

Community-driven enhancement of information ecosystems for the discovery and use of palaeontological specimen data: Cyberinfrastructure alignment workshop

Talia Karim[‡], Erica Kimmel[§], Holly Little^l, Lindsay J. Walker[¶]

[‡] University of Colorado Boulder, Boulder, United States of America

[§] independent, Sacramento, United States of America

^l Smithsonian National Museum of Natural History, Washington, DC, United States of America

[¶] University of Kansas, Lawrence, United States of America

Corresponding author: Talia Karim (talia.karim@colorado.edu)

Reviewable

v 1

Received: 19 Aug 2025 | Published: 04 Sep 2025

Citation: Karim T, Kimmel E, Little H, Walker LJ (2025) Community-driven enhancement of information ecosystems for the discovery and use of palaeontological specimen data: Cyberinfrastructure alignment workshop. Research Ideas and Outcomes 11: e169264. <https://doi.org/10.3897/rio.11.e169264>

Abstract

A two day cyberinfrastructure alignment workshop was held in May 2025 as part of the "Community-driven enhancement of information ecosystems for the discovery and use of paleontological specimen data" project, which is funded under the United States National Science Foundation (NSF) Geosciences Open Science Ecosystem (GEO OSE) programme. Participants with expertise in informatics, technical cyberinfrastructure development and management and geo- and biological sciences were invited to foster a shared and increased understanding across this broad-community of the needs for palaeontological specimen data. This report describes the activities and outcomes of the workshop and how they will contribute to final deliverables for the grant funded project.

Keywords

paleontology, palaeontology, fossil, geology, biodiversity, collection, natural history collection, specimen, cyberinfrastructure

Date and place

This workshop took place from 13-14 May 2025 at the University of Colorado Boulder (CU) and was hosted by the [CU Museum of Natural History](#).

List of participants

The twenty-four participants of this workshop (Table 1) were a mix of professionals focusing on informatics, technical development and technical product management, as well as subject matter experts from the geo- and biosciences.

Table 1. List of on-site workshop participants. Workshop organisers are indicated with an asterisk (*).		
Name	Institutional affiliation	Role or title
Adam Mansur	Smithsonian National Museum of Natural History	Data Manager of Mineral Sciences
Alex Lawrence	Smithsonian National Museum of Natural History	Data Specialist
Andrea Thomer	University of Arizona College of Information Science	Associate Professor
Ben Norton	Independent	Informatics Consultant
Carl Simpson	University of Colorado Boulder, Museum of Natural History	Assistant Professor of Geological Sciences and Curator of Invertebrate Paleontology
Casey Thater	University of Colorado Boulder, Museum of Natural History	Graduate student (museum and field studies)
Cecilie Svenningsen	Global Biodiversity Information Facility	Data Administrator
Daven Quinn	University of Wisconsin-Madison Department of Geoscience	Assistant Scientist
David Bloom	Biodiversity Information Standards (TDWG) VertNet	Chair
Ed Gilbert	University of Kansas	Symbiota Support Hub IT Management Lead + NEON
Erica Krimmel*	Independent	Information Scientist
Erin Robinson	University of Colorado Boulder Metadata Game Changers LLC	Graduate student (information science) CEO

Name	Institutional affiliation	Role or title
Geoff Ower	Illinois Natural History Survey, Species File Group	Research Programmer
Holly Little*	Smithsonian National Museum of Natural History	Informatics Manager of Paleobiology
Jack Shaw	University of Colorado Boulder	postdoc
JJ Dearborn	Smithsonian Libraries and Archives	Data Manager for the Biodiversity Heritage Library
Kit Lewers	University of Colorado Boulder	Graduate student (information science)
Laura Rocha Prado	Denver Museum of Nature & Science	Data Administrator
Lindsay Walker*	University of Kansas	Symbiota Support Hub Community Manager
Lindsey Pfeninger	University of Colorado Boulder, Museum of Natural History	Graduate student (museum and field studies)
Nancy Stevens	University of Colorado Boulder, Museum of Natural History	Director and Professor of Anthropology
Nicole McGee	University of Colorado Boulder, Museum of Natural History	Graduate student (museum and field studies)
Sharon Grant	Field Museum of Natural History	Information Systems Director
Talia Karim*	University of Colorado Boulder, Museum of Natural History	Collections Manager of Invertebrate Paleontology and Paleobotany

Introduction

This workshop is part of the "Community-driven enhancement of information ecosystems for the discovery and use of paleontological specimen data" project, which is funded under the United States National Science Foundation (NSF) [Geosciences Open Science Ecosystem \(GEO OSE\) program](#). The goal of the project is to support transformational and translational research in the geo- and biosciences by driving development in the open data landscape and by improving discoverability and use of palaeontological specimen data through community engagement and collaboration. Project personnel are actively coordinating with partners throughout the larger data ecosystem, including via two in-person workshops, of which this is the second.

