Eleodes (Blapylis) consobrina* LeConte, 1851

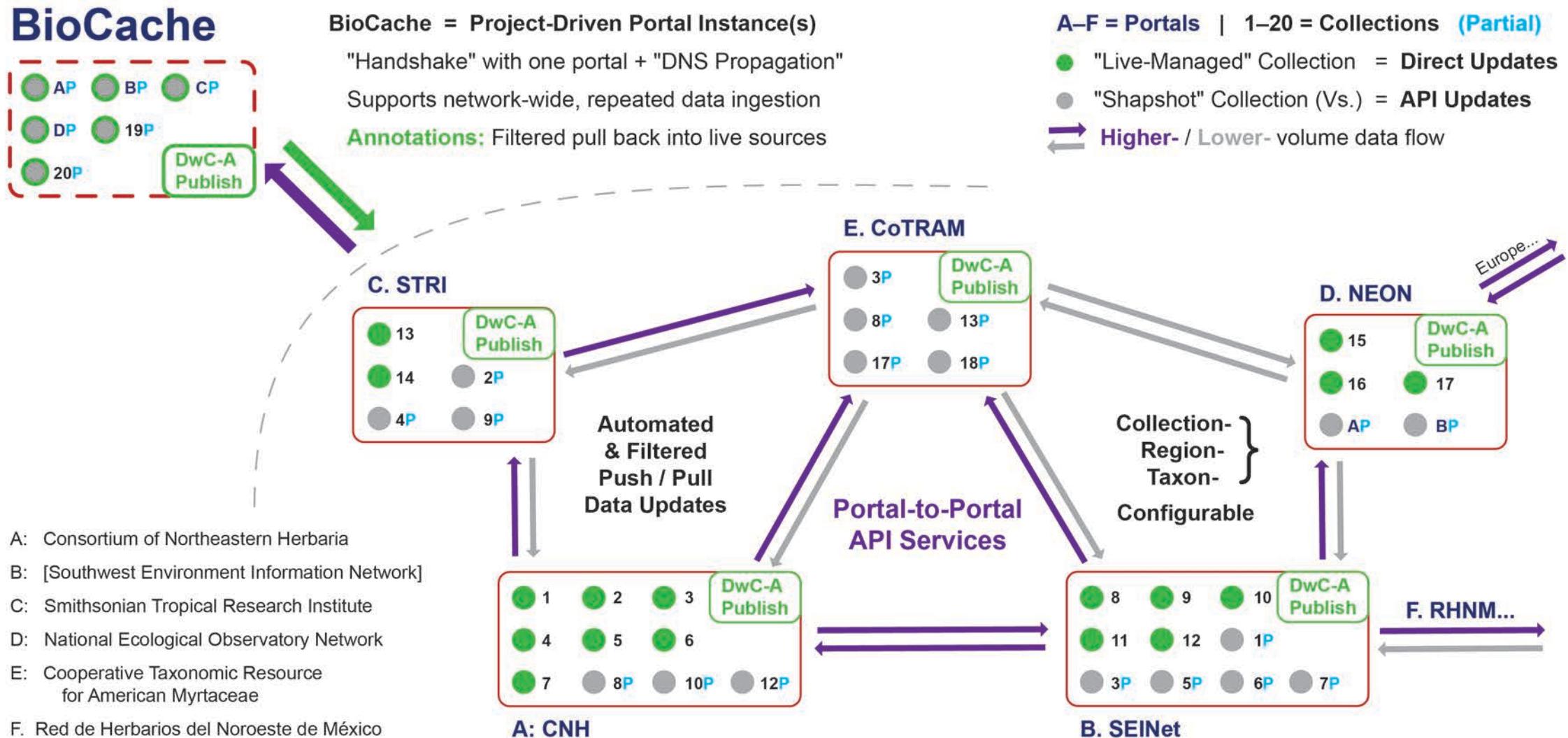
[*Eleodes consobrina* LeConte]

