



Towards a global list of accepted species VI: The Catalogue of Life checklist

Donald Hobern¹ · Saroj K. Barik² · Les Christidis³ · Stephen T. Garnett⁴ · Paul Kirk⁵ · Thomas M. Orrell⁶ · Thomas Pape⁷ · Richard L. Pyle⁸ · Kevin R. Thiele⁹ · Frank E. Zachos^{10,11,12} · Olaf Bánki¹³

Received: 18 June 2021 / Accepted: 30 August 2021 / Published online: 11 October 2021
© Gesellschaft für Biologische Systematik 2021

Abstract

Despite the important role that species names play in multiple fields, there is no globally complete list of known and described species. This lack is a result simultaneously of the complexity of planetary biodiversity, the long history of naming species in publications from all over the world, the small number of taxonomists working on many important groups, the rapid and dynamic change in knowledge for a few well-studied groups, and the limited incentives for researchers to curate such lists. Recent papers have proposed that a more formal governance mechanism is needed to assist with the translation of taxonomic knowledge to user communities. The recommendation is for the taxonomic community to assist user groups by maintaining reviewed lists that reflect as far as possible consensus among practising taxonomists and incorporating new insights and understanding as these become widely accepted within the relevant taxonomic community. The Catalogue of Life (COL) is the most significant international partnership working to deliver a list of all species by engaging a broad network of taxonomists and databases to contribute expert-curated lists for different taxonomic groups. COL, which included 1,908,823 species as of May 2021, has great experience with developing such a list across all taxonomic groups and has been modernising its processes and tools since 2017 to reflect best practice in management of large digital information assets. This paper explores the alignment between the current state and directions of COL and the initiative to improve the governance of species lists.

Keywords FAIR data · Global species databases · Biodiversity informatics · Species checklists · Catalogue of Life

This article is part of the Special Issue and Topical Collection *Towards a global list of accepted species*

✉ Donald Hobern
donald.hobern@sp2000.org

¹ Species 2000, Canberra, Australia

² CSIR-National Botanical Research Institute, Lucknow, India

³ Southern Cross University, Military Rd, East Lismore, NSW 2480, Australia

⁴ Charles Darwin University, IUBS Working Group on the Governance of Global Taxonomic Lists, Charles Darwin University, Darwin, Australia

⁵ Royal Botanic Gardens, Kew, Richmond, London, England

⁶ Integrated Taxonomic Information System, National Museum of Natural History, Smithsonian Institution, DC, Washington, USA

⁷ Statens Naturhistoriske Museum, Copenhagen, Denmark

⁸ ICZN, Bernice P. Bishop Museum, Hawaii, Honolulu, USA

⁹ Taxonomy Australia, Australian Academy of Science, Canberra, Australia

¹⁰ Natural History Museum Vienna, Vienna, Austria

¹¹ Department of Evolutionary Biology, University of Vienna, Vienna, Austria

¹² Department of Genetics, University of the Free State, Bloemfontein, South Africa

¹³ Species 2000 Naturalis, Leiden, Netherlands

The need for a comprehensive global species list

After more than a quarter of a millennium, the system of nomenclature introduced by Linnaeus for plant species in 1753 (Linnaeus, 1753) and extended to animal species in Linnaeus, 1758 remains central to all efforts to describe, document, and communicate knowledge about the world's biodiversity. Species binomials are the standard names used to refer to all described species other than viruses (which are named using a different but related system). Taxonomists also classify the new organisms they describe and arrange them in a ranked hierarchy of taxa associated with scientific names. Even with modern advances in understanding of evolution and phylogeny, this traditional name-based hierarchy of biological classification provides the primary organisational structure for accessing information on any portion of the tree of life.

Throughout the biological sciences, references to organisms are given context and meaning by applying these names. Their pervasiveness and significance are so great that non-scientists readily recognise scientific names and a large proportion of the general public knows a handful of binomials and a larger number of generic names.

Scientific names for species also play many important roles beyond biology. They are the tool used internationally for communication about species, whether for trade, conservation, biosecurity, or disease management. Governments and intergovernmental bodies rely on them when drawing up treaties, legislation, and regulations in any of these areas, including the IUCN Red Lists of Threatened Species and the CITES listings for controlling trade in endangered species; listings of notifiable pests; and identification and certification of products from agriculture, forestry, and fisheries. The importance of standardised scientific names for organisms is further increased by the global scope of many of these instruments. They also have a cultural importance by providing a means to communicate and identify species that we interact with at local and global levels. Problems arise as a result of imperfections in the management and use of these names, for example, when different names for the same species are used in different regions or by different organisations. This can obscure risks from changing pest distributions or compromise plans to conserve species across their entire range.

Scientific names arranged in a stable, well-accepted classification serve as the labels and organisational framework for the recognised units of biodiversity. They therefore provide the underpinnings for a wide variety of metrics and comparisons, particularly to evaluate the richness of biodiversity at any site (alpha diversity), support biodiversity comparisons over time and space (beta diversity),

and determine the richness of larger geographical areas (gamma diversity; Whittaker, 1972). In addition, those underpinnings can be used to assess the completeness of datasets (species accumulation curves) and estimate progress in the taxonomic endeavour itself. Metrics for these purposes could also (and for some purposes, more objectively) be constructed using DNA/RNA sequences or other biological data and clustering models or measures based on phylogenetic diversity, but scientific names remain the bridge for interpreting or communicating such units of diversity. Scientific names link past biological observations to current and future observations based on the current accepted/valid name, synonyms, and the refinement of species concepts. In doing so, they unlock broader access to information related to species (Guala, 2016).

