TITLE

A Globally Integrated Structure of Taxonomy supporting biodiversity science and conservation

AUTHOR NAMES

Emily L. Sandall[1,2,*], Aurore A. Maureaud[1,2,*], Robert Guralnick[3], Melodie A. McGeoch[4], Yanina V. Sica[1,2], Matthew S. Rogan[1,2], Douglas B. Booher[1,2], Mark John Costello[5], Robert Edwards[1,2,6], Nico Franz[7], Kate Ingenloff[1,2], Maisha Lucas[1,2], Charles J. Marsh[1,2], Jennifer McGowan[1,2,8], Stefan Pinkert[1,2,9], Ajay Ranipeta[1,2], Peter Uetz[10], John Wieczorek[11], Walter Jetz[1,2,12]

(*equal contribution)

Affiliations:

[1] Department of Ecology & Evolutionary Biology, Yale University, New Haven, CT, 06520, USA
[2] Center for Biodiversity & Global Change, Yale University, New Haven, CT, 06520, USA
[3] Department of Biodiversity Informatics, Florida Museum of Natural History, University of Florida, Gainesville, FL, USA
[4] Department of Environment and Genetics, LaTrobe University, Melbourne, Australia
[5] Faculty of Biosciences and Aquaculture, Nord University, Bodo, Norway
[6] Cleveland Museum of Natural History, Cleveland, OH, USA
[7] School of Life Sciences, Arizona State University, Tempe, AZ, USA
[8] The Nature Conservancy, Arlington, VA, USA
[9] Department of Conservation Ecology, University of Marburg, Marburg, Germany
[10] Center for Biological Data Science, Virginia Commonwealth University, Richmond, VA, USA
[11] Museum of Vertebrate Zoology, University of California, Berkeley, CA, USA
[12] E.O.Wilson Biodiversity Foundation, Durham, NC, USA

Corresponding Authors: Emily L. Sandall sandall.emily@gmail.com and Aurore A. Maureaud aurore.aqua@gmail.com

HIGHLIGHTS

- Taxonomic knowledge is a critical element to understand, catalog, and assess biodiversity and is central to measuring and achieving conservation goals, including the Post-2020 Framework of the Convention on Biological Diversity
- Taxonomy is a centuries-old discipline, and its tools, its diversity of users, and its applications are constantly expanding and evolving
- The lack of trackable and interoperable taxonomic data inhibits data integration and knowledge transfer across communities and disciplines, constraining conservation applications
- We propose a Globally Integrated Structure of Taxonomy (GIST) to increase understanding, interoperability, and interdisciplinarity across the fields of taxonomy, biodiversity, and conservation
- Funding for the implementation of the GIST should target linking data and communities across biodiversity databases

ABSTRACT

All aspects of biodiversity research, from taxonomy to conservation, rely on data associated with species names. Effective integration of names across multiple fields is paramount and depends on coordination and organization of taxonomic data. We review current efforts and find that even key applications for well-studied taxa still lack taxonomic elements required for interoperability and use. We identify opportunities offered by a metadata structure that supports improved access and integration of taxonomic backbone data, better connects taxonomic communities, and highlights broken linkages that limit the current research capacity. We recommend ways forward to improve interoperability of taxonomic data and resultant downstream use in broad biodiversity research and conservation applications.

KEYWORDS

taxonomic backbone, integrative science, data linkage, social infrastructure, biodiversity conservation

MAIN TEXT

Biodiversity and conservation sciences rely on taxonomic data

Taxonomy (see Glossary) provides the fundamental units around which we organize, assess, and mediate the components of biodiversity for research and conservation [1–6]. Both research and conservation use of taxonomic names online have expanded over the past decades. Differing needs and **values** between communities producing and using such data have led to outcomes that are centered around distinct goals [7–9]. The resulting dynamic nature, heterogeneity, and bias in taxonomic data might not be obvious to users, but can have large effects on research, cultural, and **biodiversity conservation** outcomes [10–16].

The **binomial nomenclature** came about as a standardized and shared means to reference the identity of organisms, complementing **vernacular names** and descriptions based on appearance and cultural relevance [17]. With the vast increase in the number of formally described species since Linnaeus' time, a key challenge has been to track the **species concept** for the taxon recognized, as well as its **scientific nomenclature** (both formally represented by **accepted names** and **synonyms**). These names are described and used by researchers who associate them with physical **specimens** [18] or other data and share those resources in databases and literature [19]. Each of these elements—species concept, accepted name, and synonym(s)—can be subject to revision, based on new scientific evidence or needs. Acceptance of taxonomic revisions is a scientific process with different cultures and rules. As a result, multiple taxonomic frameworks remain in use across domains of application and time periods.

This complexity is further compounded by the different types of data affiliated with names in biodiversity repositories, including spatial, functional, genetic, and physical data [5,20] (Figure 1). The **global species list** of names, higher taxonomic classification, and associated **taxonomic backbone** that references biological, genetic, and functional information to a **species** or **taxon** is the key enabler of subsequent synthesis. Research that uses species names requires linking up several data sources in supporting **integrative science** [21–25]. The ability to relate these data types to taxonomic information is essential in informing emerging interdisciplinary research and conservation applications [26,27].

Key to overcoming complexities to best use of species name resources, and achieving **taxonomic integration**, is a common structure that facilitates **interoperability** across disparate data sources. Here, we review how different objectives in taxonomy and biodiversity informatics have led to opportunities and challenges in taxonomic interoperability across data sources and types. Based on this assessment, we identify the elements necessary for a more accessible, effective, and diverse use of taxonomic data. We then tentatively combine these elements under an overarching metadata structure focused on interoperability and broad multipurpose utility, access, and longevity.

Needs for taxonomic interoperability and harmonization

Biology and biodiversity science and conservation are inherently intertwined with taxonomic data. Three examples illustrate the broad significance of taxonomic integration across data sources (Figure 1).

Human health—For zoonotic diseases in general and viruses such as SARS-CoV-2, accurate taxonomically named identities, from virus strain to host species, are key to integrating genetic, spatial and even clinical data required to assess and mitigate impact [28–30]. Quality assured taxonomic synthesis is relevant to governmental authorities across health sectors from local to national and multinational jurisdictions. This enables targeted research and communication into the origin, severity, and threats posed by such outbreaks (Figure 1A).

Species invasions—The spread of invasive species is causing long-term challenges for biodiversity and humanity. Members of *Opuntia* (Cactaceae), a widespread genus of cacti, including the common ornamental prickly pear native to the Americas are now established across continents (Figure 1B). Differences in taxonomic treatments of *Opuntia* subspecies have significantly delayed early detection and management [31], a problem that could be overcome through robust taxonomic **harmonization** and updated rapid detection tools such as field guides and electronic identification applications.

Species assessments—Each of 19,327 currently recognized butterfly species have on average six synonyms [32], although some species such as the common palearctic butterfly *Plebejus argus*, have as many as 160 [33] (Figure 1C). Thus, assessing distributions to track threat levels and population declines often requires significant efforts combining spatial data, natural history information, and taxonomic expertise.

Biodiversity studies and conservation interventions increasingly rely on more than one data source or type [21,34]. The above examples illustrate the large array of questions and integrated data usage from basic to applied that rely on the common language of taxonomy and multi-source harmonization and integration.

Diverse spectrum of taxonomic databases and communities

The current landscape of taxonomy sources in databases

Taxonomists and other key actors have addressed the need for integration through the development of taxonomic databases (Table 1). These efforts are increasingly leveraging informatics innovations, computational and storage capabilities, and novel online engagement avenues. They have catalyzed growing consensus in semantic alignment of taxonomic concepts, enhancing the potential for linking data across multiple sources [35–37]. Initiatives to standardize, maintain, and organize relevant communities around taxonomic backbones have made important progress towards this goal. Yet, these efforts often face regionally-

[38,39], taxonomically- [40–43], temporally- [44], or funding-specific [45] constraints, leading to a spectrum of longevity, interoperability, and maintenance hurdles, and stilted progress, reducing accuracy in research and conservation applications [46–48].

Broadly, we can distinguish at least three levels of origin sources of taxonomic databases. Primary source databases aim to produce a taxonomic backbone for one taxon. Successful examples include the Amphibian Species of the World, FishBase, the Reptile Database, and AviBase, which are extensive taxonomic databases of amphibians, fish, reptiles, and birds, respectively [40,48–50] (Table 1). At the level of a single taxon, database and maintenance are characterized by platforms, linking experts and the primary literature together to cover the taxonomic knowledge. Infrastructure solutions such as Symbiota and TaxonWorks provide data portals through which individual experts can synchronize changes to these databases and natural history collection information [51,52].