Palaeontology collections in the United States have spent much of the last 15 years digitising primary specimen data and sharing it via open science platforms, such as the [Global Biodiversity Information Facility \(GBIF\)](#)*¹ and [iDigBio](#). While palaeontology is an inherently interdisciplinary field spanning both the geological and biological sciences, the platforms hosting collections-based data were developed primarily for modern biological (neontological) data. As a result of the original focus on neontological data,

gaps in informatics and cyberinfrastructure obscure critically useful primary data from palaeontology collections and inhibit integration between open science resources operating in geoscience (e.g. [Paleobiology Database](#), [Neotoma Paleocology Database](#)) vs. bioscience domains. This workshop built on the results of an earlier workshop (see Karim et al. (2024)) and provided an important opportunity to connect with representatives from across the broader biodiversity informatics landscape.

Aims of the workshop

The desired outcomes of this cyberinfrastructure alignment workshop were:

- for representatives from cyberinfrastructure initiatives to better understand the user needs of stakeholders in the palaeo domain;
- for project personnel to better understand key cyberinfrastructure components such that their roles in the Palaeo Roadmap can be refined;
- for all participants to increase their awareness of and alignment between cyberinfrastructure components operating across the open data landscape.

By focusing on increasing understanding across the range of workshop participants, workshop organisers hoped to foster buy-in for work to be proposed in one of the key products resulting from this funded project: a Palaeo Roadmap (see "Envisioning a roadmap", below).

Activities

Based on positive feedback from participants of this project's first workshop, the organisers adopted a similar structure for the agenda and activities of this one. Daily morning icebreakers set an active yet informal tone and breakout group activities with changing group composition allowed participants an opportunity to converse directly with everyone attending the workshop. Substantive coffee and lunch breaks further allowed for networking and relationship-building, which are essential precursors to talking about alignment and potential future collaboration.

Lightning talks

Most workshop participants arrived knowing at least one, but not more than half of the other participants. Rather than going around the room and having people introduce themselves in sequence, each participant was asked to prepare a 5-minute lightning talk based on the prompt, "Why are you taking time to participate in this workshop?". These talks were interspersed throughout the agenda strategically alongside related discussion topics and activities and provided both memorable introductions as well as insight into the personal motivations and priorities of participants. The lightning talks also served as fodder for break and meal-time conversations, with participants observed chatting about possible collaborations or levelling-up their knowledge of new tools and projects.

Diving into the palaeo data ecosystem

One of the first activities of the workshop was to introduce participants to this project's concept of the "palaeo data ecosystem", i.e. the complex and interrelated landscape of resources that people in palaeontology collections and research use to do their work. In the schema for this ecosystem, resources are entities which have complex and non-obvious relationships with each other and one of this project's products is to map these relationships in order to provide a "finding aid"² for the community's collective knowledge. The schema is brought to life as a labelled property graph and examples (Fig. 1) were shown to workshop participants. The workshop provided an opportunity to solicit feedback about the technical structure of the schema and graph, as well as to gather more holistic ideas about connections between entities and gaps in the map.

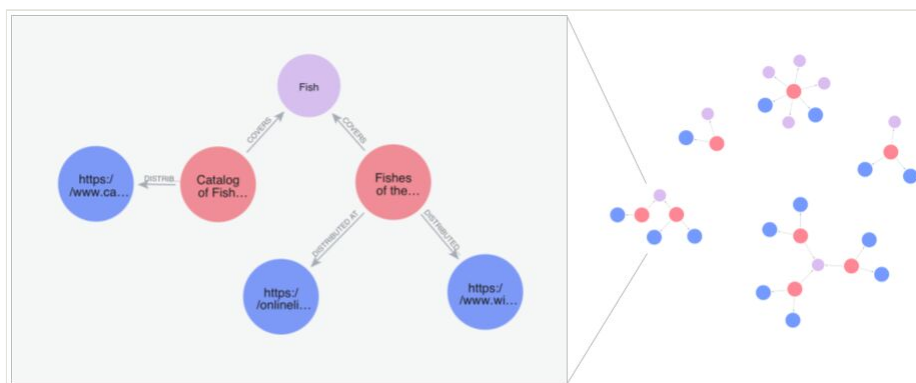


Figure 1. [doi](#)

Example section of the palaeo data ecosystem map highlighting mobilisation pipelines for resources related to fossil fish taxonomy, including those only distributed in an analogue format.

