

Conference Abstract

An Information Ecosystem Map of Resources Supporting the Mobilization and Discovery of Paleontological Specimen Data

Holly Little[‡], Talia Karim[§], Erica Krimmel[|], Lindsay J. Walker[¶]

[‡] Smithsonian National Museum of Natural History, Washington, DC, United States of America

[§] University of Colorado Museum of Natural History, Boulder, United States of America

[|] Independent, Sacramento, United States of America

[¶] Arizona State University, Tempe, United States of America

Corresponding author: Holly Little (littleh@si.edu), Talia Karim (talia.karim@colorado.edu)

Received: 10 Oct 2024 | Published: 11 Oct 2024

Citation: Little H, Karim T, Krimmel E, Walker LJ (2024) An Information Ecosystem Map of Resources Supporting the Mobilization and Discovery of Paleontological Specimen Data. Biodiversity Information Science and Standards 8: e139040. <https://doi.org/10.3897/biss.8.139040>

Abstract

Over the last decade, the United States paleontological collections community has invested heavily in the digitization of specimen-based data, including over 10 million USD funded through the [National Science Foundation's Advancing Digitization of Biodiversity Collections](#) program. Fossil specimen data—9.0 million records and counting (Global Biodiversity Information Facility 2024)—are now accessible on open science platforms such as the [Global Biodiversity Information Facility \(GBIF\)](#). However, the full potential of this data is far from realized due to fundamental challenges associated with mobilization, discoverability, and interoperability of paleontological information within the existing cyberinfrastructure landscape and data pipelines. Additionally, it can be difficult for individuals with varying expertise to develop a comprehensive understanding of the existing landscape due to its breadth and complexity. Here, we present preliminary results from a project aiming to explore how we might address these problems.

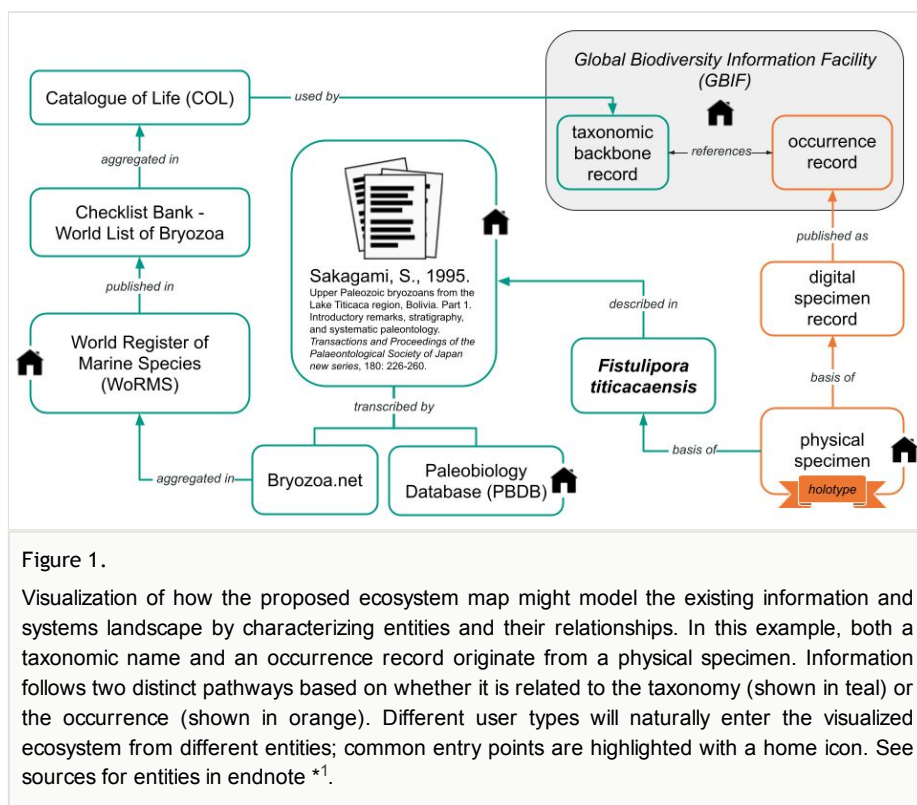
Funding from the US National Science Foundation (NSF) to the University of Colorado Museum of Natural History, Smithsonian National Museum of Natural History, and Arizona State University will result in, among other products, an “ecosystem map” for the

paleontological collections community. This map will be an information-rich visualization of entities (e.g. concepts, systems, platforms, mechanisms, drivers, tools, documentation, data, standards, people, organizations) operating in, intersecting with, or existing in parallel to our domain. We are inspired and informed by similar efforts to map the biodiversity informatics landscape (Bingham et al. 2017) and the research infrastructure landscape (Distributed System of Scientific Collections 2024), as well as by many ongoing metadata cataloging projects, e.g. [re3data](#) and the [Global Registry of Scientific Collections \(GRSciColl\)](#). Our strategy for developing this ecosystem map is to model the existing information and systems landscape by characterizing entities, e.g. potentially in a graph database as nodes with relationships to other nodes.