Secondary source taxonomic databases maximize the list of names through aggregating primary sources of names. Overarching name catalogues that function this way typically go beyond single taxa. These databases similarly aim to be authoritative sources of organism names with the intent to be regularly updated and maintained by taxonomic experts [53,54]. One of the leading initiatives is Catalogue of Life+ (COL), which relies on a group of taxonomic experts to provide updates and publish them to the catalogue [53]. COL was born out of Species 2000 and the Integrated Taxonomic Information System (ITIS) [55] and provides a model for publishing incoming data. With so many taxonomic sources (Table 1), governance and practices around taxonomy of primary and secondary databases become challenging. An international working group established within the International Union of Biological Sciences (IUBS) has laid out the issues involved to establish a dynamic but quality-assured world checklist of all species for end users [54]. COL+ thereby is an essential step in making the wider scientific community aware and supportive of a connected taxonomic backbone framework.

Tertiary source taxonomic databases build a taxonomy for the purpose of combining available biodiversity data rather than as its primary objective. Such taxonomic efforts may ‘mix and match’ between primary authorities, add further harmonization, or implement customized updates to create more comprehensive species lists and taxonomic backbones. These databases may explicitly center information around spatial or genetic data through linking multiple primary taxonomic sources and data types to support threat status, integrated map products and indicators. As the largest **biodiversity data aggregator**, the Global Biodiversity Information Facility (GBIF) [56] currently harmonizes over 2 billion occurrence records against a taxonomic backbone [57,58], informed by many dynamic taxonomic lists. Map of Life (MOL) uses a combination of taxonomies to harmonize raw names associated with several sources of species spatial data, such as GBIF, to authoritative global species lists by taxon [59].

Secondary and tertiary taxonomic databases depend on the taxonomic data from primary databases (Table 1) but are rarely fully interoperable due to separate maintenance timetables or taxonomic frameworks. Establishing operational links between data products is challenging, given the dynamic nature of taxonomic advances [21,43,60]. Efforts dedicated to interoperability and maintenance depend on communities developing and sharing each of these databases and reconciling different concepts and semantics.

Diverse communities and values around taxonomies

Collaboration is necessary in **interdisciplinary science**, including individuals and communities with diverse perspectives, values, and project emphases [61,62]. Taxonomy, biodiversity researchers and conservationists have legacies and values that position their interactions with taxonomic data. For taxonomists to successfully describe species and maintain the nomenclatural record, they center their work around legacy, history, and specialty, to create scientific knowledge and uphold the standards of their field [63]. Where appropriately incentivized, experts who assemble large-scale biodiversity resources maintain data relations, harmonization, and standards while the data itself constantly changes. Researchers relying on research-ready, taxonomically harmonized data can build on the taxon-oriented data to lead synthesis and conduct transparent analysis to make their work broadly available as part of the scientific enterprise. **Conservation managers** and **decision-makers** further enhance the value of taxonomically integrated data via informing applied local to global strategies and conservation plans [64]. Integrating perspectives across disciplines inherently brings a diversity of values in how data is ultimately structured and consumed. The consideration of these values is key for an infrastructure that successfully integrates taxonomic data, as they can be the source of interdisciplinary innovations, but also lead to misunderstandings and conflicts [12,18,65]. As a result, enabling positive outcomes for multiple types of end users is challenging but critical for collaboration, longevity, and utility of products that rely on taxonomic knowledge.

Towards effective taxonomic data interoperability

Minimum elements of an integrable taxonomic data structure

We suggest that at least six major elements are required to support successful species data synthesis and integration for a particular taxon (Figure 2) and form the basis for an overarching Globally Integrated Structure of Taxonomy (GIST). The first element, a “Globally integrated list”, comprehensively catalogs all accepted names, like a digital taxon monograph. Secondly, a “Synonyms list” directly linked to accepted names in the “Globally integrated list” matches older and divergent names in spelling, subsumed rank, or no longer valid names with current data. Third, “Authorship information” comprising the author name and year of publication associates a species name with a species concept and its synonyms. The fourth element, “Names sources and timestamps”, are the original database source name and version. It ensures reproducibility and transparency as sources and reported names change over time. Fifth, the “Species instance”, such as an observation or specimen, provides an instance of usage of the name in a data source. Finally, “Species concept in space and time” links names to dates and locations, providing the geographic context for any taxonomic name and detection of eventual needs for revision (see Glossary and Online Supplemental Information Table S2).

The interdependent nature and compound importance of GIST elements and their effects are underappreciated outside the field of taxonomy, partially due to a lack of common vocabulary between users. While important in isolation, these elements are most meaningful when considered collectively. For instance, the content of the “Synonyms List” is dependent on which source is selected as the “Globally Integrated List”. Similarly, “Species concept in space and time” may already be implicit from other elements but require further refinement to explicitly track revisions (e.g., reassignment, **splitting, lumping**). These elements enable taxonomic harmonization in regional to global datasets and help assess integration of taxa in other fields and data sources.

Globally Integrated Structure of Taxonomy

The ideal GIST would support successful data integration. At its core, the GIST builds upon existing and available taxonomic expertise [18,57,66–68] to ensure that source references and relationship among applications of species names are transparent and traceable. Most explicitly, the GIST represents: (i) definitions of metadata elements essential for effective interoperability and synthesis across databases and domains, (ii) clear terms and elements that are identifiable in taxonomic infrastructures and readily linked to Darwin Core (DwC) terms (see Table S2) [57], (iii) a proposed basis for a data standard usable between data aggregators, and (iv) a method for assessing incomplete or inaccurate information in datasets limiting innovative use of name data in taxonomy, biodiversity, and conservation.

Assessing the status of GIST

Several elements of the GIST have seen substantial development in recent years. For example, the harmonization of taxonomic names across sources and development of user-friendly tools have greatly improved interoperability [60,69–71]. Still, shortcomings in even just a single element can constrain interoperability within and across taxa. To assess the current state, we review the level of GIST coverage for nine example taxa represented in MOL taxonomies (Box 1). At present, none of the groups evaluated shows full coverage for the five evaluated elements. Except for butterflies, “Global integrated lists” of all focal taxa are seemingly well-curated (Box 1, Figure I). “Synonym lists” appear more challenging for butterflies and dragonflies, where comprehensive lists required the compilation of many sources (see Online Supplemental Information Table S3). “Authorship information” is not consistently available across sources and species, and only the reptiles received the maximum score for this element. “Name sources and timestamps” were well integrated, but the “Species concept in space and time" were poorly available and represent an avenue for improving taxonomic integration moving forward. Overall, the nine taxa assessed seem to be generally well-curated and may be considered some of the best-case scenarios regarding GIST compared to many other invertebrate taxa [72].

Examining the interoperability of commonly used data sources

We next reviewed the implications of the GIST coverage on interoperability in taxonomy, biodiversity, and conservation applications (Box 1, Figure II). Our analysis of the nine taxa identified a range of limited links across taxonomic sources, e.g., between the Open Tree of Life and mammals, birds, plants, or butterflies. Use of genetic data is hampered by the inability to match 15% of names from the National Center for Biotechnology Information (NCBI) for mammals and butterflies to MOL taxonomies. In other cases, levels of interoperability were much higher, potentially due to an upfront alignment of different efforts, e.g. where the same global species lists were already used in the Global Register of Introduced and Invasive Species (GRIIS) database [73] or the International Union for the Conservation of Nature (IUCN) Red List [74]. This is also the case for COL, for which updates and sub-list integration recently improved [54].

Overall, no taxon has resources mature enough to receive a full GIST integration score (Box 1). Inconsistency and lack of author information for birds, daisies, and amphibians (as indicated by their GIST score) limit the ability to track name identifications over time in taxonomic knowledge assessments. The number of sources required to generate a comprehensive synonym lists for butterflies and dragonflies diminish the GIST group score but does not necessarily impede subsequent integration thanks to recent efforts [32,75] (Box 1, Figure II). Exceptions that have a relatively high GIST score but are poorly integrated are crabs, while odonates, a small taxon with a relatively large research community but a lower GIST score, are exceptionally well-integrated.

