

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

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.

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.

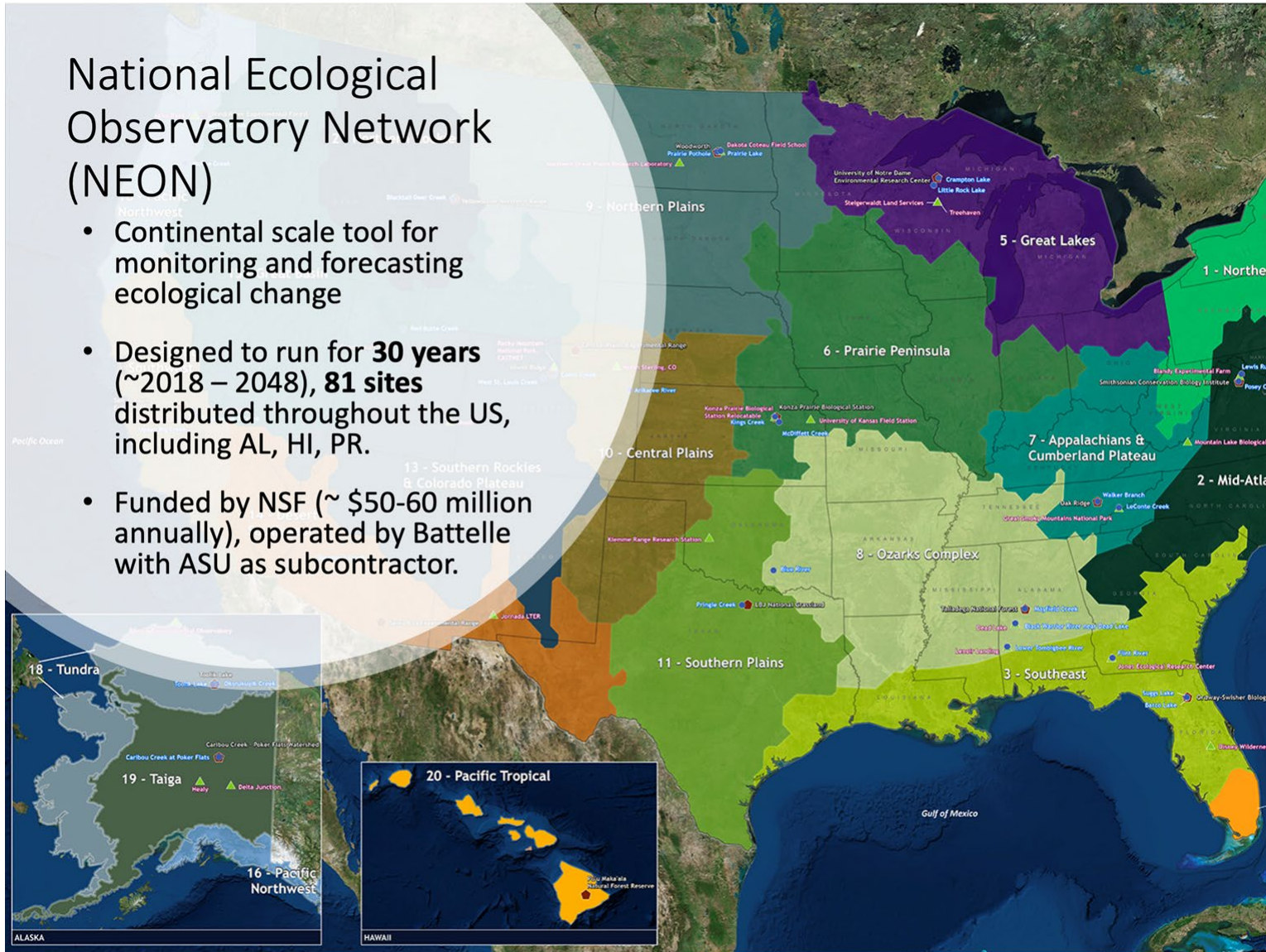
Introducing the **NEON Biorepository Data Portal** (1)

<https://biorepo.neonscience.org>

Introducing the NEON Biorepository Data Portal

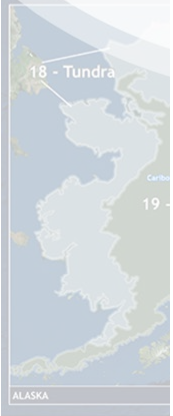
National Ecological Observatory Network (NEON)

- 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

Nati
Obs
(NEO
• Co
me
ec
• De
(~
dis
inc
• Fu
an
wi



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

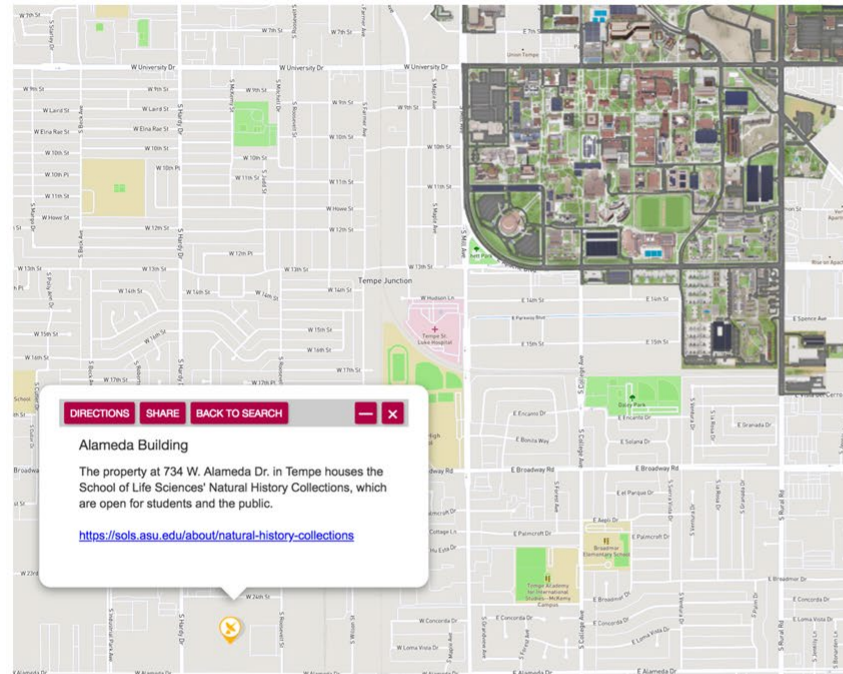


Domain facilities process samples and send them to the NEON Biorepository at ASU.



Introducing the NEON Biorepository Data Portal

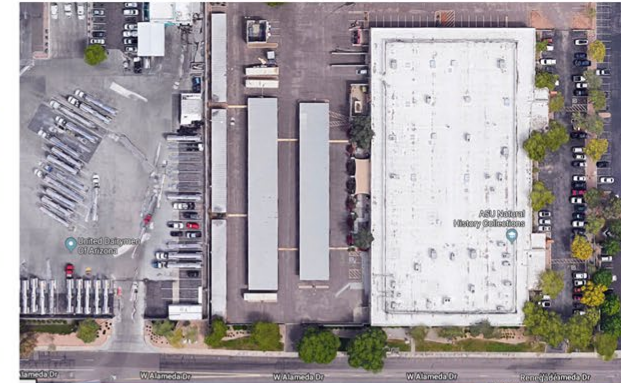
Where are we located?



734 W Alameda Drive, Tempe, AZ 85282

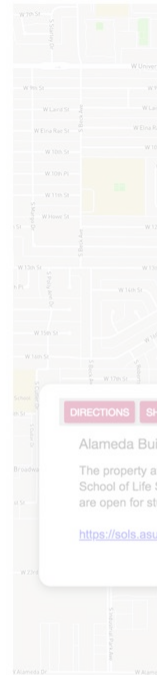
<https://www.asu.edu/map/interactive/?id=120&mrklid=66206>

- 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

Where



<https://v>

- Renov
- Open
- Free p

Who are we?



Dr. Nico Franz
Principal Investigator



Dr. Kelsey Yule
Project Manager



Dr. Laura Rocha Prado
Informatician (II)



Ed Gilbert
Informatician (I)



Laura Steger
*Environmental &
Zoological Collections
Manager*



Azhar Husain
Cryo Collections Manager



Dr. Andrew Johnston
Invertebrate Collections Manager

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

Introducing the NEON Biorepository Data Portal

ASU Biodiversity Knowledge Integration Center
Arizona State University

Nico Franz, Ed Gilbert, Azhar Husain, Andrew Johnston, Laura Rocha Prado, Laura Steger & Kelsey Yule
NEON Biorepository & Biodiversity Knowledge Integration Center – Arizona State University, Tempe, AZ, USA

An ecological collection

Sample Type	Volume/Year	Sample Type	Volume/Year
Algae - periphyton, seston, phytoplankton, diatoms		Plant	
DNA extractions (whole) ▲		Aquatic macrophytes - vouchers ▲	30
Soils ■	700	Background biomes ■	900
Whole sample ●	700	Foliage ▲	320
Macroalgae - whole sample ●	250	Tissue ▲	250
Aquatic lichens, mosses & liverworts		Voucher ■	900
Vouchers	55	Vertebrate	
Invertebrate		Fish - DNA extractions ▲	1,100
Carabids - DNA extractions ▲	4,500	Fish - fin clip ●	275
Carabids - pinned/preserved ■	8,000	Fish - voucher ■	275
Collembola - pooled ●	2,500	Insects - pinned biocoll. ●	2,000
Macroinvertebrates - pooled ●	800	Small mammal - blood ●	750
Macroinvertebrates - DNA extractions ▲	300	Small mammal - DNA extractions ▲	1,750
Frogs - pooled ●	3,750	Small mammal - fecal ●	4,000
Moose - DNA extractions ▲	1,750	Small mammal - hair/voucher ●	4,000
Moose - disease pool ▲	21,000	Small mammal - ear punch ▲	4,000
Moose - preserved/pinned ■	3,500	Small mammal - voucher ●	750
Moose - pooled ▲	3,500	Environmental	
Soil - disease pool ▲	4,500	Lichens ■	131
Soil - DNA extractions ▲	75	Soil - frozen ▲	21,000
Soil - pooled ▲	75	Soil - dry ■	375
Microbial		Particulate mass filters (PM10) ■	150
Aquatic microbes - Sierack filters ▲	850	Wet deposition ■	150
Soil microbes - DNA extractions ▲	2,500		
		Total ~104,000 samples per year	

Table 1. NEON Biorepository sample types and annual volumes.

Storage conditions:
 ● Wet (EtOH / Glutaraldehyde)
 ■ Ambient or 4° C
 ▲ Cryo (LN₂)
 ◆ -80° C / -20° C

Contact: biorepo@asu.edu
<https://biorepo.neonscience.org>

Promoting data publication

- The NEON Biorepository data portal will **foster data annotation and data publication**: Individual DwC records and research-configured, value-added DwC data packages.
- Value-added data annotations will be assignable to authors → **Authorship** of high-quality, coherently framed biodiversity datasets **incentivized and propagated** through the portal.
- Samples will become impactful through **use in question-driven research** → Portal designed to support post-original data publication sample annotations, ranging from third-party taxonomic identifications to (e.g.) trait-based data additions.

From discovery to forecasting

- The NEON Biorepository data portal (1) facilitates **discovery** of all sample-based NEON data as Darwin Core (DwC) occurrence records; and (2) uses the **Symbiota software platform** for networking biodiversity data collections (Gries et al. 2014) → NEON's unique data signals are made interoperable with those of the greater natural history collections and observational/enthusiast communities.

NEON Biorepository ↔ **Other Collections**

Highly structured, self-referential, and change-focused. Expansive taxonomic, geographic, and spatial extent.

- Integration with **discovery-focused** DwC data from iDigBio/GBIF with the **monitoring-focused** NEON Biorepository data can lead to **synergistic outcomes** for continental scale monitoring.
- Are change signals in NEON data are reflected in data from less-constrained diversity data?
- How can NEON-external data signals be leveraged to refine or refocus NEON-like efforts?

Figure 1. NEON Biorepository data portal map of field sites; showing 3 collections, ~230 species, and 8,490 occurrences of bulk carabid and pinned mosquito samples (June 5th, 2019).

Monitoring long-term ecological changes at a continental scale

Figure 2. NEON Biorepository data portal homepage.

Species checklist

Figure 3. Species checklist map of 1,022 records and 684 taxa imported from the SEINet herbarium portal (1902–2017) for the NEON Santa Rita Experimental Range field site (D14 – Desert Southwest).

How to engage

As of mid 2019, we have received 340 shipments totaling 55,300 samples. Most abundant among these are:

- 80° C-preserved soil samples (16.4k)
- EtOH-preserved arthropod "bycatch" (15k)
- EtOH-preserved, identified carabids (6.2k)
- Small mammal samples (6k)
- Pinned mosquitoes (5.5k).

- The portal is now public for anyone to explore sample availability and fitness-for-use. Contact us at biorepo@asu.edu to aid in sample discovery and for information to gain physical access to sample on- or off-site via loans (see **Sample Use Policy**).
- We are interested in engaging **prospective graduate students and/or postdoctoral researchers** inspired by our unique mission and access to NEON Biorepository samples.

