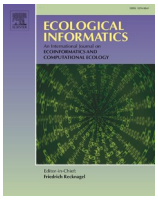
Ecological Informatics 96 (2026) 103694



Contents lists available at ScienceDirect

Ecological Informatics

journal homepage: www.elsevier.com/locate/ecolinf



Natural history societies as stewards of biodiversity data: Lessons from arachnology for improving quality, access and trust

Alexander Bach [a,*], Martina Roß-Nickoll [a], Helge Bruelheide [b,c], Aletta Bonn [c,d,e], Michael Hohner [f], Christoph Muster [g], Stefan Seifert [h], Florian Raub [i], Hubert Höfer [i]

[a] Institute for Environmental Research, RWTH Aachen University, 52074 Aachen, Germany
[b] Institute of Biology/Geobotany and Botanical Garden, Martin Luther University Halle-Wittenberg, 06108 Halle (Saale), Germany
[c] German Centre for Integrative Biodiversity Research (iDiv) Halle-Jena-Leipzig, 04103 Leipzig, Germany
[d] Helmholtz-Centre for Environmental Research (UFZ), Department Biodiversity and People, 04318 Leipzig, Germany
[e] Friedrich-Schiller-University Jena, Institute of Biodiversity, Ecology and Evolution, 07743 Jena, Germany
[f] 90425 Nürnberg, Germany
[g] Zoological Institute and Museum, University of Greifswald, 17489 Greifswald, Germany
[h] Staatliche Naturwissenschaftliche Sammlungen Bayerns, SNSB IT Center, 80638 Munich, Germany
[i] State Museum of Natural History Karlsruhe, Department Biosciences, 76133 Karlsruhe, Germany

ARTICLE INFO

Keywords:
Biodiversity
Database
Ecology
Big data

ABSTRACT

Scientific societies, particularly taxon-focused natural history societies, play a crucial role in the long-term curation, mobilization, and governance of biodiversity data. They provide both the necessary expertise and social continuity to ensure data quality, foster trust, and build community cohesion. In this paper, we highlight the contributions of a natural history society for arachnids in Central Europe, the Arachnologische Gesellschaft, by describing two complementary biodiversity data infrastructures it coordinates: the Atlas of the European Arachnids, a curated occurrence-based mapping portal, and ARAMOB, a curated research data repository for ecological community datasets. We outline how these portals were developed in close collaboration with experts, museums, and research institutions, and how they are technically linked to or embedded in national and international data infrastructures such as World Spider Catalog, NFDI4Biodiversity and GBIF. Through this case study, we show how scientific societies can function as trusted intermediaries between data producers, aggregators, and users ensuring the visibility, interpretability, and interoperability of biodiversity data. We conclude with recommendations for strengthening similar community-rooted, taxon-specific infrastructures for other organism groups.

1. Introduction

Biodiversity research depends on the mobilization, curation and integration of heterogeneous data to detect and interpret changes in species distributions, abundances and traits (Chowdhury et al., 2025; Hochkirch et al., 2021; Vargas et al., 2024). Over the past decade, global data infrastructures like the Global Biodiversity Information Facility (GBIF), have enabled access to occurrence data contributed by museums, monitoring programs, citizen scientists (Beck et al., 2013; Telenius, 2011). At the national level, initiatives such as Germany's National Research Data Infrastructure (NFDI), in particular NFDI4Biodiversity, provide the legal, technical and organizational scaffolding (e.g., persistent identifiers, metadata vocabularies, repository services) that make domain repositories interoperable and sustainable (Güntsch et al., 2024). Yet broad accessibility alone does not guarantee interpretability or fitness for purpose. For robust ecological and conservation analyses, knowing which species occurred where and when is rarely sufficient and in depth understanding of the taxa and their ecology is needed (Hochkirch et al., 2021; Ivanova and Shashkov, 2021; König et al., 2019; Löffler et al., 2021). Importantly, sound analyses of biodiversity change requires authoritative taxonomic concepts, transparent data provenance and consistent sampling metadata sustained by enduring expertise, formalized curation and stable communities of experts (Barbosa and Alves-Souza, 2025; Bayraktarov et al., 2019; Franz and Sterner, 2018; Gabrielsen, 2020; Hortal et al., 2007; Michener et al., 1997; Wieczorek et al., 2012). Taxon-focused scientific societies are uniquely positioned

* Corresponding author.
E-mail address: alexander.bach@ifer.rwth-aachen.de (A. Bach).

https://doi.org/10.1016/j.ecoinf.2026.103694
Received 19 September 2025; Received in revised form 4 March 2026; Accepted 4 March 2026
Available online 6 March 2026
1574-9541/© 2026 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

to address these needs (Corrales et al., 2025; Gardiner and Roy, 2022; Luke et al., 2023). Societies foster and retain specialized knowledge across generations of experts, something which time-limited projects cannot provide (Brar, 2024). They can maintain consensus taxonomies, curate nomenclatural changes, and cultivate a culture of review, mentoring, and stewardship. When a society undertakes data governance, it can act as a trusted intermediary between data producers, aggregators, and users: translating taxonomic updates into stable identifiers, brokering quality control and harmonizing requirements across research, conservation and policy (Franz and Sterner, 2018). It also aligns with national infrastructure services (e.g., NFDI4Biodiversity) to ensure FAIR (Stall et al., 2019) compliance and long-term operability. This role becomes especially powerful when societies build or coordinate domain-specific data infrastructures that complement global platforms rather than duplicate them by providing more depth where it matters (e.g., sampling metadata, habitat context, trait and life-history information). Arachnologists for example are well organized internationally and operate within a strong community framework. At the global level, the International Society of Arachnology brings together researchers from all continents. Regionally, a variety of national or linguistic societies such as the British Arachnological Society, the American Arachnological Society, and the Arachnologische Gesellschaft (AraGes) in German-speaking countries form dense expert networks that promote long-term collaboration. A central pillar of this global taxonomic infrastructure is the World Spider Catalog (WSC), which is self-organized and continuously updated by the community under the stewardship of the Natural History Museum Bern (Harms et al., 2022; Nentwig et al., 2015). The World Spider Catalog functions as the authoritative taxonomic reference for spiders and is universally adopted by arachnologists worldwide. It serves as a backbone for international databases such as GBIF, already demonstrating how expert-led initiatives can provide both nomenclatural stability and wide recognition.

In this paper, we use the Arachnologische Gesellschaft e.V. (AraGes), the expert community for German-speaking countries, to examine how a natural history society can structure and sustain such services and infrastructures. AraGes coordinates two complementary infrastructures. The Atlas of the European Arachnids (AraGes Atlas), a curated occurrence portal that consolidates georeferenced records under society-backed taxonomic control. The ARAMOB repository targets a distinct and crucial layer of biodiversity information: systematically collected assemblage data, where sampling effort, temporal coverage and habitat attributes are explicitly documented, so called aggregated biodiversity data (Grames et al., 2022; König et al., 2019). While other arachnological societies also maintain national recording schemes or atlases that primarily mobilize and curate occurrence records for distribution mapping, ARAMOB is designed to provide curated, analysis-ready community datasets that include standardized abundance information and sampling methodology at the study level. Together, these infrastructures address both the breadth and depth of arachnological data needs with occurrence-level discoverability for biogeographic questions and analysis-ready community datasets for ecological inference.

To our knowledge, this combination of society-led curation and interoperable community data products is rarely available in comparable scope and consistency within arachnological infrastructures elsewhere.

We examine how a taxon-focused scientific society can maintain domain-specific biodiversity infrastructures through a shared governance model involving expert curation and institutional hosting. Using the AraGes Atlas and ARAMOB as case studies, we analyze their underlying data structures, ingestion and validation workflows, and the division of responsibilities between AraGes and a federal institution. We further assess the empirical scale and demonstrate community uptake of both infrastructures, including a quantitative comparison with major public biodiversity platforms such as iNaturalist and Observation.org, and identify structural elements that may inform the development of comparable repositories for other organism groups.

As we make the case for the role of professional natural history societies, working in collaboration with museum collections and research organizations, using spiders as our exemplar in this manuscript, we also acknowledge that many other successful biodiversity data infrastructures are likewise built on public-professional partnerships. Examples include Reflora in Brazil for plant data (Pinheiro et al., 2024), SERNEC (2026) for natural history collections in the southeastern United States, and supranational European initiatives such as the European Butterfly Monitoring Scheme (Van Swaay et al., 2008).