*Eleodes (Blapylis) cooperi* Somerby and Doyen, 1976

*Eleodes (Blapylis) cordatula* Eschscholtz, 1829

# BioCache: Global access through **custom research portal instances**

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



Stay tuned, it's underway

# Ultimately: Taxonomic data intelligence for Darwin Core occurrences

The screenshot shows the RIO website interface. At the top, there's a navigation bar with links for Home, Articles, About, About Pensoft, Books, E-Books, Blog, and Journals. Below the navigation bar, a green button says "NSF Grant Proposal". To its right is the title "Research Ideas and Outcomes 2: e10610" with the DOI "https://doi.org/10.3897/rio.2.e10610" and the date "(30 Sep 2016)". Next to it is a red button labeled "Reviewable" and a dark blue button labeled "v1 CrossMark".

## Controlling the taxonomic variable: Taxonomic concept resolution for a southeastern United States herbarium portal

Nico Franz, Edward Gilbert, Bertram Ludäscher, Alan Weakley

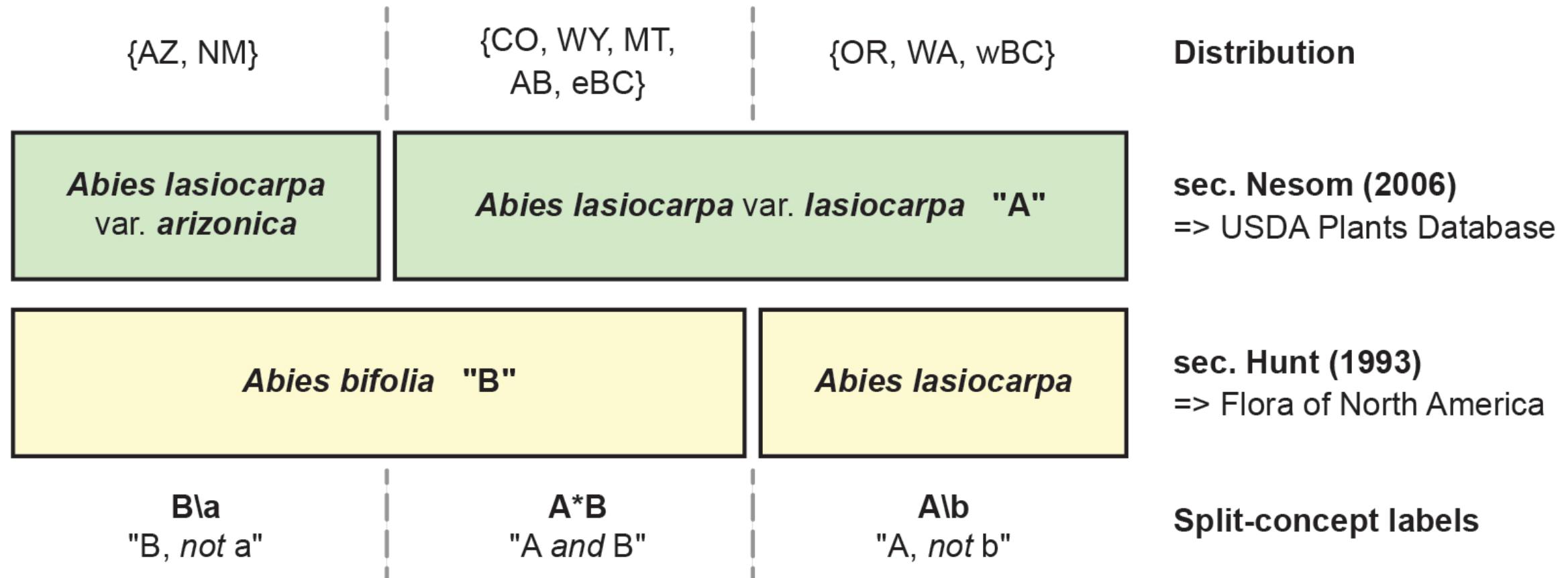
### Executive summary ▾

**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 'syntheses' 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.