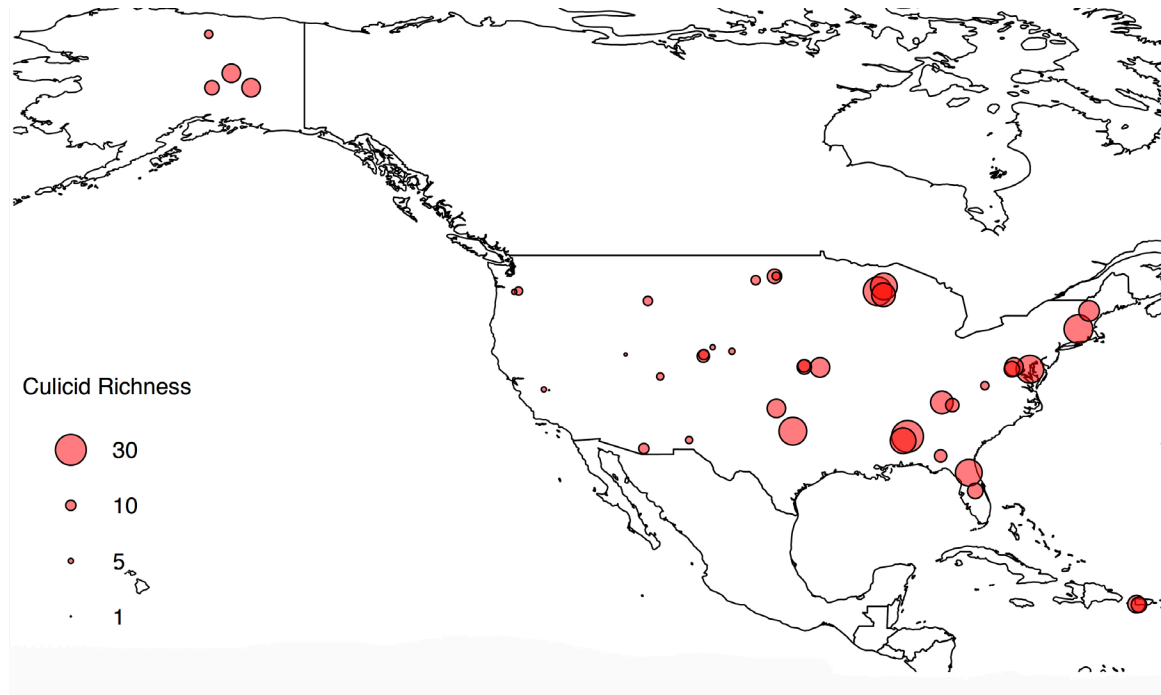
Figure 4. (A) EtOH collection; (B) -80° C samples; (C) Pinned mosquitoes.

Additional information:
 Gries, C., E. Gilbert & N. Franz. 2014. Symbiota – A virtual platform for creating voucher-based biodiversity information communities. Biodiversity Data Journal 2: e1114. <https://doi.org/10.3897/BDJ.2.e1114>
 NEON Biorepository – Sample Use Policy, Version 01, December 2018. Available at <https://www.neonscience.org/sample-use-policy>

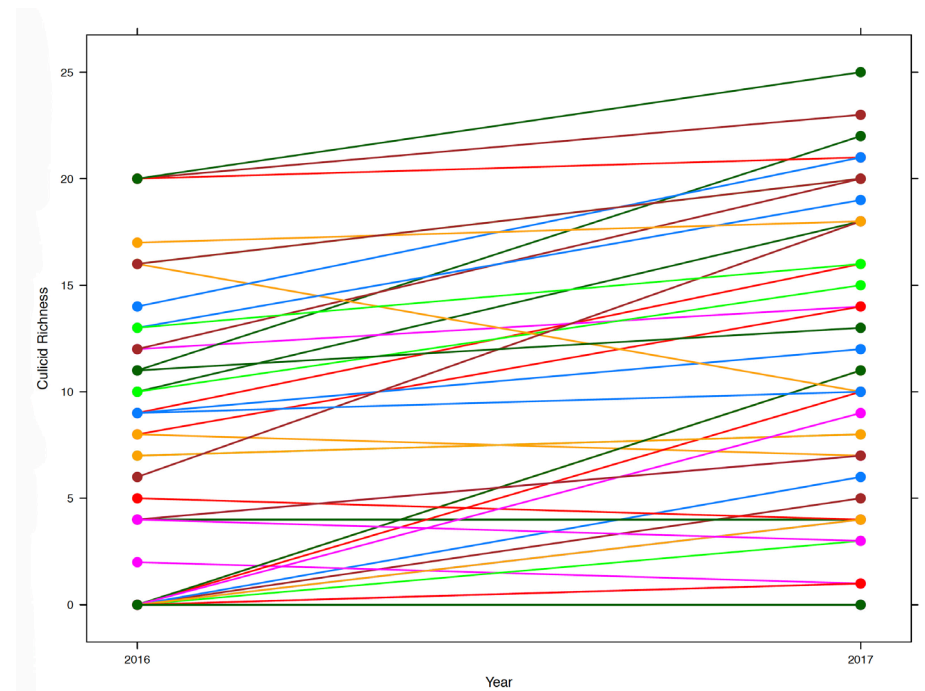
A short demonstration

<https://biorepo.neonscience.org>

[after a short BioRepo portal demo (Culicidae / SRER Plants)] *



Mosquito species richness per site, 2016–2017.



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.

Rethinking centralized biodiversity data aggregation (2)



Volume 2018
2018

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 disenfranchizing taxonomic experts
- Backbone-based data signal distortion: an example
- An objection
- Conclusion
- Acknowledgements
- Funding

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](#) ▼

 PDF  Split View  Cite  Permissions

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 process. 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.

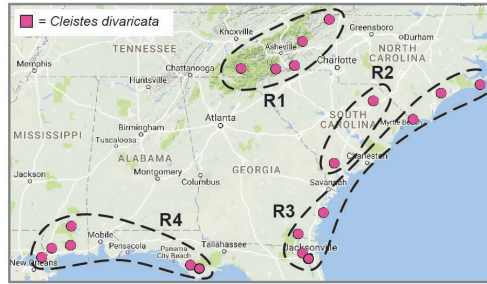
Issue Section: [Perspective/Opinion](#)

- <https://doi.org/10.1093/database/bax100>
- Critique of monolithic aggregation designs.

Alignment of 5 taxonomic schemata (1922 – present)

#	Period of use	Concept lineage 1	Concept lineage 2	Concept lineage 3	According to (sec.)
5	2009 – present	<i>Cleisteslopsis divaricata</i>	<i>Cleisteslopsis oricamporum</i>	<i>Cleisteslopsis bifaria</i>	Weakley (2015)
4	2008 – present	<i>Cleisteslopsis divaricata</i>	<i>Cleisteslopsis bifaria</i>		Kartesz (2010)
3	1993 – present	<i>Cleistes divaricata</i>	<i>Cleistes bifaria</i>		USDA Plants (2012)
2	1946 – 1993	<i>Cleistes div. var. divaricata</i>	<i>Cleistes divaricata var. bifaria</i>		Fernald (1950)
1	1922 – 1991	<i>Cleistes divaricata</i>			Radford AB (1968)

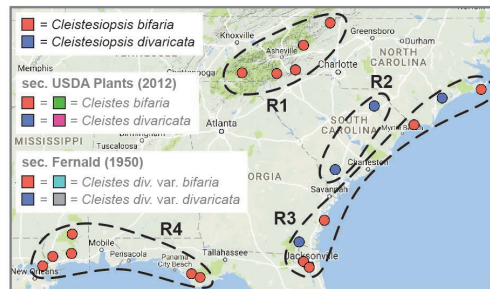
A. sec. Radford, Ahles & Bell (1968)



Distribution

	"bifaria"	"divaric."	"oricamp."
R1	o	+	o
R2	o	+	o
R3	o	+	o
R4	o	+	o

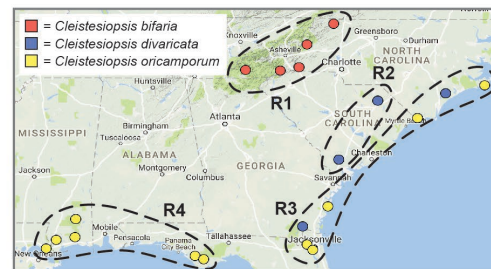
B. sec. Kartesz (2010) [BONAP]



Distribution

	bifaria	divaric.	oricamp.
R1	+	-	o
R2	-	+	o
R3	+	+	o
R4	+	-	o

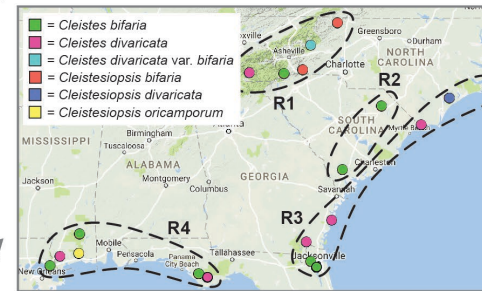
C. sec. Weakley (2015)



Distribution

	bifaria	divaric.	oricamp.
R1	+	-	-
R2	-	+	-
R3	-	+	+
R4	-	-	+

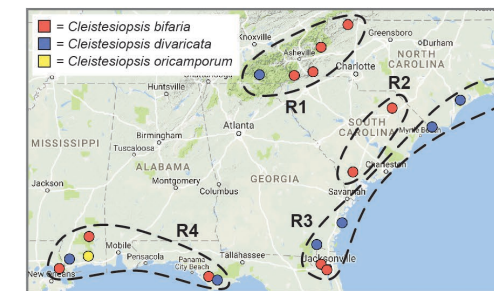
D. sec. SERNEC (2016) - RAW AGGREGATION



Distribution

	"bifaria"	"divaric."	"oricamp."
R1	(+)	(+)	-
R2	(+)	(-)	-
R3	+	+	-
R4	(+)	(+)	(+)

Data transformation to conform with single taxonomic backbone



E. sec. SERNEC (2016) - AGGREGATOR SYNTHESIS

Aggregation yields novel inferences of sympatry (R1,R4) & endemism (R2)

Distribution

	bifaria	divaric.	oricamp.
R1	(+)	(+)	-
R2	(+)	(-)	-
R3	+	+	-
R4	(+)	(+)	(+)

Data package 1
=> Ingestion

Data package 2
=> Ingestion

Data package 3
=> Ingestion

Rethinking centralized biodiversity data aggregation (2)

OXFORD ACADEMIC

DATABASE
The Journal of Biological Databases and Curation

Volumes ▾ Submit ▾ Alerts About ▾ All Database ▾

DATABASE
The Journal of Biological Databases and Curation

Volume 2018
2018

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 disenfranchizing taxonomic experts
- Backbone-based data signal distortion: an example
- An objection
- Conclusion
- Acknowledgements
- Funding

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** ▾

PDF Split View Cite Permissions

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 process. 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.

Rethinking centralized biodiversity data aggregation (2)

The screenshot shows the Oxford Academic Database website. The header includes the Oxford Academic logo and the Database title. The main content area features the article title, authors (Nico M Franz and Beckett W Sterner), publication information (Volume 2018, 2018, bax100), and a DOI link. Below the title are options for PDF, Split View, Cite, and Permissions. The abstract text is visible, discussing concerns about data quality and trust in aggregated biodiversity data. A table of contents is visible on the left side of the page.

OXFORD ACADEMIC
DATABASE
The Journal of Biological Databases and Curation

Volumes ▾ Submit ▾ Alerts About ▾ All Database ▾

DATABASE
The Journal of Biological Databases and Curation

Volume 2018
2018

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 disenfranchizing taxonomic experts
- Backbone-based data signal distortion: an example
- An objection
- Conclusion
- Acknowledgements
- Funding

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** ▾

PDF Split View Cite Permissions

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 process. 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 (2)



Volume 2018
2018

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 disenfranchizing taxonomic experts
- Backbone-based data signal distortion: an example
- An objection
- Conclusion
- Acknowledgements
- Funding

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](#) ▼

 PDF  Split View  Cite  Permissions

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 process. 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**.

Centralized biodiversity data aggregation

- Many collections remain offline.

"Hidden Data"



Individual Collections

Centralized biodiversity data aggregation

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



3A



3B



4

Individual Collections



5A



5B



6

"Hidden Data"

Individual Collections

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)

KE EMu



Multi-Collection Institution 1

Specify



Multi-Collection Institution 2



Individual Collections

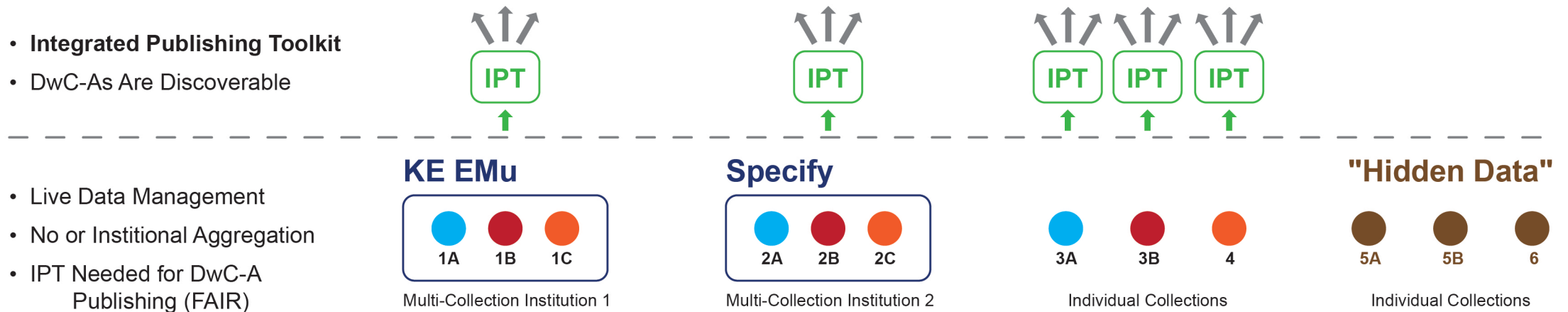
"Hidden Data"