2. Methods

2.1. Case description

2.1.1. Atlas of the European Arachnids

The Atlas of the European Arachnids succeeded the “Nachweiskarten der Spinnentiere Deutschlands” which were curated by Aloysius Staudt on spiderling.de until 2015. With the transition to https://atlas.arages.de/, responsibility moved to Michael Hohner, the platform was institutionally embedded within the Arachnologische Gesellschaft (AraGes), and its scope expanded from a Germany-focused mapping initiative to a Europe-wide, continuously updated occurrence portal for arachnids (excluding mites). Data enters the AraGes Atlas through curated imports from expert networks, museum/specimen digitization, literature extraction and citizen science observations mapped to controlled fields (Fig. 1). The AraGes Atlas is operated as a relational SQL-based information system, in which occurrence records and their associated metadata are stored in normalized tables linked through stable identifiers. Data ingestion follows a batching workflow in which validated species records are imported into the central SQL database and indexed for efficient retrieval via the public web interface. The portal provides structured access to these SQL-backed tables through server-side queries, enabling dynamic generation of species pages, distribution maps and basic ecological summary statistics. These include habitat associations, such as the frequency of records in major EUNIS habitat categories, as well as information on altitudinal ranges, phenological patterns based on observed activity periods, and method-based sampling frequencies.

2.1.2. ARAMOB: a repository for systematically collected spider community data

ARAMOB (https://aramob.de/en) is a repository for systematically sampled community datasets. The responsibility is shared between the AraGes and the State Museum of Natural History Karlsruhe. The museum provides the technical operation administration and maintenance of the public ARAMOB portal. The AraGes is responsible for scientific curation, including the coordination of expert review, taxonomic validation of submitted datasets and oversight of data ingestion workflows. The repository uses the Diversity Workbench database framework (Triebel et al., 1999), a set of open source components for building and managing biodiversity information which facilitates data curation, standardization, and accessibility (Karam et al., 2016). Ingestion pathways accept standardized campaign exports mapped to a documented schema (supplementary file S1); validation combines automated conformance checks (schema, vocabularies, internal consistency across method/effort fields) with editorial review. This approach minimizes heterogeneity and enhances data quality by enforcing strict quality criteria and data standards. ARAMOB offers an interactive web portal that allows researchers, citizen scientists, and conservation practitioners to explore the data and metadata, and to download them for own analyses. In addition, the repository supports the development of interactive explorative data analysis tools, such as ARAapp (Bach et al., 2024), which allows researchers to dynamically explore ecological patterns (e.g., habitat preferences, vertical distribution, niche positions, phenology, companion species) within spider assemblages.

The repository is configured to distribute its curated data packages

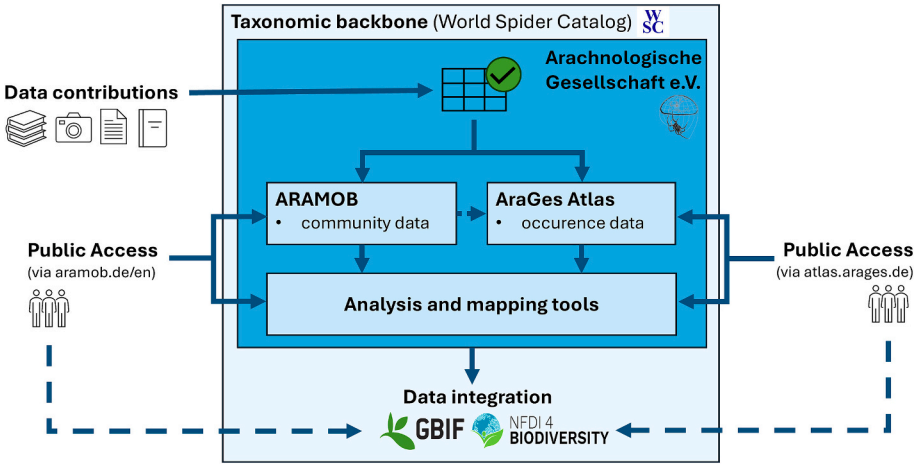


Fig. 1. Schematic overview of the data flow within the arachnological data infrastructures curated by the AraGes. Solid arrows represent the flow from data contribution to validation and subsequent assignment to one of the two infrastructures, depending on data type: ecological community data that meet defined minimum requirements are curated within ARAMOB, while species-level occurrence data are integrated into the Atlas. Both platforms feature dedicated internal tools for analysis, visualization, and mapping, which are directly powered by the curated datasets. These tools, along with the underlying data, are publicly accessible. ARAMOB publishes datasets to GBIF, and both platforms are structurally integrated into NFDI4Biodiversity which enables further public access (dashed arrows).

through established BioCASe pipelines, making the data packages accessible via major platforms such as the Global Biodiversity Information Facility (GBIF) and the NFDI4Biodiversity infrastructure, thereby contributing to international and national databases (Güntsch et al., 2024, see also Fig. 1). To support the further development and visibility of ARAMOB, the publication of dedicated data papers (e.g. Bauer et al., 2024) is actively encouraged in the open-access journal Arachnologische Mitteilungen/Arachnology Letters (Raub et al., 2023). These papers are directly linked to the corresponding datasets in the ARAMOB repository and accessible via its web portal.

2.2. Data quality control and minimal data requirements

Quality assurance follows a two-stage process that combines automated validation with expert review. On ingestion, records are mapped to controlled fields and checked for schema conformance, internal consistency and geospatial/temporal plausibility (e.g., valid coordinates and dates, method-effort rules such as positive trap numbers and time span when the pitfall trap method is specified). Taxonomic names are resolved through the World Spider Catalog API; unresolved or ambiguous names are flagged for curation. After automated checks, a plausibility assessment is conducted by community experts responsible for the respective regions, who evaluate suspicious records (e.g., outliers, unusual phenology, range extensions). In both infrastructures, records are anchored to a collection event, but the logical central unit differs between portals. The AraGes Atlas is implemented as a streamlined occurrence database in which the primary record represents an individual taxon observation. The minimal dataset here is a row defined by an accepted species identification, a single event date, geographic coordinates and a source; additional metadata (e.g. observer, remarks, determination method) can be attached but are not required for ingestion. Conceptually, event information is therefore flattened into a minimal occurrence payload that is optimized for discoverability and large-scale distribution mapping.

ARAMOB, by contrast, uses an explicitly event-centered, normalized schema based on the DiversityCollection module of Diversity Workbench. Here, the collection event is the primary entity: it stores the temporal extent of sampling (start and end date, time span), the sampling design (method and effort, such as trap type and number) and plot level context (geographic position and habitat class, e.g. EUNIS). Species information is not stored as independent occurrences but as child records attached to a given collection event via stable identifiers, each row combining a taxon identification with a non-negative integer count. This event-centric structure ensures that community composition, effort and habitat descriptors remain tightly coupled, enabling abundance-based indicators and comparative analyses across studies while still allowing richer, optional metadata to be layered on top when available (Table 1).

2.3. Citation analysis and comparative assessment of dataset scale

To quantify evidence of use, we compiled citing items for AraGes Atlas and ARAMOB and applied a temporal cutoff on 31 Dec 2024. For the AraGes Atlas, which lacks a persistent identifier, we used Harzing (2007) with the exact strings “atlas.arages.de” or “spiderling.de” in google scholar and repeated the same strings with filetype:pdf on google web. The two result sets were merged and de-duplicated. Inclusion required an explicit reference to the AraGes Atlas URL or portal name in the text, captions, or references. For ARAMOB, citing items were harvested on the same date from the GBIF dataset pages of ARAMOB-published packages (Höfer, 2025a, 2025b; Holstein, 2025), aggregated across datasets, and de-duplicated using the same rules. In both cases, only items published through 2024 were retained. These procedures likely yield conservative lower bounds because URL-based discovery without a PID and uneven index coverage can miss uses, and GBIF citation lists depend on adherence to recommended dataset citation. Curated bibliographies are provided in supplementary file S2.

To contextualize the size and taxonomic coverage of AraGes Atlas and ARAMOB, we extracted comparable metrics from two widely used public biodiversity platforms (iNaturalist, 2025; Observation

Table 1
Brief overview of minimally required data fields in AraGes Atlas and ARAMOB.

Category	Data field	AraGes Atlas	ARAMOB
Collection Event	Start date		✓
	(End) date	✓	✓
	Sampling method		✓
	Method parameters (trap type, number of traps, collecting fluid)		✓
	Source	✓	✓
Sampling Plot	Longitude	✓	✓
	Latitude	✓	✓
	EUNIS habitat		✓
Species	Identification	✓	✓
	Individual counts		✓

International, 2025); while we recognize differences in mission and curation, there are few other freely accessible datasets of comparable scale. For iNaturalist, we used the platform's search interface on 26 November 2025 and restricted results to verifiable observations of Araneae from Germany that were identified to species level. Species richness and total observation counts were taken directly from the summary counters provided by the search results page. For Observation.org, we downloaded the corresponding GBIF occurrence dataset on the same date and filtered it to German records of Araneae identified to species level. Species numbers and total individual counts were derived from this filtered dataset. The resulting metrics were summarized together with the AraGes Atlas and ARAMOB data to enable a direct comparison of species coverage and record volume across the four portals. We also compared the number of species recorded in each portal to the known 1014 spider species in Germany (Nentwig et al., 2025). This yielded a simple recognition index indicating the proportion of the national fauna represented by each platform (Alfeus et al., 2025). Values above 100% occur for the AraGes Atlas because it includes species documented as non-established introductions, which are not counted among the established native species.

