

BRIDGING THE RAUNKIÆRAN SHORTFALL THROUGH SPECIMEN-ANCHORED, AI-DERIVED TRAIT INFRASTRUCTURE

YASIN BAKIŞ

Biodiversity Research Institute, Tulane University, New Orleans, LA, USA

Abstract: Biodiversity informatics has substantially improved access to taxonomic names and species occurrences, yet functional trait knowledge remains severely constrained by the Raunkiaeran shortfall: traits are sparse, weakly standardized, rarely specimen-anchored, and poorly suited for inference under rapid global change. We argue that this limitation is not primarily a data-availability problem, but an infrastructure problem - specifically, a persistent gap between digitized biodiversity evidence and reproducible trait synthesis. Here, we outline a practical pathway for bridging this gap grounded in a coherent project lineage. FishNet 2 provides a specimen-backed foundation for geographic discovery and distributional inference. FishAIR establishes quality-aware, standardized, and AI-ready representations of specimen imagery. Biology- and hierarchy-guided machine learning, developed through BGNN and Imageomics, enables interpretable, scalable phenomic extraction. Specimen-anchored synthesis efforts, including Morphological Barcoding and the Global Fish Trait Atlas, translate extracted phenotypes into population-aware trait reference models that preserve provenance, intraspecific variation, and uncertainty. We further situate this architecture within a broader observatory framework, the Living Environmental Phenomics Observatory, and generalize its principles through AIDEN-SD as a model for interoperable, AI-ready data products. Together, this lineage reframes traits as infrastructure-derived ecological observables, enabling trait-based ecology, conservation, and global change ecology, including climate-impact assessment, at scales commensurate with rapid environmental change.

Key words: biodiversity informatics; biodiversity knowledge shortfalls; Raunkiaeran shortfall; functional traits; phenomics; biological collections; specimen-anchored data; artificial intelligence; data integration

BIODIVERSITY KNOWLEDGE SHORTFALLS REVISITED

Comprehensive biodiversity information underpins efforts to describe, monitor, and manage life on Earth, particularly under accelerating global environmental change. Biodiversity informatics has developed as a response to this challenge by mobilizing primary biodiversity data, improving access to specimen and occurrence records, and creating infrastructures for large-scale ecological and evolutionary inference (Nelson and Ellis, 2019; Page et al., 2015). Foundational work in biodiversity informatics emphasized that biological collections, taxonomic names, occurrence records, and environmental context must be organized into interoperable systems before they can support robust biodiversity synthesis (Ariño, 2010; Nelson and Ellis, 2019; Soberón and Peterson, 2004). Within this broader context, Hortal et al. (2015) articulated a framework of seven biodiversity knowledge shortfalls, including gaps related to taxonomy (Linnean), geographic distributions

(Wallacean), abundance (Prestonian), ecology (Hutchinsonian), traits and functions (Raunkiaeran), evolutionary history (Darwinian), and species interactions (Eltonian). This framework has become foundational in biodiversity informatics, providing a common language for diagnosing progress and persistent barriers (see also Chapman et al., 2024).

Over the past two decades, substantial advances have been made in addressing some of these shortfalls. Large-scale data infrastructures such as Global Biodiversity Information Facility (GBIF) and Catalogue of Life have dramatically reduced barriers (Edwards et al., 2000) to accessing taxonomic and occurrence data, contributing to meaningful progress in mitigating Linnean and Wallacean gaps. Similarly, domain-specific resources such as FishBase have consolidated decades of expert knowledge on species ecology, life history, and distribution, enabling synthesis at global scales that were previously infeasible.

While these initiatives do not eliminate all taxonomic or geographic biases, they represent transformative steps toward more complete and interoperable biodiversity information.

Parallel to advances in biodiversity digitization, trait-based ecology has emerged as a major framework for understanding community assembly, ecosystem functioning, biogeographic structure, and organismal responses to environmental change (Carmona et al., 2016; McGill et al., 2006; Violle et al., 2007). Large-scale trait infrastructures such as TRY, EltonTraits, AVONET, and FishBase have substantially expanded access to functional information across plants, mammals, birds, and fishes (“FishBase,” 2026; Kattge et al., 2020; Tobias et al., 2022; Wilman et al., 2014). These efforts have enabled major advances in functional ecology and comparative biodiversity research by standardizing trait concepts and promoting more explicit functional interpretations of organismal characteristics across ecological systems (Violle et al., 2007). At the same time, growing interest in phenomics and image-based biodiversity analysis has stimulated development of automated and semi-automated trait extraction approaches across multiple organismal systems, including plants, insects, vertebrates, and agricultural phenotyping workflows (Caranza-Rojas et al., 2017; Houle et al., 2010; Høye et al., 2021; Lürig et al., 2021; Younis et al., 2018).

Recent biodiversity cyberinfrastructure initiatives have increasingly emphasized the concept of Digital Extended Specimens (DES), in which specimens function as interconnected digital objects linked to derivative data products, annotations, environmental context, and analytical workflows (Hardisty et al., 2022; Webster, 2017). Parallel advances in biodiversity data standards, semantic interoperability, machine-actionable ecological data systems and FAIR digital infrastructure have further strengthened the capacity to integrate heterogeneous biodiversity information across collections and analytical systems (Falster et al., 2021; Walls et al., 2014; Wiczorek et al., 2012; Wilkinson et al., 2016). However, scalable infrastructures for generating reproducible, specimen-anchored functional trait representations remain comparatively underdeveloped within these emerging ecosystems.

In contrast, progress toward resolving the Raunkiæran shortfall - the lack of quantitative knowledge about organismal traits and functions - has been far more limited. Existing trait databases are typically derived from literature summaries, expert opinion, or coarse species-level generalizations (Gonçalves-Souza et al., 2023; Kattge et al., 2020; Wilman et al., 2014). Such data are often sparse, inconsistently defined, and weakly standardized across

taxa and regions. Empirical evaluations have also shown that species-averaged trait databases may poorly represent local or community-scale functional variation, particularly for environmentally plastic traits, highlighting the limitations of aggregated trait summaries for ecological inference (Cordlandwehr et al., 2013). Critically, they rarely capture intraspecific variation, geographic structure, environmental context dependency, or uncertainty (Albert et al., 2011; Maitner et al., 2023), and they are seldom linked explicitly to vouchered specimens. Recent ecological synthesis work has further emphasized that biodiversity prediction under incomplete and heterogeneous ecological data remains strongly constrained by uncertainty propagation and data integration limitations (de Bello et al., 2025). As a result, trait data remain difficult to validate, compare, or scale, limiting their usefulness for trait-based ecology, biodiversity modeling, and conservation planning. Recent biodiversity informatics research has increasingly emphasized that species-level trait summaries alone are often insufficient for robust ecological and evolutionary inference because they obscure intraspecific variation, geographic structure, sampling context, and measurement uncertainty. Emerging specimen-centered trait frameworks instead advocate preserving direct linkages between quantitative traits, individual vouchers, and associated metadata, allowing trait observations to remain dynamically interpretable, reproducible, and extensible across analytical contexts (Bakış et al., 2023; Balk et al., 2022; Jebbia et al., 2023). Within this perspective, provenance information, measurement protocols, life-stage annotations, environmental context, and specimen identifiers become essential components of trait infrastructure rather than supplementary metadata. Consequently, biodiversity informatics systems designed to support large-scale phenomic synthesis must preserve uncertainty, variability, and specimen-level traceability while enabling interoperability across collections, environmental datasets, and analytical workflows.

This imbalance is striking given the parallel growth of digitized biological collections. Natural history museums now provide access to millions of georeferenced specimens and images generated through large-scale digitization initiatives and industrial-scale collection mobilization efforts (Blagoderov et al., 2012; Nelson and Ellis, 2019; Page et al., 2015) yet these resources have not translated into commensurate advances in functional trait knowledge. Recent syntheses have highlighted the growing potential of artificial intelligence for addressing biodiversity knowledge shortfalls through computer vision, multimodal ecological modeling, automated

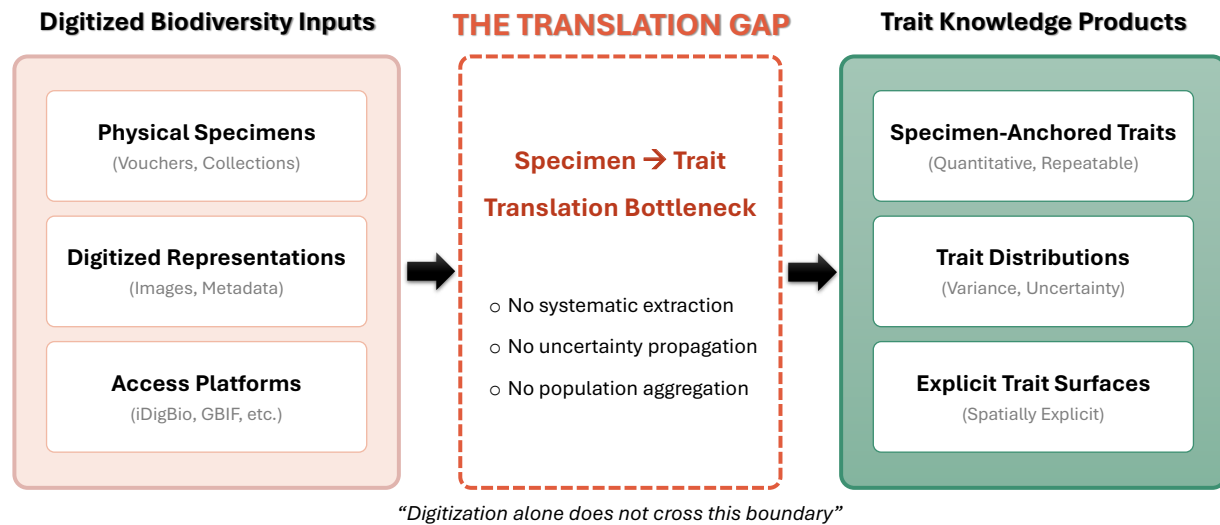


Figure 1. Biodiversity knowledge shortfalls and the specimen-to-trait translation gap.

Major biodiversity data infrastructures have substantially reduced taxonomic and distributional shortfalls by mobilizing specimen-backed occurrence data. In contrast, the Raunkiæran shortfall persists because functional traits remain weakly standardized, rarely specimen-anchored, and poorly integrated with digitized collections. The highlighted specimen-to-trait translation gap represents the structural barrier addressed in this study.

sensing, and machine-learning-assisted extraction of biological information from images, audio, genomic, and environmental datasets (Pollock et al., 2025). However, most current AI applications remain focused on species occurrence, classification, and ecological prediction rather than specimen-grounded trait translation and validation pipelines. We argue that this disconnect reflects a structural bottleneck rather than a simple data deficit. Specifically, biodiversity informatics lacks scalable, reproducible infrastructure for translating specimens into quantitative trait information (de Bello et al., 2025; Falster et al., 2021; Hardisty et al., 2022) - a gap we refer to here as the *specimen-to-trait translation gap*. Digitization efforts have succeeded in mobilizing specimens and occurrences, but without systematic mechanisms to extract, model, and validate phenotypic traits from specimen-linked digital resources (Castillo-Figueroa, 2018; Henning et al., 2018), the Raunkiæran shortfall persists and propagates into related ecological and geographic knowledge gaps.