Individual Collections

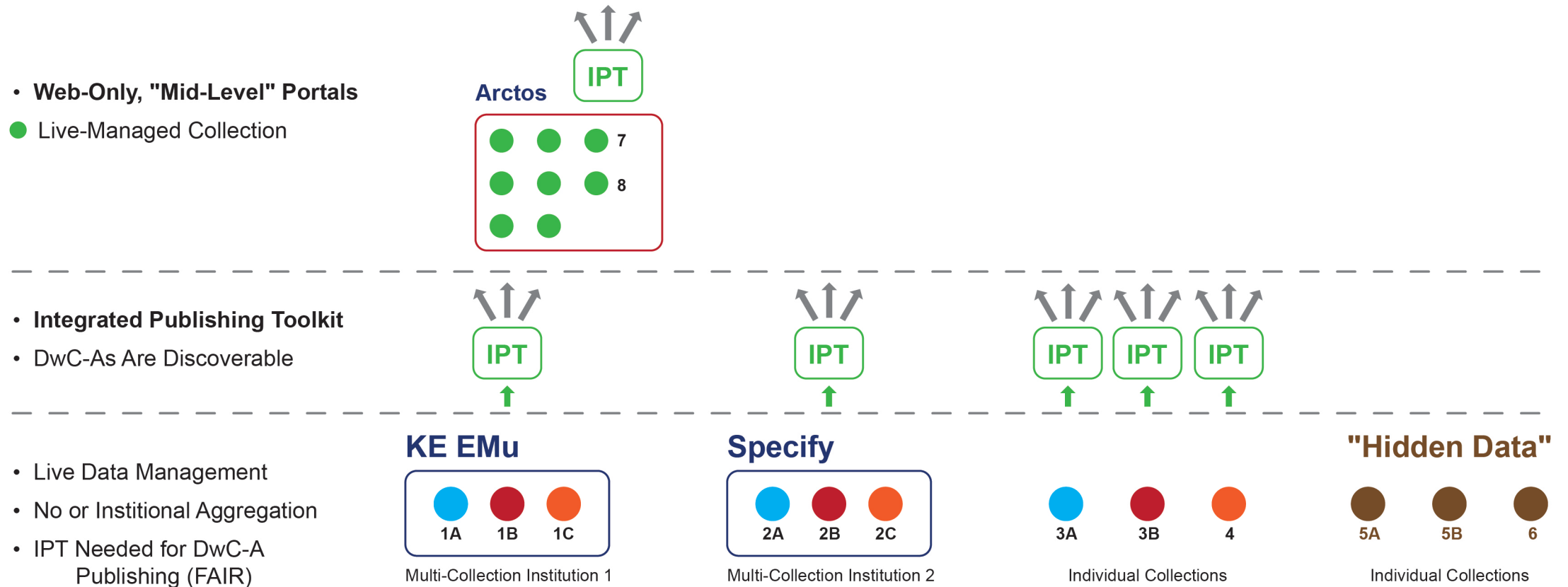
Centralized biodiversity data aggregation

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



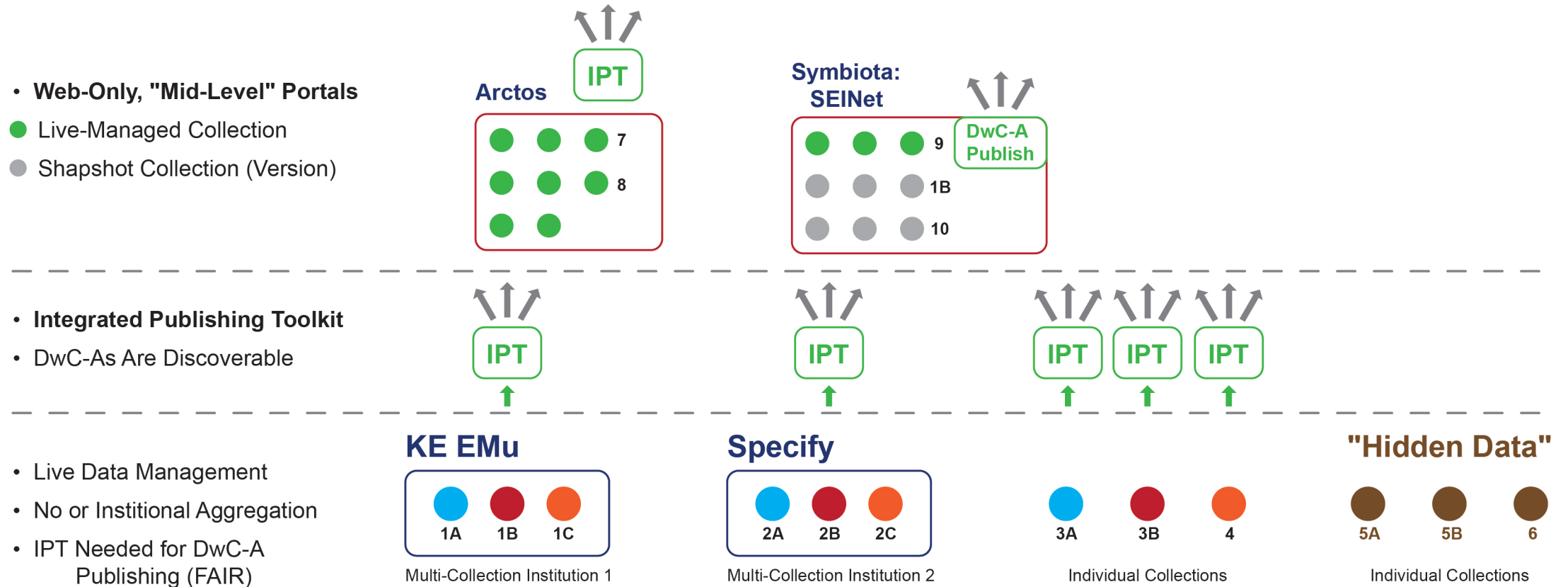
Centralized biodiversity data aggregation

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



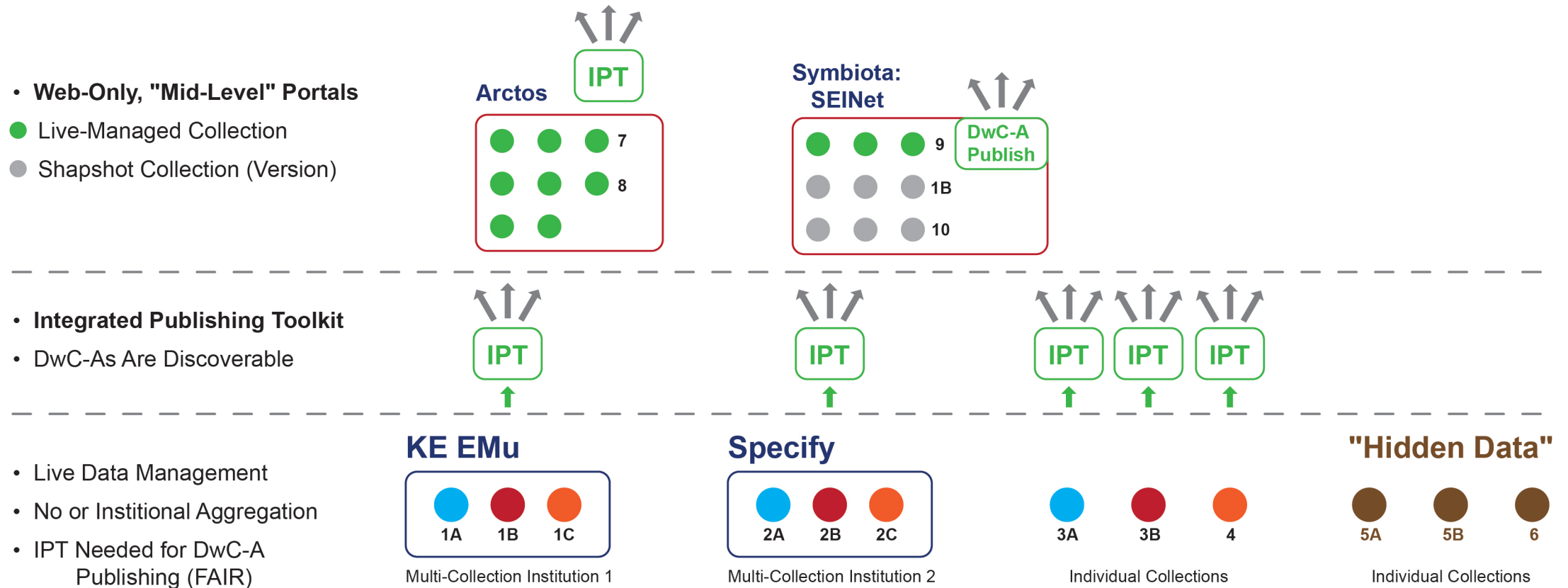
Centralized biodiversity data aggregation

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

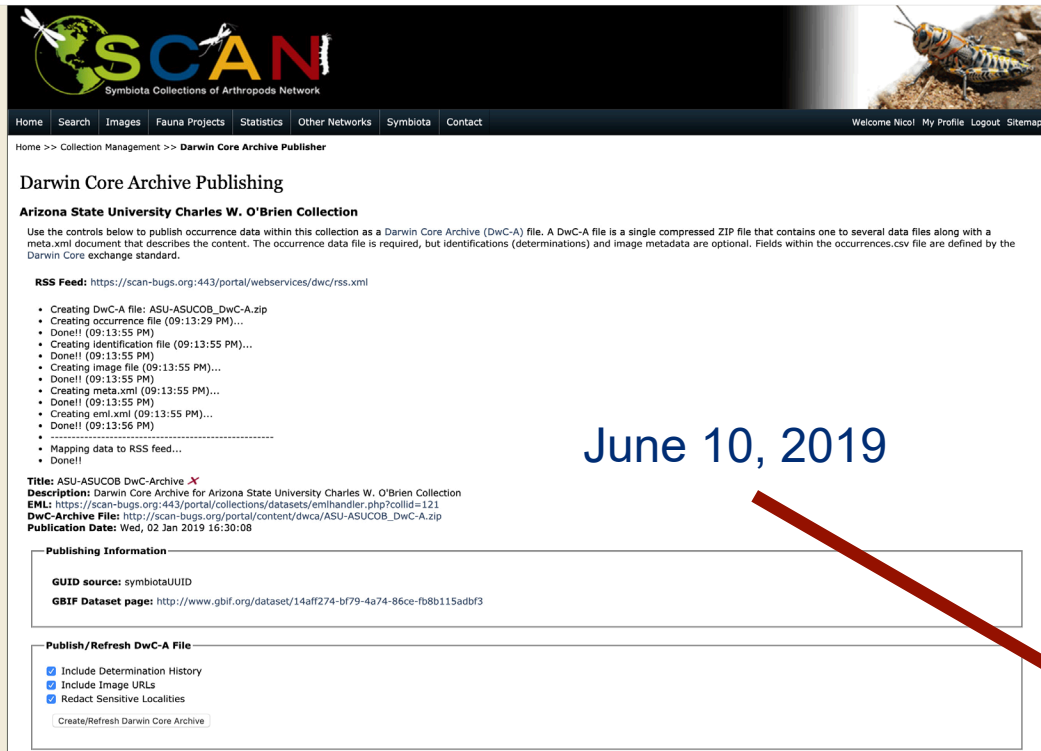


Centralized biodiversity data aggregation

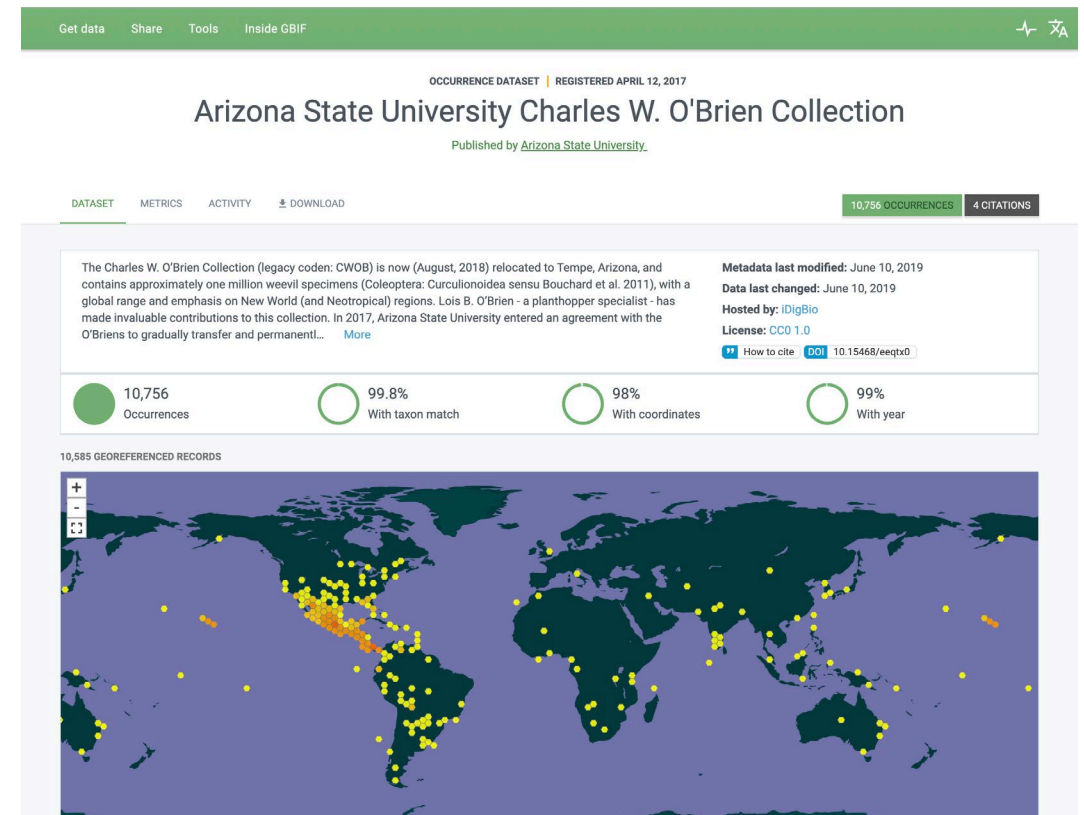
- Symbiota portals also have a custom, fully built-in, IPT-analogous "**Darwin Core Archive Publishing**" module.