3. Results

As of 26 November 2025, the Atlas of the European Arachnids contains 1089 spider species and 4,200,243 individual records. The temporal coverage extends back almost 250 years to 1757, although records prior to 1950 constitute less than 0.6% of the dataset. Most observations originate from Germany, while the AraGes Atlas includes at least single records from most other European countries.

As of 26 November 2025, ARAMOB provides community data for 1361 EUNIS habitat classified sampling sites and includes 745 species from over 500,000 individuals. The dataset is overwhelmingly method-consistent, with 98% of communities originating from pitfall-trap sampling. All specimens were identified to species level based on genital morphology.

To contextualize the magnitude of these society-led infrastructures, we extracted equivalent metrics from two widely used public biodiversity platforms (Table 2). Using consistent filtering rules, iNaturalist contained 535 verifiable German spider species and 112,898 observations, while Observation.org provided 238 species and 72,427 records. These numbers indicate that the AraGes Atlas covers more than three times as many species and over an order of magnitude more records than either citizen-science portal under comparable constraints.

Our citation search retrieved 396 citing items for the AraGes Atlas and 186 for ARAMOB published up to 2024 (Fig. 2). These include peer-reviewed journal articles, theses, institutional reports and policy documents, demonstrating that both infrastructures are used across research, conservation and applied contexts. In many cases, the portals provide the primary source of up-to-date distribution maps or analysis-ready community data for spiders, particularly in Central Europe. Together, these metrics provide quantitative evidence that the AraGes infrastructures are both substantial in content and actively used by a wider community.

Table 2
Comparative metrics for German spiders (Arachnida: Araneae) from four biodiversity data portals, summarizing species counts, total individuals/observations, and a completeness index calculated relative to the 1014 established German spider species.

Portal	Species (Araneae)	Individuals/ observations	Species completeness
AraGes Atlas	1089	4,200,243	> 100%
ARAMOB	745	500,316	73.47%
iNaturalist	535	112,898	52.76%
Observation.org	238	72,427	23.47%

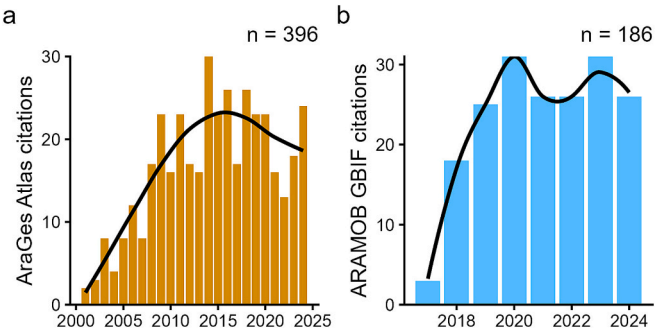


Fig. 2. Evidence of use for the two arachnological infrastructures AraGes Atlas (a) and ARAMOB (b), measured as the number of citing documents published up to 31 December 2024.

4. Discussion

Our results reveal a clear structural distinction between the two society-led infrastructures and the major public biodiversity platforms examined. The AraGes Atlas now provides more than 4.6 million occurrence records representing 1673 species, supported predominantly by expert-based genital morphological identifications, while ARAMOB contributes over 500,000 individuals across 1361 systematically sampled sites, each with complete sampling metadata. Together, these metrics demonstrate that the two portals deliver both domain breadth, through comprehensive georeferenced distribution data, and domain depth through event-centered, methodologically consistent community datasets. The AraGes has collected additional community datasets comprising more than 2.6 million individuals from 1500 plots across all regions of Germany for the upcoming trend analyses of the new Red List of German spiders (Muster et al., in prep.). These datasets have been integrated into Diversity Workbench following the procedures outlined above and will be made publicly accessible after publication.

The quantitative comparison with iNaturalist and Observation.org further underscores these structural differences. Under equivalent filtering rules, iNaturalist contained 535 verifiable German spider species and 112,898 observations, while Observation.org included 238 species and 72,427 observations. Thus, the AraGes Atlas currently covers more than three times as many species and more than an order of magnitude more individual records than the two largest public platforms combined. Importantly, these public portals rely predominantly on habitus-based identifications, as most submissions are photographic and therefore lack access to the genital structures required for reliable species-level resolution in the majority of spider taxa. As a result, their taxonomic coverage and internal consistency cannot match the rigor of expert-based morphological identifications that underpin the AraGes Atlas and ARAMOB. For downstream users, this level of prior expert validation translates into substantial time savings, as many labor-intensive steps, such as taxonomic cross-checks, plausibility controls or the removal of doubtful records, are already resolved at the point of data ingestion (Ball-Damerow et al., 2019).

These strengths reflect the society-centered model on which both infrastructures are based. The separation between AraGes Atlas (occurrence) and ARAMOB (community) delineates data granularity and metadata requirements: ARAMOB provides event-level design and effort for community inference, and AraGes Atlas supplies streamlined occurrence evidence that, when analyzed with appropriate statistical frameworks such as occupancy-detection models, can reveal broad-scale trend patterns (Bowler et al., 2021; Isaac et al., 2014; Outhwaite et al., 2019). By encoding method, effort, habitat class and plot context, they enable abundance-based indicators, alpha/beta diversity and turnover, and trait-weighted metrics, and they reduce bias relative to presence-only evidence (Engemann et al., 2015; Meyer et al., 2016; Rondinini et al., 2006). More generally, the resolution literature shows that

systematic use of aggregated data yields more complete and less biased baselines, and can supply priors to improve downstream modelling and to audit gaps in disaggregated resources (Barbosa and Alves-Souza, 2025; Jetz et al., 2012; König et al., 2019), precisely the role ARAMOB is designed to play alongside AraGes Atlas. Given the considerable heterogeneity inherent in biodiversity data collection methodologies, metadata is crucial for enabling the spatial and temporal standardization of datasets across studies, thereby enhancing the comparability and reliability of ecological analyses (Saska et al., 2021).

Framed against the Essential Biodiversity Variables (EBV) concept (Pereira et al., 2013), the two portals fill complementary slots: AraGes Atlas can primarily support Species distribution-type EBVs (locations, temporal signal, and trend directions), while ARAMOB supports abundance-based species populations EBVs and community composition EBVs (Jetz et al., 2019; Kissling et al., 2018; Proença et al., 2017).

Keeping pure (co-) occurrence databases separate from those for repeated sampling events is a common practice also in vegetation science. While the European Vegetation Archive hosts all types of vegetation descriptions based on species composition (Chytrý et al., 2016), ReSurveyEurope has a particular structure that not only describes the repeated sampling on a particular plot but also issues such as relocation inaccuracies (Knollová et al., 2024). At the same time, URL-based discoverability, heterogeneous citation practices, and incomplete logging mean that measured uptake is a lower bound; sustaining expert review across regions and generations remains a governance challenge (Coleman, 2015; Fontaine et al., 2021; Luke et al., 2023; Packer et al., 2018). Taken together, the evidence supports a view in which taxon societies act not as data silos but as domain stewards that complement global aggregators with depth where it matters: taxonomy, event context, and quality control. Alignment with external infrastructures (authoritative backbones, GBIF, NFDI services) turns curation work into reusable, citable assets and reduces transaction costs for downstream users.

4.1. Recommendations for future research data management

4.1.1. Leverage existing, proven data infrastructures

In recent years, various systems and software solutions have emerged to facilitate the efficient management of biodiversity data (Gadelha Jr et al., 2021; Hackett et al., 2019), which are now hosted by many natural history societies or other institutions (Brosens et al., 2013; Sanderson et al., 2021; Socher et al., 2014; Undap et al., 2019). These solutions streamline the workflow from field data collection to data users and employ standardized terminologies to enhance the interoperability of research and associated metadata with other database systems (Karam et al., 2016). Thus, creating an entirely new database from scratch is often unnecessary and highly labor-intensive. Instead, researchers should capitalize on established, reliable data management systems. Many of these infrastructures provide user-friendly APIs and integration interfaces with other prominent databases, such as GBIF, thereby enabling automated data dissemination and synchronization of datasets. Leveraging these proven infrastructures not only saves time and resources but also mitigates the risks associated with developing new systems, seeking overall sustainable funding for the development.

A further essential, yet often underestimated aspect of biodiversity data integration is taxonomic harmonization (Feng et al., 2022; Gadelha Jr et al., 2021; Gemeinholzer et al., 2020; Sandall et al., 2023). To ensure consistency and interoperability across datasets, it is crucial to align taxon names with authoritative taxonomic backbones. Rather than developing isolated or project-specific classifications, researchers should rely on well-established and widely accepted resources. A comprehensive overview of existing taxonomic databases and best-practice examples has recently been provided by Grenié et al. (2023). For species occurring in Germany regionalized and domain-specific taxon lists are offered via several open and freely accessible services with a dynamic data flow offered by a DWB REST Webservice for Taxon Lists as part of the Diversity Workbench services network (Seifert et al., 2015). It is delivering basic information on taxon names in use, synonyms, Red List classification and German vernacular names of a number of groups of animals, fungi and plants. The taxon list of Araneae, based on the World Spider Catalog, was compiled and is curated by the AraGes and is made visible in the ARAMOB portal.