In this paper, we revisit biodiversity knowledge shortfalls through the lens of this translation bottleneck. We synthesize recent advances in image-based phenomics, biology-guided machine learning, and specimen-centered data integration (Christin et al., 2019; Houle et al., 2010; Lürig et al., 2021; Porto et al., 2021) to argue that closing the Raunkiæran shortfall requires rethinking traits as infrastructure-derived products rather than static literature

attributes. By reframing trait knowledge as a problem of translation and integration, rather than availability alone, we outline a pathway for bridging long-standing gaps in biodiversity informatics and enabling more robust, specimen-grounded analyses of biodiversity in a changing world.

Rather than proposing additional categories of biodiversity knowledge shortfalls, we focus on translational and infrastructural mechanisms that help explain why the Raunkiæran shortfall has persisted despite large-scale digitization efforts (Gonçalves-Souza et al., 2023; Hortal et al., 2015). Figure 1 summarizes the position of the Raunkiæran shortfall within the broader landscape of biodiversity knowledge gaps, highlighting how existing infrastructures have substantially reduced taxonomic and geographic shortfalls while leaving a persistent specimen-to-trait translation gap.

THE PERSISTENT RAUNKIÆRAN SHORTFALL

Over the past two decades, biodiversity informatics and trait-based ecology have substantially expanded the availability of functional biodiversity data through global databases, digitized collections, and comparative synthesis efforts. Resources such as TRY, EltonTraits, AVONET, FishBase, and related trait initiatives have transformed access to ecological and morphological information across plants, mammals, birds, and fishes, enabling large-scale

analyses of functional diversity, ecosystem processes, and biogeographic structure (“FishBase,” 2026; Kattge et al., 2020; Tobias et al., 2022; Wilman et al., 2014). At the same time, advances in image-based phenomics, computer vision, geometric morphometrics, and automated landmarking have demonstrated increasing feasibility of extracting quantitative phenotypes directly from digitized specimens and biological imagery across multiple taxonomic groups (Houle et al., 2010; Lürig et al., 2021; Porto et al., 2021; Rolfe et al., 2021). These developments collectively represent major progress toward resolving the Raunkiæran shortfall. However, despite these advances, most existing trait infrastructures remain fragmented, weakly linked to vouchered specimens, inconsistent in their treatment of uncertainty and intraspecific variation, and difficult to integrate into interoperable biodiversity knowledge systems.

Efforts to compile organismal trait data have expanded substantially in recent decades, reflecting growing recognition of traits as central to ecological, evolutionary, and conservation analyses. Several global and regional trait resources now exist across plants, mammals, birds, fishes, and other taxa, often assembled through expert curation or synthesis of published literature. Parallel advances in image-based phenomics, automated landmarking, and AI-assisted trait extraction have further demonstrated increasing feasibility of deriving quantitative phenotypes directly from digitized specimens and biological imagery (Carranza-Rojas et al., 2017; Porto et al., 2021). These resources and approaches have enabled important comparative studies and provided foundational inputs for trait-based modeling while also motivating broader reassessment of how functional trait knowledge is generated, standardized, and synthesized across biodiversity systems (Gonçalves-Souza et al., 2023). However, despite their utility, most existing trait datasets remain literature-derived and species-averaged, and thus fall short of resolving the Raunkiæran shortfall at scales required for contemporary biodiversity science.

The persistence of this shortfall is not accidental; it reflects the historical sequencing of biodiversity infrastructures (Nelson and Ellis, 2019), in which efforts such as FishNet 2 prioritized specimen-backed geographic discovery, while dedicated mechanisms for trait translation lagged behind digitization.

A defining limitation of current trait resources is the absence of explicit specimen grounding. Trait values are typically abstracted from heterogeneous sources - field guides, taxonomic monographs, ecological studies - without direct linkage to vouchered specimens. This separation obscures provenance, prevents independent validation, and makes it difficult to reconcile conflicting trait

estimates across sources. Although recent biodiversity cyberinfrastructure initiatives such as Digital Extended Specimens have emphasized stronger integration of derivative biodiversity data products with physical collections and interoperable digital objects (Hardisty et al., 2022), specimen-level linkage remains inconsistently implemented across many existing trait infrastructures. In contrast to occurrence data, which are increasingly traceable to physical specimens through collection identifiers and biodiversity data standards (Wieczorek et al., 2012), trait data often lack comparable anchors, limiting reproducibility and long-term reliability.

Relatedly, most trait databases rely on single representative values per species, masking intraspecific variation that is often ecologically and evolutionarily meaningful. Morphological, physiological, and functional traits frequently vary across populations, environmental gradients, and life-history stages, and trait-based ecological inference increasingly depends on integrating such variation across organizational and spatial scales (Carmona et al., 2016). Species-averaged traits therefore risk misrepresenting local adaptation, phenotypic plasticity, and population-level responses to environmental change (Albert et al., 2011; Laughlin et al., 2020; Violle et al., 2012) and multidimensional functional structure within and among populations (Villéger et al., 2008). This limitation is especially problematic for analyses that seek to link traits to climate, habitat heterogeneity, or disturbance regimes, where variation within species can be as informative as differences among species.

A third, closely connected limitation is poor characterization of uncertainty. Trait values are commonly reported without measures of variance, confidence, or sampling effort, even when derived from limited or context-specific observations. Without explicit uncertainty estimates, downstream analyses may treat trait values as fixed and comparable across taxa, leading to overconfident inferences and biased model outputs (Beale and Lennon, 2012; Dormann et al., 2018; McGill et al., 2006). This problem is amplified when trait data are integrated with increasingly sophisticated statistical or mechanistic models, where uncertainty propagation is essential for robust prediction and decision-making.

Finally, existing approaches to trait compilation face fundamental challenges of scalability. Manual extraction and expert synthesis do not readily scale to the millions of specimens now digitized in natural history collections, nor do they adapt efficiently as new specimens, images, or environmental data become available. As a result, trait databases often lag behind advances in specimen digitization and remain static relative to the dynamic nature of

biodiversity data. This mismatch constrains the ability of biodiversity informatics to respond to emerging research questions and rapid environmental change.

Together, these limitations substantially constrain the use of trait data in ecology, conservation, and climate research while weakening efforts to connect biodiversity structure with ecosystem functioning and ecological processes (de Bello et al., 2010). Trait-based ecological models may overlook critical variation and uncertainty; conservation assessments may rely on oversimplified functional representations of species; and climate-impact analyses may fail to capture phenotypic responses occurring within species across space and time. As a result, the Raunkiæran shortfall persists not merely due to insufficient trait information, but because existing trait knowledge is structurally disconnected from specimens, variation, and uncertainty. Addressing this shortfall therefore requires approaches that treat traits not as static summaries, but as quantitative, specimen-anchored data products capable of scaling with biodiversity information systems. This shift enables the quantification of the Raunkiæran shortfall in terms of trait coverage, variance capture, and uncertainty propagation.

A SPECIMEN-ANCHORED SOLUTION: AI-DERIVED PHENOMICS

Several of the infrastructures discussed here represent emerging or proposed synthesis efforts that build on established components; they are presented to illustrate architectural pathways for closing the Raunkiæran shortfall rather than as fully deployed systems.

Recent advances in computer vision, phenomics, geometric morphometrics, and biology-guided machine learning have demonstrated increasing feasibility of extracting quantitative phenotypes directly from digitized specimens and biological imagery across plants, insects, vertebrates, and agricultural systems within a broader expansion of deep learning applications in ecology and biodiversity science (Carranza-Rojas et al., 2017; Christin et al., 2019; Henning et al., 2018; Houle et al., 2010; Lürig et al., 2021; Porto et al., 2021; Younis et al., 2018). These approaches include automated landmarking, image segmentation, morphological feature extraction, and AI-assisted classification pipelines capable of operating at scales not feasible through manual trait compilation alone, reflecting broader adoption of machine-learning methodologies across ecology and biodiversity science (Thessen, 2016). However, most existing systems remain taxon-specific, weakly integrated with specimen-centered biodiversity infrastructures, or limited in their treatment of provenance, uncertainty, and interoperability. The framework proposed here builds on these advances while emphasizing reproducible

specimen anchoring, uncertainty-aware synthesis, and integration across biodiversity informatics infrastructures.

In practice, the proposed framework operates by linking digitized museum specimens, associated images, and biodiversity metadata within a specimen-centered analytical workflow. FishNet 2 provides access to georeferenced specimen records and collection identifiers, while FishAIR contributes standardized and quality-aware specimen imagery suitable for AI-based analysis. Image-based phenomic methods, including segmentation, landmark extraction, and morphology inference, can then be applied directly to these images using biology-guided machine-learning approaches. Resulting measurements remain linked to individual vouchers and can subsequently be aggregated into population-level and species-level trait distributions within the Global Fish Trait Atlas framework.

A typical specimen-anchored phenomics workflow may begin with digitized specimen images linked to stable collection identifiers and associated metadata. Computer vision pipelines can then perform specimen segmentation, landmark detection, morphometric extraction, color or texture quantification, and label-associated metadata parsing directly from images or derivative media products (Lürig et al., 2021; Porto et al., 2021). Extracted measurements can subsequently be linked back to individual vouchers through persistent identifiers, enabling provenance tracking, uncertainty annotation, reproducibility, and reanalysis as analytical methods improve while also supporting broader machine-actionable biodiversity knowledge integration workflows (Bakiş et al., 2023; Balk et al., 2024; Thessen et al., 2012). Similar image-based phenotyping approaches have already been demonstrated across herbarium specimens, insect morphology, vertebrate morphometrics, and agricultural phenomics systems (Carranza-Rojas et al., 2017; Younis et al., 2018). Within the framework proposed here, these extracted phenotypes become interoperable specimen-linked data products that can support downstream ecological, evolutionary, and biodiversity informatics analyses.

Bridging the specimen-to-trait translation gap requires approaches that can operate at the scale of modern biological collections while remaining grounded in biological meaning and scientific rigor. Image-based phenomics has emerged as a promising enabling approach (Houle et al., 2010; Lürig et al., 2021) for this challenge. By leveraging digitized specimen images, phenomics seeks to extract quantitative descriptions of organismal form directly from physical vouchers, transforming static digital surrogates into sources of reproducible trait information. Unlike traditional trait compilation, which relies on manual abstraction from heterogeneous literature, image-based phenom-

ics treats specimens themselves as primary data objects for trait generation.