With the advent of the Internet and the transmission of vast and growing volumes of information on the natural world, scientific names have a fresh significance as labels for digital information. Data infrastructures rely on species names and higher classifications to organise and make biological data accessible in useful forms. Although international platforms such as the Global Biodiversity Information Facility (GBIF, 2021; species occurrence data, including specimen records), Encyclopedia of Life (EOL, 2019; species descriptions), GenBank (2021; genetic sequences), Barcode of Life Data Systems (BOLD, 2021; DNA barcode sequences), Biodiversity Heritage Library (BHL, 2021; biodiversity literature), iNaturalist (2021), and Observation International (2021; citizen science observations) vary significantly in their focus and the categories of information they manage, they all depend on an underlying taxonomic classification and knowledge of species synonyms to structure their resources. Around the world, countless government departments, non-government organisations, research groups, citizen science groups, websites, museums, herbaria, and fungaria expend resources and efforts maintaining species lists for similar purposes.

It is therefore surprising that, given the prominence and importance of scientific names, there is no comprehensive listing of these names nor an unambiguous way even to verify that a given sequence of characters is a name that has been applied to some group of organisms. Lists exist and may be of very high quality for many groups of organisms, but the gaps in available catalogues remain significant. Even more importantly, there is no fully comprehensive synopsis of how the published names relate to species in a modern classification. Linnaeus established his nomenclatural system with a catalogue of all the species known at that time, but since then the sheer volume of subsequent taxonomic research across dozens of languages and thousands of journals and other publications has hindered efforts to collate all this basic information within a single global list of named species.

The Catalogue of Life

The Catalogue of Life (COL) is the most comprehensive effort to deliver such a list. The COL partnership was established in June 2001 between Species 2000 (2019) and the Integrated Taxonomic Information System (2021; ITIS). Species 2000 was initiated in 1996 as a task group under the Taxonomic Databases Working Group (2021; TDWG, itself established in 1985 to improve sharing of biodiversity information), with sponsorship from the Committee on Data for Science and Technology (2021; CODATA), the International Union of Biological Sciences (2021; IUBS), and the International Union of Microbiological Societies (2021; IUMS). The goal of Species 2000 has been to collate a uniform and validated index to the world's known species: the Catalogue of Life. Similarly, ITIS was originally created in 1996 at the behest of the White House Subcommittee on Biodiversity and Ecosystem Dynamics and is a partnership between government agencies in the USA, Canada, and Mexico to create an easily accessible database with reliable information on global species names and their hierarchical classification.

COL works with a broad network of taxonomists and databases to construct the most complete checklist of the world's species, updated regularly (typically every month) and published in an annual edition each year since 2000 (Bisby et al., 2002; COL, 2021). The checklist is constructed by bringing together datasets identified as Global Species Databases (GSDs) for each taxonomic group. Each GSD is produced and maintained by an expert or a group of experts in the taxon in question and aims to present a comprehensive global perspective of the species included in the taxon and the names that have been used to refer to each of these species as a result of changes in taxonomic opinion over time. When COL began, species lists of this kind were available in a digital form only for a limited number of taxa, but the number of GSDs has grown each year. COL also maintains a management hierarchy (Ruggiero et al., 2015) that serves as the higher taxonomy for linking these GSDs into a single dataset. Each GSD contributes branches that connect to this common framework. This model can in part be seen as a result of the limited number of suitable checklist datasets available in a digital format before 2000. There were few areas where competing datasets needed to be considered.

When COL was first envisioned, the number of described species was estimated to be about 1.75 million (Hawksworth & Kalin-Arroyo, 1995). The first version of COL in 2000 included 220,000 species, but as of May 2021, it has grown to include more than 160 separate datasets (COL, 2021) covering a total of 1,908,823 species. The COL model has worked particularly well for moderately well-studied taxa that can be covered by small, well-respected groups of dedicated editors.

Despite this growth, major gaps have remained in the coverage of the checklist. These particularly reflect the megadiversity of some groups and the small number of taxonomists working on some of the most speciose taxa. The largest insect order (Coleoptera) includes nearly 400,000 named species (Chapman, 2009; Zhang, 2013)¹, and the four next largest (Diptera, Lepidoptera, Hymenoptera, and Hemiptera) include hundreds of thousands more. Around 200,000 historical species names for fungi that are included in the fungal nomenclator (Index Fungorum, 2021) await contemporary assessment and are not yet represented in *Species Fungorum* (Species Fungorum, 2021), the GSD for fungi. Few taxonomists work on these groups at a truly global scale; most gain deep familiarity only with the subset of the fauna or flora found in their own region. As a result, there may be no expert able to review and comment on the relationships—differences and overlaps—between regional species lists. Similar challenges exist for other large and important groups such as mites, nematodes, and many fungal taxa. Local species lists may exist at the country or regional level for some or all of these taxa, but merging these local perspectives is difficult as a result of variations in higher classification, differing views of synonymy and species delimitation (species concepts), and lack of opportunity to compare all named species within a genus or tribe. As a result, COL and its partners have struggled to develop a robust model for building and maintaining comprehensive lists of these large and difficult groups.