Advancing taxonomic integration in ecology and conservation

Our review of GIST elements documented progress in taxonomic integration (Box 1). However, differences between element scores of taxa also highlight broken links in harmonization and communication between data products. We attribute these to four main barriers (Figure 3): (i) inadequate resources allocated to fund, support and realize the benefits of interdisciplinary taxonomic initiatives [54,76,77], (ii) limited infrastructures supporting integration and links between taxonomic data [37,60,78] (iii) uncoordinated management of expertise and products with training and validation [43,76], and (iv) inadequately supported engagement around taxonomy and its uses due to poor communication and lack of working groups. Here, we provide a set of recommendations that coalesce into a path to achieving better interoperability for integrative science.

Overcoming barriers

We propose five guiding principles that leverage the GIST elements and existing efforts for ensuring open, usable, and long-term taxonomic data integration (Figure 3):

Harmonization—Considering the diversity and complexity in the landscape of taxonomic databases (Table 1), prioritizing the harmonization process and its understanding by all users is essential. While more efforts are emerging towards improving harmonization and coordination between databases [37,60,70,71,79], the GIST elements provide a standard schema for taxonomic databases. Tracking the “Species concept in space and time” GIST element represents a specific area of needed and emerging focus (Box 1, Figure I) [80,81]. Documentation of alternative names from various sources in the “Synonyms list” also represents an area of improvement to ensure globally coordinated databases within and across taxa, as recently done for butterflies and dragonflies [32,75]. Even though databases may respond to distinct codes of nomenclature [82–84], or models of governance [37,47], the GIST elements are simple enough to be transferrable across all databases and taxa and can rely on the DwC standards (see Online Supplemental Information Table S2) [57].

Transparency—To fully realize the potential of taxonomic data and ensure cohesion in subsequent uses, we need to increase its transparency [85,86]. Most lists of names and taxonomic databases are open access (especially the “Globally Integrated List” GIST element), but that is only the first step to open science, where far more can be done to improve access to methods, sources, and resources [87]. A structure supporting taxonomists and collection curators, such as Bionomia [88], improves transparency, community engagement and proper acknowledgement around the “Species instance” element. We anticipate that incorporating the elements of the GIST across databases will further facilitate the implementation of the FAIR principles (findable, accessible, interoperable, reusable) [89] by improving access to information about sources (e.g., “Names Sources & Timestamps”, “Authorship Information” elements). Moving forward, we recommend documenting how available and integrated the GIST elements are across databases and communities, expanding and following the assessment criteria in Box 1 and Online Supplemental Table S4.

Communication—While taxonomy, biodiversity and conservation sciences exchange data and information, scientific silos must be overcome particularly as they relate to regional resources and single taxon expertise [6,90]. The GIST can enhance communication as it gives a standardized vocabulary that can be readily integrated across databases, communities, and disciplines, providing terms that can be searched in disparate data types. However, this requires clear and sustained communication between actors who may be responding to different priorities and values. To achieve this, training in data standards and management, curation of metadata and proper application of taxonomic rules are some key areas to develop and are already facilitated through networks like the Integrated Digitized Biocollections annual conference (https://www.idigbio.org/) and the Biodiversity Information Standards (TDWG, https://www.tdwg.org/). A specific GIST working group could be established, following the model of the TDWG working groups, engaging with data aggregators and taxonomic databases.

Synergies—The GIST lays out a basis of standards to make biodiversity data and information readily accessible to users, most of whom are not taxonomic experts. Establishing an interdisciplinary community representing all actors central to the future of interoperable taxonomically informed projects must become a priority (Figure 3). The involvement of stakeholders from taxonomists to policymakers has developed with the IUBS Working Group on the Governance of Taxonomic Lists in relation to COL, and the Global Taxonomic Initiative (GTI) in support of the post-2020 Global Biodiversity Framework [46,77]. The GTI recommends stronger links between all stakeholders from taxonomy, biodiversity, and conservation, but it remains unclear how this immense task could be achieved over the coming decade. This requires greater awareness of user needs via dedicated and trained leaders who can navigate the networks of people working with taxonomy via collective leadership and coalition building [7,37,62,91].

Investments—Building scientific communication into outcomes requires sustained networks and informatics investment to ensure that the data are appropriately maintained and usable. Similarly to data standards in place for essential biodiversity variables, primary biodiversity data, and monitoring networks [4,76,92], we recommend a focus on incorporating and expanding the use of GIST in existing datasets. Just as researchers must submit plans for data storage and sharing in grant applications, intention in ensuring interoperability is equally critical and could be a component of publication and digital infrastructure funds. Developing funding to support the community of taxonomic data producers and users are necessary to further enhance the potential for integrative science and conservation and need to move away from the volunteer basis to ensure engagement and participation from taxonomists and beyond.

CONCLUDING REMARKS

After several centuries of naming organisms on Earth, barriers to a globally integrated assessment of the diversity of life remain (Box 1). Taxonomy is the foundation of biodiversity synthesis and conservation, and taxonomic data are central for the integration of data sources influencing research, conservation, and management practice. Growing impacts of global change on biodiversity highlight the urgency of insisting and renewing vigor in valuing, funding, and developing taxonomic integrative science and its interdisciplinary community. Important challenges need to be addressed by the scientific community to realize the full potential of taxonomic data to support biodiversity and conservation (see Outstanding Questions). Rather than fragmented data and social infrastructures [12,60,93], mechanisms for a GIST (guiding principles from Figure 3) have the potential to enhance new paradigms at the intersection of taxonomy, ecology, and conservation.

ACKNOWLEDGMENTS

This project originated from discussions in the taxon expert group hosted at the Center for Biodiversity and Global Change at Yale University, partly funded by the E.O. Wilson Foundation, including E.L.S, A.A.M., Y.V.S., M.R., K.I, D.B.B., S.P., R.E., M.L., C.J.M., A.R., and led by C.J.M for 2020-2021 and E.L.S. for 2021-2022. We thank Scott Loarie for valuable discussions, and members of the Center for Biodiversity and Global Change and the Map of Life project for general feedback during the development of this project.

REFERENCES

1. Godfray, H.C.J. (2007) Linnaeus in the information age. *Nature* 446, 259–260
2. Costello, M.J. *et al.* (2015) Conservation of biodiversity through taxonomy, data publication, and collaborative infrastructures. *Conserv. Biol.* 29, 1094–1099
3. Morrison, W.R. *et al.* (2009) The impact of taxonomic change on conservation: Does it kill, can it save, or is it just irrelevant? *Biol. Conserv.* 142, 3201–3206
4. Jetz, W. *et al.* (2019) Essential biodiversity variables for mapping and monitoring species populations. *Nat. Ecol. Evol.* 3, 539–551
5. Sigwart, J.D. *et al.* (2018) Measuring Biodiversity and Extinction—Present and Past. *Integr. Comp. Biol.* 58, 1111–1117
6. Heberling, J.M. *et al.* (2021) Data integration enables global biodiversity synthesis. *Proc. Natl. Acad. Sci.* 118
7. Wüster, W. *et al.* (2021) Confronting taxonomic vandalism in biology: conscientious community self-organization can preserve nomenclatural stability. *Biol. J. Linn. Soc.* 133, 645–670
8. Hey, J. *et al.* (2003) Understanding and confronting species uncertainty in biology and conservation. *Trends Ecol. Evol.* 18, 597–603
9. Stropp, J. *et al.* (2022) Taxonomic uncertainty and the challenge of estimating global species richness. *J. Biogeogr.* 49, 1654–1656
10. Clark, J.A. and May, R.M. (2002) Taxonomic Bias in Conservation Research. *Science* 297, 191–192
11. Isaac, N.J.B. *et al.* (2004) Taxonomic inflation: its influence on macroecology and conservation. *Trends Ecol. Evol.* 19, 464–469
12. Garnett, S.T. and Christidis, L. (2017) Taxonomy anarchy hampers conservation. *Nature* 546, 25–27
13. Vogel Ely, C. *et al.* (2017) Implications of poor taxonomy in conservation. *J. Nat. Conserv.* 36, 10–13
14. Correia, R.A. *et al.* (2018) Nomenclature instability in species culturomic assessments: Why synonyms matter. *Ecol. Indic.* 90, 74–78
15. McGowan, J. *et al.* (2020) Conservation prioritization can resolve the flagship species conundrum. *Nat. Commun.* 11, 994