Recent advances in biology- and hierarchy-guided machine learning make large-scale specimen-to-trait translation tractable without sacrificing biological meaning. Within this lineage, Biology-Guided Neural Networks (BGNN) operationalize the use of taxonomic and phylogenetic structure as inductive bias, ensuring that phenomic inference respects evolutionary relationships rather than treating traits as independent features (Elhamod et al., 2022). In parallel, Imageomics emphasizes modular, reproducible workflows that integrate structured biological knowledge with machine learning to support benchmarking, interpretation, and method evolution. Together, these approaches transform automated image analysis from a black-box exercise into a governed, evolution-consistent phenomic engine capable of generating biologically interpretable traits at collection scale.

At its core, image-based phenomics enables the systematic measurement of morphological features - such as size, shape, proportions, and structural relationships - across large numbers of specimens. When applied to digitized collections, this approach allows trait data to scale with specimen availability rather than expert time. Importantly, because traits are derived directly from images linked to vouchered material, they retain explicit provenance and can be re-evaluated as methods, interpretations, or taxonomic concepts evolve. In this sense, phenomics offers a pathway for aligning trait knowledge with the same principles of traceability and reproducibility that now underpin occurrence data.

Recent advances in biology-guided and knowledge-guided machine learning (Elhamod et al., 2023; Reichstein et al., 2019; Roscher et al., 2020) further strengthen the potential of phenomics as a translation infrastructure. Rather than treating trait extraction as a purely data-driven computer vision problem, these approaches incorporate biological constraints, anatomical expectations, phylogenetic information, or ecological priors into model structure and interpretation, paralleling broader efforts to integrate deep learning with process understanding in environmental and Earth-system sciences (Reichstein et al., 2019). Biological guidance may include anatomical relationships, symmetry, or known scaling laws, while knowledge-guided frameworks incorporate taxonomic context or ecological expectations. By constraining learning with domain knowledge, such models improve interpretability, reduce biologically implausible outputs, and facilitate integration between machine learning predictions and existing ecological understanding, aligning with broader efforts to-

ward explainable machine learning for scientific discovery (Roscher et al., 2020).

Equally critical is the explicit treatment of specimen provenance, repeatability, and uncertainty. Specimen-anchored phenomics allows each trait measurement to be associated with a specific voucher, locality, and imaging context, enabling rigorous quality control and comparative analyses. Repeated measurements across specimens, populations, or imaging workflows provide empirical estimates of variance, while statistical modeling can be used to characterize uncertainty arising from sampling, measurement error, or model inference. This contrasts sharply with most existing trait resources, where uncertainty is implicit or absent and cannot be propagated into downstream analyses. Figure 2 situates specimen-anchored phenomics within a coherent infrastructure lineage, illustrating how successive projects collectively operationalize specimen-to-trait translation at scale.

Positioning AI-derived phenomics as a solution to the specimen-to-trait translation gap also requires careful clarification of its role relative to traditional taxonomy and systematics. Phenomics does not replace expert taxonomic knowledge or biological interpretation. Instead, it functions as translation infrastructure, converting specimen-level evidence into quantitative representations that can be interrogated, validated, and synthesized by domain experts. Taxonomy provides the conceptual framework within which traits gain meaning, while phenomics supplies scalable measurements that complement and extend expert knowledge. When embedded within specimen-centered workflows, AI becomes a tool for amplifying, rather than supplanting, biological expertise.

Viewed in this light, AI-derived phenomics represents a critical missing layer in biodiversity informatics: an operational bridge between collections and functional knowledge. By enabling specimen-grounded, uncertainty-aware trait extraction at scale, phenomics directly addresses the structural limitations that have sustained the Raunkiærán shortfall and its downstream effects. In the next section, we illustrate how this approach can be operationalized within an integrated biodiversity infrastructure to generate population- and species-level trait reference models.

CASE STUDY: THE GLOBAL FISH TRAIT ATLAS

The Global Fish Trait Atlas (GFTA) provides a concrete illustration of how specimen-anchored phenomics can bridge persistent biodiversity knowledge shortfalls when embedded within an integrated informatics framework. Rather than functioning as a standalone database as in Page et al. (2015), GFTA operates as a translation

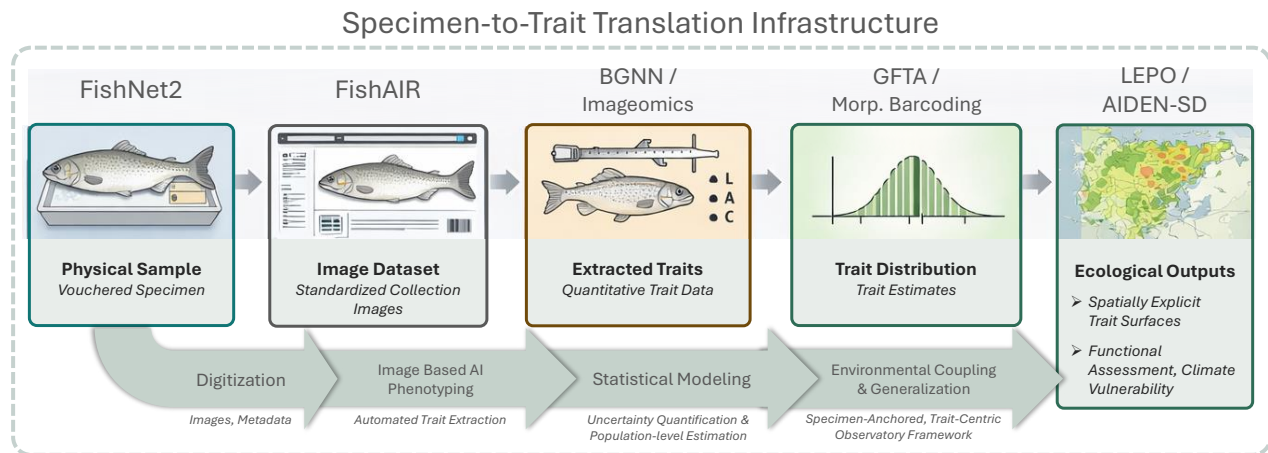


Figure 2. Project lineage enabling specimen-anchored trait synthesis.

A coherent infrastructure lineage connects digitized specimens to quantitative trait knowledge. FishNet 2 provides specimen-backed geographic context; FishAIR establishes quality-aware, AI-ready image representations; BGNN and Imageomics enable biology-guided phenomic extraction; Morphological Barcoding and the Global Fish Trait Atlas synthesize specimen-anchored trait distributions; and LEPO extends these components into an environmental observatory framework. Together, these elements operationalize specimen-to-trait translation at scale.

layer that connects museum specimens, quantitative phenotypes, and species-level ecological knowledge into a coherent analytical system consistent with emerging Digital Extended Specimen concepts and interoperable biodiversity cyberinfrastructure frameworks (Hardisty et al., 2022; Webster, 2017). In practice, this integration is enabled by linking large-scale museum specimen networks (e.g., “FishNet2,” 2026), image-based phenomic pipelines developed for digitized collections (e.g., “FishAIR,” 2026), and species-level ecological syntheses (e.g., “FishBase,” 2026) within a common specimen-centered reference framework.

A defining feature of the Global Fish Trait Atlas is that all traits are anchored to individual vouchered specimens. This contrasts with many existing trait databases that summarize traits as species-level averages derived from heterogeneous literature sources. Specimen anchoring allows traits to be represented as distributions that capture intraspecific variation, population-level structure, and geographic gradients, rather than single point estimates. For global change ecology, this distinction is critical: phenotypic responses to environmental change often manifest as shifts in variance, plasticity, or spatial patterning that may be obscured when traits are reduced to species-level averages (Violle et al., 2007).

These specimen records originate from multiple aggregation contexts, including domain-specific networks (e.g., FishNet2, FishAIR, FishBase), regional collection consortia associated with applied monitoring efforts such as the Great Lakes Invasives Network (“GLIN,” 2026),

and national digitization infrastructures such as iDigBio (“iDigBio,” 2026) and GBIF (“GBIF,” 2026). These specimen records establish the spatial, temporal, and taxonomic scaffolding necessary for reproducible trait inference and interoperable integration across distributed biodiversity infrastructures using shared standards and persistent identifiers (Wieczorek et al., 2012). Digitized specimen images associated with these records are then used as inputs for image-based phenomic pipelines that extract quantitative morphological measurements directly from physical vouchers (Bakış et al., 2023). In parallel, species-level ecological and life-history information contributes contextual knowledge that supports interpretation and synthesis of trait patterns across environments and taxa. The atlas thus emerges not from any single data source, but from the alignment of specimens, phenotypes, and ecological knowledge within a shared reference frame.

The current implementation of GFTA operates at a scale sufficient to demonstrate both feasibility and scientific value without relying on exhaustive coverage. On the order of hundreds of thousands of specimens, representing tens of thousands of fish species across freshwater and marine systems, have been processed to generate high-dimensional morphological trait data. Such scale is made possible by existing specimen aggregation and digitization efforts, combined with automated phenomic pipelines that operate directly on collection images. Importantly, these traits are not reduced to single species means. Instead, they retain links to individual specimens, allowing aggregation to population-level and species-level summaries while

preserving information on variance and sampling context. This order-of-magnitude scale is large enough to reveal emergent patterns, yet small enough to remain transparent and extensible as collections and digitization efforts continue to grow.

Within this framework, population-level trait distributions arise naturally from repeated measurements across specimens collected in different locations, habitats, and environmental conditions. When specimens are grouped by geography or environmental context, trait distributions reveal gradients rather than fixed values, enabling analyses of phenotypic structure across space. These distributions can be projected geographically to explore how morphology varies along climatic, hydrological, or biogeographic axes, providing a functional complement to traditional species distribution data. In this way, GFTA makes visible a dimension of biodiversity - spatially explicit trait variation - that has remained largely inaccessible in existing informatics systems.

The integration of specimen-anchored traits with ecological and environmental context enables new classes of research questions. For example, climate-trait coupling can be examined by relating morphological variation to temperature regimes, flow conditions, or other environmental drivers inferred from specimen localities. Such analyses move beyond asking where species occur to how their phenotypes vary in response to environment. Similarly, quantitative trait distributions provide tools for detecting cryptic diversity and specimen misidentification, as populations or individuals that deviate systematically from expected phenotypic patterns can be flagged for further taxonomic investigation. In these cases, phenomics does not replace taxonomic expertise but offers a scalable means of prioritizing specimens and lineages for closer scrutiny.

As a case study, GFTA demonstrates how specimen-anchored, AI-derived phenomics can be operationalized within an integrated infrastructure to transform collections into dynamic sources of trait knowledge. By linking individual specimens to quantitative traits and ecological context, the atlas illustrates a pathway for converting digitized biodiversity data into functional, spatially explicit, and uncertainty-aware representations. This approach highlights how addressing the specimen-to-trait translation gap can simultaneously advance trait-based ecology, climate-change research, and systematic biology.

QUANTIFYING PROGRESS ON THE RAUNKIÆRAN SHORTFALL

A central challenge in addressing biodiversity knowledge shortfalls is determining not only whether progress has been made, but how such progress can be measured.