Other challenges have been well described in other papers in this series and reflect the state of taxonomic research for better studied groups. For vertebrates and some plant groups, the level of study and the number of active taxonomists has led to the construction of multiple alternative checklists for the same group. Whereas limitations in resources and expertise have hampered COL's efforts to include lists of Hymenoptera or Lepidoptera, the relative excess of research in these other groups has made selection of the most appropriate list contentious, leading to fragmentation and uncertainty for user communities.

The embryonic state of biodiversity informatics when COL was originally established has also left other challenges that now need to be resolved. When COL initially engaged GSDs to contribute to the checklist, digital solutions did not exist for open access licensing of data. This left COL with a legacy of contributor agreements that do not reflect contemporary best practice. By contrast, data sharing within GBIF (2014a, 2014b) is now based around Creative Commons licences and data citation practices based on digital object

¹ Coleoptera 392,415 species, Diptera 160,591 species, Lepidoptera 158,570 species, Hymenoptera 155,517 species, Hemiptera 104,165 species

identifiers (DOIs), and COL is in the process of updating the licence arrangements with each GSD.

The processes used by COL for constructing the checklist have also, until recently, relied heavily on labour-intensive intervention by the editors using ageing software and tools. Thus, while GBIF and other infrastructures depend on COL as the core for their taxonomic data management, the gaps in coverage have forced each of these partners to develop bespoke processes and tools to handle names that are not yet included in COL. These efforts lead to inconsistencies in function, capabilities and content between databases that would ideally be interoperable.

In 2015, a summit was held involving many of the major biodiversity initiatives—COL, ITIS, EOL, BHL, GBIF, BOLD, and others. Each of these programs has independent objectives, structures, and assets and its own history of cooperation, information sharing, and tool developments. These parties together recognised the need and benefits for future collaboration in developing shared solutions and services and in investigating coordinated strategies for sustainable funding for these services. The group coalesced around building a shared Catalogue of Life. The long-term benefits of coordinating these efforts would be a common resource (and shared rewards and risks) that would ultimately lead to a collaboration and a mutually beneficial and durable relationship.

In 2017, in response to this and other challenges, COL, GBIF, Naturalis Biodiversity Center, and other partners received support from the Netherlands Ministry of Education, Science, and Culture through the Netherlands Biodiversity Information Facility to develop a single shared catalogue using the expert-curated data from the COL GSDs in combination with other datasets and automated processes contributed by GBIF. By unifying the efforts and deliverables of these two leading biodiversity data infrastructures, this project has opened the door to building an open, shared, and sustainable consensus taxonomy that can support data management within and between biodiversity information initiatives. All pre-existing infrastructure has been rebuilt, including web services, the portal and the software for the assembly of the catalogue, and its editorial work. The new COL portal, API, and COL ChecklistBank were made public in December 2020 (COL, 2021) and are hosted by GBIF.

Evolving directions

The redevelopment of COL and its services builds on lessons learned over more than two decades and on active experience in the development of human, data, and software solutions to biodiversity informatics challenges. COL's culture and tools are evolving to improve support both for taxonomists as the producers and curators of taxonomic knowledge and for the agencies, infrastructures, and other users that

depend on its services. This process includes several inter-related transformations.

Engaging with the whole taxonomic community

The GSD model reflects both the small number of dedicated contributors that existed when Species 2000 began its work and the immature state of open data solutions at the time. Many GSDs were developed and maintained by one or a few closely collaborating taxonomists maintaining the data in their own spreadsheets or local database files. Users and other interested researchers typically emailed these editors with suggested edits or corrections, with updates published to COL on variable schedules.

This model has several weaknesses. In the best cases, highly dedicated, responsive, and collegiate editors have maintained a resource that is highly valued by the international taxonomic community. However, this places heavy reliance on the efforts of a few people and cannot guarantee that the views of the wider taxonomic community are reflected. As noted above, for the most hyperdiverse groups, no individual is likely to be able to lay the foundations for a comprehensive global list, while for the best studied groups, multiple competing perspectives and classifications may exist, and significant community facilitation is required to deliver a list that represents a realistic consensus or working compromise.

Since 2000, web-based creation and curation of collaborative products has become common practice across many communities, including for software development, wiki-based authoring of content (most significantly the Wikimedia platforms), and tools for simultaneous live editing of documents. All these developments have helped to establish new paradigms for distributed teams to develop shared products and to respond to the simultaneous drive for FAIR (Findable, Accessible, Interoperable and Reusable) data, open licencing, and reuse of foundational datasets. Since its inception in 1996, ITIS.gov has demonstrated the value of a model that receives stable long-term funding to deliver persistent, resolvable, free, and open data. Meanwhile, taxonomic communities such as those that make up the World Register of Marine Species (2021; WoRMS) and the various Species File groups (Species File Group, 2013) demonstrate how an international editorial model can support the continuous curation of a GSD without reliance on a single individual or closed circle of collaborators.