16. Montoya, R.D. (2022) *Power of position: classification and the biodiversity sciences*, The MIT Press
17. Winston, J.E. (2018) Twenty-First Century Biological Nomenclature—The Enduring Power of Names. *Integr. Comp. Biol.* 58, 1122–1131
18. Thomson, S.A. *et al.* (2018) Taxonomy based on science is necessary for global conservation. *PLOS Biol.* 16, e2005075
19. Agapow, P. *et al.* (2004) The Impact of Species Concept on Biodiversity Studies. *Q. Rev. Biol.* 79, 161–179
20. König, C. *et al.* (2019) Biodiversity data integration—the significance of data resolution and domain. *PLOS Biol.* 17, e3000183
21. Feng, X. *et al.* (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
22. Bisby, F.A. (2000) The Quiet Revolution: Biodiversity Informatics and the Internet. *Science* 289, 2309–2312
23. Miller, J.T. and Jolley-Rogers, G. (2014) Correcting the disconnect between phylogenetics and biodiversity informatics. *Zootaxa* 3754, 195–200
24. Deans, A.R. *et al.* (2012) Time to change how we describe biodiversity. *Trends Ecol. Evol.* 27, 78–84
25. Riedel, A. *et al.* (2013) Integrative taxonomy on the fast track - towards more sustainability in biodiversity research. *Front. Zool.* 10, 15
26. Mace, G.M. (2004) The Role of Taxonomy in Species Conservation. *Philos. Trans. Biol. Sci.* 359, 711–719
27. van Klink, R. *et al.* (2022) Emerging technologies revolutionise insect ecology and monitoring. *Trends Ecol. Evol.* 37, 872–885
28. Gorbalenya, A.E. *et al.* (2020) The species Severe acute respiratory syndrome-related coronavirus: classifying 2019-nCoV and naming it SARS-CoV-2. *Nat. Microbiol.* 5, 536–544
29. Forster, P. *et al.* (2020) Phylogenetic network analysis of SARS-CoV-2 genomes. *Proc. Natl. Acad. Sci.* 117, 9241–9243
30. Wang, C. *et al.* (2020) The establishment of reference sequence for SARS-CoV-2 and variation analysis. *J. Med. Virol.* 92, 667–674
31. Smith, G. *et al.* (2011) South Africa’s ongoing Opuntia Mill. (Cactaceae) problem: the case of O. microdasys (Lehm.) Pfeiff. *Bradleya* 29
32. Pinkert, S. *et al.* (2022) Global geographical and latitudinal variation in butterfly species richness captured through a comprehensive country-level occurrence database. *Glob. Ecol. Biogeogr.* 31, 830–839
33. Macgregor, C.J. *et al.* (2019) Climate-induced phenology shifts linked to range expansions in species with multiple reproductive cycles per year. *Nat. Commun.* 10, 4455

34. Edwards, J.L. *et al.* (2000) Interoperability of Biodiversity Databases: Biodiversity Information on Every Desktop. *Science* 289, 2312–2314
35. Franz, N.M. *et al.* (2016) Two Influential Primate Classifications Logically Aligned. *Syst. Biol.* 65, 561–582
36. Sen, A. *et al.* (2020) The Automated Taxonomic Concept Reasoner. *Biodivers. Inf. Sci. Stand.* 4, e59074
37. Sterner, B.W. *et al.* (2020) Decentralized but Globally Coordinated Biodiversity Data. *Front. Big Data* 3, 519133
38. de Jong, Y. *et al.* (2015) PESI - a taxonomic backbone for Europe. *Biodivers. Data J.* DOI: 10.3897/BDJ.3.e5848
39. Bénichou, L. *et al.* (2018) Consortium of European Taxonomic Facilities (CETAF) best practices in electronic publishing in taxonomy. *Eur. J. Taxon.* DOI: 10.5852/ejt.2018.475
40. Lepage, D. *et al.* (2014) Avibase – a database system for managing and organizing taxonomic concepts. *ZooKeys* 420, 117–135
41. McClure, C.J.W. *et al.* (2020) Towards reconciliation of the four world bird lists: hotspots of disagreement in taxonomy of raptors. *Proc. R. Soc. B Biol. Sci.* 287, 20200683
42. Wilson, D.E. and Reeder, D.M. (2005) *Mammal Species of the World: A Taxonomic and Geographic Reference*, ((3rd edn)), Johns Hopkins University Press, Baltimore, Maryland
43. Costello, M.J. *et al.* (2013) Global Coordination and Standardisation in Marine Biodiversity through the World Register of Marine Species (WoRMS) and Related Databases. *PLOS ONE* 8, e51629
44. Kindt, R. (2020) WorldFlora: An R package for exact and fuzzy matching of plant names against the World Flora Online taxonomic backbone data. *Appl. Plant Sci.* 8, e11388
45. Britz, R. *et al.* (2020) Funding, training, permits—the three big challenges of taxonomy. *Megataxa* 1
46. Conix, S. *et al.* (2021) Towards a global list of accepted species III. Independence and stakeholder inclusion. *Org. Divers. Evol.* 21, 631–643
47. Lien, A.M. *et al.* (2021) Towards a global list of accepted species IV: Overcoming fragmentation in the governance of taxonomic lists. *Org. Divers. Evol.* 21, 645–655
48. Uetz, P. *et al.* (2021) A Quarter Century of Reptile and Amphibian Databases. *Herpetol. Rev.* 52, 246–255
49. Froese, R. and Pauly, D. (2022) FishBase. World Wide Web electronic publication.[Online]. Available: https://www.fishbase.se/summary/citation.php. [Accessed: 15-Aug-2022]
50. Frost, D.R. (1999) Amphibian Species of the World: an Online ReferenceAmerican Museum of Natural History

51. Dmitriev, D. *et al.* (2022) TaxonWorks as a Tool for Managing Large Biodiversity Projects. *Biodivers. Inf. Sci. Stand.* 6, e93668
52. Gries, C. *et al.* (2014) Symbiota – A virtual platform for creating voucher-based biodiversity information communities. *Biodivers. Data J.* 2, e1114
53. Banki, O. *et al.* (2019) Catalogue of Life Plus: A collaborative project to complete the checklist of the world’s species. *Biodivers. Inf. Sci. Stand.* 3, e37652
54. Hobern, D. *et al.* (2021) Towards a global list of accepted species VI: The Catalogue of Life checklist. *Org. Divers. Evol.* 21, 677–690
55. (2022) Integrated Taxonomic Information System (ITIS)[Online]. Available: http://www.itis.gov. [Accessed: 05-Oct-2022]
56. Global Biodiversity Information Facility (GBIF)[Online]. Available: https://www.gbif.org/. [Accessed: 05-Oct-2022]
57. Wieczorek, J. *et al.* (2012) Darwin Core: An Evolving Community-Developed Biodiversity Data Standard. *PLOS ONE* 7, e29715
58. Leonelli, S. (2013) Classificatory Theory in Biology. *Biol. Theory* 7, 338–345
59. Jetz, W. *et al.* (2012) Integrating biodiversity distribution knowledge: toward a global map of life. *Trends Ecol. Evol.* 27, 151–159
60. Grenié, M. *et al.* Harmonizing taxon names in biodiversity data: A review of tools, databases and best practices. *Methods Ecol. Evol.* n/a
61. Rolin, K. (2015) Values in Science: The Case of Scientific Collaboration. *Philos. Sci.* 82, 157–177
62. Manolis, J.C. *et al.* (2009) Leadership: a New Frontier in Conservation Science. *Conserv. Biol.* 23, 879–886
63. Wheeler, Q.D. *et al.* (2004) Taxonomy: Impediment or Expedient? *Science* 303, 285–285
64. Kühl, H.S. *et al.* (2020) Effective Biodiversity Monitoring Needs a Culture of Integration. *One Earth* 3, 462–474
65. Raposo, M.A. *et al.* (2017) What really hampers taxonomy and conservation? A riposte to Garnett and Christidis. *Zootaxa* 4317, 179–184
66. Pyle, R.L. (2016) Towards a Global Names Architecture: The future of indexing scientific names. *ZooKeys* DOI: 10.3897/zookeys.550.10009
67. Guralnick, R.P. *et al.* (2007) Towards a collaborative, global infrastructure for biodiversity assessment. *Ecol. Lett.* 10, 663–672
68. Sigovini, M. *et al.* (2016) Open Nomenclature in the biodiversity era. *Methods Ecol. Evol.* 7, 1217–1225
69. Norman, K.E.A. *et al.* (2020) taxadb: A high-performance local taxonomic database interface. *Methods Ecol. Evol.* 11, 1153–1159
70. Chamberlain, S.A. and Szöcs, E. (2013) taxize: taxonomic search and retrieval in RF1000Research
71. Barve, V. (2020) Taxotools: Tools to handle taxonomic lists. R package.Zenodo