When traits are treated as static, literature-derived attributes, assessment of completeness is largely qualitative (Meyer et al., 2016) and difficult to formalize. In contrast, specimen-anchored trait frameworks enable explicit, quantitative indicators of progress toward resolving the Raunkiærán shortfall by making trait knowledge measurable, repeatable, and scalable.

One immediate change introduced by specimen-anchored traits is the ability to quantify intraspecific variation. Rather than representing species by single trait values, repeated measurements across specimens reveal multidimensional trait distributions that capture population-level structure, geographic differentiation, and phenotypic spread (Villéger et al., 2008). The presence of well-characterized within-species variance marks a qualitative shift in functional knowledge: trait values become empirical distributions rather than abstractions. From the perspective of the Raunkiærán shortfall, progress can thus be measured by the proportion of taxa for which trait variance, rather than point estimates, is available.

A second advance is the incorporation of uncertainty-aware trait summaries. Because specimen-anchored traits are derived from multiple observations, uncertainty arising from sampling, measurement error, and model inference and analytical provenance can be explicitly estimated and propagated through reproducible provenance-aware workflows (“PROV Model Primer,” 2013). Trait summaries can therefore include measures of confidence, variance, or credibility intervals, enabling more robust downstream analyses. The availability of explicit uncertainty transforms trait data from descriptive inputs into statistically defensible quantities, reducing the risk of overconfident inference in ecological and conservation models.

Specimen anchoring also enables the generation of spatially explicit trait surfaces. When trait measurements are linked to georeferenced specimens, traits can be mapped across environmental and geographic gradients, revealing patterns of functional variation that parallel - but extend beyond - species distributions. Such spatial trait representations provide a functional counterpart to occurrence maps and allow direct assessment of how phenotypes vary across landscapes, climates, and habitats. The emergence of spatially explicit trait information represents a concrete indicator of progress toward integrating trait and geographic knowledge domains.

Building on these advances, several indicators can be proposed to assess trait-knowledge completeness in a systematic way. These include: the number of species with specimen-anchored trait distributions rather than single values; the average number of specimens contributing to

trait estimates per species or region; the extent to which uncertainty is quantified and reported; and the geographic coverage of trait data relative to known species ranges. Together, these indicators provide a multidimensional view of trait knowledge that can be tracked over time as collections, digitization, and analytical capacity expand.

Operationalizing specimen-anchored trait infrastructures at scale will also require interoperable and machine-actionable data models capable of linking specimens, traits, provenance, uncertainty, and environmental context across distributed biodiversity systems. Emerging semantic and linked-data approaches, including biodiversity ontologies, Darwin Core extensions, and modular schema frameworks such as LinkML, provide promising foundations for representing these relationships in reproducible and AI-ready forms (Moxon et al., 2026; Walls et al., 2014; Wieczorek et al., 2012). Although development of a formal schema is beyond the scope of this paper, interoperability-aware trait representation will likely be essential for scalable synthesis across collections, taxa, and analytical workflows.

Importantly, such indicators also inform prioritization of digitization and sampling efforts. By identifying taxa, regions, or environments where trait coverage remains sparse or uncertainty remains high, specimen-anchored frameworks can guide strategic investments in imaging, digitization, and collection. Rather than treating all specimens or taxa as equally deficient, this approach supports targeted actions that maximize gains in functional knowledge. In this way, progress toward closing the Raunkiæran shortfall becomes not only measurable, but actionable.

By reframing trait completeness as an outcome of specimen coverage, variance estimation, uncertainty characterization, and spatial representation, specimen-anchored phenomics provides a pathway for moving beyond qualitative assessments of trait gaps. Quantifying progress in these terms allows biodiversity informatics to track how effectively functional knowledge is being generated and to align future data mobilization efforts with the goal of resolving long-standing trait shortfalls.

IMPLICATIONS FOR BIODIVERSITY INFORMATICS AND CONSERVATION

The emergence of specimen-anchored trait data has broad implications for biodiversity informatics and applied conservation science. By grounding traits in physical vouchers and linking them to geography, environment, and uncertainty, this approach reshapes how functional biodiversity is represented, analyzed, and used for decision-making. Specimen-anchored phenomics is most effective when embedded within existing biodiversity infra-

structures - leveraging specimen aggregation networks, phenomic pipelines, and curated species knowledge rather than duplicating them. Such specimen-anchored trait infrastructures also provide an evidence base for prioritizing digitization, imaging, and sampling efforts where functional knowledge gaps are greatest.

Reshaping trait-based ecology

Trait-based ecology has traditionally relied on coarse, species-averaged trait values (McGill et al., 2006) to infer ecological strategies, community assembly, and ecosystem functioning. Specimen-anchored traits fundamentally alter this paradigm by enabling analyses that incorporate intraspecific variation, population structure, and environmental context. Rather than treating species as homogeneous functional units, ecological models can account for phenotypic diversity within species and across landscapes, improving the ability to evaluate functional responses to environmental disturbance and ecological change (Mouillot et al., 2013). This shift improves the realism of trait-based inference, particularly in systems where local adaptation and phenotypic plasticity play key roles. As a result, functional diversity can be assessed not only across taxa, but also within them, offering a more nuanced understanding of ecological processes and the relationships between biodiversity, ecosystem functioning, and environmental change (de Bello et al., 2010).

Advancing climate vulnerability assessments

Climate vulnerability assessments increasingly seek to incorporate functional traits (Pacifiçi et al., 2015) as indicators of sensitivity, exposure, and adaptive capacity. However, assessments based on static or literature-derived trait values risk oversimplifying organismal responses to environmental change. Specimen-anchored traits enable climate-trait coupling, allowing phenotypic variation to be analyzed directly in relation to temperature, hydrology, or other environmental gradients. This supports vulnerability assessments that reflect observed phenotypic responses rather than assumed species-level characteristics. In doing so, specimen-anchored phenomics provides a pathway for identifying populations or regions where functional traits suggest heightened sensitivity or resilience to climate change.

Informing conservation planning

For conservation planning, access to spatially explicit and uncertainty-aware trait information enables more targeted and defensible prioritization. Functional traits influence species' responses to habitat alteration, exploitation, and environmental stressors, yet trait uncertainty and vari-

ability are rarely considered explicitly in conservation decision frameworks. Specimen-anchored trait distributions allow conservation planners to evaluate not only which species are at risk, but how functional variation is distributed across regions and populations. This perspective can inform actions such as identifying climate refugia, prioritizing areas for protection based on functional diversity, or targeting monitoring efforts where phenotypic responses are poorly characterized.

Complementarity with existing biodiversity infrastructures

Importantly, specimen-anchored trait frameworks are not intended to replace existing biodiversity data infrastructures, but to extend and complement them. Global platforms such as Global Biodiversity Information Facility (GBIF) have transformed access to occurrence and distribution data, while knowledge bases such as FishBase provide authoritative syntheses of species ecology and life history. Specimen-anchored phenomics builds on these foundations by adding a missing analytical layer that translates specimens into quantitative traits. Occurrence data provide spatial context, species knowledge provides ecological meaning, and specimen-anchored traits supply functional measurements that can be integrated across systems.

By operating as a translation layer rather than a replacement, specimen-anchored trait infrastructures strengthen the overall biodiversity informatics ecosystem. They leverage existing investments in collections, digitization, and species knowledge, while enabling new forms of synthesis that were previously unattainable. In this way, addressing the Raunkiæran shortfall through specimen-anchored phenomics enhances the interoperability and analytical power of biodiversity data as a whole, supporting more informed ecological research and conservation action in a rapidly changing world.

TOWARD A LIVING ENVIRONMENTAL PHENOMICS OBSERVATORY

The integration of specimen-anchored trait synthesis with environmental context motivates a shift from static trait compilations toward a living observatory model. The Living Environmental Phenomics Observatory (LEPO) extends specimen-centered phenomics by coupling reproducibly derived traits to environmental covariates using specimen locality and time as anchors. This architecture enables direct examination of how phenotypes vary across climatic, hydrologic, and land-use gradients, and how those patterns change through time.

By treating traits as dynamic ecological observables rather than fixed species attributes, LEPO supports inference about phenotypic plasticity, sensitivity, and resilience under global change. Richer trait distributions improve niche characterization and provide a foundation for trait-mediated interaction hypotheses, contributing to progress on Hutchinsonian and Eltonian knowledge gaps. Importantly, LEPO is not a speculative construct, but a logical integration of existing components - FishNet 2, FishAIR, and specimen-anchored trait synthesis - within a reproducible, governance-aware cyberinfrastructure framework.

Although fishes provide a tractable and well-developed case study for specimen-anchored phenomics due to extensive digitization efforts, large image repositories, and established biodiversity infrastructures, the broader framework proposed here is not inherently fish-specific. Comparable opportunities increasingly exist across herbarium collections, insect imaging systems, vertebrate morphology repositories, agricultural phenomics platforms, and other digitized biological collections (Carranza-Rojas et al., 2017; Rolfe et al., 2021; Younis et al., 2018). Different organismal systems will require different extraction strategies, trait vocabularies, and analytical workflows; however, the underlying principles of specimen anchoring, provenance-aware trait generation, uncertainty tracking, and interoperable digital integration are broadly transferable across biodiversity domains.

The same principles generalize beyond ichthyology. Through efforts such as AI-Ready Data Ecosystem for Networking of Scientific Discovery (AIDEN-SD), interoperable, AI-ready workflows and data products can support cross-domain synthesis, enabling trait-based inference across taxa and ecosystems in a manner consistent with the scale and urgency of global change.

CONCLUSIONS

The persistent Raunkiæran shortfall reflects not only missing measurements, but the absence of infrastructure capable of translating digitized biodiversity evidence into reproducible, specimen-anchored trait knowledge. While biodiversity informatics has dramatically improved access to names and occurrences, functional traits have remained disconnected from vouchers, uncertainty, and population-level variation.

We outline a practical route forward grounded in a coherent project lineage (Figure 2). FishNet 2 provides a specimen-backed backbone for geographic discovery. FishAIR establishes quality-aware, AI-ready representations of specimen imagery. BGNN and Imageomics

operationalize biology- and hierarchy-guided phenomic extraction within reproducible workflows. Morphological Barcoding and the Global Fish Trait Atlas synthesize extracted phenotypes into specimen-anchored trait reference models that preserve provenance, variation, and uncertainty. LEPO extends this architecture into a continuously updating observatory aligned with global change ecology, while AIDEN-SD generalizes its principles for interoperable, AI-ready data products across domains.

Some components of this infrastructure lineage are already operational (e.g., FishNet 2, FishAIR, BGNN, and Imageomics), whereas others represent emerging synthesis efforts or architectural frameworks currently under development (e.g., GFTA, LEPO, and AIDEN-SD). Together, they illustrate one possible pathway toward specimen-anchored trait synthesis within a broader and rapidly evolving biodiversity informatics landscape.

Reframing traits as infrastructure-derived ecological observables enables measurable progress on functional knowledge at scales required for forecasting, conservation planning, and climate-impact assessment. The future of biodiversity science may increasingly be shaped not only by the volume of digitized data, but by whether we build durable translation infrastructures capable of converting specimens into actionable ecological knowledge under accelerating global change.