This introduction to the palaeo data ecosystem was followed by an activity in which participants were divided into two groups, provided with a fossil specimen and asked to outline a pathway for mobilising data about the specimen (Fig. 2). A group discussion followed this activity and participants were asked to consider:

- What types of relationships exist between entities? Which are you most interested in?
- Where do you wish data were more integrated?
- Where are the gaps in this map?
- Are there properties of a given entity that you would find particularly useful?
- How might we make this map most useful to you? Considering scope, accessibility etc...

Groups were influenced by the perspectives and previous experiences their members brought to the table, with both groups needing to spend time agreeing on a conceptual model before sketching out actual connections or data pathways. Much of the post-

activity discussion centred around how data modelling and mobilisation are constrained by specimen cataloguing practices, which are complicated by legacy data. Participants recognised that the problems related to legacy data are not unique to palaeontology collections. Another constraint that emerged from this activity was the frequent lack of digital accessibility for both primary specimen data and the types of authoritative resources (e.g. field notes, taxonomic lists) that contextualise and extend primary specimen data. The groups independently recognised a divide in the ability to mobilise data related to modern geography (e.g. collecting locality – relatively easy) versus geologic context (e.g. stratigraphy of the collecting event – relatively challenging). This divide parallels the general bias of specimen data mobilisation platforms towards neontological collections. There was also significant discussion about the need to bring cultural considerations into the ecosystem map, especially those framed by the CARE Principles for Indigenous Data Governance*⁴ and NAGPRA*⁵. Participants noted the importance of including these considerations in the ecosystem map as a way to help promote this work more broadly.

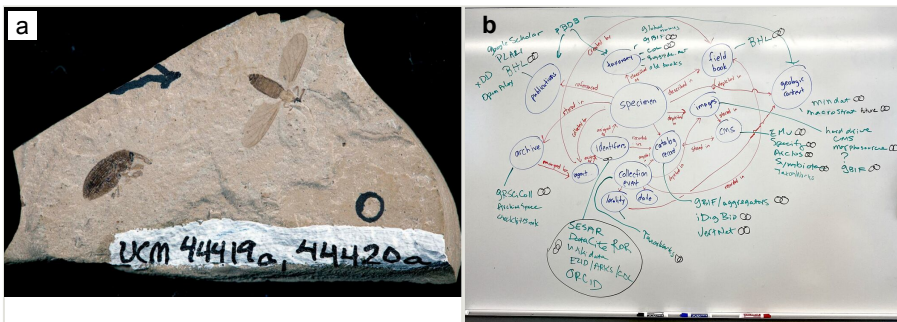
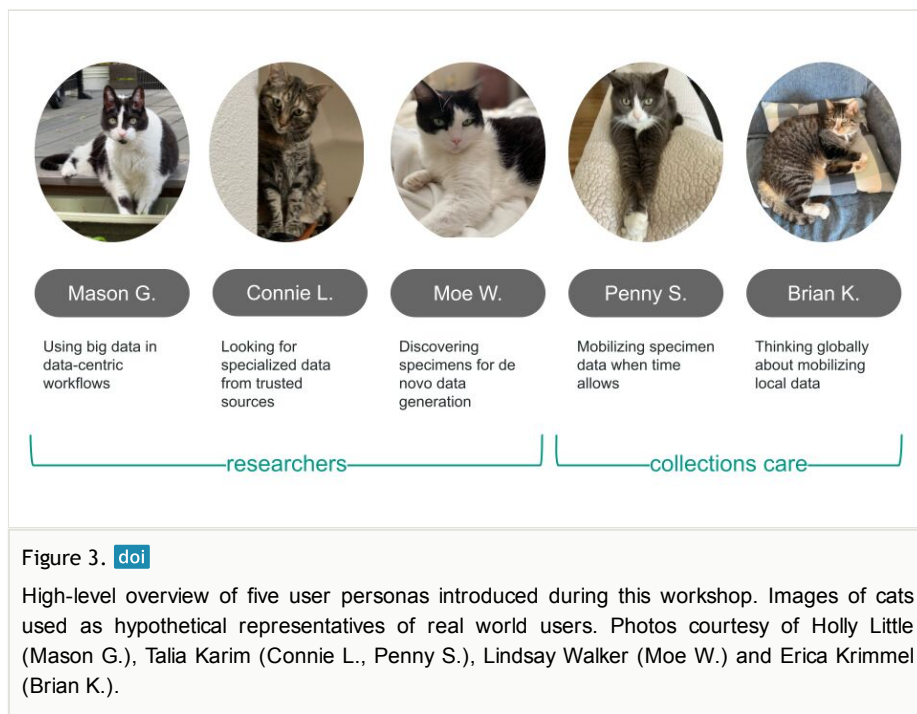