The ecosystem map will enable us to provide guidance for communities working *across* different sectors of the landscape, promoting a shared understanding of the ecosystem that everyone works in together. We can also use the map to identify points of entry and engagement at various stages of the paleontological data process, and to engage diverse members *within* the paleontological community. We see three primary user types for this map: people new(er) to the community, people with expertise in a subset of the community, and people working to integrate initiatives and systems across communities. Each of these user types needs tailored access to the ecosystem map and its community knowledge. By promoting shared knowledge with the map, users will be able to identify their own space within the ecosystem and the connections or partnerships that they can utilize to expand their knowledge or resources, relieving the burden on any single individual to hold a comprehensive understanding.

For example, the flow of taxonomic information between publications, collections, digital resources, and biodiversity aggregators is not straightforward or easy to understand. A person with expertise in collections care may want to use the ecosystem map to understand why taxonomic identifications associated with their specimen occurrence records are showing up incorrectly when published to GBIF. We envision that our final ecosystem map will visualize the flow of taxonomic information and how it is used to interpret specimen occurrence data, thereby highlighting to this user where problems may be happening and whom to ask for help in addressing them (Fig. 1).

Ultimately, development of this map will allow us to identify mobilization pathways for paleontological data, highlight core cyberinfrastructure resources, define cyberinfrastructure gaps, strategize future partnerships, promote shared knowledge, and engage a broader array of expertise in the process. Contributing domain-based evidence FAIRly*² requires expertise that bridges the content (e.g. paleontology) and the mechanics (e.g. informatics). By centering the role of humans in open science cyberinfrastructure throughout our process, we hope to develop systems that create and sustain such expertise.



Keywords

paleontology, data ecosystem

Presenting author

Holly Little

Presented at

SPNHC-TDWG 2024

Funding program

United States National Science Foundation (NSF) Geosciences Open Science Ecosystem (GEO OSE) Award #s: [2324688](#), [2324689](#), [2324690](#).

Grant title

Collaborative Research: GEO OSE Track 1: Community-driven enhancement of information ecosystems for the discovery and use of paleontological specimen data.

Conflicts of interest

The authors have declared that no competing interests exist.

References

- Bingham H, Doudin M, Weatherdon L, Despot-Belmonte K, Wetzel F, Groom Q, Lewis E, Regan E, Appeltans W, Güntsch A, Mergen P, Agosti D, Penev L, Hoffmann A, Saarenmaa H, Geller G, Kim K, Kim H, Archambeau A, Häuser C, Schmeller D, Geijzendorffer I, García Camacho A, Guerra C, Robertson T, Runnel V, Valland N, Martin C (2017) The Biodiversity Informatics Landscape: Elements, Connections and Opportunities. Research Ideas and Outcomes 3 <https://doi.org/10.3897/rio.3.e14059>
- Distributed System of Scientific Collections (2024) Research Infrastructure Landscape. <https://www.dissco.eu/dissco/ri-landscape/>. Accessed on: 2024-7-12.
- Global Biodiversity Information Facility (2024) GBIF Occurrence Download. <https://doi.org/10.15468/dl.w8gnuh>. Accessed on: 2024-7-11.

Endnotes

*1 Entity sources, often including a unique identifier embedded in the URL:

- Bryozoa.net, <https://bryozoa.net/annual/taxa1995.html>
- PBDB, https://paleobiodb.org/classic/basicTaxonInfo?taxon_no=165842
- WoRMS, <https://www.marinespecies.org/aphia.php?p=taxdetails&id=1403202>
- World List of Bryozoa, <https://doi.org/10.48580/dfrdl-3bb>
- Checklist Bank, <https://www.checklistbank.org/dataset/294826/taxon/867QJ>
- Catalogue of Life, <https://www.catalogueoflife.org/data/taxon/867QJ>
- GBIF taxonomic backbone record, <https://www.gbif.org/species/8496836>
- GBIF occurrence record, <https://www.gbif.org/occurrence/4899188245>
- digital specimen record, https://db.kahaku.go.jp/webmuseum_en/detail?cls=col_c3_01&pkey=NMNS%20PA,18102,

*2 FAIRly, i.e. in a manner that is Findable, Accessible, Interoperable, and Reusable. See [FAIR Principles](#) for more.