Example of Symbiota's DwC-A Publishing module (SCAN: ASUCOB)



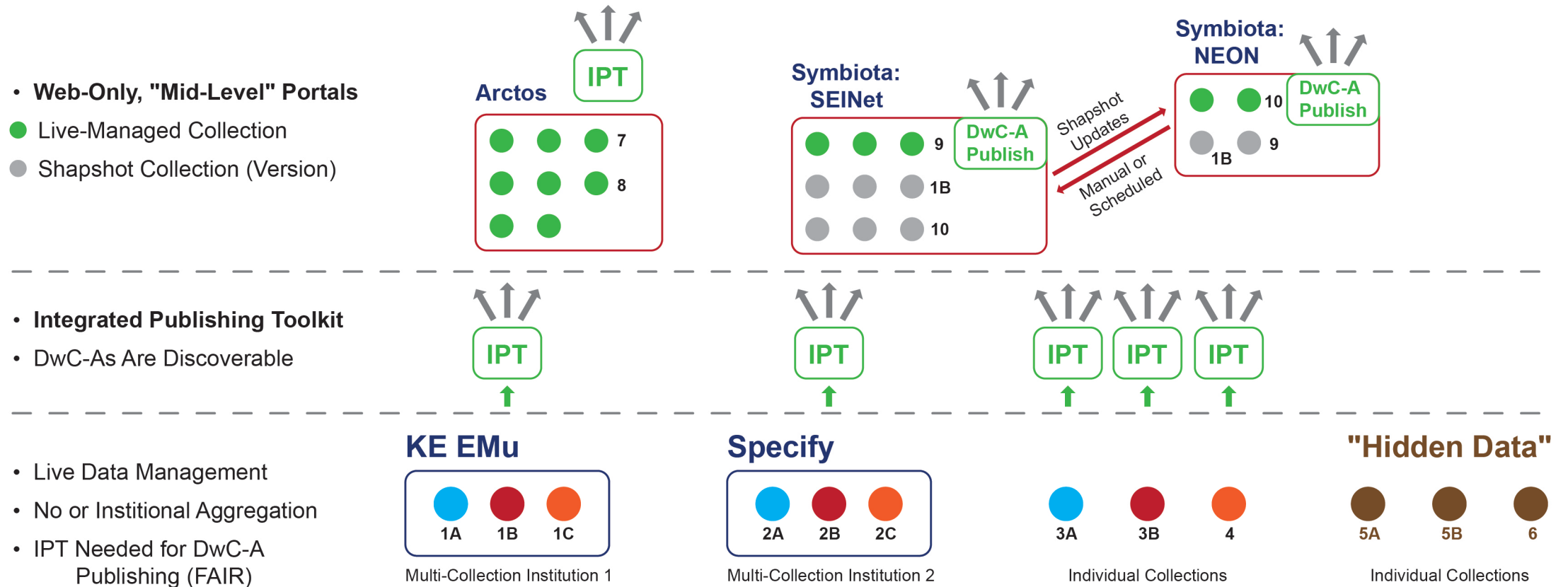
The screenshot shows the 'Darwin Core Archive Publishing' page for the 'Arizona State University Charles W. O'Brien Collection'. It includes a navigation bar with 'Home', 'Search', 'Images', 'Fauna Projects', 'Statistics', 'Other Networks', 'Symbiota', and 'Contact'. The main content area features a list of publishing progress items, a date stamp 'June 10, 2019', and a 'Publishing Information' section with fields for GUID source, GBIF dataset page, and options to include determination history, image URLs, and redact sensitive localities. A red arrow points from the date stamp to the right-hand page.



The screenshot shows the GBIF Occurrence Dataset page for the 'Arizona State University Charles W. O'Brien Collection'. It features a green header with 'Get data', 'Share', 'Tools', and 'Inside GBIF'. The main content area includes a description of the collection, metadata last modified on June 10, 2019, and a license of CC0 1.0. Below this are four circular progress indicators showing 10,756 occurrences, 99.8% with taxon match, 98% with coordinates, and 99% with year. At the bottom, there is a world map titled '10,585 GEOREFERENCED RECORDS' with yellow dots indicating specimen locations.

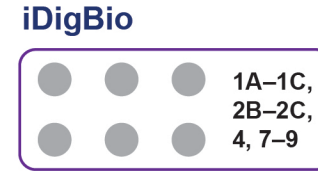
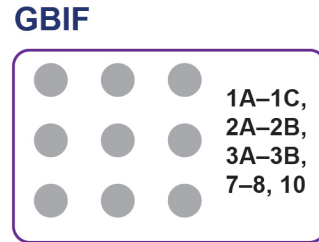
Centralized biodiversity data aggregation

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

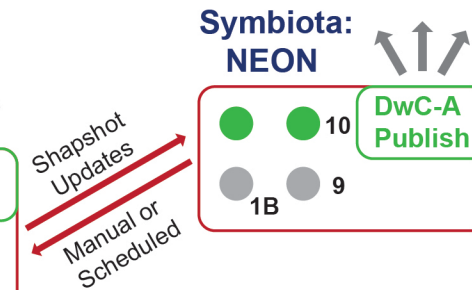
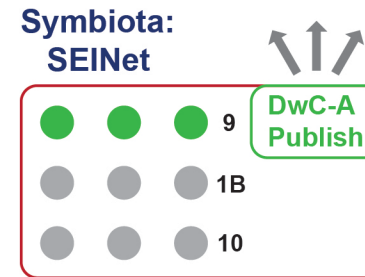
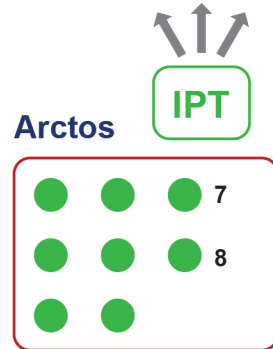


Highest-level aggregators typically only support collection snapshots!

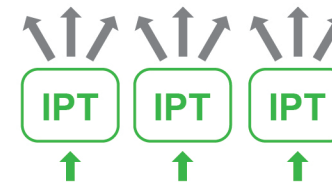
- Continental to Global Aggregators
- Overlapping Sets of Snapshots



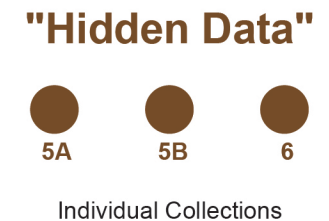
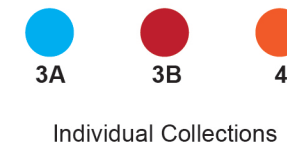
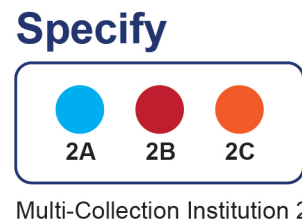
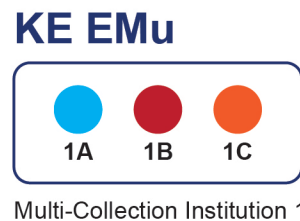
- Web-Only, "Mid-Level" Portals
- Live-Managed Collection
- Snapshot Collection (Version)



- Integrated Publishing Toolkit
- DwC-As Are Discoverable



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

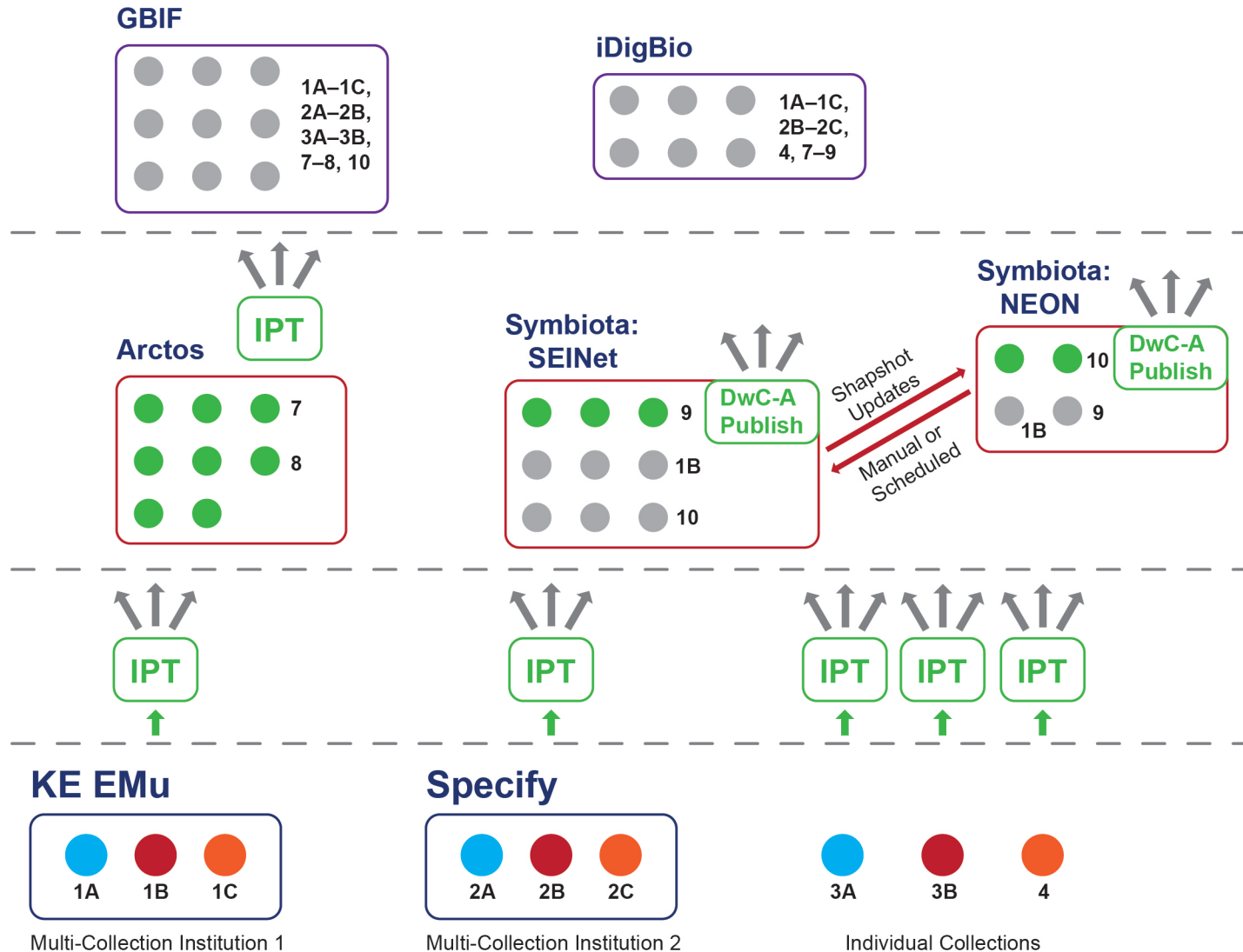


This hierarchy sustains an imbalance in directional data flow: Annotations on global datasets are hard to pull downwards.

- Global Discovery
- Data Annotation
- Published Analysis

Primary Flow of New Data / Annotations

- Expert Community Access and QC Engagement



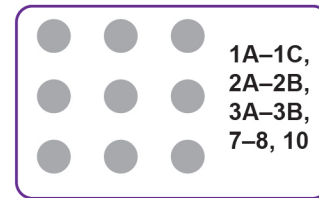
Moreover, by the time we reach the top, most **experts/enthusiasts no longer feel at home** (cf. [Wenger 2000](#)).

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

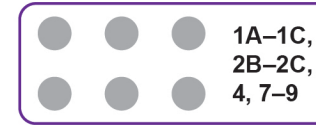
Communities of Practice



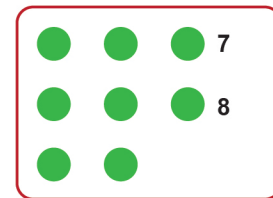
GBIF



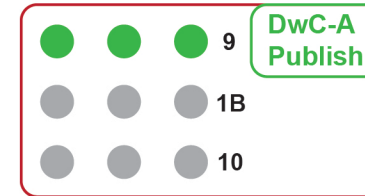
iDigBio



Arctos

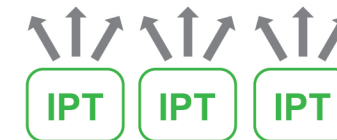
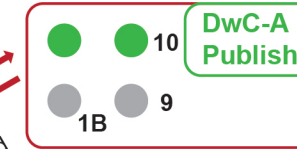


Symbiota: SEINet



Snapshot Updates
Manual or Scheduled

Symbiota: NEON



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

KE EMu



Multi-Collection Institution 1

Specify



Multi-Collection Institution 2



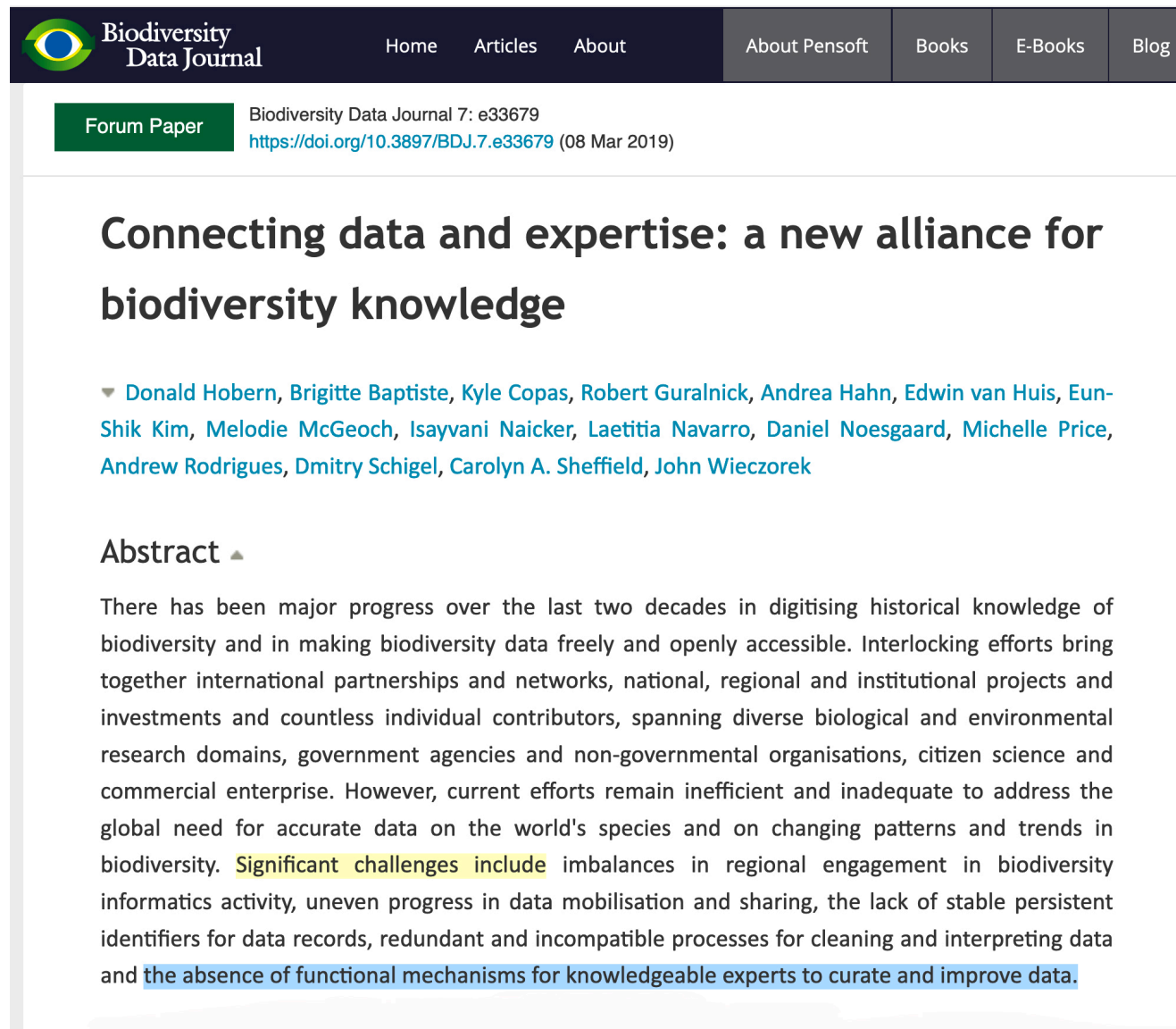
Individual Collections

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 image shows a screenshot of a web page from Biodiversity Data Journal. The page features a dark blue header with the journal's logo and navigation links. Below the header, there is a green box indicating the article type as a 'Forum Paper'. The article title is 'Connecting data and expertise: a new alliance for biodiversity knowledge', and the authors are listed as 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, and John Wieczorek. The abstract section begins with a paragraph discussing the progress in digitising historical knowledge and the challenges in biodiversity data management.