4.1.2. Tailor your data management approach – one size does not fit all

Establishing two separate data environments for heterogeneous occurrence data and for systematic assemblage and time series data turned out to be a practical approach for managing biodiversity data based on their origin: a comprehensive repository versus a quality-for-use ensured subset. The AraGes Atlas is designed as an all-inclusive repository. It integrates every record available, whether opportunistically collected or methodically sampled. In contrast, ARAMOB represents a refined subset of the AraGes Atlas data. Here, only data derived from systematic sampling events accompanied by standardized metadata are incorporated to overcome data resolution problems (König et al., 2019). In other words, instead of privileging one data type over another, the AraGes Atlas hosts all records, while ARAMOB focuses on those meeting clearly defined criteria. Regarding metadata and data standardization, the implementation of the Diversity Workbench database framework for arachnological data provides a robust foundation for capturing data complexity using standardized ontologies and terminologies, ensuring interoperability for global data synthesis. Determining the necessary metadata also highlights the importance of collaborating closely with experts in the field to identify the relevant data based on the method and species group. Implementing a hierarchical quality structure could also be useful in allowing for the inclusion of all data but prioritizing their quality by separating them into different levels. Here, the metrics developed by Veiga et al. (2017) could be used to support scientists with easy-to-use indicators when selecting their data.

4.1.3. Establish automated data pipelines

Investing in the development of robust data pipelines is essential for modern biodiversity data management. Internal data pipelines can automatically extract, clean, and transform raw structured biodiversity data into analysis-ready formats (Mathew et al., 2014). To facilitate this process, easy-to-use quality criteria must be developed to assist users in assessing data quality. Data quality assurance encompasses the establishment of data standards and minimum requirements, including relevant metadata (Löffler et al., 2021; Michener et al., 1997), to facilitate cross-study analysis. This requires further methodological standardization approaches, such as those of Saska et al. (2021) for pitfall trap data, to establish criteria allowing cross-study analysis of systematical collected data. ARAMOB's use of Diversity Workbench exemplifies how automated workflows can minimize manual errors and reduce the time required for data curation. Internal SQL-query pipelines are used for error checking, but also for data inspection and analyses. On the other hand, external pipelines are designed to automatically export curated datasets to global data infrastructures such as GBIF and NFDI services (Novotný et al., 2022).

This ensures that high-quality data reaches a broad audience and is continuously updated, supporting timely research and policy-making decisions.

4.1.4. Institutional partnerships

Establishing long-term institutional partnerships among academic institutions, museums and research organizations to share responsibility for data curation and storage is indispensable. Importantly, the effective technical provision of these repositories requires dedicated institutions with appropriate budgets (Costello et al., 2014; Hochkirch et al., 2021). It is important to recognize, that successful database infrastructure governance can take on different forms, where the lead can be taken by natural history societies, museum collections or research institutes (Eversole et al., 2019; Kühl et al., 2020; Tewksbury et al., 2014). In most

cases, professional societies play a key role with their in-depth taxon expertise and volunteer base, as, at least in Europe, most species records originate not from academic collections, and these cannot be curated by academic scientists or museum curators alone, either. However, as we make the point here, while the role of natural history societies is crucial, it is often overlooked or taken for granted. We argue that given the accelerating species loss as well as loss of species knowledge, their role needs to be strengthened and supported also by funding organizations or institutional partnerships (Costello et al., 2014; Feng et al., 2022). In our case, responsibilities are deliberately divided: the AraGes acts as the domain-specific authority, coordinating species experts who provide taxonomic validation, review incoming datasets and ensure the scientific quality of the curated records. The State Museum of Natural History Karlsruhe, as a museum and permanently funded institution, provides the technical and administrative backbone, including server operation, database maintenance and administration. In this context, explicitly addressing the data's hosting location, ideally within broader infrastructures such as the NFDI4Biodiversity, further underlines the commitment to sustainable and transparent data management. Importantly, these partnerships do not only serve data integrity. They also foster social cohesion within scientific communities and provide platforms for training and mentoring the next generation of experts, skills that are increasingly scarce yet vital for biodiversity research (Páll-Gergely et al., 2024).

4.1.5. User-friendly data portals: simplicity and intuitive use as core principles

User-friendly data portals connected to natural history societies and existing communities are essential in lowering barriers to data sharing and promoting widespread adoption among researchers and citizen scientists alike (Boyle, 2013; Enke et al., 2012). Our web portals have been designed with simplicity and intuitive use as core principles. Consequently, to encourage data sharing among researchers, the submission process must be as simple and intuitive as possible. Time efficiency is crucial and one of the main reasons why researchers often refrain from sharing their data (Tenopir et al., 2011). Conversely, data access solutions must also be designed for ease of use.

Moreover, portal operators must maintain close ties with user communities to ensure that these platforms continuously meet evolving research and community needs (Sterner et al., 2020). By actively engaging with networks and natural history societies such as AraGes and collaborating with research institutions, portal operators can gain critical user feedback and foster a necessary environment of mutual trust and cooperation (Gabrielsen, 2020). This close relationship not only drives further innovation in portal functionalities but also strengthens the overall data-sharing ecosystem, ultimately maximizing the utility and adoption of our biodiversity data platforms.

5. Conclusions

The case of the arachnological data portals illustrates how taxon-specific infrastructures can effectively combine expert-driven data curation with user-oriented access and analysis tools. Importantly, the breadth and depth documented here are enabled by sustained, society-based governance that ensures continuous expert stewardship over time. AraGes provides the long-term expert network and mandate to a) maintain consensus taxonomic concepts and identification standards, b) coordinate regional specialists for plausibility review, and c) enforce community-agreed minimum metadata requirements for analysis-ready assemblage data. In the German-speaking region, there is currently no alternative, comparably curated spider database for research-grade ecological analyses in Germany outside the AraGes-led infrastructures. This reflects the absence of an institution other than the professional society that continuously provides coordinated expert curation, standard setting, and long-term editorial responsibility. Rooted in long-standing scientific natural history societies, both portals provide both reliable technical frameworks, and also foster identity, continuity, and trust within their respective expert communities. Looking beyond arachnology, similar approaches can be developed for other taxonomic groups. In fact, several initiatives have already adopted this approach (Engel et al., 2023). These efforts show that taxon-focused portals are particularly well suited to ensure high-quality data entry and long-term curation because they operate within close-knit networks of taxonomic specialists who possess the necessary domain knowledge. Furthermore, these platforms serve as social and professional hubs, reinforcing community cohesion and motivating participation through recognition and peer engagement. To realize their full potential, such portals should be closely linked to national and international biodiversity data infrastructures such as NFDI4Biodiversity or GBIF. This integration enables broader data discoverability and facilitates cross-taxon analyses, allowing researchers to study ecological interactions, co-occurrence patterns, and large-scale biodiversity trends across taxonomic boundaries.

Yet despite their crucial role, the contributions of natural history societies are often taken for granted. Beyond research, the society-curated nature of these infrastructures is directly relevant for biodiversity management: expert-validated distribution evidence and standardized community datasets constitute core inputs for regional and national Red List assessments and for regulatory reporting workflows that require traceable taxonomic concepts and documented sampling context.

Maintaining cost-free access to high-quality biodiversity data, as practiced by many taxon-focused communities, requires sustained institutional backing and recognition since many of these societies operate without dedicated staff or long-term financial support.

Ultimately, taxon-specific data portals represent a critical layer in the global biodiversity data architecture. They ensure that biodiversity data is not only available, but also curated, meaningful, and trustworthy. Strengthening these community-rooted infrastructures across taxonomic domains is essential for advancing biodiversity science and evidence-based conservation in the decades to come.

CRediT authorship contribution statement

Alexander Bach: Writing – original draft, Visualization, Methodology, Conceptualization. **Martina Roß-Nickoll:** Writing – review & editing, Project administration, Methodology, Investigation, Conceptualization. **Helge Bruelheide:** Writing – review & editing, Methodology, Conceptualization. **Aletta Bonn:** Writing – review & editing, Methodology, Conceptualization. **Michael Hohner:** Writing – review & editing, Software, Methodology. **Christoph Muster:** Writing – review & editing, Resources, Methodology, Conceptualization. **Stefan Seifert:** Writing – review & editing, Software, Conceptualization. **Florian Raub:** Writing – original draft, Visualization, Software, Methodology, Investigation, Data curation, Conceptualization. **Hubert Höfer:** Writing – original draft, Resources, Project administration, Methodology, Funding acquisition, Data curation, Conceptualization.