72. Troudet, J. *et al.* (2017) Taxonomic bias in biodiversity data and societal preferences. *Sci. Rep.* 7, 9132

73. Pagad, S. *et al.* (2018) Introducing the Global Register of Introduced and Invasive Species. *Sci. Data* 5, 170202

74. The IUCN Red List of Threatened Species. *IUCN Red List of Threatened Species*. [Online]. Available: https://www.iucnredlist.org/en. [Accessed: 05-Oct-2022]

75. Sandall, E.L. *et al.* (2022) Country-level checklists and occurrences for the world's Odonata (dragonflies and damselflies). *J. Biogeogr.* 49, 1586–1598

76. Pagad, S. *et al.* (2022) Country Compendium of the Global Register of Introduced and Invasive Species. *Sci. Data* 9, 391

77. Abrahamse, T. *et al.* (2021) *The Global Taxonomy Initiative in Support of the Post-2020 Global Biodiversity Framework.*, Secretariat of the Convention of Biological Diversity

78. Padial, J.M. *et al.* (2010) The integrative future of taxonomy. *Front. Zool.* 7, 16

79. Costello, M.J. *et al.* (2018) Sustainable Biodiversity Databasing: International, Collaborative, Dynamic, Centralised. *Trends Ecol. Evol.* 33, 803–805

80. Upham, N. *et al.* (2022) Extended Taxonomic Curation: Moving beyond species lists to linking species data. *Biodivers. Inf. Sci. Stand.* 6, e93670

81. Pyle, R. *et al.* (2022) Modeling Taxon Concepts: A new approach to an old problem. *Biodivers. Inf. Sci. Stand.* 6, e93927

82. International Commission on Zoological Nomenclature *et al.*, eds. (1999) *International code of zoological nomenclature: Code international de nomenclature zoologique*, (4th ed.), International Trust for Zoological Nomenclature, c/o Natural History Museum

83. Turland, N. *et al.*, eds. (2018) *International Code of Nomenclature for algae, fungi, and plants*, 159, Koeltz Botanical Books

84. IPNI (2022) International Plant Names Index[Online]. Available: https://www.ipni.org/. [Accessed: 21-Oct-2022]

85. Hobern, D. *et al.* (2019) Connecting data and expertise: a new alliance for biodiversity knowledge. *Biodivers. Data J.* 7, e33679

86. Marinho, L.C. and Beech, E. (2020) How phantom databases could contribute to conservation assessments. *Sci. Nat.* 107, 21

87. Gallagher, R.V. *et al.* (2020) Open Science principles for accelerating trait-based science across the Tree of Life. *Nat. Ecol. Evol.* 4, 294–303

88. Bionomia[Online]. Available: https://bionomia.net/. [Accessed: 20-Oct-2022]

89. Wilkinson, M.D. *et al.* (2016) The FAIR Guiding Principles for scientific data management and stewardship. *Sci. Data* 3, 160018

90. Hochkirch, A. *et al.* (2021) A strategy for the next decade to address data deficiency in neglected biodiversity. *Conserv. Biol.* 35, 502–509

91. Clark, S.G. and Wallace, R.L. (2015) Integration and interdisciplinarity: concepts, frameworks, and education. *Policy Sci.* 48, 233–255

92. Kissling, W.D. *et al.* (2018) Building essential biodiversity variables (EBVs) of species distribution and abundance at a global scale. *Biol. Rev.* 93, 600–625

93. Jackson, M.D. *et al.* (2017) Taxonomy is not beholden to its dependencies: a rebuttal to Garnett and Christidis. *PeerJ Prepr.* 5:e3060v1
94. Burgin, C.J. *et al.* (2018) How many species of mammals are there? *J. Mammal.* 99, 1–14
95. Cronin, M.A. *et al.* (2013) Genetic Variation and Differentiation of Bison (Bison bison) Subspecies and Cattle (Bos taurus) Breeds and Subspecies. *J. Hered.* 104, 500–509
96. Uetz, P. (2022) The Reptile Database[Online]. Available: http://www.reptile-database.org/. [Accessed: 15-Aug-2022]
97. Odonata Central - World Odonata List[Online]. Available: https://www.odonatacentral.org/app/#/wol/. [Accessed: 11-Oct-2022]
98. Freiberg, M. *et al.* (2020) LCVP, The Leipzig catalogue of vascular plants, a new taxonomic reference list for all known vascular plants. *Sci. Data* 7, 416
99. Borsch, T. *et al.* (2020) World Flora Online: Placing taxonomists at the heart of a definitive and comprehensive global resource on the world’s plants. *TAXON* 69, 1311–1341
100. WFO (2022): World Flora Online.[Online]. Available: http://www.worldfloraonline.org/. [Accessed: 05-Oct-2022]
101. Plants of the World Online | Kew Science. *Plants of the World Online*. [Online]. Available: https://powo.science.kew.org/. [Accessed: 05-Oct-2022]
102. Banki, O. *et al.* (2019) Catalogue of Life Plus: A collaborative project to complete the checklist of the world’s species. *Biodivers. Inf. Sci. Stand.* DOI: http://dx.doi.org/10.3897/biss.3.37652
103. WoRMS Editorial Board (2022) World Register of Marine Species.VLIZ
104. Parr, C.S. *et al.* (2014) The Encyclopedia of Life v2: Providing Global Access to Knowledge About Life on Earth. *Biodivers. Data J.* 2, e1079
105. Ocean Biodiversity Information System[Online]. Available: https://obis.org/. [Accessed: 05-Oct-2022]
106. Loarie, S. Introducing Taxon Frameworks. *iNaturalist*. [Online]. Available: https://www.inaturalist.org/blog/19977-introducing-taxon-frameworks. [Accessed: 05-Oct-2022]
107. Schoch, C.L. *et al.* (2020) NCBI Taxonomy: a comprehensive update on curation, resources and tools. *Database* 2020
108. Rees, J. and Cranston, K. (2017) Automated assembly of a reference taxonomy for phylogenetic data synthesis. *Biodivers. Data J.* 5, e12581
109. Palomares, M.L.D. and Pauly, D. (2022) SeaLifeBase. World Wide Web electronic publication.[Online]. Available: www.sealifebase.org; version (04/2022). [Accessed: 27-Oct-2022]
110. Weigelt, P. *et al.* (2020) GIFT – A Global Inventory of Floras and Traits for macroecology and biogeography. *J. Biogeogr.* 47, 16–43

FIGURES

Figure 1. Research themes and examples with associated data types relying on taxonomic integration. Innermost ring (dark blue): main data categories. Center ring (lighter blue gradient): data applications (foundational to applied) across the four data categories. Outer ring (gray gradient; from taxonomy to biodiversity to conservation): example research questions and applications (from taxonomy, biodiversity, and conservation). Arrows on the outermost edge of the rings denotes a linkage with one of three examples (A, B, C), illustrating how integration facilitates a transparent connection between primary data, biodiversity analysis and practice and could avoid problems downstream: (A) SARS-CoV-2 global designation and identification. (B) Management of invasive plants in the genus *Opuntia*. (C) Spatial range comparison of the butterfly *Plebejus argus*, characterized by 160 synonyms. Photo sources and credits are documented in Online Supplemental Information Table S6.