Biodiversity Data Journal Home Articles About About Pensoft Books E-Books Blog

Forum Paper 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

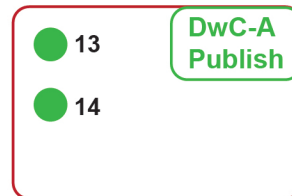
De-centralized, but global

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

A–F = Portals | 1–20 = Collections

● "Live-Managed" Collection

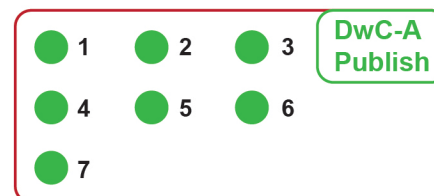
C. STRI



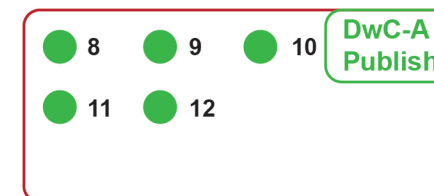
A: Consortium of Northeastern Herbaria

B: [Southwest Environment Information Network]

C: Smithsonian Tropical Research Institute



A: CNH



B. SEINet

De-centralized, but global

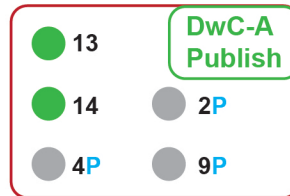
- **Partial, relevant *collection* snapshot subsets** are represented.

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

● "Live-Managed" Collection

● "Shapshot" Collection (Vs.)

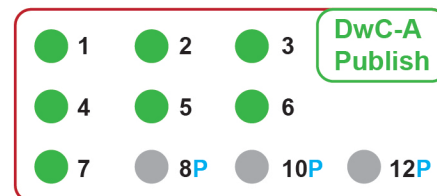
C. STRI



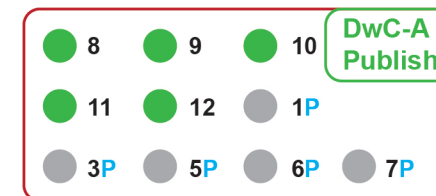
A: Consortium of Northeastern Herbaria

B: [Southwest Environment Information Network]

C: Smithsonian Tropical Research Institute



A: CNH



B. SEINet

De-centralized, but global

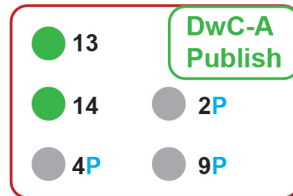
- Even **partial**, relevant *portal* snapshot **subsets** are ingestible, with provenance.

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

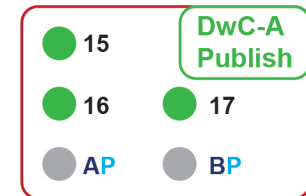
● "Live-Managed" Collection

● "Shapshot" Collection (Vs.)

C. STRI



D. NEON

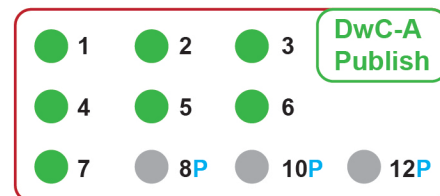


A: Consortium of Northeastern Herbaria

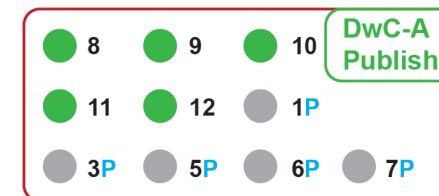
B: [Southwest Environment Information Network]

C: Smithsonian Tropical Research Institute

D: National Ecological Observatory Network



A: CNH



B. SEINet

De-centralized, but global

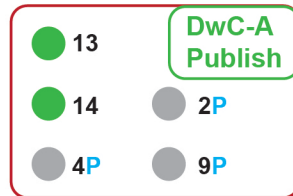
- Some research-themed portals may only include partial collection snapshots.

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

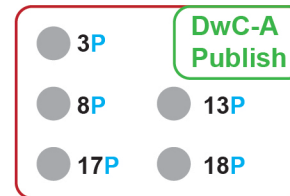
● "Live-Managed" Collection

● "Shapshot" Collection (Vs.)

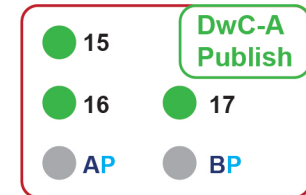
C. STRI



E. CoTRAM



D. NEON



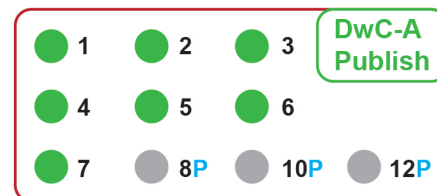
A: Consortium of Northeastern Herbaria

B: [Southwest Environment Information Network]

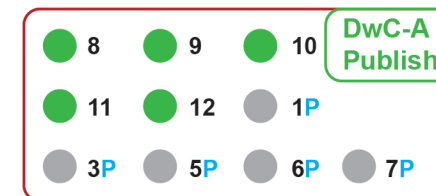
C: Smithsonian Tropical Research Institute

D: National Ecological Observatory Network

E: Cooperative Taxonomic Resource
for American Myrtaceae



A: CNH



B. SEINet

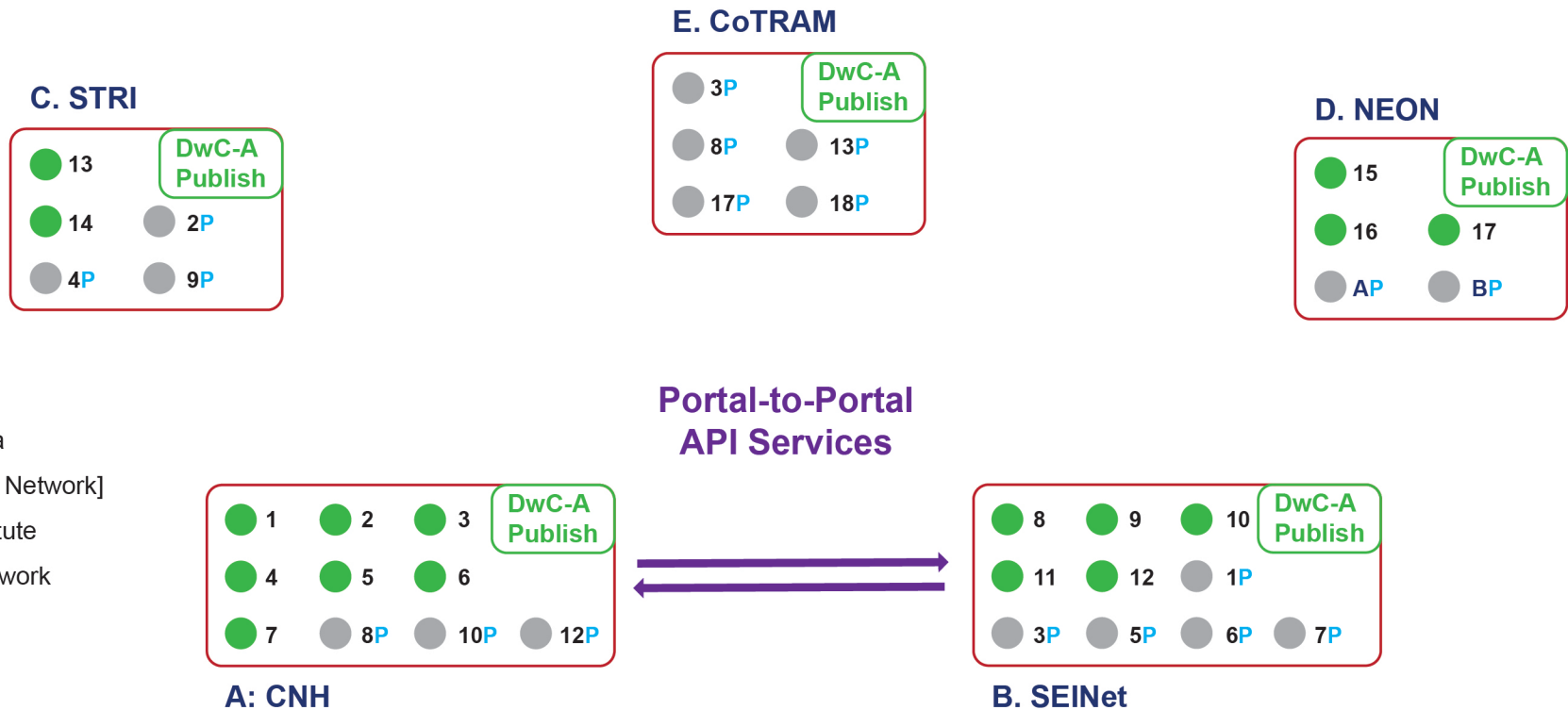
De-centralized, but global

- 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



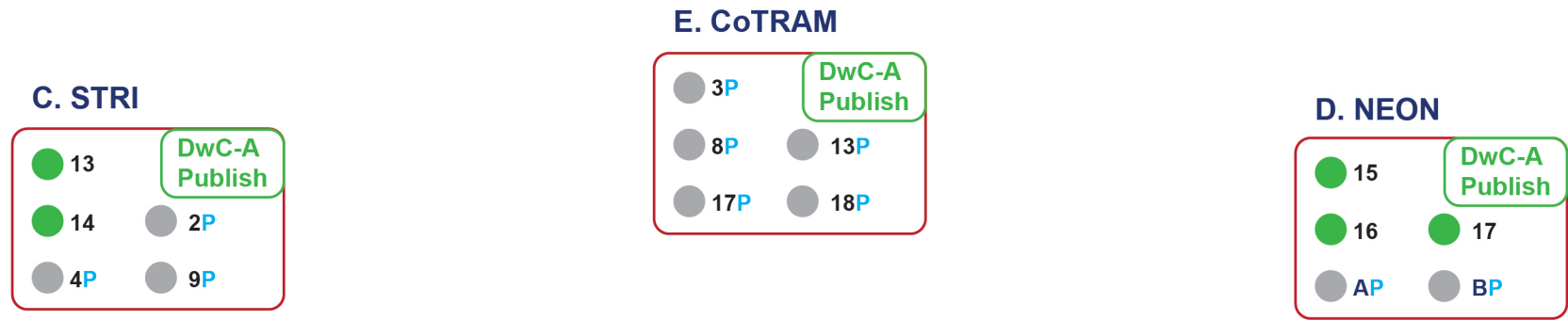
- A: Consortium of Northeastern Herbaria
- B: [Southwest Environment Information Network]
- C: Smithsonian Tropical Research Institute
- D: National Ecological Observatory Network
- E: Cooperative Taxonomic Resource for American Myrtaceae