ACKNOWLEDGMENTS

The author thanks Dr. Henry Bart and Dr. Brian Sidlauskas for collaboration and for creating an institutional environment that supported the development of specimen-based cyberinfrastructure and phenomics-enabled biodiversity research.

This research was supported in part by the U.S. National Science Foundation (NSF) through awards supporting the FishNet 2, Biology-Guided Neural Networks (BGNN), and Imageomics initiatives (NSF award numbers: FishNet 2 - DBI 2031693, BGNN - OAC 1940322, Imageomics - OAC 2118240). Additional support was provided through related NSF-funded efforts focused on specimen digitization, image-based phenomics, and biodiversity data integration.

Competing Interests: The authors have declared that no competing interests exist.

REFERENCES

Albert, C.H., Grassein, F., Schurr, F.M., Vieilledent, G., Violle, C., 2011. When and how should intraspecific variability be considered in trait-based plant ecology? *Perspect. Plant Ecol. Evol. Syst.* 13, 217–225. <https://doi.org/10.1016/j.ppees.2011.04.003>

Ariño, A.H., 2010. Approaches to estimating the universe of natural history collections data. *Biodiversity Informatics* 7, 81–92. <https://doi.org/10.17161/BI.V7I2.3991>

Bakiş, Y., Wang, X., Altıntaş, B., Jebbia, D., Bart Jr., H., 2023. On image quality metadata, FAIR in ML, AI-readiness and reproducibility: Fish-AIR example. *Biodiversity Information Science and Standards* 7, e112178. <https://doi.org/10.3897/biss.7.112178>

Balk, M.A., Bradley, J., Maruf, M., Altıntaş, B., Bakiş, Y., Bart, H.L., Breen, D., Florian, C.R., Greenberg, J., Karpatne, A., Karnani, K., Mabee, P., Pepper, J., Jebbia, D., Tabarin, T., Wang, X., Lapp, H., 2024. A FAIR and modular image-based workflow for knowledge discovery in the emerging field of imageomics. *Methods Ecol. Evol.* 15, 1129–1145. <https://doi.org/10.1111/2041-210X.14327>

Balk, M.A., Deck, J., Emery, K.F., Walls, R.L., Reuter, D., LaFrance, R., Arroyo-Cabrales, J., Barrett, P., Blois, J., Boileau, A., Brenskelle, L., Cannarozzi, N.R., Cruz, J.A., Dávalos, L.M., de la Sancha, N.U., Gyawali, P., Hantak, M.M., Hopkins, S., Kohli, B., King, J.N., Koo, M.S., Lawing, A.M., Machado, H., McCrane, S.M., McLean, B., Morgan, M.E., Pilaar Birch, S., Reed, D., Reitz, E.J., Sewnath, N., Upham, N.S., Villaseñor, A., Yohe, L., Davis, E.B., Guralnick, R.P., 2022. A solution to the challenges of interdisciplinary aggregation and use of specimen-level trait data. *iScience* 25, 105101. <https://doi.org/10.1016/j.isci.2022.105101>

Beale, C.M., Lennon, J.J., 2012. Incorporating uncertainty in predictive species distribution modelling. *Philosophical Transactions of the Royal Society B* 367, 247–258. <https://doi.org/10.1098/RSTB.2011.0178>

Blagoderov, V., Kitching, I., Livermore, L., Simonsen, T., Smith, V., 2012. No specimen left behind: industrial scale digitization of natural history collections. *Zookeys* 209, 133–146. <https://doi.org/10.3897/zookeys.209.3178>

Carmona, C.P., de Bello, F., Mason, N.W.H., Lepš, J., 2016. Traits without borders: Integrating functional diversity across scales. *Trends Ecol. Evol.* 31, 382–394. <https://doi.org/10.1016/j.tree.2016.02.003>

Carranza-Rojas, J., Goeau, H., Bonnet, P., Mata-Montero, E., Joly, A., 2017. Going deeper in the automated identification of herbarium specimens. *BMC Evol. Biol.* 17, 181. <https://doi.org/10.1186/s12862-017-1014-z>

Castillo-Figueroa, D., 2018. Beyond specimens: linking biological collections, functional ecology and biodiversity conservation. *Rev. Peru. Biol.* 25, 343–348. <https://doi.org/10.15381/rpb.v25i3.14246>

- Chapman, M., Goldstein, B.R., Schell, C.J., Brashares, J.S., Carter, N.H., Ellis-Soto, D., Faxon, H.O., Goldstein, J.E., Halpern, B.S., Longdon, J., Norman, K.E.A., O'Rourke, D., Scoville, C., Xu, L., Boettiger, C., 2024. Biodiversity monitoring for a just planetary future. *Science* 383, 34–36. <https://doi.org/10.1126/science.adh8874>
- Christin, S., Hervet, É., Lecomte, N., 2019. Applications for deep learning in ecology. *Methods Ecol. Evol.* 10, 1632–1644. <https://doi.org/10.1111/2041-210X.13256>
- Cordlandwehr, V., Meredith, R.L., Ozinga, W.A., Bekker, R.M., van Groenendael, J.M., Bakker, J.P., 2013. Do plant traits retrieved from a database accurately predict on-site measurements? *Journal of Ecology* 101, 662–670. <https://doi.org/10.1111/1365-2745.12091>
- de Bello, F., Fischer, F.M., Puy, J., Shipley, B., Verdú, M., Götzenberger, L., Lavorel, S., Moretti, M., Wright, I.J., Berg, M.P., Carmona, C.P., Cornelissen, J.H.C., Dias, A.T.C., Gibb, H., Lepš, J., Madin, J.S., Majeková, M., Pausas, J.G., Segrestin, J., Sobral, M., Zanne, A.E., Garnier, E., 2025. Raunkiæran shortfalls: Challenges and perspectives in trait-based ecology. *Ecol. Monogr.* 95, e70018. <https://doi.org/10.1002/ecm.70018>
- de Bello, F., Lavorel, S., Díaz, S., Harrington, R., Cornelissen, J.H.C., Bardgett, R.D., Berg, M.P., Cipriotti, P., Feld, C.K., Hering, D., Martins da Silva, P., Potts, S.G., Sandin, L., Sousa, J.P., Storkey, J., Wardle, D.A., Harrison, P.A., 2010. Towards an assessment of multiple ecosystem processes and services via functional traits. *Biodivers. Conserv.* 19, 2873–2893. <https://doi.org/10.1007/s10531-010-9850-9>
- Dormann, C.F., Bobrowski, M., Dehling, D.M., Harris, D.J., Hartig, F., Lischke, H., Moretti, M.D., Pagel, J., Pinkert, S., Schleuning, M., Schmidt, S.I., Sheppard, C.S., Steinbauer, M.J., Zeuss, D., Kraan, C., 2018. Biotic interactions in species distribution modelling: 10 questions to guide interpretation and avoid false conclusions. *Global Ecology and Biogeography* 27, 1004–1016. <https://doi.org/10.1111/geb.12759>
- Edwards, J.L., Lane, M.A., Nielsen, E.S., 2000. Interoperability of biodiversity databases: biodiversity information on every desktop. *Science* 289, 2312–2314. <https://doi.org/10.1126/SCIENCE.289.5488.2312>
- Elhamod, M., Diamond, K.M., Maga, A.M., Bakis, Y., Bart, H.L., Mabee, P., Dahdul, W., Leipzig, J., Greenberg, J., Avants, B., Karpatne, A., 2022. Hierarchy-guided neural network for species classification. *Methods Ecol. Evol.* 13, 642–652. <https://doi.org/10.1111/2041-210X.13768>
- Elhamod, M., Khurana, M., Manogaran, H.B., Uyeda, J.C., Balk, M.A., Dahdul, W., Bakis, Y., Bart, H.L., Mabee, P.M., Lapp, H., Balhoff, J.P., Charpentier, C., Carlyn, D., Chao, W.-L., Stewart, C. V., Rubenstein, D.I., Berger-Wolf, T., Karpatne, A., 2023. Discovering novel biological traits from images using phylogeny-guided neural networks, in: *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*. ACM, New York, NY, USA, pp. 3966–3978. <https://doi.org/10.1145/3580305.3599808>
- Falster, D., Gallagher, R., Wenk, E.H., Wright, I.J., In-diarto, D., Andrew, S.C., Baxter, C., Lawson, J., Allen, S., Fuchs, A., Monro, A., Kar, F., Adams, M.A., Ahrens, C.W., Alfonzetti, M., Angevin, T., Apgaua, D.M.G., Arndt, S., Atkin, O.K., Atkinson, J., Auld, T., Baker, A., von Balthazar, M., Bean, A., Blackman, C.J., Bloomfield, K., Bowman, D.M.J.S., Bragg, J., Brodribb, T.J., Buckton, G., Burrows, G., Caldwell, E., Camac, J., Carpenter, R., Catford, J.A., Cawthray, G.R., Cernusak, L.A., Chandler, G., Chapman, A.R., Cheal, D., Cheesman, A.W., Chen, S.-C., Choat, B., Clinton, B., Clode, P.L., Coleman, H., Cornwell, W.K., Cosgrove, M., Crisp, M., Cross, E., Crous, K.Y., Cunningham, S., Curran, T., Curtis, E., Daws, M.I., DeGabriel, J.L., Denton, M.D., Dong, N., Du, P., Duan, H., Duncan, D.H., Duncan, R.P., Duretto, M., Dwyer, J.M., Edwards, C., Esperon-Rodriguez, M., Evans, J.R., Everingham, S.E., Farrell, C., Firn, J., Fonseca, C.R., French, B.J., Frood, D., Funk, J.L., Geange, S.R., Ghannoum, O., Gleason, S.M., Gosper, C.R., Gray, E., Groom, P.K., Grootemaat, S., Gross, C., Guerin, G., Guja, L., Hahs, A.K., Harrison, M.T., Hayes, P.E., Henery, M., Hochuli, D., Howell, J., Huang, G., Hughes, L., Huisman, J., Ilic, J., Jagdish, A., Jin, D., Jordan, G., Jurado, E., Kanowski, J., Kassel, S., Kellermann, J., Kenny, B., Kohout, M., Kooyman, R.M., Kotowska, M.M., Lai, H.R., Laliberté, E., Lambers, H., Lamont, B.B., Lanfear, R., van Langevelde, F., Laughlin, D.C., Laugier-Kitchener, B.-A., Laurance, S., Lehmann, C.E.R., Leigh, A., Leishman, M.R., Lenz, T., Lepschi, B., Lewis, J.D., Lim, F., Liu, U., Lord, J., Lusk, C.H., Macinnis-Ng, C., McPherson, H., Magallón, S., Manea, A., López-Martinez, A., Mayfield, M., McCarthy, J.K., Meers, T., van der Merwe, M., Metcalfe, D.J., Milberg, P., Mokany, K., Moles, A.T., Moore, B.D., Moore, N., Morgan, J.W., Morris, W., Muir, A., Munroe, S., Nicholson, Á., Nicolle, D., Nicotra, A.B., Niinemets, Ü., North, T., O'Reilly-Nugent, A., O'Sullivan, O.S., Oberle, B.,