Figure 2. Participants were asked to consider a specimen image and out sketch data mobilisation pathways, especially considering primary specimen data: taxonomy, identifiers, locality, collector(s), stratigraphy and publications:
 a: Image of a fossil specimen (UCM 44419). CC BY 4.0 University of Colorado Museum of Natural History Paleontology Section; [doi](#)
 b: One of the group’s resulting diagrams. Photo courtesy of Erica Krimmel. [doi](#)

Experiencing the ecosystem through user personas

Now having a basic understanding of the palaeo data ecosystem, participants were asked to explore it through the lens of users who conduct palaeontology research and who care for palaeontology collections. A user persona is a semi-fictional, composite character created, based on real data—here, a series of user interviews, plus holistic insights from this project’s first workshop (Karim et al. 2024). User personas are commonly used in product and service design, where they help designers understand user needs, behaviour and goals in order to guide development. For this project, five user personas (Fig. 3) were created to represent two broad categories of palaeontology data stakeholders: researchers (three personas) and collections care professionals (two

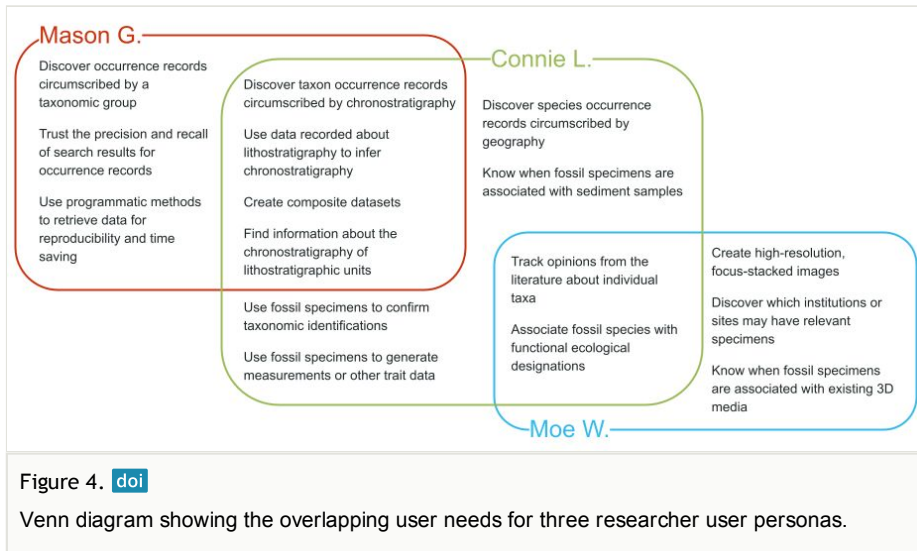
personas). The personas were developed to define explicit user needs, explore users' relationships to resources across the data ecosystem and identify high-impact areas of improvement, i.e. those which would benefit multiple users or disciplines beyond palaeontology. Additionally, these user personas also represent end-users of tools maintained and developed by many of the workshop's participants.



After introducing the concept of user personas and the five individual ones created for the project, participants were split into groups for an activity playfully named "LARP-ing in the land of paleo data".³ This activity consisted of two rounds, with the first focused on researcher user personas and the second focused on collections care professionals. For each round, participants broke into groups (one per user persona) and then were asked to brainstorm how to accomplish the user's stated goal, being specific about what resources the user needed and where the user might experience expressways (wins) and detours (challenges). The objective of this activity was for workshop participants to gain a better understanding of how these users experience the current palaeo data ecosystem.