These developments make it possible and desirable for COL to oversee an opening up and expansion of the communities that maintain each GSD sector. It is important to recognise and credit the significant contributions that have been made by past and current contributors to each sector, but a more open model will help COL to be a trusted and responsive resource for the whole taxonomic community.

This will require better tools and associated social networking solutions that can support contributions from all interested taxonomists and assist them with resolving differing perspectives.

Flexibility for different taxonomic groups

For more than 20 years, there has been concern over the “taxonomic impediment” (Cresswell & Bridgewater, 2000), the shortfall between available taxonomic expertise and the needs of society to name and identify organisms. It is largely true that the size of the group and the challenges of its taxonomy are inversely related to the number of qualified taxonomists dedicated to that group (Coleman, 2015; Fisher et al., 2011; Lücking, 2020). For example, relative to the number of species in the taxonomic group, many more taxonomists work on plant and vertebrate groups than on mites, nematodes, parasitic wasps, or fungi. Yet these latter four groups are of significant importance in understanding ecological complexity or in managing agricultural and natural systems. The rise of genetic approaches (DNA sequencing) is promising as they help to overcome the small size of the organisms and make the challenge easier to tackle. Despite, or possibly because, theoretical and conceptual research issues still remain to be addressed (methods, operational delimitation of species, management of proxy-identifiers, etc.), these methods will hopefully attract more young scientists.

This has consequences for species lists at both ends of the scale of expertise. For the best-studied groups, the taxonomic workforce has described all known species, and much effort is expended mapping taxon concepts at fine scale and debating the boundaries between closely related species or the appropriate taxonomic rank for well-defined forms. A result is that there may be multiple well-curated global lists of the species for such groups, each reflecting the taxonomic judgement of a subset of the research community. For birds, mammals, and a few other groups, this diversity of opinion has been the major challenge in establishing a reference list of species for wider use. However, for birds, there is now an initiative through the International Ornithological Union to consolidate the competing lists into a single global list (McClure et al., 2020).

On the other hand, the paucity of taxonomists working on hyperdiverse taxa has multiple compounding implications. No taxonomist is likely to be familiar with all species in one of these taxa, even within a single country, or be in a position to review all species within such a taxon at the global scale. This means that the best global lists may be only partially synonymised mosaics based on regional perspectives. The scale of the task makes it difficult even to begin construction of comprehensive lists, and overworked taxonomists may not be able to prioritise the task.

The variation relates not only to the relative size of the workforce needed to deal with each taxon but also to the governance models that already exist for different taxa, or could do so in the foreseeable future.

In all cases, it is important to document clearly what expert base has contributed to each GSD.

Within microbiology, challenges around species recognition and documentation resulted in most older published names not being safely resolvable. As a result, bacterial nomenclature was completely revised in 1980 (Sneath, 2005), with the publication of the Approved Lists of Bacterial Names as a new starting-point for bacterial taxonomy and with updates published regularly through the *International Journal of Systematic and Evolutionary Microbiology* (Microbiology Society, 2021) as the journal of record for publication of novel microbial taxa. The International Committee on Taxonomy of Viruses (2021; ICTV) similarly exercises control over the publication of new viral names.

This level of formalisation does not yet exist for most other taxonomic groups, although the International Commission on Zoological Nomenclature (2021; ICZN) has established the List of Available Names (LAN) as a process (ICZN, n.d.), and this has been used to standardise rotifer nomenclature (ICZN, 2019). This is in part a reflection of the number of workers active on eukaryotic organisms and the challenges of securing the necessary international agreements. However, major collaborations exist around many eukaryotic groups to develop and curate shared information resources on species and their names. Some of these, such as the World Register of Marine Species (2021b) and the World Flora Online (WFO; WoRMS, 2021a; Borsch, et al., 2020), have models that allow interested taxonomists working on the group to become part of the editorial team. For some other taxonomic groups, this editorial work may currently be carried out by a single individual, reflecting either a legacy of decades of assuming this responsibility or the low number of taxonomists able to work on a given list (e.g., Eschmeyer’s Catalog of Fishes; Fricke et al., 2021). Many of these efforts are maintained using personal spreadsheets or databases and are at risk of loss when those individuals stop working or should their software be compromised.

These sociological aspects are very important. The variation is a product of multiple factors:

- The size of the taxonomic group
- The number of active taxonomists
- The level of public awareness and interest in the group
- The extent to which international communication about the group is considered necessary by the broader society (species that are pests and cause diseases are more critical than obscure benign species)

COL is responding to these variations in the size of the taxonomic workforce and their associated governance arrangements by maintaining a flexible approach to working with researchers to secure the best possible list for each taxon. Where significant maturity and organisation of effort exists, COL has no intention to duplicate tools and services but will build on these efforts. Where less advanced communities exist (as with some of the current GSDs that rely on the work of small numbers of individuals), COL is providing the tools and support to assist with producing the best possible comparable results. Where no community yet exists (often because of the scale of the challenge), COL will provide the most complete automated solutions possible and innovate in the development of interfaces that enable interested parties to contribute even at the level of small fixes to individual records. Over time, as the existing context becomes more and more complete and major errors and issues are addressed, this flexibility should enable communities for all taxa to develop and become mature.