- Onoda, Y., Ooi, M.K.J., Osborne, C.P., Paczkowska, G., Pekin, B., Guilherme Pereira, C., Pickering, C., Pickup, M., Pollock, L.J., Poot, P., Powell, J.R., Power, S.A., Prentice, I.C., Prior, L., Prober, S.M., Read, J., Reynolds, V., Richards, A.E., Richardson, B., Roderick, M.L., Rosell, J.A., Rossetto, M., Rye, B., Rymer, P.D., Sams, M.A., Sanson, G., Sauquet, H., Schmidt, S., Schönenberger, J., Schulze, E.-D., Sendall, K., Sinclair, S., Smith, B., Smith, R., Soper, F., Sparrow, B., Standish, R.J., Staples, T.L., Stephens, R., Szota, C., Taseski, G., Tasker, E., Thomas, F., Tissue, D.T., Tjoelker, M.G., Tng, D.Y.P., de Tombeur, F., Tomlinson, K., Turner, N.C., Veneklaas, E.J., Venn, S., Veski, P., Vlasveld, C., Vorontsova, M.S., Warren, C.A., Warwick, N., Weerasinghe, L.K., Wells, J., Westoby, M., White, M., Williams, N.S.G., Wills, J., Wilson, P.G., Yates, C., Zanne, A.E., Zemunik, G., Ziemińska, K., 2021. AusTraits, a curated plant trait database for the Australian flora. *Sci. Data* 8, 254. <https://doi.org/10.1038/s41597-021-01006-6>
- FishAIR [WWW Document], 2026. URL <https://fishair.org/> (accessed 5.24.26).
- FishBase [WWW Document], 2026. URL <https://www.fishbase.org/> (accessed 5.24.26).
- FishNet2 [WWW Document], 2026. URL <https://fishnet2.net> (accessed 5.29.26).
- GBIF [WWW Document], 2026. URL <https://www.gbif.org/> (accessed 5.29.26).
- GLIN [WWW Document], 2026. URL <https://www.glin.org/glin> (accessed 5.29.26).
- Gonçalves-Souza, T., Chaves, L.S., Boldorini, G.X., Ferreira, N., Gusmão, R.A.F., Perônico, P.B., Sanders, N.J., Teresa, F.B., 2023. Bringing light onto the Raunkiær shortfall: A comprehensive review of traits used in functional animal ecology. *Ecol. Evol.* 13, e10016. <https://doi.org/10.1002/ece3.10016>
- Hardisty, A.R., Ellwood, E.R., Nelson, G., Zimkus, B., Buschbom, J., Addink, W., Rabeler, R.K., Bates, J., Bentley, A., Fortes, José A B, Hansen, S., Macklin, J.A., Mast, A.R., Miller, J.T., Monfils, A.K., Paul, D.L., Wallis, E., Webster, M., 2022. Digital extended specimens: Enabling an extensible network of biodiversity data records as integrated digital objects on the Internet. *Bioscience* 72, 978–987. <https://doi.org/10.1093/biosci/biac060>
- Henning, T., Plitzner, P., Güntsch, A., Berendsohn, W.G., Müller, A., Kilian, N., 2018. Building compatible and dynamic character matrices – Current and future use of specimen-based character data. *Bot. Lett.* 165, 352–360. <https://doi.org/10.1080/23818107.2018.1452791>
- Hortal, J., De Bello, F., Diniz-Filho, J.A.F., Lewinsohn, T.M., Lobo, J.M., Ladle, R.J., 2015. Seven shortfalls that beset large-scale knowledge of biodiversity. *Annu. Rev. Ecol. Evol. Syst.* 46, 523–549. <https://doi.org/10.1146/ANNUREV-ECOL-SYS-112414-054400/1>
- Houle, D., Govindaraju, D.R., Omholt, S., 2010. Phenomics: The next challenge. *Nat. Rev. Genet.* 11, 855–866. <https://doi.org/10.1038/nrg2897>
- Høye, T.T., Ärje, J., Bjerger, K., Hansen, O.L.P., Iosifidis, A., Leese, F., Mann, H.M.R., Meissner, K., Melvad, C., Raitoharju, J., 2021. Deep learning and computer vision will transform entomology. *Proceedings of the National Academy of Sciences USA* 118, e2002545117. <https://doi.org/10.1073/pnas.2002545117>
- iDigBio [WWW Document], 2026. URL <https://www.idigbio.org> (accessed 5.29.26).
- Jebbia, D., Wang, X., Bakis, Y., Bart, H.L., Greenberg, J., 2023. Toward a flexible metadata pipeline for fish specimen images, in: *Communications in Computer and Information Science*. Springer Science and Business Media Deutschland GmbH, pp. 175–190. https://doi.org/10.1007/978-3-031-39141-5_15
- Kattge, J., Bönsch, G., Díaz, S., Lavorel, S., Prentice, I.C., Leadley, P., Tautenhahn, S., Werner, G.D.A., Aakala, T., Abedi, M., Acosta, A.T.R., Adamidis, G.C., Adamson, K., Aiba, M., Albert, C.H., Alcántara, J.M., Alcázar, C., Aleixo, I., Ali, H., Amiaud, B., Ammer, C., Amoroso, M.M., Anand, M., Anderson, C., Anten, N., Antos, J., Apgaua, D.M.G., Ashman, T., Asmara, D.H., Asner, G.P., Aspinwall, M., Atkin, O., Aubin, I., Baastrup-Spohr, L., Bahalkeh, K., Bahn, M., Baker, T., Baker, W.J., Bakker, J.P., Baldocchi, D., Baltzer, J., Banerjee, A., Baranger, A., Barlow, J., Barneche, D.R., Baruch, Z., Bastianelli, D., Battles, J., Bauerle, W., Bauters, M., Bazzato, E., Beckmann, M., Beeckman, H., Beierkuhnlein, C., Bekker, R., Belfry, G., Belluau, M., Beloiu, M., Benavides, R., Benomar, L., Berdugo-Lattke, M.L., Berenguer, E., Bergamin, R., Bergmann, J., Bergmann Carlucci, M., Berner, L., Bernhardt-Römermann, M., Bigler, C., Bjorkman, A.D., Blackman, C., Blanco, C., Blonder, B., Blumenthal, D., Bocanegra-González, K.T., Boeckx, P., Bohlman, S., Böhning-Gaese, K., Boissvert-Marsh, L., Bond, W., Bond-Lamberty, B., Boom, A., Boonman, C.C.F., Bordin, K., Boughton, E.H., Boukili, V., Bowman, D.M.J.S., Bravo, S., Brendel, M.R., Broadley, M.R., Brown, K.A., Bruehlheide, H.,

- Brumnich, F., Bruun, H.H., Bruy, D., Buchanan, S.W., Bucher, S.F., Buchmann, N., Buitenwerf, R., Bunker, D.E., Bürger, J., Burrascano, S., Burslem, D.F.R.P., Butterfield, B.J., Byun, C., Marques, M., Scalón, M.C., Caccianiga, M., Cadotte, M., Cailleret, M., Camac, J., Camarero, J.J., Company, C., Campetella, G., Campos, J.A., Cano-Arboleda, L., Canullo, R., Carbognani, M., Carvalho, F., Casanoves, F., Castagneyrol, B., Catford, J.A., Cavender-Bares, J., Cerabolini, B.E.L., Cervellini, M., Chacón-Madrigal, E., Chapin, K., Chapin, F.S., Chelli, S., Chen, S., Chen, A., Cherubini, P., Chianucci, F., Choat, B., Chung, K., Chytrý, M., Ciccarelli, D., Coll, L., Collins, C.G., Conti, L., Coomes, D., Cornelissen, J.H.C., Cornwell, W.K., Corona, P., Coyea, M., Craine, J., Craven, D., Cromsigt, J.P.G.M., Cseceserits, A., Cufar, K., Cuntz, M., da Silva, A.C., Dahlin, K.M., Dainese, M., Dalke, I., Dalle Fratte, M., Dang-Le, A.T., Danihelka, J., Dannonura, M., Dawson, S., de Beer, A.J., De Frutos, A., De Long, J.R., Dechant, B., Delagrange, S., Delpierre, N., Derroire, G., Dias, A.S., Diaz-Toribio, M.H., Dimitrakopoulos, P.G., Dobrowolski, M., Doktor, D., Dřevojan, P., Dong, N., Dransfield, J., Dressler, S., Duarte, L., Ducouret, E., Dullinger, S., Durka, W., Duursma, R., Dymova, O., E-Vojtkó, A., Eckstein, R.L., Ejtehadi, H., Elser, J., Emilio, T., Engemann, K., Erfanian, M.B., Erfmeier, A., Esquivel-Muelbert, A., Esser, G., Estiarte, M., Domingues, T.F., Fagan, W.F., Fagúndez, J., Falster, D.S., Fan, Y., Fang, J., Farris, E., Fazlioglu, F., Feng, Y., Fernandez-Mendez, F., Ferrara, C., Ferreira, J., Fidelis, A., Finegan, B., Firn, J., Flowers, T.J., Flynn, D.F.B., Fontana, V., Forey, E., Forgiarini, C., François, L., Frangipani, M., Frank, D., Frenette-Dussault, C., Freschet, G.T., Fry, E.L., Fyllas, N.M., Mazzochini, G.G., Gachet, S., Gallagher, R., Ganade, G., Ganga, F., García-Palacios, P., Gargaglione, V., Garnier, E., Garrido, J.L., de Gasper, A.L., Gea-Izquierdo, G., Gibson, D., Gillison, A.N., Girollo, A., Glasenhardt, M., Gleason, S., Gliesch, M., Goldberg, E., Gödel, B., Gonzalez-Akre, E., Gonzalez-Andujar, J.L., González-Melo, A., González-Robles, A., Graae, B.J., Granda, E., Graves, S., Green, W.A., Gregor, T., Gross, N., Guerin, G.R., Günther, A., Gutiérrez, A.G., Haddock, L., Haines, A., Hall, J., Hambuckers, A., Han, W., Harrison, S.P., Hattingh, W., Hawes, J.E., He, T., He, P., Heberling, J.M., Helm, A., Hempel, S., Hentschel, J., Hérault, B., Hereş, A., Herz, K., Heuertz, M., Hickler, T., Hietz, P., Higuchi, P., Hipp, A.L., Hiron, A., Hock, M., Hogan, J.A., Holl, K., Honnay, O., Hornstein, D., Hou, E., Hough-Snee, N., Hovstad, K.A., Ichie, T., Igić, B., Illa, E., Isaac, M., Ishihara, M., Ivanov, L., Ivanova, L., Iversen, C.M., Izquierdo, J., Jackson, R.B., Jackson, B., Jactel, H., Jagodzinski, A.M., Jandt, U., Jansen, S., Jenkins, T., Jentsch, A., Jespersen, J.R.P., Jiang, G., Johansen, J.L., Johnson, D., Jokela, E.J., Joly, C.A., Jordan, G.J., Joseph, G.S., Junaedi, D., Junker, R.R., Justes, E., Kabzems, R., Kane, J., Kaplan, Z., Kattenborn, T., Kavelenova, L., Kearsley, E., Kempel, A., Kenzo, T., Kerkhoff, A., Khalil, M.I., Kinlock, N.L., Kissling, W.D., Kitajima, K., Kitzberger, T., Kjeller, R., Klein, T., Kleyer, M., Klimešová, J., Klipel, J., Kloeppel, B., Klotz, S., Knops, J.M.H., Kohyama, T., Koike, F., Kollmann, J., Komac, B., Komatsu, K., König, C., Kraft, N.J.B., Kramer, K., Kreft, H., Kühn, I., Kumarathunge, D., Kuppler, J., Kurokawa, H., Kurosawa, Y., Kuyah, S., Laclau, J., Lafleur, B., Lallai, E., Lamb, E., Lamprecht, A., Larkin, D.J., Laughlin, D., Le Bagousse-Pinguet, Y., le Maire, G., le Roux, P.C., le Roux, E., Lee, T., Lens, F., Lewis, S.L., Lhotsky, B., Li, Y., Li, X., Lichstein, J.W., Liebergesell, M., Lim, J.Y., Lin, Y., Linares, J.C., Liu, C., Liu, D., Liu, U., Livingstone, S., Llusà, J., Lohbeck, M., López-García, Á., Lopez-Gonzalez, G., Lososová, Z., Louault, F., Lukács, B.A., Lukeš, P., Luo, Y., Lussu, M., Ma, S., Maciel Rabelo Pereira, C., Mack, M., Maire, V., Mäkelä, A., Mäkinen, H., Malhado, A.C.M., Mallik, A., Manning, P., Manzoni, S., Marchetti, Z., Marchino, L., Marcilio-Silva, V., Marcon, E., Marignani, M., Markesteijn, L., Martin, A., Martínez-Garza, C., Martínez-Vilalta, J., Mašková, T., Mason, K., Mason, N., Massad, T.J., Masse, J., Mayrose, I., McCarthy, J., McCormack, M.L., McCulloh, K., McFadden, I.R., McGill, B.J., McPartland, M.Y., Medeiros, J.S., Medlyn, B., Meerts, P., Mehrabi, Z., Meir, P., Melo, F.P.L., Mencuccini, M., Meredieu, C., Messier, J., Mészáros, I., Metsaranta, J., Michaletz, S.T., Michelaki, C., Migalina, S., Milla, R., Miller, J.E.D., Minden, V., Ming, R., Mokany, K., Moles, A.T., Molnár, A., Molofsky, J., Molz, M., Montgomery, R.A., Monty, A., Moravcová, L., Moreno-Martínez, A., Moretti, M., Mori, A.S., Mori, S., Morris, D., Morrison, J., Mucina, L., Mueller, S., Muir, C.D., Müller, S.C., Munoz, F., Myers-Smith, I.H., Myser, R.W., Nagano, M., Naidu, S., Narayanan, A., Natesan, B., Negoita, L., Nelson, A.S., Neuschulz, E.L., Ni, J., Niedrist, G., Nieto, J., Niinemets, Ü., Nolan, R., Nottebrock, H., Nouvellon, Y., Novakovskiy, A., Nystuen, K.O., O'Grady, A., O'Hara, K., O'Reilly-Nugent, A., Oakley, S., Oberhuber, W., Ohtsuka, T., Oliveira, R., Öllner, K., Olson, M.E., Onipchenko, V., Onoda, Y., Onstein, R.E., Ordóñez, J.C., Osada, N., Ostonen,