During the post-activity discussion, participants shared thoughts both about the specific user goals and the bigger-picture systemic conditions affecting these users' ability to make progress towards their goals. Considering the researcher personas, workshop participants concluded that each of the three had fundamentally unique data needs (Fig. 4) and that lumping "palaeo researchers" into a single category can create false impressions. They also noted how difficult it is for researchers to pull together disparate

resources in the current palaeo data ecosystem and that this difficulty constrains the types of questions a researcher can explore. For the two collections care personas, participants felt their needs were more similar to each other, but that the variability in institutional support and priorities may impact their individual ability to participate in and contribute to digitisation efforts and the palaeo data ecosystem.



For all personas, workshop participants noted opportunities to improve the user experience by providing better tools, offering more training and documentation and creating richer human networks to ease navigation of the ecosystem. Participants also noted opportunities to upskill users themselves in order to raise the bar for constructing community resources with the lowest common denominator in mind. This led into a discussion of who incentivises or influences the personas – for example, a university's demand for tenure may incentivise the researcher to work on topics that can lead to more publications more quickly; or grant-making requirements may influence how the collections professional shares data. Such systemic drivers could be opportunities to promote best practices if they were more integrated in the palaeo data ecosystem. For example, a researcher today may access specimen-based palaeo data, but then not publish their own version of the data (enhanced for research use) in a discoverable way because the publisher does not require, much less facilitate, doing so. If publisher expectations evolved, researchers would be incentivised as well.

Discussion at the meta-level recognised that, even for workshop participants who represent large, mature products, getting feedback from users is a challenge. Participants noted that many projects they had worked on did not have well-integrated mechanisms for feedback to provide information for current or future development. It was also recognised that the creation and nurturing of such mechanisms takes time and resources that are not always available on grant-funded or otherwise term-limited projects, making it difficult to implement a responsive feedback loop as part of the product development.

Envisioning a roadmap

This activity started with an overview of the project team's vision for the Palaeo Roadmap – a final product of this project that defines a framework for designing future work to improve the discoverability and use of collections-based palaeontological data. Participants were introduced to key elements of the Roadmap (Fig. 5): a description of stakeholders and audiences; what the existing palaeo data ecosystem looks like; actionable recommendations for expanding and improving this existing ecosystem; and a vision for the future of collections-based palaeo data, including community and sustainability considerations. Although the initial version of the Roadmap will be a static document, it will be released in a format that can be versioned and maintained by the community as a living document that evolves alongside research needs and the cyberinfrastructure landscape. With this context in mind, workshop participants were then asked to consider and discuss:

Outline of Paleo Roadmap Vision		
<p>Includes:</p> <ul style="list-style-type: none"> • Requirements and recommendations • Current landscape supporting FAIR palaeontological data 	<p>Must be:</p> <ul style="list-style-type: none"> • Actionable by key collaborators • Represent requirements of diverse end-user audiences 	<p>Community Guide:</p> <ul style="list-style-type: none"> • Creates framework to build scalable future projects, initiatives, and proposals

Figure 5. [doi](#)

Outline of Palaeo Roadmap vision illustrating the planned approach and structure of the document at the end of the workshop.

- How can the Roadmap be useful for guiding future development or identifying development priorities?
- Are there technical elements that need to be present for the Roadmap to be actionable? Or a structure that you've seen elsewhere and liked?
- What might make the Roadmap a tool you can use within your context?
- What else do you need to know in order to understand our vision for the Roadmap?

The discussion was productive and led to many good suggestions, especially about how to get community buy-in for the Roadmap. Participants suggested that mapping general recommendations to concrete tasks would be helpful when considering potential implementation, as would calling out specific organisations or initiatives by name when relevant. Framing concrete tasks in the context of the larger vision or objective could provide easy evidence for another organisation to include related work in their own strategic plan. There was broad consensus on the importance of having relevant parts of

the Roadmap included in the formal strategic planning documents of organisations that plan their work this way. The group underscored that using examples throughout the Roadmap was a crucial communication tool. Another suggestion for building community buy-in was to go grassroots and encourage members of the [Paleo Data Working Group](#) to approach their institutional administrators, using the Roadmap to start a conversation about opportunities and challenges at the institutional and/or collection level. Participants proposed including language in the Roadmap that readers could use in funding proposals for their own related work, as well as identifying areas where projects with intersecting interests might want to collaborate in seeking funding.