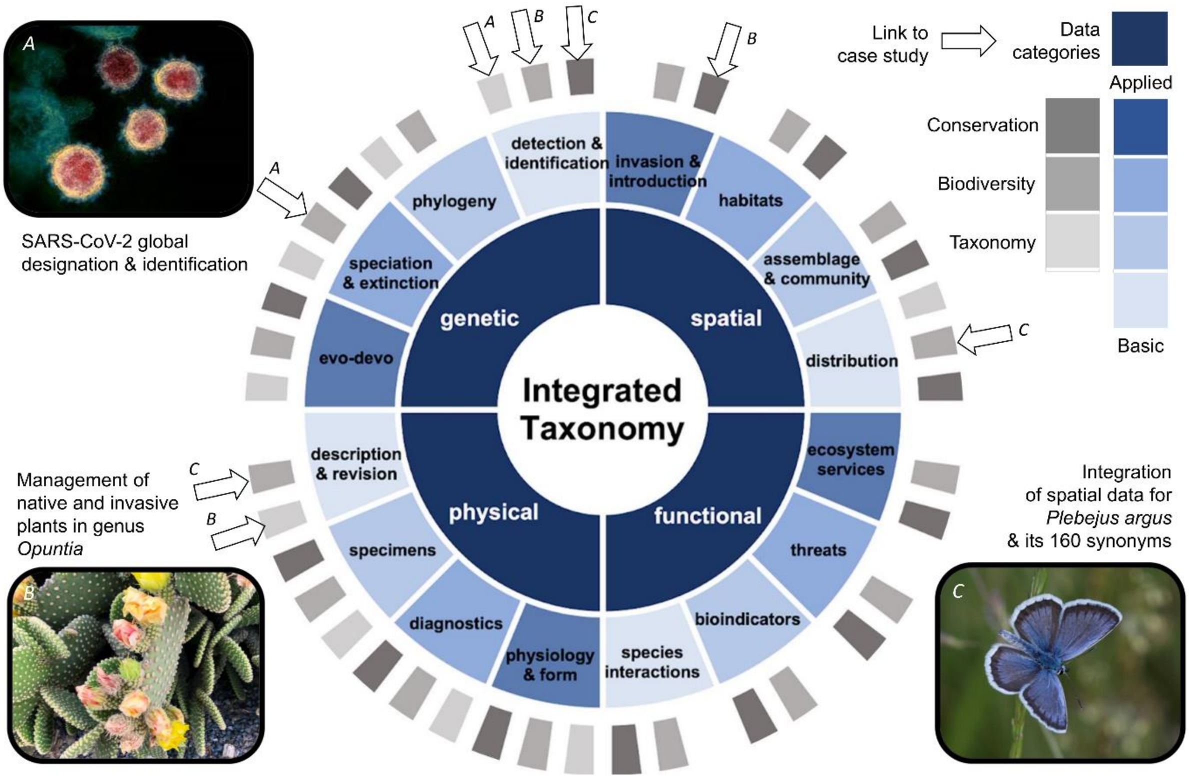


Figure 2. Foundational elements of the Globally Integrated Structure of Taxonomy (GIST). Illustration of each GIST element using *Bison bison* as an example. Accepted as a species in Mammal Species of the World [94] version 3 from 2005, with subspecies synonyms lumped based on phylogenetic evidence [95], there are occurrence records in GBIF as of 25th February 2022 under three scientific names: *Bison bison bison*, *Bison bison athabascae,* and *Bison bison*. See also Online Supplemental Information Table S2, and Online Supplemental Information Table S6 for photo sources and credits.

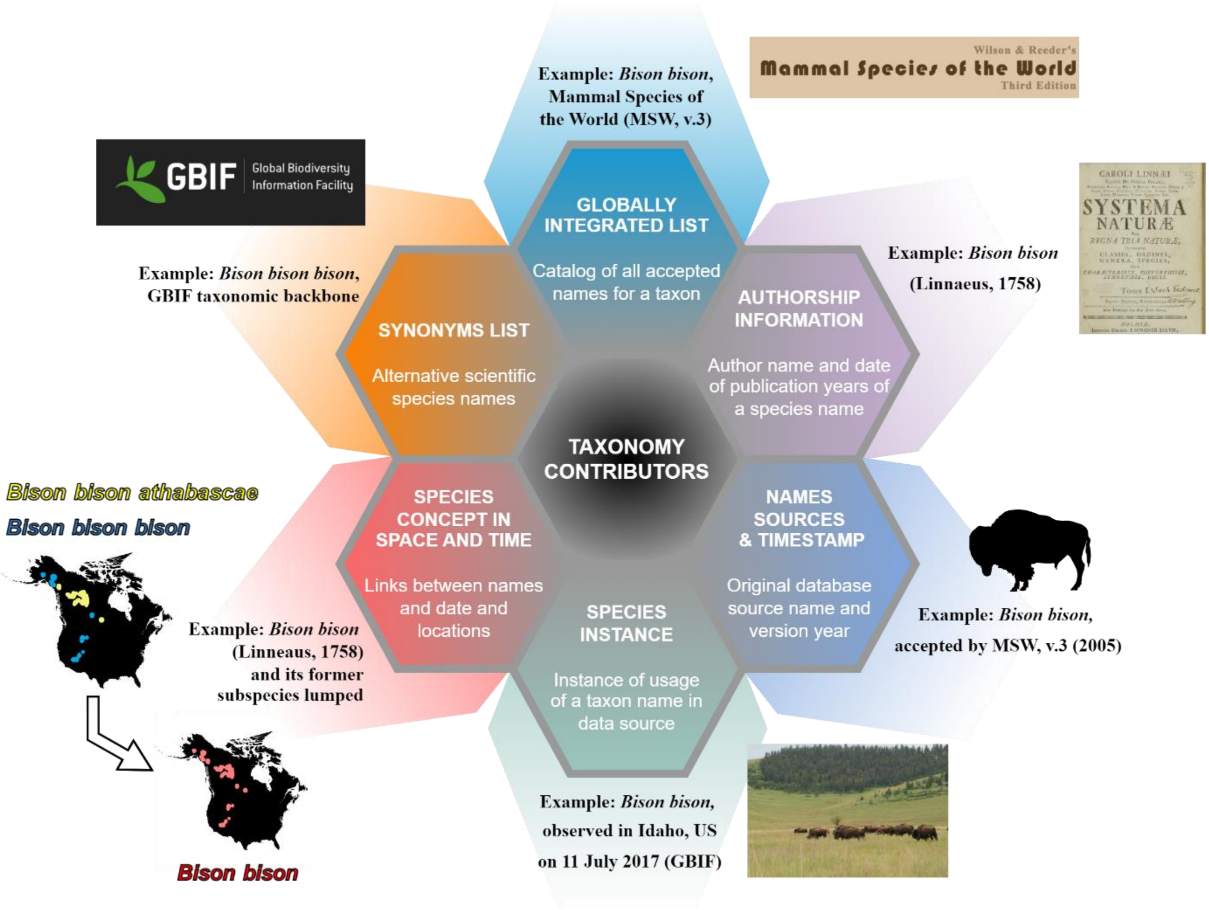
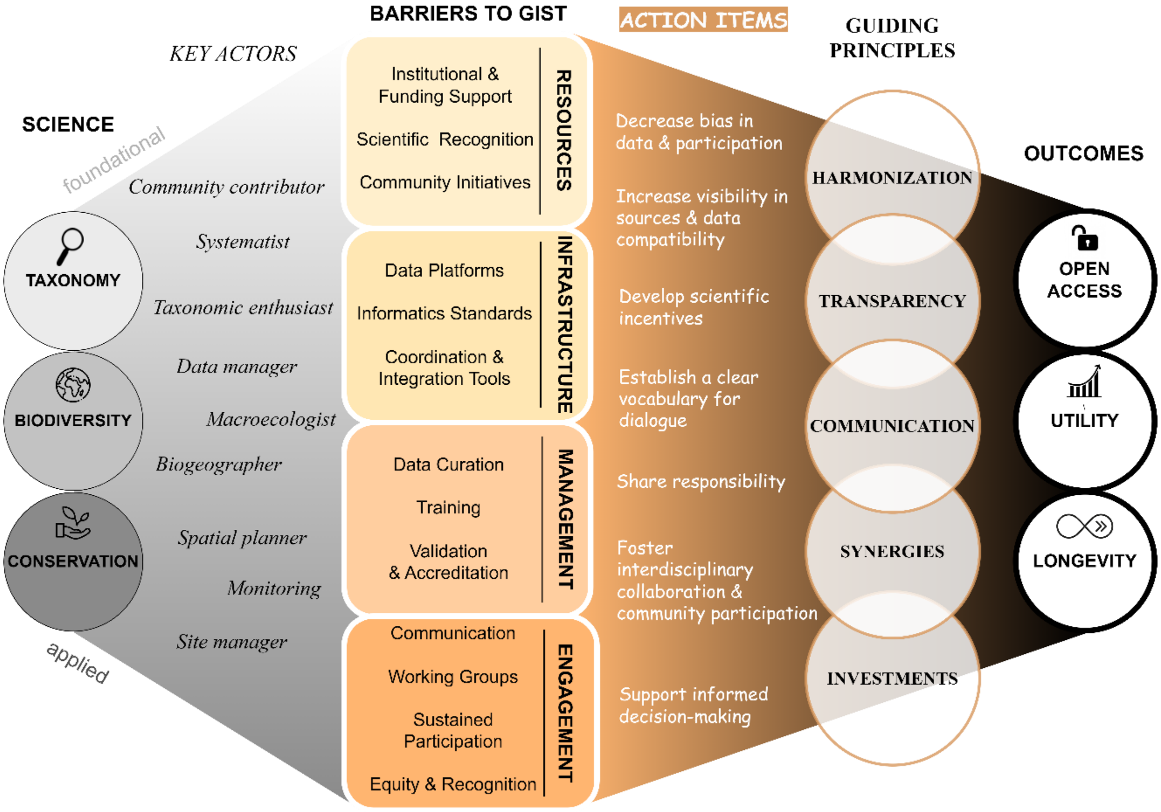


Figure 3. Recommendations for a standardized and user-friendly infrastructure of taxonomy. Key actors from foundational to applied sciences (left circles) are acting to overcome the main challenges to a GIST (center boxes) by following principles ultimately leading to better outcomes of access, utility, and longevity of taxonomy.