- I., Ottaviani, G., Otto, S., Overbeck, G.E., Ozinga, W.A., Pahl, A.T., Paine, C.E.T., Pakeman, R.J., Papatgeorgiou, A.C., Parfionova, E., Pärtel, M., Patacca, M., Paula, S., Paule, J., Pauli, H., Pausas, J.G., Peco, B., Penuelas, J., Perea, A., Peri, P.L., Petisco-Souza, A.C., Petraglia, A., Petritan, A.M., Phillips, O.L., Pierce, S., Pillar, V.D., Pisek, J., Pomogaybin, A., Poorter, H., Portsmouth, A., Poschlod, P., Potvin, C., Pounds, D., Powell, A.S., Power, S.A., Prinzing, A., Puglielli, G., Pyšek, P., Raavel, V., Rammig, A., Ransijn, J., Ray, C.A., Reich, P.B., Reichstein, M., Reid, D.E.B., Réjou-Méchain, M., de Dios, V.R., Ribeiro, S., Richardson, S., Riibak, K., Rillig, M.C., Riviera, F., Robert, E.M.R., Roberts, S., Robroek, B., Roddy, A., Rodrigues, A.V., Rogers, A., Rollinson, E., Rolo, V., Römermann, C., Ronzhina, D., Roscher, C., Rosell, J.A., Rosenfield, M.F., Rossi, C., Roy, D.B., Royer-Tardif, S., Rüger, N., Ruiz-Peinado, R., Rumpf, S.B., Rusch, G.M., Ryo, M., Sack, L., Saldaña, A., Salgado-Negret, B., Salguero-Gomez, R., Santa-Regina, I., Santacruz-García, A.C., Santos, J., Sardans, J., Schamp, B., Scherer-Lorenzen, M., Schleuning, M., Schmid, B., Schmidt, M., Schmitt, S., Schneider, J. V., Schowanek, S.D., Schrader, J., Schrod, F., Schuldt, B., Schurr, F., Selaya Garvizu, G., Semchenko, M., Seymour, C., Sfair, J.C., Sharpe, J.M., Sheppard, C.S., Sheremetiev, S., Shiodera, S., Shipley, B., Shovon, T.A., Siebenkäs, A., Sierra, C., Silva, V., Silva, M., Sitzia, T., Sjöman, H., Slot, M., Smith, N.G., Sodhi, D., Soltis, P., Soltis, D., Somers, B., Sonnier, G., Sørensen, M.V., Sosinski, E.E., Soudzilovskaia, N.A., Souza, A.F., Spasojevic, M., Sperandii, M.G., Stan, A.B., Stegen, J., Steinbauer, K., Stephan, J.G., Sterck, F., Stojanovic, D.B., Strydom, T., Suarez, M.L., Svenning, J., Svitková, I., Svitok, M., Svoboda, M., Swaine, E., Swenson, N., Tabarelli, M., Takagi, K., Tappeiner, U., Tarifa, R., Tauugourdeau, S., Tavanoglu, C., te Beest, M., Tedersoo, L., Thiffault, N., Thom, D., Thomas, E., Thompson, K., Thornton, P.E., Thuiller, W., Tichý, L., Tissue, D., Tjoelker, M.G., Tng, D.Y.P., Tobias, J., Török, P., Tarin, T., Torres-Ruiz, J.M., Tóthmérész, B., Treurnicht, M., Trivel-lone, V., Trollet, F., Trotsiuk, V., Tsakalos, J.L., Tsi-ripidis, I., Tysklind, N., Umehara, T., Usoltsev, V., Vadeboncoeur, M., Vaezi, J., Valladares, F., Vamosi, J., van Bodegom, P.M., van Breugel, M., Van Cleemput, E., van de Weg, M., van der Merwe, S., van der Plas, F., van der Sande, M.T., van Kleunen, M., Van Meerbeek, K., Vanderwel, M., Vanselow, K.A., Vårhammar, A., Varone, L., Vasquez Valderrama, M.Y., Vassilev, K., Vellend, M., Veneklaas, E.J., Verbeeck, H., Verheyen, K., Vibrans, A., Vieira, I., Villacís, J., Violle, C., Vivek, P., Wagner, K., Waldram, M., Waldron, A., Walker, A.P., Waller, M., Walther, G., Wang, H., Wang, F., Wang, W., Watkins, H., Watkins, J., Weber, U., Weedon, J.T., Wei, L., Weigelt, P., Weiher, E., Wells, A.W., Wellstein, C., Wenk, E., Westoby, M., Westwood, A., White, P.J., Whitten, M., Williams, M., Winkler, D.E., Winter, K., Womack, C., Wright, I.J., Wright, S.J., Wright, J., Pinho, B.X., Ximenes, F., Yamada, T., Yamaji, K., Yanai, R., Yankov, N., Yguel, B., Zanini, K.J., Zanne, A.E., Zelený, D., Zhao, Y., Zheng, Jingming, Zheng, Ji, Ziemińska, K., Zirbel, C.R., Zizka, G., Zo-Bi, I.C., Zotz, G., Wirth, C., 2020. TRY plant trait database – enhanced coverage and open access. *Glob. Chang. Biol.* 26, 119–188. <https://doi.org/10.1111/gcb.14904>
- Laughlin, D.C., Gremer, J.R., Adler, P.B., Mitchell, R.M., Moore, M.M., 2020. The net effect of functional traits on fitness. *Trends Ecol. Evol.* 35, 1037–1047. <https://doi.org/10.1016/j.tree.2020.07.010>
- Lürrig, M.D., Donoughe, S., Svensson, E.I., Porto, A., Tsuboi, M., 2021. Computer vision, machine learning, and the promise of phenomics in ecology and evolutionary biology. *Front. Ecol. Evol.* 9, 642774. <https://doi.org/10.3389/fevo.2021.642774>
- Maitner, B., Gallagher, R., Svenning, J., Tietje, M., Wenk, E.H., Eiserhardt, W.L., 2023. A global assessment of the Raunkiærán shortfall in plants: geographic biases in our knowledge of plant traits. *New Phytologist* 240, 1345–1354. <https://doi.org/10.1111/nph.18999>
- McGill, B.J., Enquist, B.J., Weiher, E., Westoby, M., 2006. Rebuilding community ecology from functional traits. *Trends Ecol. Evol.* 21, 178–185. <https://doi.org/10.1016/j.tree.2006.02.002>
- Meyer, C., Weigelt, P., Kreft, H., 2016. Multidimensional biases, gaps and uncertainties in global plant occurrence information. *Ecol. Lett.* 19, 992–1006. <https://doi.org/10.1111/ELE.12624>
- Mouillot, D., Graham, N.A.J., Villéger, S., Mason, N.W.H., Bellwood, D.R., 2013. A functional approach reveals community responses to disturbances. *Trends Ecol. Evol.* 28, 167–177. <https://doi.org/10.1016/j.tree.2012.10.004>
- Moxon, S.A.T., Solbrig, H., Harris, N.L., Kalita, P., Miller, M.A., Patil, S., Schaper, K., Bizon, C., Caufield, J.H., Cuesta, S.C., Cox, C., Dekervel, F., Dooley, D.M., Duncan, W.D., Fliss, T., Gehrke, S., Graefe, A.S.L., Hegde, H., Ireland, A.J., Jacobsen, J.O.B., Krishnamurthy, M., Kroll, C., Linke, D., Ly, R., Matent-