Introducing the Palaeo Data Knowledge Hub

The Palaeo Data Knowledge Hub*⁷ was presented as a solution to host the lightweight content for the Palaeo Data Working Group community of practice and as a discovery hub of resources for anyone producing, managing or utilising specimen-based palaeo data. The Knowledge Hub is envisioned as being adaptable to other domains beyond palaeontology and as a tool for ongoing engagement and continuous knowledge-sharing across stakeholder communities. Workshop participants were given an overview of the information architecture and current technical structure for the Knowledge Hub, followed by a demonstration of the live website. They were then asked to spend time exploring the site on their own computers and to provide feedback via an EasyRetro board. Feedback was very helpful and ranged from the mundane-but-essential (e.g. pointing out a broken navigation link) to the technical (e.g. suggestions for mobile browser access) and the intellectual (e.g. adjustments to the information architecture). In general, participants liked the premise of the Knowledge Hub as well as the simplicity of the current site design. They had constructive criticism for improving on this design and they also wondered if user personas could be integrated into the information architecture. In addition to feedback about the site design, this activity resulted in a list of contents that workshop participants would like to either discover or share via the Knowledge Hub.

Future roads, Take me home

The final structured activity of the workshop asked participants to think broadly and without constraints about what might be possible for fossil data given limitless resources. Everyone remained together for this activity and discussed:

- If given infinite funding and human power, what is one cyberinfrastructure partnership or enhancement you would prioritise in the next 2 years? In the next 5 years?
- Beyond funding and human power, what else promotes effective life cycles for cyberinfrastructure resources?
- What drives success in a partnership between cyberinfrastructure resources?

As people were talking, workshop organisers visualised the conversation on a whiteboard (Fig. 6). Several themes emerged early on across the prompts, including the

Unconference

The workshop wrapped up with an unconference, where ideas from beyond the agenda were given the spotlight. Participants recorded potential unconference topics throughout the two days of the workshop by writing on sticky notes and placing them in an "idea parking lot". During the break prior to the unconference, workshop organisers grouped and categorised the parking lot ideas and then participants voted for their preferences. Based on broad interest across topics, the unconference time was split in half to accommodate two rounds of breakout groups, with participants joining groups, based on personal interest. The first round of topics included "data sources", "sharing best practices" and "vocabularies". In the second round, groups discussed "capacity", "product lifecycle" and "citations and metrics". Highlights from each topic are summarised here:

- **Data sources.** The origin of this topic came from participants wondering where to find or store data about radiocarbon or isotopic analyses associated with fossils. People in the group shared information about specific repositories and, more generally, discussed needs and strategies related to publishing and (re)using data across domains. Issues noted included managing data for destructive analyses, providing credit to the full spectrum of roles in the data lifecycle and balancing best practices in action vs. theoretical standards.
- **Sharing best practices.** Scoped as best practices for data management and mobilisation, including specific data storage solutions and data standards. As one group member put it, "How do I learn how to go from a standard to implementation of that standard?". Discussion centred on the need for community-written and reviewed documentation, controlled vocabularies etc. to help bridge the "last mile"⁶ of implementation. Such documentation can be difficult to maintain alongside iterative and ongoing development, though versioning can be a helpful technique. The group also talked about strategies for optimising relational databasing to one's domain (e.g. a specific palaeo collection), while still being able to share to an interoperable layer (e.g. a data aggregator).
- **Vocabularies.** The group focused on the jurisdictional/governance challenges of having authoritative vocabularies and the necessity of community curation to ensure a vocabulary meets real data needs. As part of this, the group talked about what makes something a controlled vocabulary and who has responsibility to host/make it available.
- **Capacity.** Discussion here honed in on the idea of interdisciplinary training pipelines as a way to increase capacity across job roles, as well as how to design roles that attract "unicorn" workers, i.e. those who come to a job with extensive cross-training and a broad set of skills. The group stressed the need for identifying who the right communities and people are to involve in creating this vision of interdisciplinary training pipelines.
- **Product life cycle.** Scoped to focus on roles that help this life cycle concept run smoothly and using experience from the Biodiversity Heritage Library as grounding for the conversation. Group members emphasised the importance of a "product owner" who converts user requests into development tickets and can

estimate how extensive the work is; this was not a familiar role to most workshop participants. Discussion also included how language used in the software development industry (e.g. “owner” and “product”) is hard to sell to people working in museums and academia because of associations with capitalism and how this is indicative of the greater challenge of bringing product life cycle expertise into these communities.