# Alignment by Alan Weakley (<http://herbarium.unc.edu/flora.htm>)



# Ultimately: Taxonomic data intelligence for Darwin Core occurrences

The screenshot shows the RIO website interface. At the top, there's a navigation bar with links for Home, Articles, About, About Pensoft, Books, E-Books, Blog, and Jour. Below the navigation is a teal banner for an NSF Grant Proposal titled "Research Ideas and Outcomes 2: e10610" with the URL <https://doi.org/10.3897/rio.2.e10610> (30 Sep 2016). To the right of the banner are buttons for "Reviewable" and "v1" along with a CrossMark logo. The main content area features the title "Controlling the taxonomic variable: Taxonomic concept resolution for a southeastern United States herbarium portal" and the authors' names: Nico Franz, Edward Gilbert, Bertram Ludäscher, Alan Weakley. Below the title is an "Executive summary" section containing detailed text about the project's goals and methods.

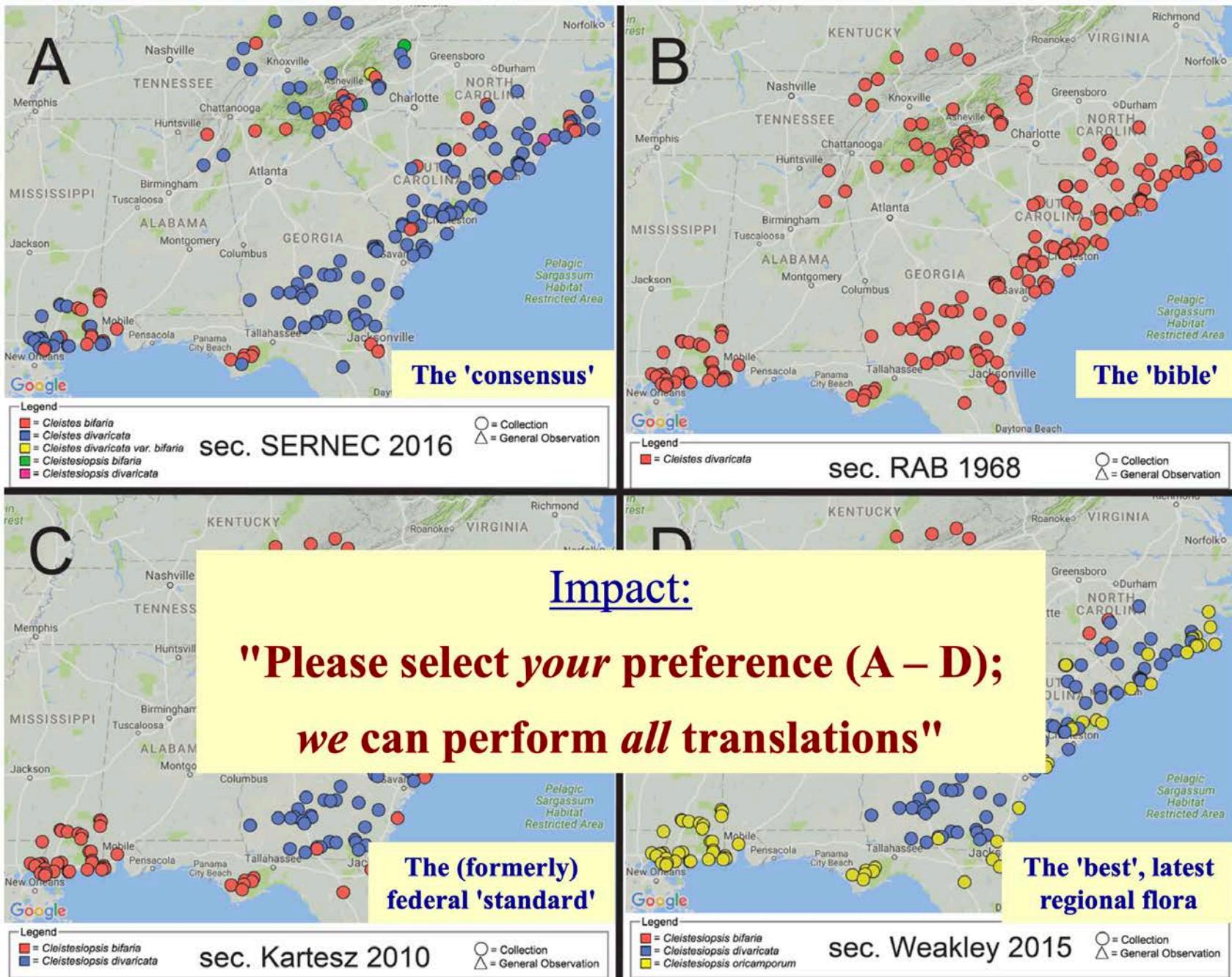
- <https://doi.org/10.3897/rio.2.e10610>
- Engaging expert/enthusiast communities  
↔ **need for pluralism and democracy**  
for and **among taxonomic perspectives**  
in biodiversity data aggregation designs.
- Spatial reasoning tools (RCC-5) can help  
attain consistent and comprehensive  
**taxonomic meaning mappings** ↔  
**intelligence** for data integration across  
evolving or conflicting views.