- zoglu, N., Overton, J.A., Saunders, J.L., Unni, D.R., Vaidya, G., Vierdag, W.-M.A.M., LinkML Community Contributors, Bruskiewich, R.M., Carbon, S., Cavanna, E., Chandonia, J.-M., Cholia, S., Dichter, B., Eloë-Fadrosh, E.A., Emonet, V., Essaid, S., Fellows Yates, J.A., Flack, J., Ghosh, S.S., G.-G.D., Jarecka, D., Jiao, D., Joachimiak, M.P., Korolev, V., Lapkin, V., McLoughlin, N., Miller, S.D., Milton, M., Moore, J., Munoz-Torres, M., Nichols, B.N., Reese, J.T., Savage, V., Stroemert, P., Teoh, J., Thessen, A., To, I., Trivedi, P., Vialard, V., Whetzel, T., Ruebel, O., Chute, C.G., Brush, M.H., Haendel, M.A., Mungall, C.J., 2026. LinkML: an open data modeling framework. *Gigascience* 15, 1–13. <https://doi.org/10.1093/GIGASCIENCE/GIAF152>
- Nelson, G., Ellis, S., 2019. The history and impact of digitization and digital data mobilization on biodiversity research. *Philosophical Transactions of the Royal Society B* 374. <https://doi.org/10.1098/RSTB.2017.0391/42184>
- Pacifici, M., Foden, W.B., Visconti, P., Watson, J.E.M., Butchart, S.H.M., Kovacs, K.M., Scheffers, B.R., Hole, D.G., Martin, T.G., Akçakaya, H.R., Corlett, R.T., Huntley, B., Bickford, D., Carr, J.A., Hoffmann, A.A., Midgley, G.F., Pearce-Kelly, P., Pearson, R.G., Williams, S.E., Willis, S.G., Young, B., Rondinini, C., 2015. Assessing species vulnerability to climate change. *Nat. Clim. Chang.* 5, 215–224. <https://doi.org/10.1038/nclimate2448>
- Page, L.M., Macfadden, B.J., Fortes, J.A., Soltis, P.S., Riccardi, G., 2015. Digitization of Biodiversity Collections Reveals Biggest Data on Biodiversity. *Bio-science* 65, 841–842. <https://doi.org/10.1093/BIOSCI/BIV104>
- Pollock, L.J., Kitzes, J., Beery, S., Gaynor, K.M., Jarzyna, M.A., Mac Aodha, O., Meyer, B., Rolnick, D., Taylor, G.W., Tuia, D., Berger-Wolf, T., 2025. Harnessing artificial intelligence to fill global shortfalls in biodiversity knowledge. *Nature Reviews Biodiversity* 1, 166–182. <https://doi.org/10.1038/s44358-025-00022-3>
- Porto, A., Rolfe, S., Maga, A.M., 2021. ALPACA: A fast and accurate computer vision approach for automated landmarking of three-dimensional biological structures. *Methods Ecol. Evol.* 12, 2129–2144. <https://doi.org/10.1111/2041-210X.13689>
- PROV Model Primer [WWW Document], 2013. URL <https://www.w3.org/TR/prov-primer/> (accessed 5.24.26).
- Reichstein, M., Camps-Valls, G., Stevens, B., Jung, M., Denzler, J., Carvalhais, N., Prabhat, 2019. Deep learning and process understanding for data-driven Earth system science. *Nature* 566, 195–204. <https://doi.org/10.1038/s41586-019-0912-1>
- Rolfe, S., Pieper, S., Porto, A., Diamond, K., Winchester, J., Shan, S., Kirveslahti, H., Boyer, D., Summers, A., Maga, A.M., 2021. SlicerMorph: An open and extensible platform to retrieve, visualize and analyse 3D morphology. *Methods Ecol. Evol.* 12, 1816–1825. <https://doi.org/10.1111/2041-210X.13669>
- Roscher, R., Bohn, B., Duarte, M.F., Garcke, J., 2020. Explainable machine learning for scientific insights and discoveries. *IEEE Access* 8, 42200–42216. <https://doi.org/10.1109/ACCESS.2020.2976199>
- Soberón, J., Peterson, A.T., 2004. Biodiversity informatics: managing and applying primary biodiversity data. *Philosophical Transactions of the Royal Society B* 359, 689–698. <https://doi.org/10.1098/RSTB.2003.1439>
- Thessen, A., 2016. Adoption of machine learning techniques in ecology and earth science. *One Ecosystem* 1, e8621. <https://doi.org/10.3897/oneeco.1.e8621>
- Thessen, A.E., Cui, H., Mozzherin, D., 2012. Applications of natural language processing in biodiversity science. *Adv. Bioinformatics* 2012, 391574. <https://doi.org/10.1155/2012/391574>
- Tobias, J.A., Sheard, C., Pigot, A.L., Devenish, A.J.M., Yang, J., Sayol, F., Neate-Clegg, M.H.C., Alioravainen, N., Weeks, T.L., Barber, R.A., Walkden, P.A., MacGregor, H.E.A., Jones, S.E.I., Vincent, C., Phillips, A.G., Marples, N.M., Montaña-Centellas, F.A., Leandro-Silva, V., Claramunt, S., Darski, B., Freeman, B.G., Bregman, T.P., Cooney, C.R., Hughes, E.C., Capp, E.J.R., Varley, Z.K., Friedman, N.R., Korntheuer, H., Corrales-Vargas, A., Trisos, C.H., Weeks, B.C., Hanz, D.M., Töpfer, T., Bravo, G.A., Remeš, V., Nowak, L., Carneiro, L.S., Moncada R., A.J., Matysioková, B., Baldassarre, D.T., Martínez-Salinas, A., Wolfe, J.D., Chapman, P.M., Daly, B.G., Sorensen, M.C., Neu, A., Ford, M.A., Mayhew, R.J., Fabio Silveira, L., Kelly, D.J., Annorbah, N.N.D., Pollock, H.S., Grabowska-Zhang, A.M., McEntee, J.P., Carlos T. Gonzalez, J., Meneses, C.G., Muñoz, M.C., Powell, L.L., Jamie, G.A., Matthews, T.J., Johnson, O., Brito, G.R.R., Zyskowski, K., Crates, R., Harvey, M.G., Jurado Zevallos, M., Hosner, P.A., Bradfer-Lawrence, T., Maley, J.M., Stiles, F.G., Lima, H.S., Provost, K.L., Chibesa, M., Mashao, M., Howard, J.T., Mlamba, E., Chua, M.A.H., Li, B., Gómez, M.I., García, N.C., Päckert, M., Fuchs, J., Ali, J.R., Derryberry, E.P., Carlson, M.L., Urriza, R.C., Brzeski, K.E., Prawiradilaga, D.M., Rayner, M.J., Miller, E.T., Bowie, R.C.K., Lafontaine, R., Scofield, R.P.,

- Lou, Y., Somarathna, L., Lepage, D., Illif, M., Neuschulz, E.L., Templin, M., Dehling, D.M., Cooper, J.C., Pauwels, O.S.G., Analuddin, K., Fjeldsø, J., Seddon, N., Sweet, P.R., DeClerck, F.A.J., Naka, L.N., Brawn, J.D., Aleixo, A., Böhning-Gaese, K., Rahbek, C., Fritz, S.A., Thomas, G.H., Schleuning, M., 2022. AVONET: morphological, ecological and geographical data for all birds. *Ecol. Lett.* 25, 581–597. <https://doi.org/10.1111/ele.13898>
- Villéger, S., Mason, N.W.H., Mouillot, D., 2008. New multidimensional functional diversity indices for a multifaceted framework in functional ecology. *Ecology* 89, 2290–2301. <https://doi.org/10.1890/07-1206.1>
- Violle, C., Enquist, B.J., McGill, B.J., Jiang, L., Albert, C.H., Hulshof, C., Jung, V., Messier, J., 2012. The return of the variance: Intraspecific variability in community ecology. *Trends Ecol. Evol.* 27, 244–252. <https://doi.org/10.1016/j.tree.2011.11.014>
- Violle, C., Navas, M., Vile, D., Kazakou, E., Fortunel, C., Hummel, I., Garnier, E., 2007. Let the concept of trait be functional! *Oikos* 116, 882–892. <https://doi.org/10.1111/j.0030-1299.2007.15559.x>
- Walls, R.L., Deck, J., Guralnick, R., Baskauf, S., Beaman, R., Blum, S., Bowers, S., Buttigieg, P.L., Davies, N., Endresen, D., Gandolfo, M.A., Hanner, R., Janning, A., Krishtalka, L., Matsunaga, A., Midford, P., Morrison, N., Tuama, É.Ó., Schildhauer, M., Smith, B., Stucky, B.J., Thomer, A., Wieczorek, J., Whitacre, J., Wooley, J., 2014. Semantics in support of biodiversity knowledge discovery: An introduction to the biological collections ontology and related ontologies. *PLoS One* 9, e89606. <https://doi.org/10.1371/JOURNAL.PONE.0089606>
- Webster, M.S., 2017. The Extended Specimen, The Extended Specimen: Emerging Frontiers in Collections-Based Ornithological Research. CRC Press. <https://doi.org/10.1201/9781315120454>
- Wieczorek, J., Bloom, D., Guralnick, R., Blum, S., Döring, M., Giovanni, R., Robertson, T., Vieglais, D., 2012. Darwin Core: An evolving community-developed biodiversity data standard. *PLoS One* 7, e29715. <https://doi.org/10.1371/JOURNAL.PONE.0029715>
- Wilkinson, M.D., Dumontier, M., Aalbersberg, I.J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., da Silva Santos, L.B., Bourne, P.E., Bouwman, J., Brookes, A.J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C.T., Finkers, R., Gonzalez-Beltran, A., Gray, A.J.G., Groth, P., Goble, C., Grethe, J.S., Heringa, J., 't Hoen, P.A.C., Hooft, R., Kuhn, T., Kok, R., Kok, J., Lusher, S.J., Martone, M.E., Mons, A., Packer, A.L., Persson, B., Rocca-Serra, P., Roos, M., van Schaik, R., Sansone, S.-A., Schultes, E., Sengstag, T., Slater, T., Strawn, G., Swertz, M.A., Thompson, M., van der Lei, J., van Mulligen, E., Velterop, J., Waagmeester, A., Wittenburg, P., Wolstencroft, K., Zhao, J., Mons, B., 2016. The FAIR guiding principles for scientific data management and stewardship. *Sci. Data* 3, 160018. <https://doi.org/10.1038/sdata.2016.18>
- Wilman, H., Belmaker, J., Simpson, J., de la Rosa, C., Rivadeneira, M.M., Jetz, W., 2014.
- Younis, S., Weiland, C., Hoehndorf, R., Dressler, S., Hickler, T., Seeger, B., Schmidt, M., 2018. Taxon and trait recognition from digitized herbarium specimens using deep convolutional neural networks. *Bot. Lett.* 165, 377–383. <https://doi.org/10.1080/23818107.2018.1446357>