- **Citations and metrics.** The group shared ideas on novel ways to promote and track specimen data use (e.g. click-through rates of resolvable specimen identifiers, data mining in the literature, providing structured metadata) and talked about what would enable more integrated tracking (e.g. repository identifiers, instrument vocabularies for instrument usage). Discussion followed about how to incentivise/enforce citation best practices and how to deal with artificial intelligence both as a creator to cite and as an operator that needs to use citations. Members of the group agreed that, even under ideal circumstances, it is difficult to include everyone who ought to be credited.

Participant feedback

Participant feedback was overwhelmingly positive, both via candid remarks and a post-workshop survey. Although response to the survey was poor (30%), everyone who did respond said they either “strongly agreed” or “agreed” that: (1) The objectives of the workshop were clear; (2) There were sufficient opportunities for questions and interaction; (3) This workshop was relevant to their work; (4) This workshop was worth their time; and (5) They felt comfortable participating in this workshop environment.

Participants reported that the most useful thing they gained from the workshop were the connections and networking opportunities, as well as the general overview of challenges in the palaeo data community. One remarked that they “expected everything to be already figured out” when coming to this workshop – the organisers hope they were able to put an optimistic lens on how this is not the case! One suggestion for improvement from a couple of survey responses was to include a more structured workshop summary and next steps at the end of day two. Other survey respondents said they left wanting more wrap-up or full circle discussion of certain topics; such a session might have addressed this as well. In general, the workshop structure of seeding content, taking a brief break and then following up with an activity related to that content was successful. Several survey respondents indicated that they particularly enjoyed the user personas LARP-ing activity.

Key outcomes and discussions

Several overarching themes emerged over the course of this two-day workshop and provided an important perspective for the project team to incorporate into the Palaeo Roadmap.

Human capacity came up in multiple contexts, especially thinking about the professional development burden placed on individuals who work in palaeo research and/or collections care. These individuals are critical for providing subject matter expertise to cyberinfrastructure projects. However, doing so can be a major capacity challenge when contributing requires understanding that bridges the content (e.g. palaeontology) and the mechanics (e.g. informatics). Few institutions or cyberinfrastructure projects have staff in roles dedicated to being this bridge; where these roles do exist, for example, as a collections-focused information science position at a museum or as user support staff with a science background at a data aggregator, they underscore how valuable the role is. These roles help fast-track understanding for subject matter experts, thus protecting their capacity to be experts in their particular niche within palaeo research or collections care. In parallel, it is important to recognise the motivations of experts and to explore where change might be best incentivised on the system level, for example, using publisher requirements to affect research practices.

Human capacity is also affected by *capacity and sustainability at the project level*. Across workshop participants, the group had significant experience managing projects of various scales and levels of technical complexity. Project governance and structure came up repeatedly as a high-impact area for improvement and participants recognised that, as projects emerge from academia and transition into production, they often need to be more intentional about scoping themselves for sustainable success. As one participant noted, "We are all our own worst enemies because we aren't narrowing the scope enough to be actionable". Setting the right scope further allows projects to build relationships with and complement each other. Coming full circle, as important as building and maintaining relationships between projects is, this is another area where individual human capacity is a challenge. Many projects in the palaeo data ecosystem are small and supported by a minimal crew who lack the time necessary for meaningful collaboration with those outside their immediate working sphere.

As challenging as it is at all scales, *not addressing the challenge of capacity is costly*. Technical resources and their associated organisational knowledge are expensive to develop and maintain, particularly in the intersectional and interdisciplinary landscape of palaeo data. The group discussed the idea of calculating "replacement cost" for some of the projects represented. As the keepers of such knowledge, individual people are truly the centre of not only research and collections (a theme that emerged from this project's 2024 workshop engaging stakeholders in the palaeo research community), but also of cyberinfrastructure projects, which frequently rely on "unicorn" developers. Part of "cyberinfrastructure alignment" in the palaeo data community needs to be figuring out how to support the researchers, collections care professionals, unicorn developers and others who collectively cultivate organisational knowledge so that they have the potential to recognise opportunities for aligning projects and people on more technical levels. In doing so, technical resources may be used more cost effectively and for wider audiences.