Automating construction

Until now GSDs have been contributing to COL using a variety of data formats, reflecting the capacity and technical support available to each set of authors. This has included spreadsheets in a variety of formats, database exports, and, in some rare cases, word processor documents. As a result, COL has historically been obliged to carry out bespoke processing to prepare datasets for inclusion in the checklist. Updated versions of the same dataset have required the same processing. Therefore, processes have often been labour-intensive, and there has been limited scope to maintain stable identifier schemes for names and species. Even when datasets are presented to COL in standardised formats, special processing has sometimes been required to handle known issues related to problematic data records or to exclude portions of the data from the COL Checklist that may be supplied by another GSD. Reliance on a complex multi-stage human editorial process introduces risk of error and introduction or re-introduction of mistakes between releases.

More recently, there has been a growing shift across the community towards managing most GSDs in well-structured databases that support COL-compliant data exports rather than word processor or spreadsheet tools. At the same time, the new COL infrastructure includes tools that simplify and allow for automation of the construction process for the checklist. These factors will combine to improve the stability, quality, and sustainability of COL for the future. Increased standardisation of data inputs makes it simpler to detect changes in the data and to maintain stable identifiers between releases. Improved and iterative curation processes will also open the door for COL to work with authors to

integrate names and content harvested automatically from new publications. For example, Pensoft (2021) publications deposit summary datasets in GBIF, containing the information necessary to add new names and species to COL. Similarly, Plazi digests both current and older publications to generate similar data for historical taxon treatments. These shifts to greater standardisation and automation allow COL to process additions and corrections much more rapidly, even offering immediate feedback to contributors on issues detected during import of their datasets. Such standardisation and ongoing improvements to COL software will also make it possible to be more transparent regarding the origin of each data element and the basis on which each name is accepted or treated as a synonym in each version of the checklist.

Stabilising identifiers

A major challenge associated with constructing large datasets from heterogeneous, primarily text-based resources is in the reliable recognition of the “same” data record presented from different sources or from the same source over time. Addressing this challenge requires several related issues to be resolved.

There must first be agreement and clarity around a well-defined concept of what each data record in COL should represent. As discussed in Pyle et al., 2021, a list may enumerate known (accepted) species, or it may enumerate species names that have been published in accordance with one of the nomenclatural codes, or names that may not have been so published but that have been widely used (including vernacular names), or alternative variations in the spelling and formatting of known names, or even known usages of any scientific name in the literature. Any or all of these could be collected into datasets, and appropriate rules could be followed in each case to determine whether a new data record represents a new instance that should be added to the dataset. Depending on the choices made, *Felis concolor* Linnaeus, 1771 may or may not be the same name as *Puma concolor* (Linnaeus, 1771). Both these representations derive from the same action by Linnaeus when he gave the first of these names to the species. The subsequent move from the genus *Felis* to the genus *Puma* added information on the relationships and classification for this species, but did not in itself change the scope, i.e., the set of organisms, to which it related. Nevertheless, an information service needs somehow to recognise both these name strings and to be able to relate them together.

For COL, two separate classes of data are of primary importance: published scientific names and accepted species (as far as possible based on credible community consensus). For scientific names, GSDs should aim to include every published name of species rank and ideally infraspecific names

(although for some groups many of these are obscure and perhaps never reused, so each GSD makes its own judgement regarding the effort to devote to completeness). Other names (including vernacular names) may be added, but these should be clearly marked as separate from published scientific names. Over time, COL should ideally expand to make it possible to review alternative classifications based on different taxonomic viewpoints, all composed using the same scientific name records. In practice, the initial focus is on completeness and developing a single species checklist that includes all published scientific names at species rank and an accepted name for each species. Fossil species are eligible for inclusion, but again must be clearly separated from living species. COL aims to deliver a comprehensive and complete dataset for both published scientific names and accepted species, through the efforts of the GSD communities and through partnerships with the major nomenclatural databases (*ZooBank*, *International Plant Names Index*, *Index Fungorum*).

Secondly, it must be possible to determine when two data records both relate to the same published scientific name despite differences in spelling or form. Binomial nomenclature leads to many situations where this decision may be contested. As another example, the “botanical” (covering plants, algae, and fungi) and zoological codes have both required gender agreement between each species epithet and the associated genus, but most taxonomists working with Lepidoptera have abandoned this distinction and now use the spelling of the epithet in the original publication without consideration for gender agreement. For example, Fabricius named a plume moth *Pterophorus leucodactylus* in 1794. It was subsequently moved to a different genus and known, with gender agreement, as *Megalorhipida leucodactyla* (Fabricius, 1794). With the abandonment of gender agreement for moth names, the preferred form is now *Megalorhipida leucodactylus* (Fabricius, 1794). For different purposes, these could be treated as non-essential variation within a single name, or as two names that each need to be recognised.

Similarly, if species are under consideration rather than their names, perspectives may differ on whether addition of new synonyms or changes in taxonomic placement alter the species concept sufficiently for it to be considered separate and in need of its own distinct identifier.