Declaration of generative AI and AI-assisted technologies in the writing process

During the preparation of this work the authors used ChatGPT in order to improve the readability and language of the manuscript. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the content of the published article.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

We sincerely thank everyone who has contributed observations to the Atlas of European Arachnids and all researchers and institutions who have shared curated datasets with the ARAMOB repository. We are especially grateful to the many volunteers and community members who dedicate their time and expertise to advancing arachnological knowledge and biodiversity data quality through their continued engagement. Our thanks also go to the editors of the Arachnologische Mitteilungen, especially Tobias Bauer, for their long-standing commitment to maintaining a platform for high-quality arachnological publishing and knowledge exchange. Their collective contributions form the foundation of the infrastructures presented in this paper. The ARAMOB research project was initially supported by the German Research Foundation (DFG: 16372061). The AraGes use case is supported within the project "Establishment of the National Research Data Infrastructure (NFDI)" in the consortium NFDI4Biodiversity (DFG: 442032008).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ecoinf.2026.103694.

Data availability

All data used in this study are provided as Supplementary Material. In addition, the complete dataset together with the full analysis code required to reproduce the figure is publicly available in a project GitHub repository: https://github.com/alexander-bach/stewarding-biodiversity-data.

References

Alfeus, M., Irish, J., Birkhofer, K., 2025. Recognition and completeness metrics from iNaturalist and GBIF can inform future citizen science and research projects: a case study on arthropods in Namibia. Biodivers. Conserv. 34, 467–480. https://doi.org/10.1007/s10531-024-02981-z.

Bach, A., Raub, F., Höfer, H., Ottermanns, R., Roß-Nickoll, M., 2024. ARAapp: filling gaps in the ecological knowledge of spiders using an automated and dynamic approach to analyze systematically collected community data. Database 2024, baae004. https://doi.org/10.1093/database/baae004.

Ball-Damerow, J.E., Brenskelle, L., Barve, N., Soltis, P.S., Sierwald, P., Bieler, R., LaFrance, R., Ariño, A.H., Guralnick, R.P., 2019. Research applications of primary biodiversity databases in the digital age. PLoS One 14, e0215794.

Barbosa, W.L., Alves-Souza, S.N., 2025. Data quality issues in data used in species distribution models: a systematic literature review. Eco. Inform. 91, 103378. https://doi.org/10.1016/j.ecoinf.2025.103378.

Bauer, T., Höfer, H., Schirmel, J., 2024. Spider assemblages (Arachnida: Araneae) in urban grassland patches in Karlsruhe (Baden-Württemberg, Germany). Arachnol. Mitt. 68. https://doi.org/10.30963/aramit6801.

Bayraktarov, E., Ehmke, G., O'Connor, J., Burns, E.L., Nguyen, H.A., McRae, L., Possingham, H.P., Lindenmayer, D.B., 2019. Do big unstructured biodiversity data mean more knowledge? Front. Ecol. Evol. 7, 1–5. https://doi.org/10.3389/fevo.2018.00239.

Beck, J., Ballesteros-Mejia, L., Nagel, P., Kitching, I.J., 2013. Online solutions and the "Wallacean shortfall": what does GBIF contribute to our knowledge of species' ranges? Divers. Distrib. 19, 1043–1050. https://doi.org/10.1111/ddi.12083.

Bowler, D.E., Eichenberg, D., Conze, K., Suhling, F., Baumann, K., Benken, T., Bönsel, A., Bittner, T., Drews, A., Günther, A., Isaac, N.J.B., Petzold, F., Seyring, M., Spengler, T., Trockur, B., Willigalla, C., Bruelheide, H., Jansen, F., Bonn, A., 2021. Winners and losers over 35 years of dragonfly and damselfly distributional change in Germany. Divers. Distrib. 27, 1353–1366. https://doi.org/10.1111/ddi.13274.

Boyle, J., 2013. Biology must develop its own big-data systems. Nature 499, 7. https://doi.org/10.1038/499007a.

Brar, G., 2024. Nurturing entomological talent by bridging generational gaps through mentorship. Am. Entomol. 70, 23. https://doi.org/10.1093/ae/tmae052.

Brosens, D., Vankerkhoven, F., Ignace, D., Wegnez, P., Noé, N., Heughebaert, A., Bortels, J., Dekoninck, W., 2013. FORMIDABEL: the Belgian ants database. Zookeys 306, 59–70. https://doi.org/10.3897/zookeys.306.4898.

Chowdhury, S., Bowler, D.E., Boutaud, E., Bleich, O., Bruelheide, H., Buse, J., Engel, T., Gebert, J., Grescho, V., Gürlich, S., Harry, I., Jansen, F., Klenke, R.A., Van Klink, R., Winter, M., Bonn, A., 2025. Widespread decline of ground beetles in Germany. Divers. Distrib. 31, e70112. https://doi.org/10.1111/ddi.70112.

Chytrý, M., Hennekens, S.M., Jiménez-Alfaro, B., Knollová, I., Dengler, J., Jansen, F., Landucci, F., Schaminée, J.H.J., Aćić, S., Agrillo, E., Ambarlı, D., Angelini, P., Apostolova, I., Attorre, F., Berg, C., Bergmeier, E., Biurrun, I., Botta-Dukát, Z., Brisse, H., Campos, J.A., Carlón, L., Čarni, A., Casella, L., Csiky, J., Ćušterevska, R., Dajić Stevanović, Z., Danihelka, J., De Bie, E., De Ruffray, P., De Sanctis, M., Dickoré, W.B., Dimopoulos, P., Dubyna, D., Dziuba, T., Ejrnæs, R., Ermakov, N., Ewald, J., Fanelli, G., Fernández-González, F., FitzPatrick, Ú., Font, X., García-Mijangos, I., Gavilán, R.G., Golub, V., Guarino, R., Haveman, R., Indreica, A., Işık Gürsoy, D., Jandt, U., Janssen, J.A.M., Jiroušek, M., Kącki, Z., Kavgacı, A., Kleikamp, M., Kolomiychuk, V., Krstivojević Ćuk, M., Krstonošić, D., Kuzemko, A., Lenoir, J., Lysenko, T., Marcenò, C., Martynenko, V., Michalcová, D., Moeslund, J.E., Onyshchenko, V., Pedashenko, H., Pérez-Haase, A., Peterka, T., Prokhorov, V., Rašomavičius, V., Rodríguez-Rojo, M.P., Rodwell, J.S., Rogova, T., Ruprecht, E., Rūsiņa, S., Seidler, G., Šibík, J., Šilc, U., Škvorc, Ž., Sopotlieva, D., Stančić, Z., Svenning, J., Swacha, G., Tsiripidis, I., Turtureanu, P.D., Uğurlu, E., Uogintas, D., Valachovič, M., Vashenyak, Y., Vassilev, K., Venanzoni, R., Virtanen, R., Weekes, L., Willner, W., Wohlgemuth, T., Yamalov, S., 2016. European vegetation archive (EVA): an integrated database of European vegetation plots. Appl. Veg. Sci. 19, 173–180. https://doi.org/10.1111/avsc.12191.

Coleman, C.O., 2015. Taxonomy in times of the taxonomic impediment – examples from the community of experts on amphipod crustaceans. J. Crustac. Biol. 35, 729–740. https://doi.org/10.1163/1937240X-00002381.

Corrales, C., Bacela-Spychalska, K., Buzan, E., Ekrem, T., Ferreira, S., Goodall-Copestake, W., Van Ommen Kloeke, E., Hollingsworth, P.M., Bourlat, S.J., 2025. The role of community science in DNA -based biodiversity monitoring. Mol. Ecol. 34, e70100. https://doi.org/10.1111/mec.70100.

Costello, M.J., Appeltans, W., Bailly, N., Berendsohn, W.G., de Jong, Y., Edwards, M., Froese, R., Huettmann, F., Los, W., Mees, J., Segers, H., Bisby, F.A., 2014. Strategies for the sustainability of online open-access biodiversity databases. Biol. Conserv. 173, 155–165. https://doi.org/10.1016/j.biocon.2013.07.042.

Engel, T., Chowdhury, S., Friedrichs-Manthey, M., von Gönner, J., Herrmann, T., Klenke, R., Koch Sheard, J., Peters, B., Bonn, A., 2023. Digitalisierung in citizen science und naturschutz – anwendungsbeispiele aus der Praxis. Nat. Landsch. 98, 319–329. https://doi.org/10.19217/NuL2023-06-07.

Engemann, K., Enquist, B.J., Sandel, B., Boyle, B., Jørgensen, P.M., Morueta-Holme, N., Peet, R.K., Violle, C., Svenning, J., 2015. Limited sampling hampers "big data" estimation of species richness in a tropical biodiversity hotspot. Ecol. Evol. 5, 807–820. https://doi.org/10.1002/ece3.1405.

Enke, N., Thessen, A., Bach, K., Bendix, J., Seeger, B., Gemeinholzer, B., 2012. The user's view on biodiversity data sharing - investigating facts of acceptance and requirements to realize a sustainable use of research data. Eco. Inform. 11, 25–33. https://doi.org/10.1016/j.ecoinf.2012.03.004.