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



Portal-to-Portal
API Services



- 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

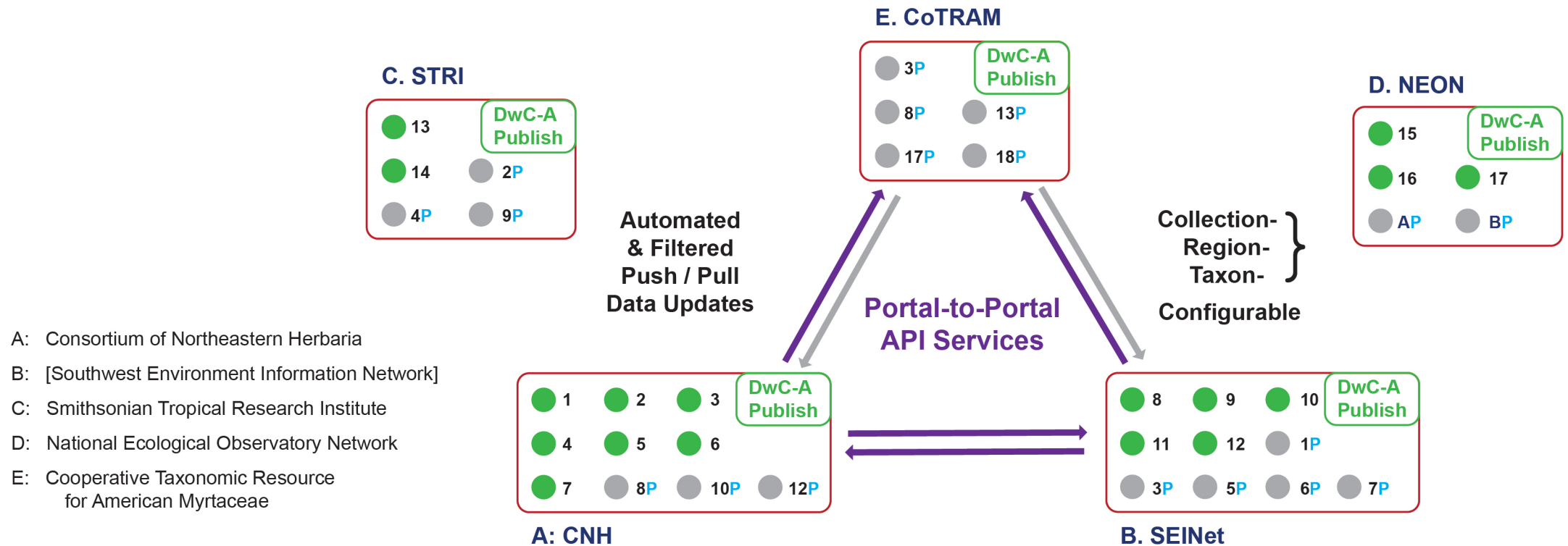
De-centralized, but global

- 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



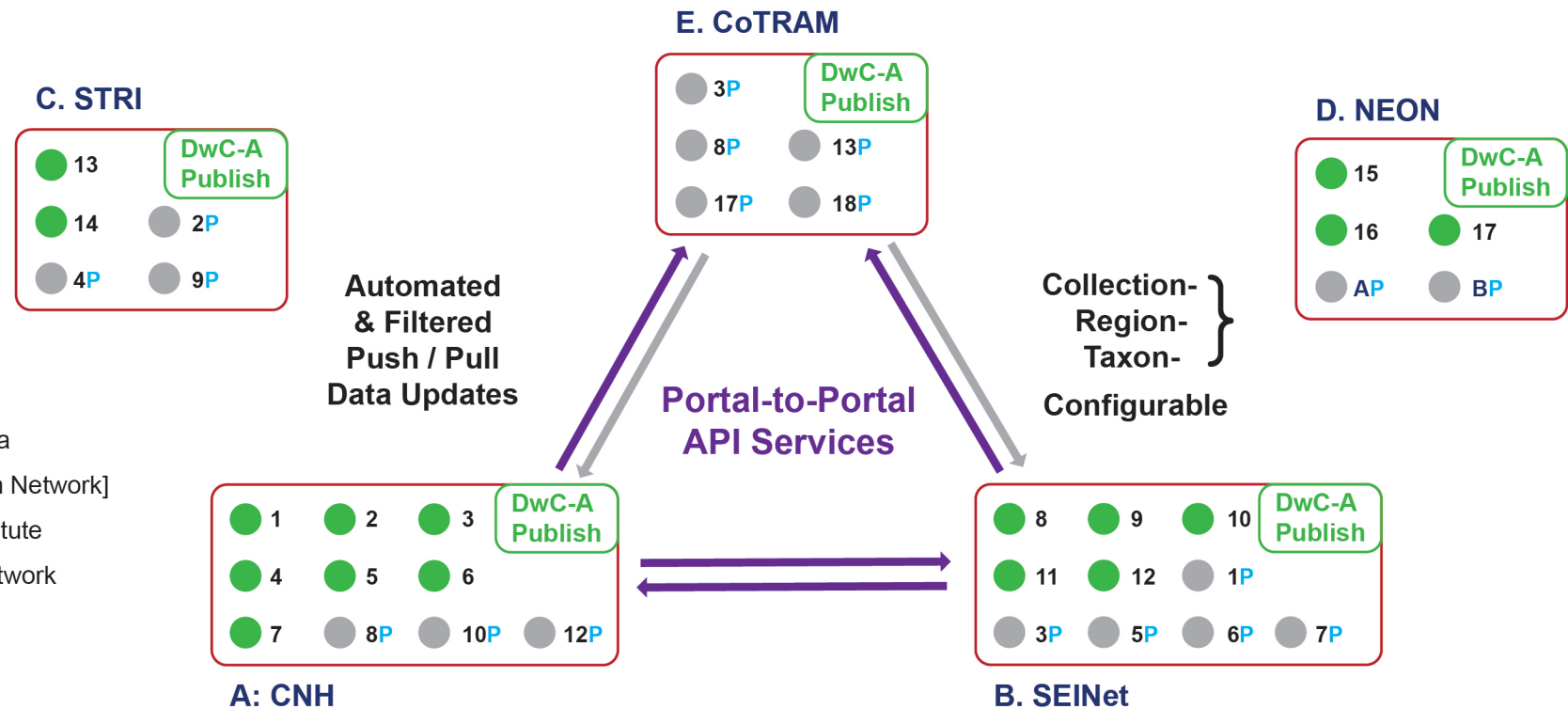
De-centralized, but global

- 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

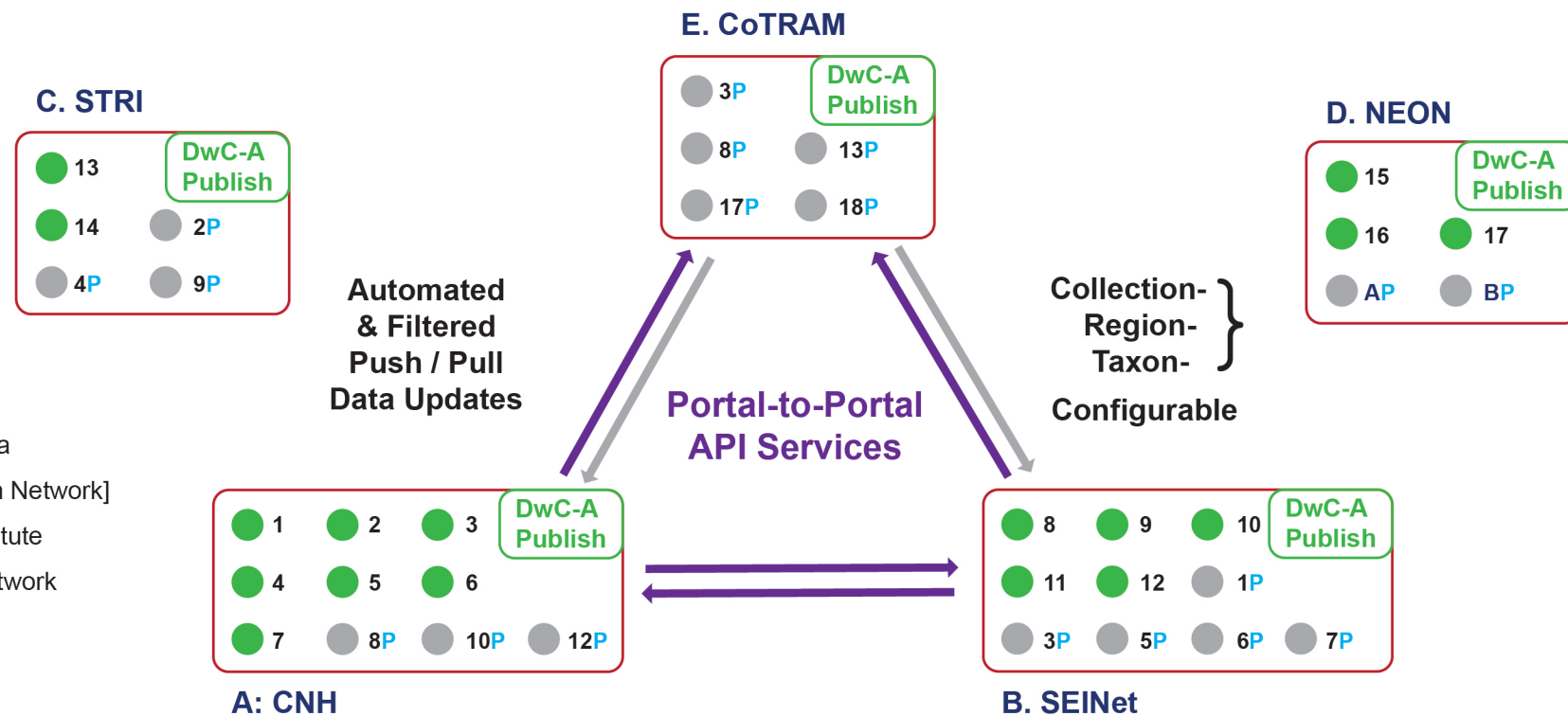


- A: Consortium of Northeastern Herbaria
- B: [Southwest Environment Information Network]
- C: Smithsonian Tropical Research Institute
- D: National Ecological Observatory Network
- E: Cooperative Taxonomic Resource for American Myrtaceae

De-centralized, but global

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

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

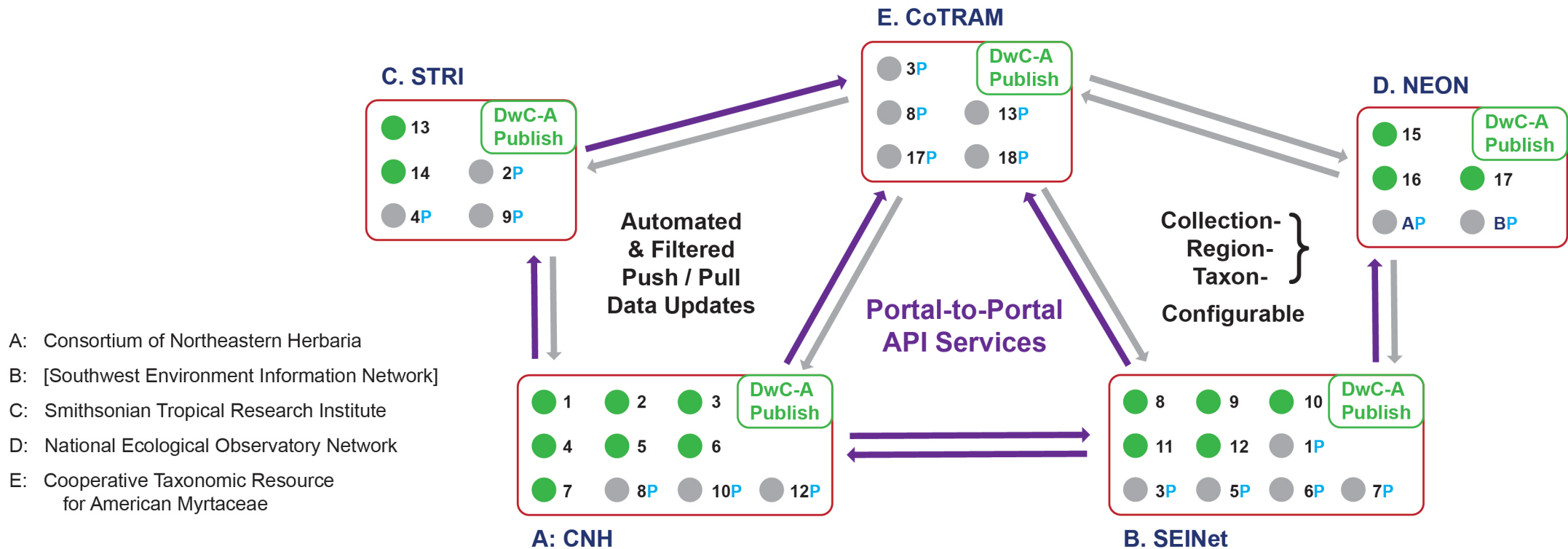
- The de-centralized network is **broadly extensible** between closely (high data flow) or remotely (low data flow) related communities.

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

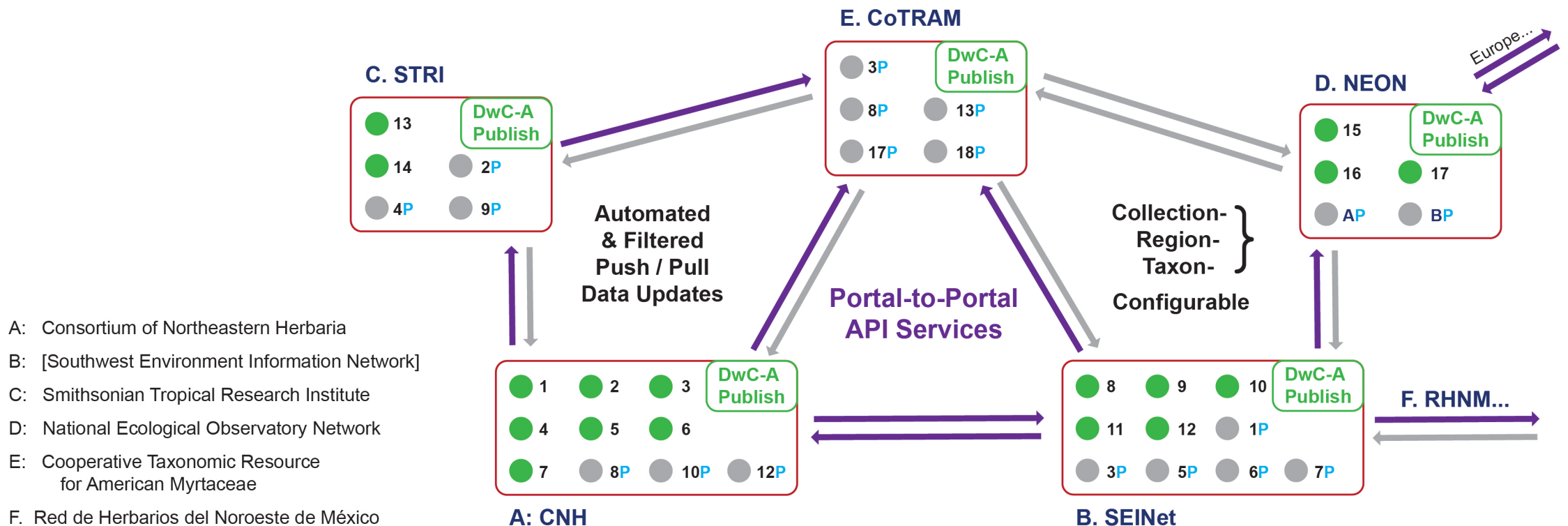
- On the basis of a shared API service culture, a de-centralized data portal network can potentially grow to attain **global coverage**.