Thirdly, once a set of data records is recognised as referring to the same published scientific name or a set of scientific names is recognised as referring to a single species, there is a need for a reliable way to refer to this name or species and to this set of records. In the digital context, this is a fundamental informatics challenge. Assigning and maintaining stable digital identifiers is hard, particularly when all the underlying source information may change over time. The way that these identifiers should behave is determined by

the first and second issues, but the practical implementation depends on a trusted party managing the mapping between datasets and between versions of the same dataset and on this party providing robust services that have the expected behaviour under all conditions.

COL has invested significant resources in the past towards maintaining stable identifiers for species records across multiple editions of the checklist. Most significantly, COL experimented with Life Science Identifiers (LSIDs) for the species in annual checklist versions. These efforts have been compromised by the variation in technical capabilities of GSDs and the high level of manual processing required to construct the checklist. As the COL Checklist is increasingly constructed automatically from stable well-managed GSDs, each with their own internal practices to manage local record identifiers, it will be possible to associate all names with stable identifiers and to implement policies that determine when species circumscriptions have changed sufficiently to require new identifiers. COL will use linked open data principles to enable users and software to follow these changes. Stable identifiers, particularly DOIs for checklist versions and for the source GSDs, will also bring benefits through enabling correct citation of names and concepts from any publication or digital resource.

Expanding and enriching the data

The primary responsibility for COL is to manage and organise the species checklists that taxonomists create for different taxonomic groups so that these can be accessed and interpreted consistently by users and contribute to an integrated list of all species. Establishing a comprehensive and well-reviewed checklist and classification of all species and maintaining it over time as knowledge grows will be an invaluable service to many communities. This therefore remains the priority for COL.

However, many stakeholders have legitimate reasons to access species information using alternative names, classifications, and even non-Linnaean identifiers (e.g., species proxy identifiers from DNA-based identifications, such as the Barcode of Life Database Barcode Index Numbers (BIN) and the UNITE (2020) molecular OTU identifiers for Fungi).

The data standards and software developed by COL for sharing and integrating checklist data have the potential to readily support publication of a range of species lists, including national or regional species lists (which may be important for legislative purposes and which may adopt different names, synonymy, and classification from the COL Checklist), lists of protected or introduced species (such as the Red Lists and the Global Register of Introduced and Invasive Species), or more specialised local or thematic lists. COL offers its ChecklistBank as a home for publishing these datasets. Publishing to ChecklistBank brings multiple

advantages. The datasets become citable and downloadable in various standard formats. ChecklistBank performs a range of integrity checks on taxonomic data and reports these to assist authors with corrections. Species lists inside ChecklistBank can be accessed remotely through the COL API and through reusable web interface components. Additionally, names added to ChecklistBank in this way can be made accessible to GSD authors and the appropriate nomenclator if they are not already included in COL.

ChecklistBank includes tools that are used for constructing the COL Checklist itself. These tools simplify development and maintenance of a checklist from multiple sources and enable editors to locate duplicate names or incomplete information. Publishing component datasets to ChecklistBank makes it possible for checklists other than the COL Checklist itself (e.g., national or regional species lists) to be constructed, published, and maintained. As COL enhances its own tools (e.g., automating addition of newly published names from the taxonomic journals), these functions will also become available to other users of ChecklistBank.

COL will also build on work by GBIF to integrate DNA-based identifier schemes such as BINs and UNITE identifiers into the hierarchy of the COL Checklist. This will allow such identifiers to be used intelligently in conjunction with classical Linnaean species nomenclature.

Although the COL Checklist itself represents just the nomenclatural and taxonomic framework for biodiversity information, COL identifiers and services will make it possible to link other information, such as distributional data from GBIF, literature from BHL, and trait information from the Encyclopedia of Life (2019), and other biodiversity information systems such as FishBase (Froese & Pauly, 2021) and SeaLifeBase (Palomares & Pauly, 2021).

Duplication of effort

Efforts to list scientific names and species include the work of taxonomists constructing global, regional, or local catalogues; governments and intergovernmental agencies maintaining lists for conservation, biosecurity, or trade; nomenclators building reference lists of published names; thesauri and vocabularies for cataloguing literature; and data infrastructures seeking to organise other digital information.

There is significant duplication of effort across these parties. Basic information on published names may be transcribed, verified, and databased by each initiative. Taxonomists may contribute both to national information systems and to global checklist datasets. There is significant scope for streamlining this activity, in particular by ensuring that each new nomenclatural act is captured once and shared in a structured digital format with all who need the information. Increasingly publishers such as Pensoft and aggregators such as Plazi (2017) are delivering streams of updates that fulfil

this need. An integrated approach is required to accelerate access and uptake of these data.

Until now, GBIF has built on versions of the COL Checklist to develop an automatically constructed taxonomic backbone including all the names required to organise GBIF data. Unifying this activity with the development and maintenance of the COL Checklist allows each of these names to be treated as an extended unreviewed component within a single checklist. This will simplify use of the COL Checklist for any parties that need to handle names that are not yet included. All such names can be flagged as unreviewed and may be excluded from any presentation of the checklist but can still have stable identifiers that will continue to resolve once a taxonomic community has made a decision on its position. Integrating all of these names into a single system also simplifies workflows for communities to learn of these names and handle them within their own datasets.