Eversole, C.B., Powell, R.L., Lizarro, D.E., Moreno, F., Calderon Vaca, G., Aparicio, J., Crocker, A.V., 2019. Introduction of a novel natural history collection: a model for global scientific collaboration and enhancement of biodiversity infrastructure with a focus on developing countries. Biodivers. Conserv. 28, 1921–1931. https://doi.org/10.1007/s10531-019-01765-0.

Feng, X., Enquist, B.J., Park, D.S., Boyle, B., Breshears, D.D., Gallagher, R.V., Lien, A., Newman, E.A., Burger, J.R., Maitner, B.S., Merow, C., Li, Y., Huynh, K.M., Ernst, K., Baldwin, E., Foden, W., Hannah, L., Jørgensen, P.M., Kraft, N.J.B., Lovett, J.C., Marquet, P.A., McGill, B.J., Morueta-Holme, N., Neves, D.M., Núñez-Regueiro, M.M., Oliveira-Filho, A.T., Peet, R.K., Pillet, M., Roehrdanz, P.R., Sandel, B., Serra-Diaz, J. M., Šímová, I., Svenning, J., Violle, C., Weitemier, T.D., Wiser, S., López-Hoffman, L., 2022. A review of the heterogeneous landscape of biodiversity databases: opportunities and challenges for a synthesized biodiversity knowledge base. Glob. Ecol. Biogeogr. 31, 1242–1260. https://doi.org/10.1111/geb.13497.

Fontaine, C., Fontaine, B., Prévot, A.-C., 2021. Do amateurs and citizen science fill the gaps left by scientists? Curr. Opin. Insect Sci. 46, 83–87. https://doi.org/10.1016/j.cois.2021.03.001.

Franz, N.M., Sterner, B.W., 2018. To increase trust, change the social design behind aggregated biodiversity data. Database 2018, bax100. https://doi.org/10.1093/database/bax100.

Gabrielsen, A.M., 2020. Openness and trust in data-intensive science: the case of biocuration. Med. Health Care Philos. 23, 497–504. https://doi.org/10.1007/s11019-020-09960-5.

Gadelha Jr, L.M.R., de Siracusa, P.C., Dalcin, E.C., da Silva, L.A.E., Augusto, D.A., Krempser, E., Affe, H.M., Costa, R.L., Mondelli, M.L., Meirelles, P.M., Thompson, F., Chame, M., Ziviani, A., de Siqueira, M.F., 2021. A survey of biodiversity informatics: concepts, practices, and challenges. WIREs Data Min. Knowl. Discov. 11, e1394. https://doi.org/10.1002/widm.1394.

Gardiner, M.M., Roy, H.E., 2022. The role of community science in entomology. Annu. Rev. Entomol. 67, 437–456. https://doi.org/10.1146/annurev-ento-072121-075258.

Gemeinholzer, B., Vences, M., Beszteri, B., Bruy, T., Felden, J., Kostadinov, I., Miralles, A., Nattkemper, T.W., Printzen, C., Renz, J., Rybalka, N., Schuster, T., Weibulat, T., Wilke, T., Renner, S.S., 2020. Data storage and data re-use in taxonomy—the need for improved storage and accessibility of heterogeneous data. Org. Divers. Evol. 20, 1–8. https://doi.org/10.1007/s13127-019-00428-w.

Grames, E.M., Montgomery, G.A., Boyes, D.H., Dicks, L.V., Forister, M.L., Matson, T.A., Nakagawa, S., Prendergast, K.S., Taylor, N.G., Tingley, M.W., Wagner, D.L., White, T.E., Woodcock, P., Elphick, C.S., 2022. A framework and case study to systematically identify long-term insect abundance and diversity datasets. Conserv. Sci. Pract. 4, e12687. https://doi.org/10.1111/csp2.12687.

Grenié, M., Berti, E., Carvajal-Quintero, J., Dädlow, G.M.L., Sagouis, A., Winter, M., 2023. Harmonizing taxon names in biodiversity data: a review of tools, databases and best practices. Methods Ecol. Evol. 14, 12–25. https://doi.org/10.1111/2041-210X.13802.

Güntsch, A., Overmann, J., Ebert, B., Bonn, A., Le Bras, Y., Engel, T., Hovstad, K.A., Lange Canhos, D.A., Newman, P., Van Ommen Kloeke, E., Ratcliffe, S., Le Roux, M.,