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



BioCache = Project-Driven Portal Instance(s)

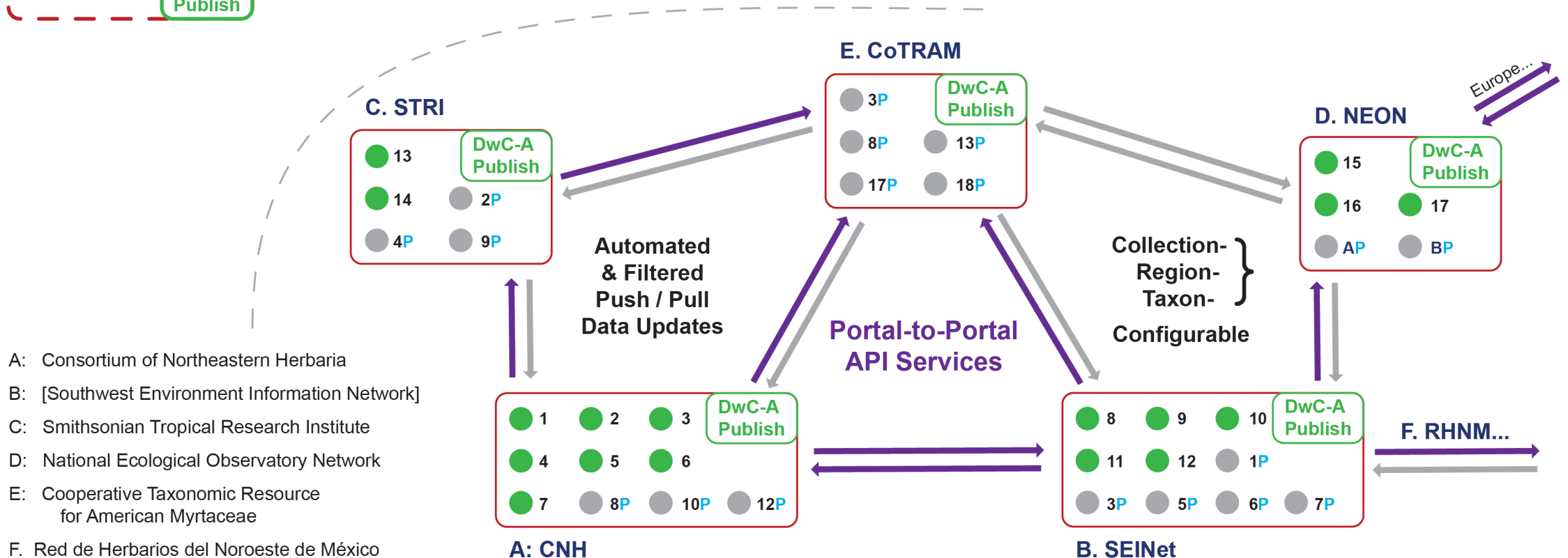
"Handshake" with one portal + "DNS Propagation"

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

● "Live-Managed" Collection = Direct Updates

● "Snapshot" Collection (Vs.) = API Updates

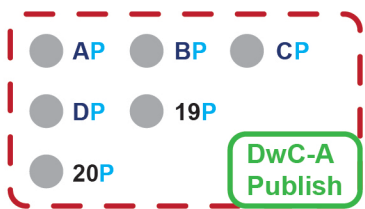
➡ Higher- / Lower- volume data flow



BioCache: Global access through custom research portal instances

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

BioCache



BioCache = Project-Driven Portal Instance(s)

"Handshake" with one portal + "DNS Propagation"

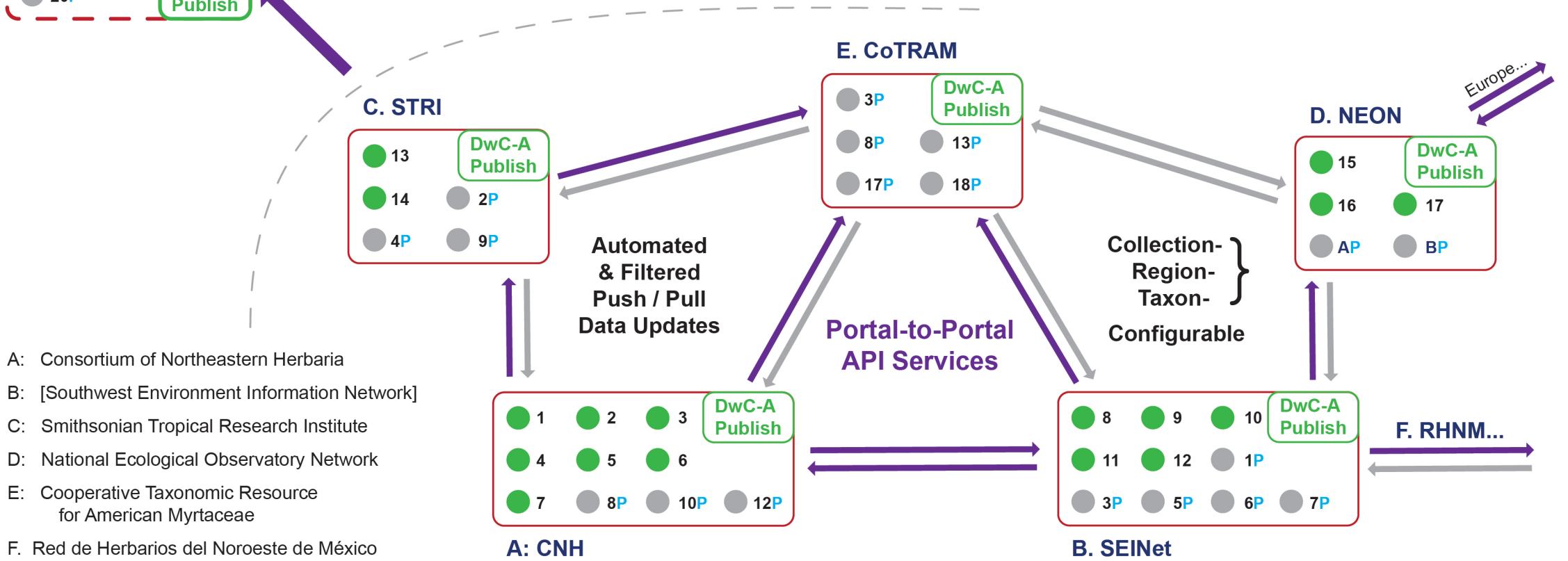
Supports network-wide, repeated data ingestion

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

● "Live-Managed" Collection = Direct Updates

● "Snapshot" Collection (Vs.) = API Updates

➡ Higher- / Lower- volume data flow

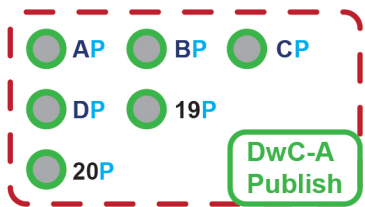


- A: Consortium of Northeastern Herbaria
- B: [Southwest Environment Information Network]
- C: Smithsonian Tropical Research Institute
- D: National Ecological Observatory Network
- E: Cooperative Taxonomic Resource for American Myrtaceae
- F: Red de Herbarios del Noroeste de México

BioCache: Global access through custom research portal instances

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

BioCache



BioCache = Project-Driven Portal Instance(s)

"Handshake" with one portal + "DNS Propagation"

Supports network-wide, repeated data ingestion

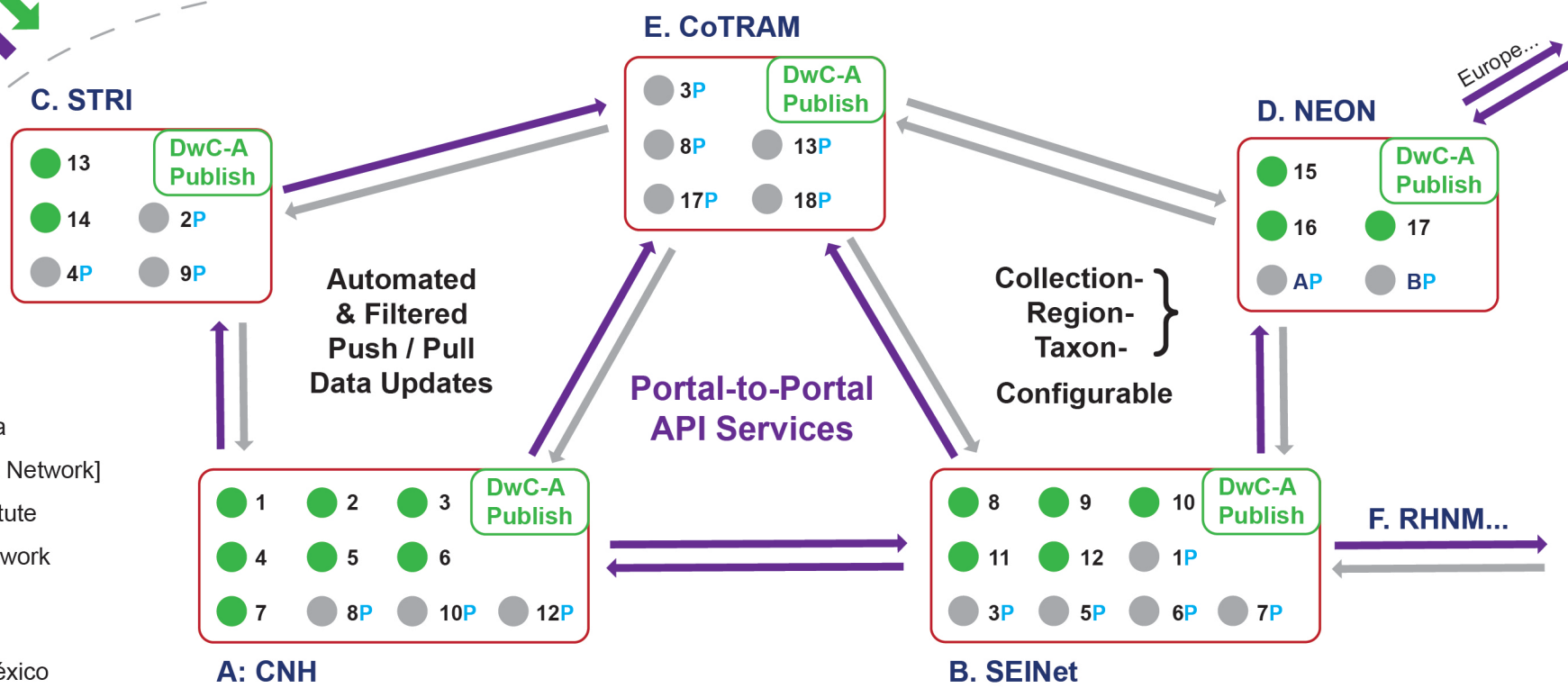
Annotations: Filtered pull back into live sources

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

A: CNH

B. SEINet

Stay tuned, it's underway

Taxonomic data intelligence for Darwin Core occurrences (3)



Home Articles About About Pensoft Books E-Books Blog Jour

NSF Grant Proposal Research Ideas and Outcomes 2: e10610
<https://doi.org/10.3897/rio.2.e10610> (30 Sep 2016)

Reviewable v1 

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

▼ Nico Franz, Edward Gilbert, Bertram Ludäscher, Alan Weakley

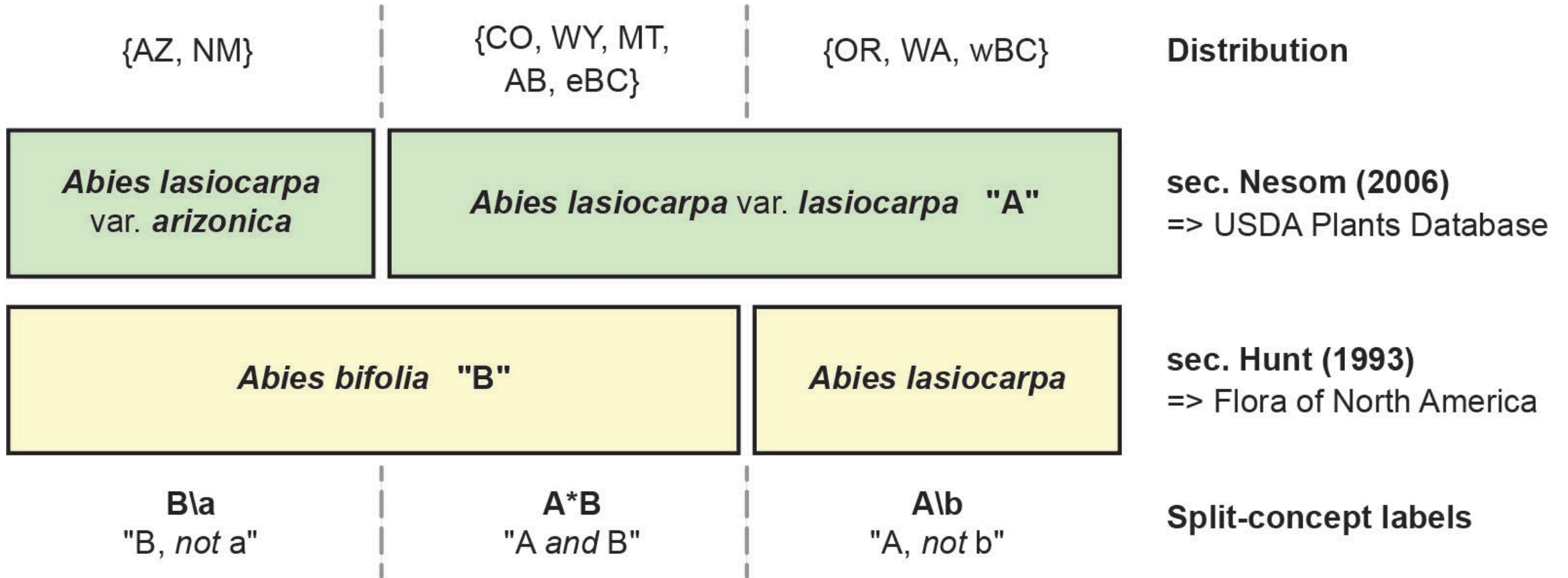
Executive summary ▲

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>)



Taxonomic data intelligence for Darwin Core occurrences (3)

 Home Articles About About Pensoft Books E-Books Blog Jour

NSF Grant Proposal Research Ideas and Outcomes 2: e10610
<https://doi.org/10.3897/rio.2.e10610> (30 Sep 2016)

Reviewable v1 

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.
- Spatial reasoning tools (RCC–5) can help attain consistent and comprehensive **taxonomic meaning mappings** ↔ **intelligence** for data integration across evolving or conflicting views.