# Ultimately: Taxonomic data intelligence for Darwin Core occurrences

The screenshot shows the RIO website interface. At the top, there's a navigation bar with links for Home, Articles, About, About Pensoft, Books, E-Books, Blog, and Jour. Below the navigation is a teal banner for an NSF Grant Proposal titled "Research Ideas and Outcomes 2: e10610" with the URL <https://doi.org/10.3897/rio.2.e10610> (30 Sep 2016). To the right of the banner are buttons for "Reviewable" and "v1" along with a CrossMark logo. The main content area features the title "Controlling the taxonomic variable: Taxonomic concept resolution for a southeastern United States herbarium portal" and the authors' names: Nico Franz, Edward Gilbert, Bertram Ludäscher, Alan Weakley. Below the title is an "Executive summary" section containing detailed text about the project's goals and methods.

- <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,**  
in relation to "the taxonomic variable".

"Controlling the taxonomic variable"



Source: Franz et al. 2016. Controlling the taxonomic variable: [...]. RIO Journal. doi:10.3897/rio.2.e10610

Our project just (re)started, but look!

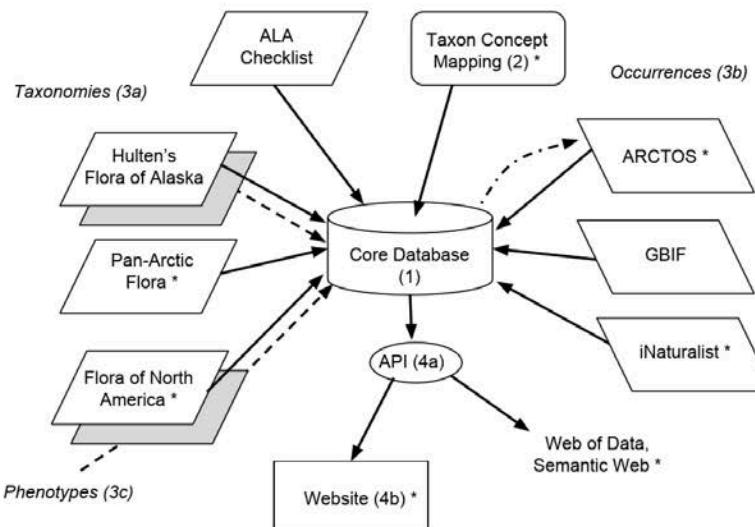
# "Taxonomically intelligent data integration for a new **Flora of Alaska**"

## Blog 1: Project roadmap

Posted by [Cam](#) 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:



- NSF DBI 1759964
- PIs Ickert-Bond & Webb
- Reconciling Hulten, FNA & Pan-Arctic Flora
- See <http://alaskaflora.org/>

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](#),

# iNaturalist is mostly there already (eBird is too)

## Taxon Framework Relationships

About Taxon Frameworks

Taxon Frameworks help us describe 'what we mean' by a taxon. Taxon Framework Relationships are mappings between taxa on iNaturalist.org (internal taxa) and taxa on sources (external taxa) associated with a Taxon Framework. [Read more](#)

**Relationship**

match  one-to-one  alternate position  
 many-to-many  many-to-one  
 one-to-many  not external  
 not internal

**Taxon framework taxon**

any

**Internal taxon (including descendants)**

Type taxon name

**External taxon (including children)**

**Internal taxon rank**

any

**External taxon rank**

any

**Added by**

Start typing someone's name

**Active**

Either

**Filter** **Clear filters**

**Taxon Framework Relationship 332324**

Covered by a taxon framework for Class Magnoliopsida Sourced to [Plants of the World Online](#)

Relationship: Match

Plants of the World Online		iNaturalist.org
Species <i>Trapa hyrcana</i> (parent: Genus <i>Trapa</i> )	Match	Species <i>Trapa hyrcana</i> (parent: Genus <i>Trapa</i> )

**Taxon Framework Relationship 332323**

Covered by a taxon framework for Class Lycopodiopsida Sourced to [Plants of the World Online](#)

Relationship: Deviation

Plants of the World Online		iNaturalist.org
Species <i>Lycopodium obtusifolium</i> (parent: )	One-to-one	Species <i>Phlegmariurus obtusifolius</i> (parent: Genus <i>Phlegmariurus</i> )

We follow PPG I in recognizing *Phlegmariurus* as a segregate of *Huperzia*.

**Taxon Framework Relationship 332322**

Covered by a taxon framework for Class Lycopodiopsida Sourced to [Plants of the World Online](#)

Relationship: Deviation

Plants of the World Online		iNaturalist.org
----------------------------	--	-----------------

[https://www.inaturalist.org/taxon\\_framework\\_relationships](https://www.inaturalist.org/taxon_framework_relationships)

# iNaturalist is mostly there already (eBird is too)

**iNaturalist**  Explore Community More

[« Back to Taxon Framework Relationship](#)

## Taxon Framework Relationship 71572

[View taxonomy details for Ageroniini](#)

Covered by a taxon framework for Family Nymphalidae Sourced to [Markku Savela's Lepidoptera and Some Other Life Forms](#)