TABLE

Table 1: Overview of the three levels of taxonomic databases and their respective goals and sources. The list is non-exhaustive, with many more taxonomic-oriented databases and projects that exist and cover a wider range of taxa. Outcomes are described using the following acronyms: taxonomic classification (TC), ecological description (ED), trait data (TD), spatial data (SD), genetic data (GD), phylogenetic data (PD), citizen participation (CP), decision-making (DM), species protection (SP). * indicates a database belonging to primary and secondary database levels.

Database	Taxonomic scope	Source(s) of name data	Database dependency	Outcomes	Year created	Refs
Primary database producing novel backbones						
Amphibian Species of the World (ASW)	Amphibians			TC	1980	[50]
FishBase	Marine and fresh. fishes	Catalog of Fishes		TC, ED, TD	1987	[49]
Reptile Database	Reptiles			TC, ED, TD	1995	[48,96]
AviBase	Birds			TC	2003	[40]
World Odonata List (WOL)	Odonates			TC	2005	[97]
Mammal Diversity Database (MDD)	Mammals			TC	2018	[94]
Leipzig Catalog of Vascular Plants (LCVP)	Plants	13 sources incl. POWO, COL, ITIS	Multiple	TC	2020	[98]
World Flora Online (WFO)	Plants	The Plant List		TC	2012	[99,100]
Plants of the World Online (POWO)	Plants	World Checklist of Selected Plant Families	One	TC	2017	[101]

Secondary databases combining primary taxonomic lists, linking primary databases						
Integrated Taxonomic Information System (ITIS)*	Any taxa	Incl. WOL	Multiple	TC	1996	[55]
Catalogue of Life (COL+)	Any taxa	165 sources	Multiple	TC	2001	[102]
World Register of Marine Species (WoRMS)*	Marine, select fresh. and terr. taxa	European Register of Marine Species, FishBase	Multiple	TC	2007	[103]
Encyclopedia of Life (EoL)	Any taxa	712 datasets	Multiple	TC	2008	[104]
Tertiary databases whose primary focus is not taxonomic but aggregating biodiversity data						
Global Biodiversity Information Facility (GBIF)	Any taxa	eBird, iNaturalist, COL	Multiple	SD	1999	[56]
Ocean Biodiversity Information System (OBIS)	Any marine taxa	WoRMS	One	SD	2000	[105]
iNaturalist	Any taxa	25, incl. WOL, WoRMS, FishBase, ASW, IUCN, Reptile Database, POWO	Multiple	CP, SD, ED, DM	2008	[106]
Map of Life (MOL)	Select verts., plants, inverts.	Incl. AviBase, ASW, WOL, Reptile Database, WoRMS, GBIF, IUCN	Multiple	SD, ED, DM	2012	[59]
GenBank (NCBI)	Any taxa	>20, incl. MDD, ASW, WoRMS,	Multiple	GD	1979	[107]

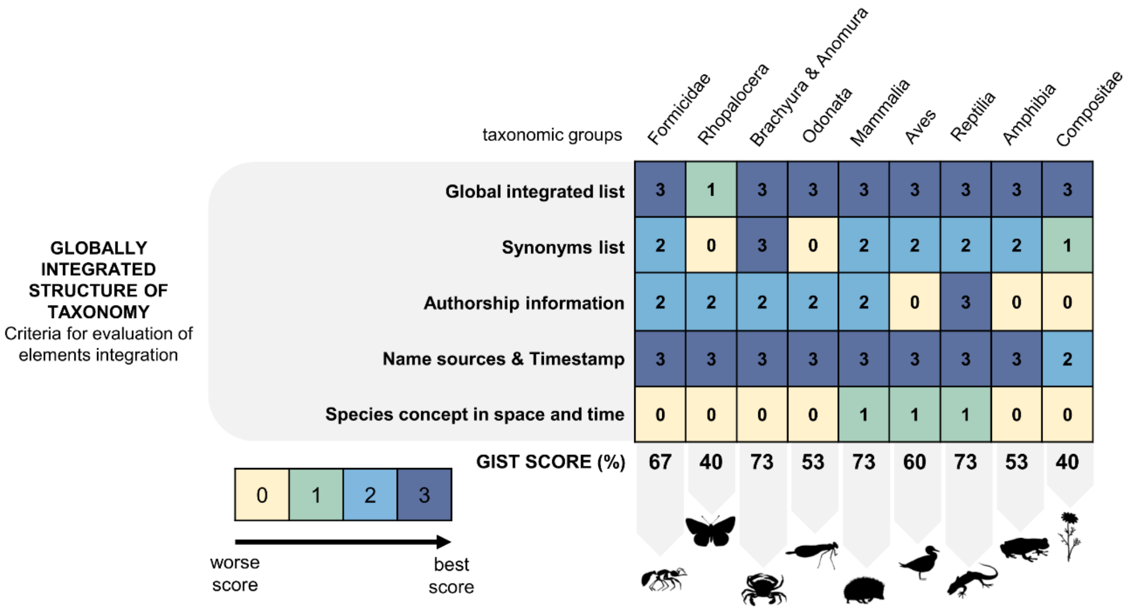
		ITIS, WFO, Reptile Database				
Open Tree of Life	Any taxa	10 incl. WoRMS, NCBI, GBIF	Multiple	PD	2015	[108]
SeaLifeBase	Marine and fresh. inverts.	WoRMS	One	ED, TD	2005	[109]
Global Inventory of Floras and Traits (GIFT)	Plants	The plant list iPlant	Multiple	TD	2020	[110]
International Union for the Conservation of Nature Red List of Threatened Species (IUCN)	Some animals, fungi, plant taxa	Unknown	Likely Multiple	SP, DM	1964	[74]
Global Register of Introduced and Invasive Species (GRIIS)	Any taxa	GBIF, MOL	Multiple	ED, DM, SP	2006	[73,76]

TEXT BOX

Box 1. Assessment of GIST coverage and interoperability for nine example taxa.

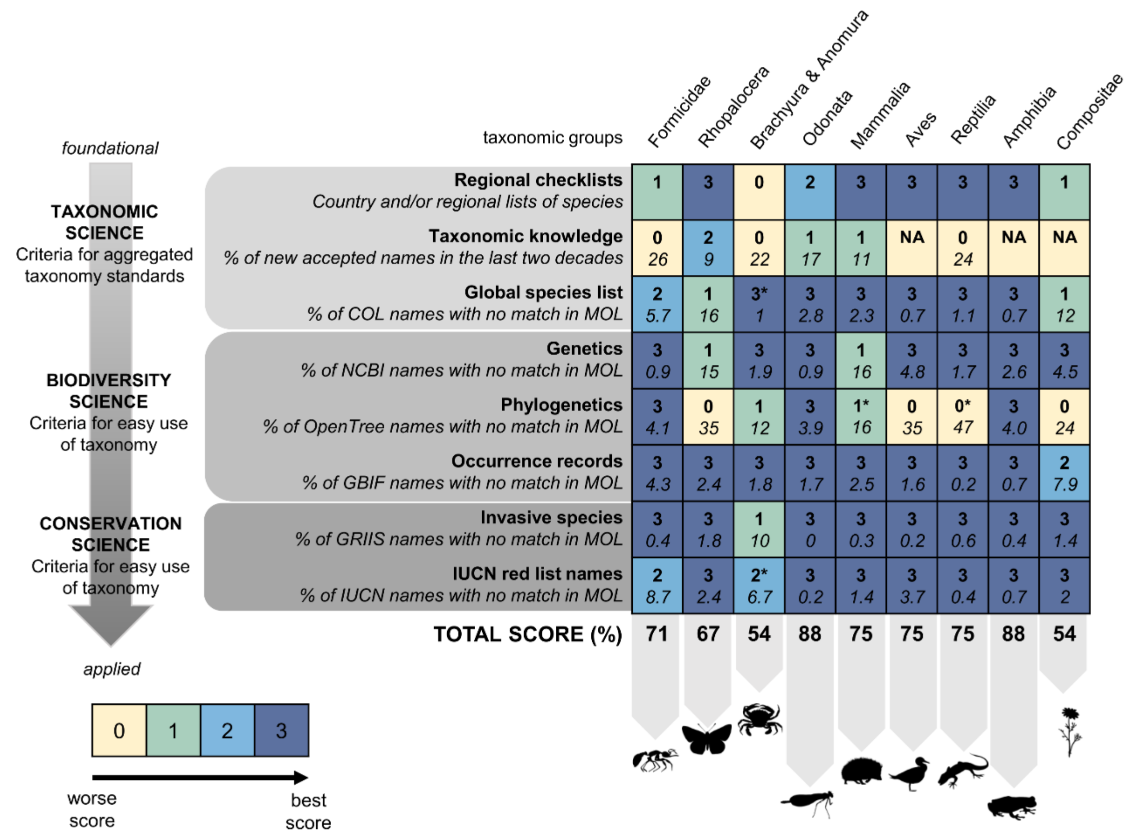
We use the Map of Life [59] taxonomic database to evaluate availability and data interoperability of the GIST elements.