Taxonomic data intelligence for Darwin Core occurrences (3)

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

▼ Nico Franz, Edward Gilbert, Bertram Ludäscher, Alan Weakley

Executive summary ▲

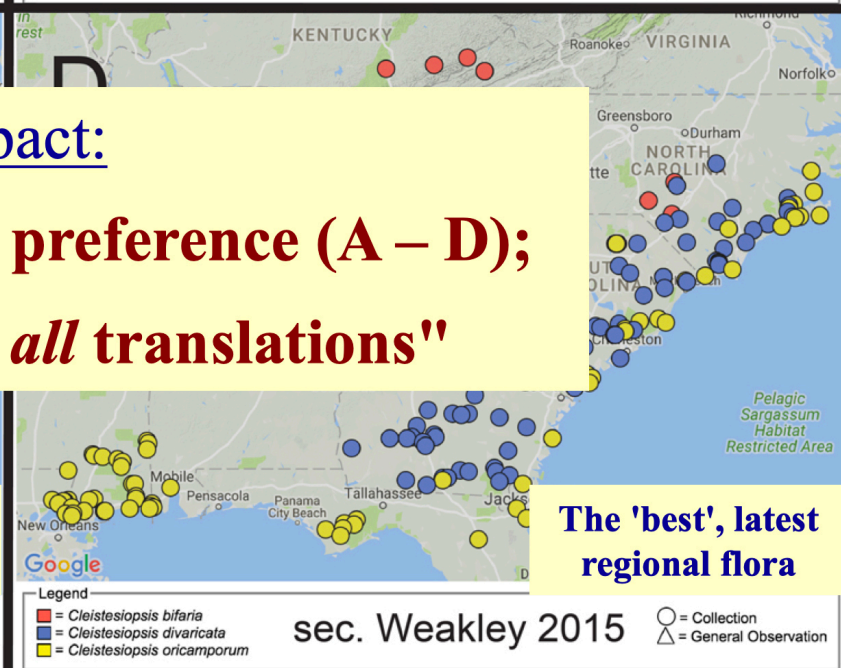
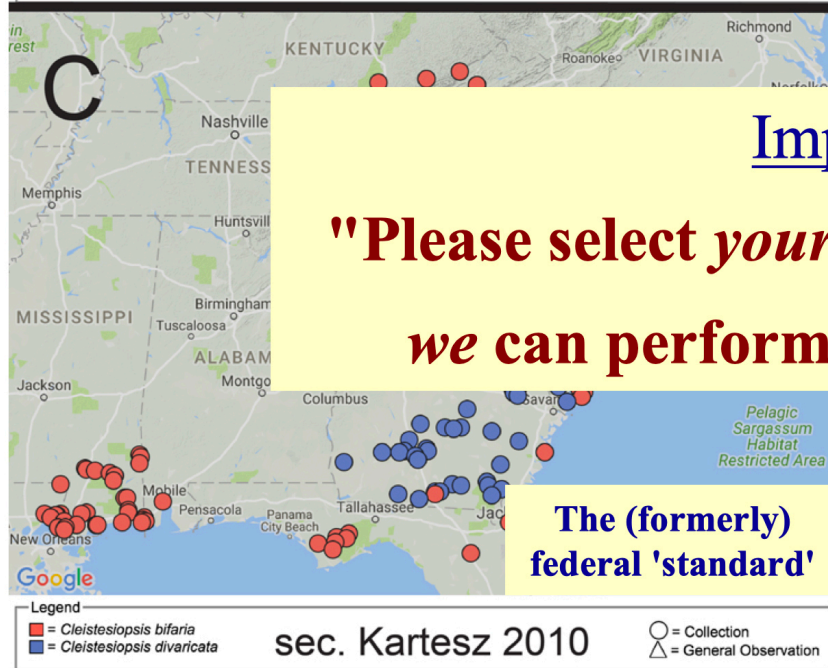
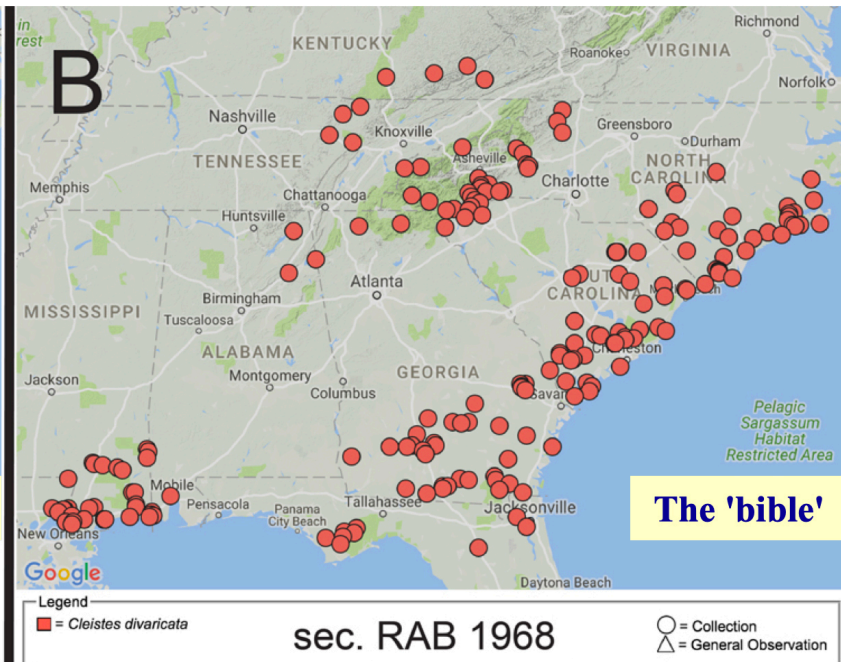
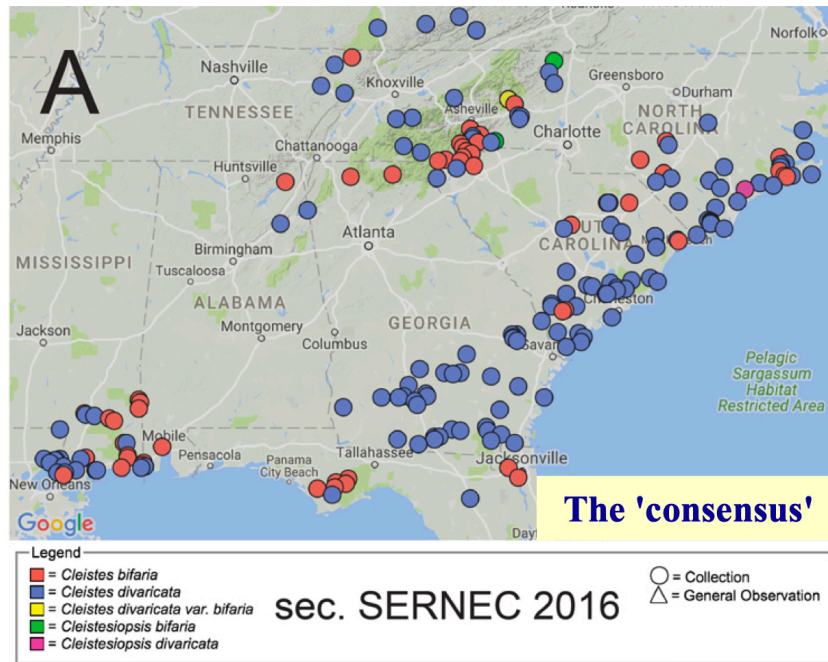
Overview. Taxonomic names are imperfect identifiers of specific and sometimes conflicting taxonomic perspectives in aggregated biodiversity data environments. The inherent ambiguities of names can be mitigated using syntactic and semantic conventions developed under the taxonomic concept approach. These 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.
- 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"



Impact:

**"Please select *your* preference (A – D);
we can perform *all* translations"**

Project remains pending, but look!

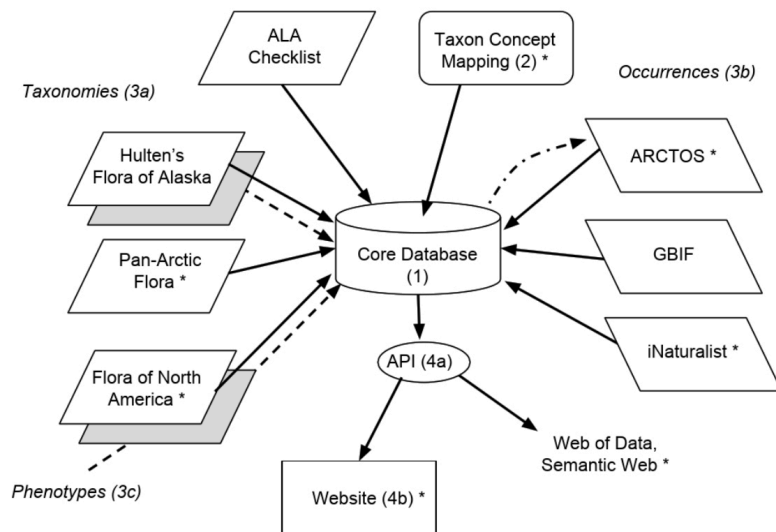
"Taxonomically intelligent data integration for a new **Flora of Alaska**"

Blog 1: Project roadmap

Posted by [Cam](#) on 2018-08-17

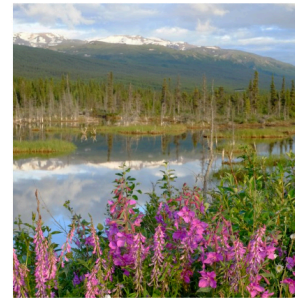
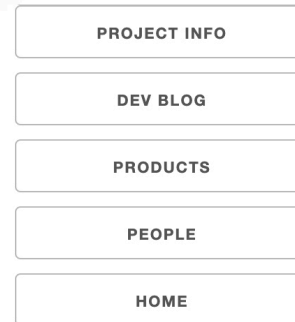
Hi! Welcome to this development blog, this website, and this project. Steffi and I were so pleased to get this NSF funding — many, many thanks to the ABI panelists, the four reviewers, and to NSF and US taxpayers.

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:



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., [4store](#), [Allegrograph](#),



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

iNaturalist is mostly there already

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 Trapa hyrcana (parent: Genus Trapa)	Match	Species Trapa hyrcana (parent: Genus Trapa)

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 Lycopodium obtusifolium (parent:)	One-to-one	Species Phlegmariurus obtusifolius (parent: Genus Phlegmariurus)

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

iNaturalist is mostly there already



[« Back to Taxon Framework Relationship](#)

[View taxonomy details for Ageroniini](#)

Taxon Framework Relationship 71572

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 Epicaliina (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)

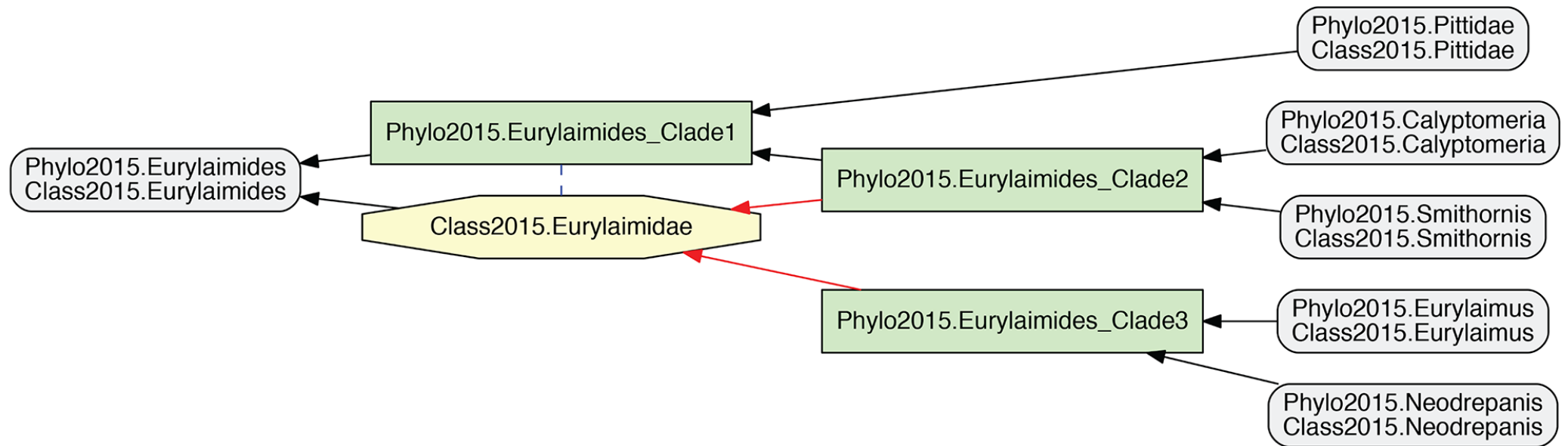
iNat splits Catonephelini/ina from Epicaliini/ina and has Eubagini, Epicaliini, Callicorini, Ageroniini. Epiphilini as tribes sibling to Biblidini rather than as subtribes within Biblidini

Downstream deviations for Epicaliini 1

https://www.inaturalist.org/taxon_framework_relationships

Recent addition: Concept alignment for phylogenomic trees

Nodes	
congruent	6
Phylo2015	3
Class2015	1
Edges	
overlaps	1
is_a (inferred)	2
is_a (input)	8



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>