Technology is currently under-leveraged in the palaeo data ecosystem, with plenty of concrete opportunities to make substantive improvements. For example, by expanding global coverage for accessible data about lithostratigraphy or putting geospatial

coordinates on palaeo specimen collecting locations whenever feasible. Refreshingly, no one at the workshop felt that palaeo collections care professionals should be solely responsible for providing research-ready data. In fact, there was discussion about the importance of creating technical systems to allow for duality: data in its original format provided alongside interpretations of that data tailored for specific purposes, similar to calls from the biodiversity collections community for new data flows to realize the "Extended Specimen" concept (Lendemer et al. 2020). New strategies for scoping, integration across projects and identifying technical gaps with proposed solutions could enable improved distribution of effort and resources across the cyberinfrastructure landscape and availability of critical data, ultimately leading to a more comprehensive and supported landscape.

This cyberinfrastructure alignment workshop laid the groundwork for future collaborations by successfully helping participants better understand the needs and roles of various stakeholders and by increasing participants' awareness of cyberinfrastructure components operating across the open data landscape. The workshop organisers believe that such understanding and awareness is critical for building alignment between players in the palaeo data ecosystem, beginning here with the participants of this workshop. Alignment includes having not only an overarching and shared understanding of challenges, but also the willingness to consider actionable recommendations with end-users in mind. Although the question remains as to how to sustain momentum for such work in the long-term, this project's Palaeo Roadmap will be a tool for facilitating near-term alignment with an eye to the future. Workshop participants thought that the Roadmap has an ambitious vision, but were overall supportive and also believed it could be an effective way to advocate for the needs of the palaeo data community, particularly if it aligned with strategic frameworks and funding/work priorities of specifically targeted projects. The organisers gained important perspectives from attendees and look forward to continuing conversations with this group throughout the process of finalising the inaugural version of the Palaeo Roadmap.

Acknowledgements

Thank you sincerely to all those who participated in this workshop and made it the success that it was! Extra thanks to Alex Lawrence for co-facilitating and to Nicole McGee, Lindsey Pfenninger, Casey Thater, Jerah Brewster and Sam Eads for their help with lunchtime logistics.

Funding program

This workshop is part of the project, "Community-driven enhancement of information ecosystems for the discovery and use of paleontological specimen data" which is funded by the NSF GEO OSE programme under the following awards: [2324688](#), [2324689](#), [2324690/2525603](#).

Hosting institution

University of Colorado Boulder

Conflicts of interest

The authors have declared that no competing interests exist.

References

- Karim T, Krimmel E, Little H, Walker L (2024) Community-driven enhancement of information ecosystems for the discovery and use of paleontological specimen data: Stakeholder engagement workshop. *Research Ideas and Outcomes* 10 <https://doi.org/10.3897/rio.10.e134840>
- Lendemer J, Thiers B, Monfils AK, Zaspel J, Ellwood ER, Bentley A, LeVan K, Bates J, Jennings D, Contreras D, Lagomarsino L, Mabee P, Ford LS, Guralnick R, Gropp RE, Revelez M, Cobb N, Seltmann K, Aime MC (2020) The Extended Specimen Network: A Strategy to Enhance US Biodiversity Collections, Promote Research and Education. *BioScience* 70 (1): 23-30. <https://doi.org/10.1093/biosci/biz140>

Supplementary material

Suppl. material 1: Editable version of Figure 6b [doi](#)

Authors: Erica Krimmel

Data type: diagram

[Download file](#) (1.24 MB)

Endnotes

- *1 As of this writing (2025-05-21), there are 9,748,143 occurrence records in the GBIF data portal where *basisOfRecord* = "FossilSpecimen".
- *2 A finding aid is traditionally "a description that typically consists of contextual and structural information about an archival resource" (<https://dictionary.archivists.org/entry/finding-aid.html>).
- *3 LARP-ing, or "Live Action Role-Playing," is a style of game where players physically get into character. In the United States, LARPing frequently involves fantasy battles and dozens if not hundreds of participants. No fantasy battles were staged during the course of this workshop.
- *4 CARE Principles (Collective benefit, Authority to control, Responsibility and Ethics) from the Global Indigenous Data Alliance: <https://www.gida-global.org/care>

- *5 Native American Graves Protection and Repatriation Act (NAGPRA): <https://www.nps.gov/subjects/nagpra/index.htm>
- *6 The "last mile" problem is a concept commonly discussed in fields like transportation planning, telecommunications and supply chain management. It refers to the fact that the last stage of a journey is typically the most expensive because the end-points are dispersed and economies of scale are no longer effective.
- *7 The Palaeo Data Knowledge Hub is actively in development and will be accessible in September 2025 at: <https://paleo-data.github.io/>.