Figure I. Evaluation of the GIST elements for nine taxa. Taxa were graded for each element according to a specific criteria/metric (italic). The length of arrows at the bottom are proportional to the GIST score attributed to each taxon.



taxonomic groups	Formicidae	Rhopalocera	Brachyura & Anomura	Odonata	Mammalia	Aves	Reptilia	Amphibia	Compositae
Global integrated list	3	1	3	3	3	3	3	3	3
Synonyms list	2	0	3	0	2	2	2	2	1
Authorship information	2	2	2	2	2	0	3	0	0
Name sources & Timestamp	3	3	3	3	3	3	3	3	2
Species concept in space and time	0	0	0	0	1	1	1	0	0
GIST SCORE (%)	67	40	73	53	73	60	73	53	40

Figure II. Evaluation of name data integration in biodiversity and conservation for nine taxa. Following the approach in Figure I, we evaluated linkages and data availability for an entire taxon from taxonomy, biodiversity, and conservation sciences. Arrow lengths at the bottom of the matrix are proportional to the score attributed to each taxon. '*' specifies when the taxon could only be partly matched due to data availability or inability to access part of the data. 'NA' specifies when the taxon could not be assessed because of lack of integration of the GIST elements.



taxonomic groups	Formicidae	Rhopalocera	Brachyura & Anomura	Odonata	Mammalia	Aves	Reptilia	Amphibia	Compositae
Regional checklists *Country and/or regional lists of species*	1	3	0	2	3	3	3	3	1
Taxonomic knowledge *% of new accepted names in the last two decades*	0 *26*	2 *9*	0 *22*	1 *17*	1 *11*	NA	0 *24*	NA	NA
Global species list *% of COL names with no match in MOL*	2 *5.7*	1 *16*	3* *1*	3 *2.8*	3 *2.3*	3 *0.7*	3 *1.1*	3 *0.7*	1 *12*
Genetics *% of NCBI names with no match in MOL*	3 *0.9*	1 *15*	3 *1.9*	3 *0.9*	1 *16*	3 *4.8*	3 *1.7*	3 *2.6*	3 *4.5*
Phylogenetics *% of OpenTree names with no match in MOL*	3 *4.1*	0 *35*	1 *12*	3 *3.9*	1* *16*	0 *35*	0* *47*	3 *4.0*	0 *24*
Occurrence records *% of GBIF names with no match in MOL*	3 *4.3*	3 *2.4*	3 *1.8*	3 *1.7*	3 *2.5*	3 *1.6*	3 *0.2*	3 *0.7*	2 *7.9*
Invasive species *% of GRIIS names with no match in MOL*	3 *0.4*	3 *1.8*	1 *10*	3 *0*	3 *0.3*	3 *0.2*	3 *0.6*	3 *0.4*	3 *1.4*
IUCN red list names *% of IUCN names with no match in MOL*	2 *8.7*	3 *2.4*	2* *6.7*	3 *0.2*	3 *1.4*	3 *3.7*	3 *0.4*	3 *0.7*	3 *2*
TOTAL SCORE (%)	71	67	54	88	75	75	75	88	54

Methods

Each taxonomy was compiled thanks to the expertise of scientists working with taxonomy integrated into Map of Life. For each taxon group, we selected and integrated available up-to-date sources (see Online Supplemental Information Text S1 and Table S3 for methods and sources).

In Figure I, we attributed grades (0–3) according to specific qualitative and semi-quantitative criteria (see Online Supplemental Information Table S4) and derived overall ranking from the percentage of the sum score relative to potential maximum score.

In Figure II, we assessed taxonomic data interoperability in different fields for the same taxa. We selected key databases from taxonomic science (COL), biodiversity science (NCBI, Open Tree of Life, GBIF), and conservation science and practice (GRIIS, IUCN red list assessment; see Table 1 for details on these data sources, and Online Supplemental Information Text S2 for methods on data processing). Taxa are graded (0–3) in each category according to the name linkage between databases and attributed a percentage based on the sum score relative to potential maximum score. See Online Supplemental Information Table S5 for details on grading criteria.

GLOSSARY

Accepted name. The scientific name of a taxon that has been formally validated and approved by scientific experts, with a published description, related to an identifiable species concept with a clear lineage and a known type specimen.

Binomial nomenclature. System of naming species using two Latin terms, genus (rank above species) and specific epithet.

Biodiversity data aggregator. A digital platform for collecting and sharing biodiversity data.

Biodiversity conservation. Scientific discipline and practice for maintaining and protecting natural resources and ecosystems.

Conservation manager. An individual responsible for actions in an organization aiming at the protection of the environment, landscape, seascape, biodiversity, and/or wildlife.

Decision-maker. An individual responsible for making strategic decisions based on multiple variables and dependent on the amount of information available.

Global species list. List of accepted names covering an entire taxonomic group defined by all species contained in a taxonomic rank (family, order, class, kingdom). It can be a preferred taxonomic authority or a compilation of accepted names in absence of a defined authority.

Harmonization. Process of joining and integrating data from multiple sources to make a unified dataset.

Integrative science. Science that brings together multiple disciplines, taxonomic groups, spatial, temporal, and organizational scales, and/or communities, and allows exploring and testing new paradigms to transform current practices.

Interdisciplinary science. Science related to more than one discipline.

Interoperability. Ability for databases or systems to exchange information without effort from the end user.

Scientific nomenclature. Recognized scientific names of organisms, typically a binomial name including genus and species.

Species. Group of organisms that can be considered one taxonomic unit, typically as the lowest taxonomic rank that has an accepted name.

Species concept. Description of delineating traits that represents a taxonomic unit and can change over time with new data or specimen evidence.

Species splitting and lumping. Used in the context of changing application of a species name due to varying taxonomic opinion, whereby a species name is divided into several names or several names are grouped into one name. This is distinct from the process of adding new species or synonyms.

Specimen. Physical example of an organism.

Synonym. Alternative names to the accepted names. These names are other names referring to a species concept. In a taxonomic backbone scheme, these names are appended to the species list as a “child” term to accepted names when clear matching can be done with accepted names.

Taxonomic backbone. A data structure for matching taxonomic synonyms to accepted names, within a hierarchy.

Taxonomic integration. Integrative science focused on new inferences between taxonomies, or between taxonomies and other products or disciplines.

Taxon. A term denoting a commonly recognized unit or collective of organisms. Also called taxonomic group.

Taxonomist. An individual who identifies, classifies, or describes taxa.

Taxonomy. Science of the classification of organisms.

Values. The moral, societal, or epistemic basis for actions.

Vernacular name. A common, non-scientific name for an organism, which may be regional or draw on the features of an organism.

OUTSTANDING QUESTIONS

- The GIST is foundational to enhance the access, utility, and longevity of databases and infrastructures based on names (Figure 3). Who will fund and support its implementation? Where should it be implemented and by who?
- We emphasized the need for better interoperability with a matching analysis between taxonomic backbones of multiple databases. If upscaled, how can we ensure that interoperability assessments using GIST as a basis are expanded and validated across taxonomically informed databases and experts?
- We described the expertise needed in relation to GIST, but how can we ensure those capacities are built and maintained, and are widely accessible for taxonomists, biodiversity researchers, and conservation practitioners?
- This review focused on the importance of connecting databases and communities working on taxonomy that differ in terms of their values and priorities. How can we commonly initiate and fund interdisciplinary leadership to navigate this complexity for improved and maintained utility, recognition, and access? How can we train young leaders to be capable to communicate between diverse communities and taxonomy users?