COL recognises the need both for a comprehensive nomenclator (a nomenclatural catalogue containing all published scientific names) and for a comprehensive taxonomic checklist (assigning all published names and combinations to an appropriate position in a list of all species). The data required to build the nomenclator is reused in the taxonomic checklist. Another goal for COL is therefore to streamline its relationship with the nomenclators for different taxonomic kingdoms so that a single authoritative data record is maintained for every published name and is available for use without re-entry by any user or infrastructure.

Research data linkages

COL's primary mission is still to deliver a comprehensive list of all the world's species. The partnership with GBIF not only positions COL as part of the world's most significant biodiversity data infrastructure but also lays the foundation for the expert-curated COL Checklist to become the core of a larger data management framework for names, species and classifications.

As a global infrastructure, COL is positioned to provide the services that link species names with other resources, including the nomenclators, published literature, and GBIF data, but also bridging to communities and data resources that seek to use genomics solutions to map species diversity, particularly efforts such as the Barcode of Life Data Systems (BOLD), the global database for DNA barcodes and for the DNA-based specimen clusters identified using the Barcode Index Number (BIN) system, and UNITE which constructs molecular operational taxonomic units (mOTUs) for fungi. These platforms offer important and complementary views of diversity in many of the groups where COL is lacking comprehensive species lists and where much global biodiversity remains unnamed. GBIF is already including BINs and UNITE mOTUs in its taxonomic framework, and COL

will partner to build tools that support navigation between scientific names and DNA-based identifiers.

An important area for cross-linkage will be with the phylogenetic representations offered particularly by Open Tree of Life. COL and Open Tree of Life are complementary and address differing needs. Phylogenetic representations may respond rapidly to new insights into evolutionary history and offer scope for computational analyses. Representations based on scientific names relate more readily to published literature and other information sources and are more intuitive for many users. Open Tree of Life depends on good sources for information on published names, and COL GSDs will benefit from growth in phylogenetic understanding (Open Tree of Life, 2021).

Another important aspect of modern digital research infrastructures is seamless linkage between research data and the uses of these data. As GBIF's management of species occurrence data demonstrates, centralised services and standardised use of DOIs greatly simplify citation by users of data and enable data contributors and funders to trace usage and impact of their datasets. This helps source institutions to demonstrate the value of their work. COL is working with GBIF to deliver citation and usage tracking services for the COL Checklist and for all datasets in COL ChecklistBank.

Response to the IUBS working group vision

Following the publication of Garnett and Christidis (2017) and subsequent debate, the International Union of Biological Sciences (IUBS) funded a project including workshops to explore governance of species lists. The first of these workshops, in Darwin, Australia in February 2020, led to the publication of Garnett et al. (2020), outlining ten principles for creating a single authoritative list of the world's species, supported by the establishment of an IUBS Working Group. Olaf Bánki and Donald Hobern represented COL in this process.

These principles focus on the translation of taxonomic knowledge into products that serve the needs of policy and society. They explicitly seek to maintain the independence of taxonomy as a science from external non-scientific influence on its conclusions. The principles aim to ensure that taxonomists are empowered to publish the results of their research without influence based on the likely impact of this research on conservation or development activity and without any limitations on the ability to develop new concepts regarding species boundaries and classifications.

Garnett et al. (2020) therefore recognises that there is no requirement for a governance mechanism for the processes of taxonomy itself, but there is such a requirement for the processes that mediate taxonomy to wider audiences with a particular set of requirements, specifically

those with an operational need for aggregated and integrated species lists that synthesise the understanding of taxonomists and deliver an appropriate and authoritative snapshot of current knowledge.

Under this model, the process requires consideration of three distinct tiers of activities, each focused on a different aspect of the work and each supported by a different governance model:

- **Nomenclature:** The nomenclatural commissions and codes that manage the rules and mechanics that pertain to the publication of new scientific names, while necessarily rather legalistic, provide consistency and clarity around the names themselves, independently of how these are subsequently interpreted taxonomically. A set of global databases (nomenclators) exists to catalogue these names for different kingdoms (International Plant Names Index (2021) for plants, ZooBank (2021) for animals, Index Fungorum for fungi, etc.). These nomenclators make no judgment on the taxonomic status of the associated species or other taxa.
- **Taxonomy:** The work of taxonomic revision and species description is a scientific activity that is normally (but not necessarily) moderated through peer review processes and more generally through the degree to which other taxonomists adopt and build upon published treatments. This is a slow and continuous process and is not always readily translated into a simple summary view for wider use. Species lists of various kinds, including globally complete lists, are an important output from taxonomic research, forming the core of COL and representing the major means by which taxonomies are communicated to users of taxonomic research.
- **Policy and wider society:** Species names are important to governments; multilateral environmental agreements; including the Convention on Biological Diversity (CBD; 2021), the Convention on International Trade in Endangered Species of Wild Fauna and Flora (CITES; 2021) and the Convention on the Conservation of Migratory Species of Wild Animals (CMS; 2021), and the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services (IPBES; 2021); conservation agencies; and a wide range of other stakeholders, particularly in legal and treaty contexts. For these uses, scientific credibility remains important, but uptake is also affected by the need for stability and the processes needed to maintain alignment with legislative tools. This is the level at which the recommendations from Garnett et al. (2020) come into play, focusing as they do on the creation of criteria by which these stakeholders can trust and confidently adopt the outputs from taxonomy, including products such as global species lists from COL.