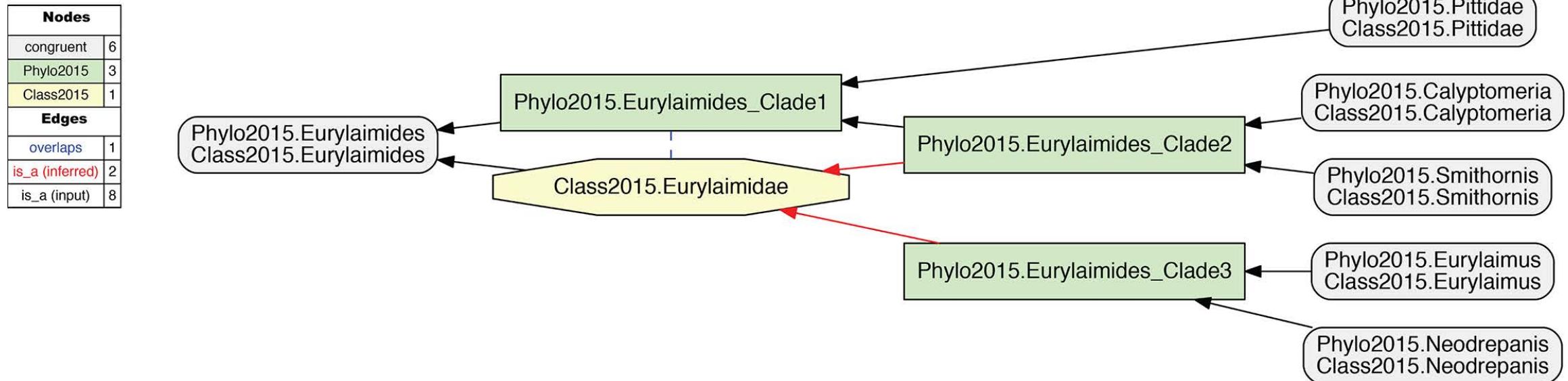
Relationship: Deviation			
Markku Savela's Lepidoptera and Some Other Life Forms			iNaturalist.org
Subtribe Eubagina (parent: Tribe Biblidini)	Many-to-many	Tribe Ageroniini (parent: Subfamily Biblidinae)	
Subtribe Epicalina (parent: Tribe Biblidini)		Tribe Epicaliini (parent: Subfamily Biblidinae)	
Subtribe Callicorina (parent: Tribe Biblidini)		Tribe Callicorini (parent: Subfamily Biblidinae)	
Subtribe Ageroniina (parent: Tribe Biblidini)		Tribe Biblidini (parent: Subfamily Biblidinae)	
Subtribe Epiphilina (parent: Tribe Biblidini)		Tribe Catonephelini (parent: Subfamily Biblidinae)	
Tribe Biblidini (parent: Subfamily Biblidinae)		Tribe Eubagini (parent: Subfamily Biblidinae)	
		Tribe Epiphilini (parent: Subfamily Biblidinae)	

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

Downstream deviations for Epicaliini 1

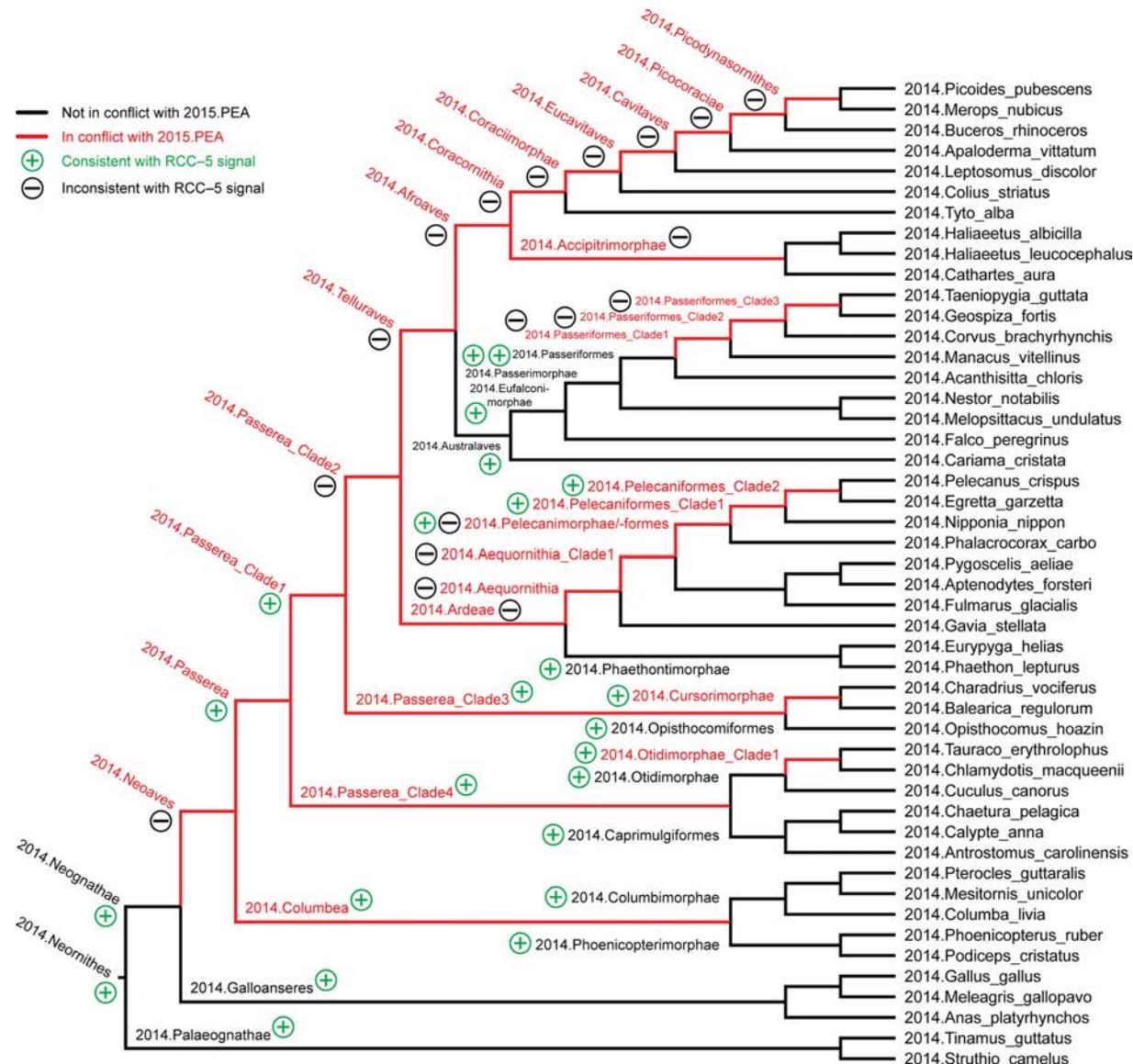
[https://www.inaturalist.org/taxon\\_framework\\_relationships](https://www.inaturalist.org/taxon_framework_relationships)