Smith, V.S., Triebel, D., Fichtmueller, D., Luther, K., 2024. National biodiversity data infrastructures: ten essential functions for science, policy, and practice. BioScience 75, 139–151. https://doi.org/10.1093/biosci/biae109.
Hackett, R.A., Belitz, M.W., Gilbert, E.E., Monfils, A.K., 2019. A data management workflow of biodiversity data from the field to data users. Appl. Plant Sci. 7, e11310. https://doi.org/10.1002/aps3.11310.
Harms, D., Nentwig, W., Gloor, D., Harvey, M.S., 2022. Introducing the world Arachnida catalog: the new research environment for (almost all) arachnid orders. J. Arachnol. 50. https://doi.org/10.1636/JoA-S-21-063.
Harzing, A.W., 2007. Publish or Perish. available from. https://harzing.com/resources/publish-or-perish.
Hochkirch, A., Samways, M.J., Gerlach, J., Böhm, M., Williams, P., Cardoso, P., Cumberlidge, N., Stephenson, P.J., Seddon, M.B., Clausnitzer, V., Borges, P.A.V., Mueller, G.M., Pearce-Kelly, P., Raimondo, D.C., Danielczak, A., Dijkstra, K.B., 2021. A strategy for the next decade to address data deficiency in neglected biodiversity. Conserv. Biol. 35, 502–509. https://doi.org/10.1111/cobi.13589.
Höfer, H., 2025a. The Environmental Sample Collection of Spiders at the Staatliches Museum für Naturkunde Karlsruhe. https://doi.org/10.15468/PR26SZ.
Höfer, H., 2025b. The Spider Collection at the Staatliches Museum für Naturkunde Karlsruhe. https://doi.org/10.15468/FMXX2I.
Holstein, J., 2025. The ARAMOB Collection at the Staatliches Museum für Naturkunde Stuttgart. https://doi.org/10.15468/8AERJ4.
Hortal, J., Lobo, J.M., Jiménez-Valverde, A., 2007. Limitations of biodiversity databases: case study on seed-plant diversity in Tenerife, Canary Islands. Conserv. Biol. 21, 853–863. https://doi.org/10.1111/j.1523-1739.2007.00686.x.
iNaturalist, 2025. iNaturalist.org [WWW Document]. URL. https://www.inaturalist.org (accessed 20.12.25).
Isaac, N.J.B., Van Strien, A.J., August, T.A., De Zeeuw, M.P., Roy, D.B., 2014. Statistics for citizen science: extracting signals of change from noisy ecological data. Methods Ecol. Evol. 5, 1052–1060. https://doi.org/10.1111/2041-210X.12254.
Ivanova, N.V., Shashkov, M.P., 2021. The possibilities of GBIF data use in ecological research. Russ. J. Ecol. 52, 1–8. https://doi.org/10.1134/S1067413621010069.
Jetz, W., McPherson, J.M., Guralnick, R.P., 2012. Integrating biodiversity distribution knowledge: toward a global map of life. Trends Ecol. Evol. 27, 151–159. https://doi.org/10.1016/j.tree.2011.09.007.
Jetz, W., McGeoch, M.A., Guralnick, R., Ferrier, S., Beck, J., Costello, M.J., Fernandez, M., Geller, G.N., Keil, P., Merow, C., Meyer, C., Muller-Karger, F.E., Pereira, H.M., Regan, E.C., Schmeller, D.S., Turak, E., 2019. Essential biodiversity variables for mapping and monitoring species populations. Nat. Ecol. Evol. 3, 539–551. https://doi.org/10.1038/s41559-019-0826-1.
Karam, N., Müller-Birn, C., Gleisberg, M., Fichtmüller, D., Tolksdorf, R., Güntsch, A., 2016. A terminology service supporting semantic annotation, integration, discovery and analysis of interdisciplinary research data. Datenbank-Spektrum 16, 195–205. https://doi.org/10.1007/s13222-016-0231-8.
Kissling, W.D., Ahumada, J.A., Bowser, A., Fernandez, M., Fernández, N., García, E.A., Guralnick, R.P., Isaac, N.J.B., Kelling, S., Los, W., McRae, L., Mihoub, J.-B., Obst, M., Santamaria, M., Skidmore, A.K., Williams, K.J., Agosti, D., Amariles, D., Arvanitidis, C., Bastin, L., De Leo, F., Egloff, W., Elith, J., Hobern, D., Martin, D., Pereira, H.M., Pesole, G., Peterseil, J., Saarenmaa, H., Schigel, D., Schmeller, D.S., Segata, N., Turak, E., Uhlir, P.F., Wee, B., Hardisty, A.R., 2018. Building essential biodiversity variables (EBVs) of species distribution and abundance at a global scale. Biol. Rev. 93, 600–625. https://doi.org/10.1111/brv.12359.
Knollová, I., Chytrý, M., Bruelheide, H., Dullinger, S., Jandt, U., Bernhardt-Römermann, M., Biurrun, I., De Bello, F., Glaser, M., Hennekens, S., Jansen, F., Jiménez-Alfaro, B., Kadaš, D., Kaplan, E., Klinkovská, K., Lenzner, B., Pauli, H., Sperandii, M.G., Verheyen, K., Winkler, M., Abdaladze, O., Aćić, S., Acosta, A.T.R., Alignier, A., Andrews, C., Arlettaz, R., Attorre, F., Axmanová, I., Babbi, M., Baeten, L., Baran, J., Barni, E., Benito-Alonso, J., Berg, C., Bergamini, A., Berki, I., Boch, S., Bock, B., Bode, F., Bonari, G., Boublík, K., Britton, A.J., Brunet, J., Bruzzaniti, V., Buholzer, S., Burrascano, S., Campos, J.A., Carlsson, B., Carranza, M. L., Černý, T., Charmillot, K., Chiarucci, A., Choler, P., Chytrý, K., Corcket, E., Csecserits, A., Cutini, M., Czarniecka-Wiera, M., Danihelka, J., De Francesco, M.C., De Frenne, P., Di Musciano, M., De Sanctis, M., Deák, B., Decocq, G., Dembicz, I., Dengler, J., Di Cecco, V., Dick, J., Diekmann, M., Dierschke, H., Dirnböck, T., Doerfler, I., Doležal, J., Döring, U., Durak, T., Dwyer, C., Ejrnæs, R., Ermakova, I., Erschbamer, B., Fanelli, G., Fernández-Calzado, M., Fickert, T., Fischer, A., Fischer, M., Foremnik, K., Frouz, J., García-González, R., García-Magro, D., García-Mijangos, I., Gavilán, R.G., Germ, M., Ghosn, D., Gigauri, K., Gizela, J., Golob, A., Golub, V., Gómez-García, D., Gowing, D., Grytnes, J., Güler, B., Gutiérrez-Girón, A., Haase, P., Haider, S., Hájek, M., Halassy, M., Harásek, M., Härdtle, W., Heinken, T., Hester, A., Humbert, J., Ibáñez, R., Illa, E., Jaroszewicz, B., Jensen, K., Jentsch, A., Jiroušek, M., Kalníková, V., Kanka, R., Kapfer, J., Kazakis, G., Kermavnar, J., Kesting, S., Khanina, L., Kindermann, E., Kotrík, M., Koutecký, T., Kozub, Ł., Kuhn, G., Kutnar, L., La Montagna, D., Lamprecht, A., Lenoir, J., Lepš, J., Leuschner, C., Lorite, J., Madsen, B., Ugarte, R.M., Malicki, M., Maliniemi, T., Máliš, F., Maringer, A., Marrs, R., Matesanz, S., Metze, K., Meyer, S., Millett, J., Mitchell, R.J., Moeslund, J.E., Moiseev, P., Di Cella, U.M., Mudrák, O., Müller, F., Müller, N., Naaf, T., Nagy, L., Napoleone, F., Nascimbene, J., Navrátilová, J., Ninot, J.M., Niu, Y., Normand, S., Ogaya, R., Onipchenko, V., Orczewska, A., Ortmann-Ajkai, A., Pakeman, R.J., Pardo, I., Pätsch, R., Peet, R.K., Penuelas, J., Peppler-Lisbach, C., Pérez-Hernández, J., Pérez-Haase, A., Petraglia, A., Petřík, P., Pielech, R., Piórkowski, H., Pladevall-Izard, E., Poschlod, P., Prach, K., Praleskouskaya, S., Prokhorov, V., Provoost, S., Puşcaş, M., Pustková, Š., Randin, C. F., Rašomavičius, V., Reczyńska, K., Rédei, T., Řehounková, K., Richner, N., Risch, A. C., Rixen, C., Rosbakh, S., Roscher, C., Rosenthal, G., Rossi, G., Rötzer, H., Roux, C., Rumpf, S.B., Ruprecht, E., Rūsiņa, S., Sanz-Zubizarreta, I., Schindler, M., Schmidt, W., Schories, D., Schrautzer, J., Schubert, H., Schuetz, M., Schwabe, A., Schwaiger, H., Schwartze, P., Šebesta, J., Seiler, H., Šilc, U., Silva, V., Šmilauer, P., Šmilauerová, M., Sperle, T., Stachurska-Swakoń, A., Stanik, N., Stanisci, A., Steffen, K., Storm, C., Stroh, H.G., Sugorkina, N., Świerkosz, K., Świerszcz, S., Szymura, M., Teleki, B., Thébaud, G., Theurillat, J., Tichý, L., Treier, U.A., Turtureanu, P.D., Ujházy, K., Ujházyová, M., Ursu, T.M., Uziębło, A.K., Valkó, O., Van Calster, H., Van Meerbeek, K., Vandevoorde, B., Vandvik, V., Varricchione, M., Vassilev, K., Villar, L., Virtanen, R., Vittoz, P., Voigt, W., Von Hessberg, A., Von Oheimb, G., Wagner, E., Walther, G., Wellstein, C., Wesche, K., Wilhelm, M., Willner, W., Wipf, S., Wittig, B., Wohlgemuth, T., Woodcock, B.A., Wulf, M., Essl, F., 2024. ReSurveyEurope: a database of resurveyed vegetation plots in Europe. J. Veg. Sci. 35, e13235. https://doi.org/10.1111/jvs.13235.
König, C., Weigelt, P., Schrader, J., Taylor, A., Kattge, J., Kreft, H., 2019. Biodiversity data integration—the significance of data resolution and domain. PLoS Biol. 17, e3000183. https://doi.org/10.1371/journal.pbio.3000183.
Kühl, H.S., Bowler, D.E., Bösch, L., Bruelheide, H., Dauber, J., Eichenberg, David., Eisenhauer, N., Fernández, N., Guerra, C.A., Henle, K., Herbinger, I., Isaac, N.J.B., Jansen, F., König-Ries, B., Kühn, I., Nilsen, E.B., Pe'er, G., Richter, A., Schulte, R., Settele, J., van Dam, N.M., Voigt, M., Wägele, W.J., Wirth, C., Bonn, A., 2020. Effective biodiversity monitoring needs a culture of integration. One Earth 3, 462–474. https://doi.org/10.1016/j.oneear.2020.09.010.
Löffler, F., Wesp, V., König-Ries, B., Klan, F., 2021. Dataset search in biodiversity research: do metadata in data repositories reflect scholarly information needs? PLoS One 16, e0246099.
Luke, S.H., Roy, H.E., Thomas, C.D., Tilley, L.A.N., Ward, S., Watt, A., Carnaghi, M., Jaworski, C.C., Tercel, M.P.T.G., Woodrow, C., Aown, S., Banfield-Zanin, J.A., Barnsley, S.L., Berger, I., Brown, M.J.F., Bull, J.C., Campbell, H., Carter, R.A.B., Charalambous, M., Cole, L.J., Ebejer, M.J., Farrow, R.A., Fartyal, R.S., Grace, M., Highet, F., Hill, J.K., Hood, A.S.C., Kent, E.S., Krell, F., Leather, S.R., Leybourne, D. J., Littlewood, N.A., Lyons, A., Matthews, G., Mc Namara, L., Menéndez, R., Merrett, P., Mohammed, S., Murchie, A.K., Noble, M., Paiva, M., Pannell, M.J., Phon, C., Port, G., Powell, C., Rosell, S., Sconce, F., Shortall, C.R., Slade, E.M., Sutherland, J.P., Weir, J.C., Williams, C.D., Zielonka, N.B., Dicks, L.V., 2023. Grand challenges in entomology: priorities for action in the coming decades. Insect Conserv. Divers. 16, 173–189. https://doi.org/10.1111/icad.12637.
Mathew, C., Güntsch, A., Obst, M., Vicario, S., Haines, R., Williams, A.R., de Jong, Y., Goble, C., 2014. A semi-automated workflow for biodiversity data retrieval, cleaning, and quality control. Biodivers. Data J. 2, e4221.
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.
Michener, W.K., Brunt, J.W., Helly, J.J., Kirchner, T.B., Stafford, S.G., 1997. Nongeospatial metadata for the ecological sciences. Ecol. Appl. 7, 330–342. https://doi.org/10.1890/1051-0761(1997)007%255B0330:NMFTES%255D2.0.CO;2.
Nentwig, W., Gloor, D., Kropf, C., 2015. Spider taxonomists catch data on web. Nature 528, 479. https://doi.org/10.1038/528479a.
Nentwig, W., Blick, T., Gloor, D., Hänggi, A., Kropf, C., 2025. Spinnen Europas [WWW Document]. URL. www.araneae.unibe.ch (accessed 06.02.25).
Novotný, P., Seifert, S., Rohn, M., Diewald, W., Štech, M., Triebel, D., 2022. Software infrastructure and data pipelines established for technical interoperability within a cross-border cooperation for the flora of the Bohemian forest. Biodivers. Data J. 10, e87254. https://doi.org/10.3897/BDJ.10.e87254.
Observation International, 2025. Observation.org [WWW Document]. URL. https://www.observation.org (accessed 22.12.25).
Outhwaite, C.L., Powney, G.D., August, T.A., Chandler, R.E., Rorke, S., Pescott, O.L., Harvey, M., Roy, H.E., Fox, R., Roy, D.B., Alexander, K., Ball, S., Bantock, T., Barber, T., Beckmann, B.C., Cook, T., Flanagan, J., Fowles, A., Hammond, P., Harvey, P., Hepper, D., Hubble, D., Kramer, J., Lee, P., MacAdam, C., Morris, R., Norris, A., Palmer, S., Plant, C.W., Simkin, J., Stubbs, A., Sutton, P., Telfer, M., Wallace, I., Isaac, N.J.B., 2019. Annual estimates of occupancy for bryophytes, lichens and invertebrates in the UK, 1970–2015. Sci. Data 6, 259. https://doi.org/10.1038/s41597-019-0269-1.
Packer, L., Monckton, S.K., Onuferko, T.M., Ferrari, R.R., 2018. Validating taxonomic identifications in entomological research. Insect Conserv. Divers. 11, 1–12. https://doi.org/10.1111/icad.12284.
Páll-Gergely, B., Krell, F.-T., Ábrahám, L., Bajomi, B., Balog, L.E., Boda, P., Csuzdi, C., Dányi, L., Fehér, Z., Hornok, S., Horváth, A., Kóbor, P., Koczor, S., Kontschán, J., Kovács, P., Kovács, T., Lukátsi, M., Majoros, G., Murányi, D., Németh, T., Pernecker, B., Puskás, G., Rózsa, L., Soltész, Z., Szita, É., Szűts, T., Tóth, B., Tőke, A., Vas, Z., Zsuga, K., Zsupos, V., Csabai, Z., Móra, A., 2024. Identification crisis: a fauna-wide estimate of biodiversity expertise shows massive decline in a central European country. Biodivers. Conserv. 33, 3871–3903. https://doi.org/10.1007/s10531-024-02934-6.
Pereira, H.M., Ferrier, S., Walters, M., Geller, G.N., Jongman, R.H.G., Scholes, R.J., Bruford, M.W., Brummitt, N., Butchart, S.H.M., Cardoso, A.C., Coops, N.C., Dulloo, E., Faith, D.P., Freyhof, J., Gregory, R.D., Heip, C., Höft, R., Hurtt, G., Jetz, W., Karp, D.S., McGeoch, M.A., Obura, D., Onoda, Y., Pettorelli, N., Reyers, B., Sayre, R., Scharlemann, J.P.W., Stuart, S.N., Turak, E., Walpole, M., Wegmann, M., 2013. Essential biodiversity variables. Science 339, 277–278. https://doi.org/10.1126/science.1229931.
Pinheiro, F.D.C., Forzza, R.C., Leitman, P.M., Prado, J., 2024. The REFLORA program: implementation, repatriation, and creation of the REFLORA virtual herbarium as a tool for biodiversity studies. Biota Neotrop. 24, e20241701. https://doi.org/10.1590/1676-0611-bn-2024-1701.
Proença, V., Martin, L.J., Pereira, H.M., Fernandez, M., McRae, L., Belnap, J., Böhm, M., Brummitt, N., García-Moreno, J., Gregory, R.D., Honrado, J.P., Jürgens, N.,

Opige, M., Schmeller, D.S., Tiago, P., Van Swaay, C.A.M., 2017. Global biodiversity monitoring: from data sources to essential biodiversity variables. Biol. Conserv. 213, 256–263. https://doi.org/10.1016/j.biocon.2016.07.014.

Raub, F., Bach, A., Bauer, T., Höfer, H., 2023. Editorial. Data paper publication in Arachnologische Mitteilungen – goals, review criteria, editorial procedures, format, data management and mobilization. Arachnol. Mitt. 66, iii–iv. https://doi.org/10.30963/aramit6612.

Rondinini, C., Wilson, K.A., Boitani, L., Grantham, H., Possingham, H.P., 2006. Tradeoffs of different types of species occurrence data for use in systematic conservation planning. Ecol. Lett. 9, 1136–1145. https://doi.org/10.1111/j.1461-0248.2006.00970.x.

Sandall, E.L., Maureaud, A.A., Guralnick, R., McGeoch, M.A., Sica, Y.V., Rogan, M.S., Booher, D.B., Edwards, R., Franz, N., Ingenloff, K., Lucas, M., Marsh, C.J., McGowan, J., Pinkert, S., Ranipeta, A., Uetz, P., Wieczorek, J., Jetz, W., 2023. A globally integrated structure of taxonomy to support biodiversity science and conservation. Trends Ecol. Evol. 38, 1143–1153. https://doi.org/10.1016/j.tree.2023.08.004.

Sanderson, C., Braby, M.F., Bond, S., 2021. Butterflies Australia: a national citizen science database for monitoring changes in the distribution and abundance of Australian butterflies. Austral Entomol. 60, 111–127. https://doi.org/10.1111/aen.12513.

Saska, P., Makowski, D., Bohan, D.A., van der Werf, W., 2021. The effects of trapping effort and sources of variability on the estimation of activity-density and diversity of carabids in annual field crops by pitfall trapping; a meta-analysis. Entomol. Gen. 41, 553–566. https://doi.org/10.1127/entomologia/2021/1211.

Seifert, S., Weiss, M., Triebel, D., 2015. The Diversity Workbench Service Cloud – Regionalised and Domain-Specific Taxon Lists.

SERNEC, 2026. Data Portal [WWW Document]. URL. http://sernecportal.org/index.php (accessed 27.12.26).

Socher, S., Ahlmer, W., Weibulat, T., 2014. Flora von Bayern Datenbanken in der diversity Workbench: Konzepte und Datenflüsse. Ber. Bayer. Bot. Ges. 84, 131–134.

Stall, S., Yarmey, L., Cutcher-Gershenfeld, J., Hanson, B., Lehnert, K., Nosek, B., Parsons, M., Robinson, E., Wyborn, L., 2019. Make scientific data FAIR. Nature 570, 27–29. https://doi.org/10.1038/d41586-019-01720-7.

Sterner, B.W., Gilbert, E.E., Franz, N.M., 2020. Decentralized but globally coordinated biodiversity data. Front. Big Data 3, 519133. https://doi.org/10.3389/fdata.2020.519133.

Telenius, A., 2011. Biodiversity information goes public: GBIF at your service. Nord. J. Bot. 29, 378–381. https://doi.org/10.1111/j.1756-1051.2011.01167.x.

Tenopir, C., Allard, S., Douglass, K., Aydinoglu, A.U., Wu, L., Read, E., Manoff, M., Frame, M., 2011. Data sharing by scientists: practices and perceptions. PLoS One 6, e21101. https://doi.org/10.1371/journal.pone.0021101.

Tewksbury, J.J., Anderson, J.G.T., Bakker, J.D., Billo, T.J., Dunwiddie, P.W., Groom, M. J., Hampton, S.E., Herman, S.G., Levey, D.J., Machnicki, N.J., Del Rio, C.M., Power, M.E., Rowell, K., Salomon, A.K., Stacey, L., Trombulak, S.C., Wheeler, T.A., 2014. Natural history's place in science and society. BioScience 64, 300–310. https://doi.org/10.1093/biosci/biu032.

Triebel, D., Hagedorn, G., Rambolt, G., 1999. Diversity Workbench – A Virtual Research Environment for Building and Accessing Biodiversity and Environmental Data. [WWW Document]. URL. http://www.diversityworkbench.net (accessed 23.01.23).

Undap, N., Papu, A., Schillo, D., Ijong, F.G., Kaligis, F., Lepar, M., Hertzer, C., Böhringer, N., König, G.M., Schäberle, T.F., Wägele, H., 2019. First survey of Heterobranch Sea slugs (Mollusca, Gastropoda) from the island Sangihe, North Sulawesi, Indonesia. Diversity 11, 170. https://doi.org/10.3390/d11090170.

Van Swaay, C.A.M., Nowicki, P., Settele, J., Van Strien, A.J., 2008. Butterfly monitoring in Europe: methods, applications and perspectives. Biodivers. Conserv. 17, 3455–3469. https://doi.org/10.1007/s10531-008-9491-4.

Vargas, C.A., Bottin, M., Sarkinen, T., Richardson, J.E., Celis, M., Villanueva, B., Sanchez, A., 2024. How to fill the biodiversity data gap: is it better to invest in fieldwork or curation? Plant Divers. 46, 39–48. https://doi.org/10.1016/j.pld.2023.06.003.

Veiga, A.K., Saraiva, A.M., Chapman, A.D., Morris, P.J., Gendreau, C., Schigel, D., Robertson, T.J., 2017. A conceptual framework for quality assessment and management of biodiversity data. PLoS One 12, e0178731. https://doi.org/10.1371/journal.pone.0178731.

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. https://doi.org/10.1371/journal.pone.0029715.