# Recent addition: Concept alignment for phylogenomic trees



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

- "Synthesis"  
OpenTree  
vs. RCC-5



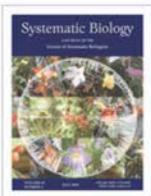
## Hopeful conclusions

Biodiversity informatics need not bet against systematics' very significant future impact on restructuring biodiversity knowledge; instead we can **incentivize experts/enthusiasts** in publishing high-quality, "data-intelligent" biodiversity data products.

If you have the passion and/or stomach for that continuously biodiversity data-restructuring systematics research future, **join us now as an early adopter of taxonomic concepts and RCC–5 relationships!**

# Acknowledgments

- **#Biodiversity Next 2019 organizers.**
- ASU Biocollections & BioKIC team; particularly Atriya Sen.
- 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>



Volume 65, Issue 4  
July 2016

## Article Contents

Abstract

METHODS

RESULTS

DISCUSSION

CONCLUSIONS

SUPPLEMENTARY MATERIAL

## Two Influential Primate Classifications Logically Aligned

Nico M. Franz , Naomi M. Pier, Deeann M. Reeder, Mingmin Chen, Shizhuo Yu, Parisa Kianmajd, Shawn Bowers, Bertram Ludäscher [Author Notes](#)

*Systematic Biology*, Volume 65, Issue 4, July 2016, Pages 561–582, <https://doi.org/10.1093/sysbio/syw023>

Published: 22 March 2016 [Article history](#) ▾



PDF

Split View

Cite

Permissions

### Abstract

Classifications and phylogenies of perceived natural entities change in the light of new evidence. Taxonomic changes, translated into Code-compliant names, frequently lead to name:meaning dissociations across succeeding treatments. Classification standards such as the *Mammal Species of the World* (MSW) may experience significant levels of taxonomic change from one edition to the next, with potential costs to long-term, large-scale information integration. This circumstance challenges the biodiversity and phylogenetic data communities to express taxonomic congruence and incongruence in ways that both humans and machines can process, that is, to logically represent taxonomic alignments across multiple classifications. We demonstrate that such alignments are feasible for two classifications of primates corresponding to the second and third MSW editions. Our approach has three main components: (i) use of taxonomic concept labels, that