COL has historically focused on supporting the second of these tiers and on delivering a product that reflects the contributions and expertise of the taxonomic community. The IUBS WG focuses attention on the requirements to ensure that such products fully meet the needs of the third tier.

The IUBS WG has recognised that “current systems provide a solid foundation for the creation and management of global lists, with the COL Consortium having a version that is progressing rapidly” (Garnett et al., 2020; p. 8). There is no need to reinvent all the processes and solutions that have been developed by COL and its biodiversity informatics partners. However, there are opportunities to align existing efforts within COL to modernise its processes, particularly to address the “FAIR”ness and relevance of the COL Checklist through the new COL infrastructure project, with the recommendations of the IUBS WG. The recommendations offered by the IUBS WG serve as outline benchmarks for COL to clarify expectations for its contributors and for the characteristics of the infrastructure it offers.

The community of COL contributors needs to be supported and assisted in its work, but at the same time, COL as a whole needs to evolve further so that it can serve as a global infrastructure that delivers a comprehensive checklist. The Catalogue of Life Global Team has recognised a number of areas for which criteria and policies can be used to support the development of the checklist and to guide improvement in the coverage, quality, “FAIR”ness, and community basis of the resulting product. These criteria and policies will serve multiple purposes within COL:

- Guide new and existing contributors as they plan or improve their data curation and publishing efforts.
- Provide a basis for COL to review and evaluate datasets for possible inclusion as part of the COL Checklist.
- Assist COL in selecting the optimal dataset where multiple alternatives exist.
- Form a roadmap for COL to support enhancement of all sector datasets.
- Provide metrics for measuring progress in developing the COL Checklist.
- Demonstrate that the COL Checklist has robust scientific support and is trustworthy.

Over time, as the community matures, the detail associated with these criteria can change, and the minimum thresholds can be raised. It is important to acknowledge that taxonomic communities will not all face the same challenges in seeking to meet these goals. For some taxa, the nomenclatural foundations may be well established and fully catalogued, but a number of competing taxonomic viewpoints may need to be resolved. At the other end of the spectrum, there will be taxa for which no expert today can offer a truly global

and consistent perspective on the number of species within widely distributed and hyperspeciose genera.

For each sector of the checklist, the criteria can support evaluation of the following key questions:

- Does a candidate dataset for a sector meet the minimal criteria?
- If there are multiple alternative candidate datasets for the same sector, does one of these more fully meet the criteria for adoption? Is it possible to help the developers of these alternatives to cooperate to develop a unified product?
- What support (investments, tools, etc.) is needed to help the developers of each sector move beyond minimal criteria towards the goal?

Consistent with the IUBS Working Group principles, COL is establishing the following criteria for measuring progress, with a minimum requirement specified for each GSD:

1. Coverage/comprehensiveness

Species lists need to aim for completeness in regard to all known species and to all the names that have been published and used to refer to these species.

- **Goal:** Each sector includes all published names and all currently accepted species for its given group.
- **Minimum:** Each sector is represented by a high-quality dataset with an associated path for remaining gaps to be addressed.

2. Richness

Species lists should enable users to interpret all binomial combinations used to refer to any species and to access basic information on the species.

- **Goal:** Each sector includes all significant combinations and synonyms for the group and should be open to supporting other useful standard elements valued by taxonomists and others working with the group (types, distribution, etc.).
- **Minimum:** Each sector has the mechanism to add significant combinations and synonyms to ensure the dataset meets user needs.

3. Scope

Each taxonomic group should be the focus of a community of expertise that takes responsibility for the group as a whole.

- **Goal:** Each sector covers the entire global biota for a higher taxon included in the COL Management Classification.
- **Minimum:** The taxonomic community for each sector takes responsibility for a taxon not otherwise included in COL and commits to collaborate with other communities to complete a higher taxon included in the COL Management Classification.

4. Nomenclatural consistency

The nomenclatural status for each included name should be clear and supported by references.

- **Goal:** Each sector follows a clear, appropriate, and well-documented model for determining what names should be included, indicates the nomenclatural status for each such name, and collaborates with COL's efforts to unify management of nomenclatural data with the datasets managed by global nomenclators for each group.
- **Minimum:** Each sector provides a nomenclatural status for all names included.

5. Taxonomic consistency

Species lists should as far as possible reflect consistent synoptic assessments of the whole group (i.e. where the same authorities provide a judgment on all species, rather than where a list is constructed from multiple sources via literature review or automated tools).

- **Goal:** Each sector offers a taxonomically consistent view of accepted taxa included at each rank—i.e. the listed set of child taxa for any taxon is based on references to appropriate revisionary work rather than simply aggregation of published names.
- **Minimum:** Each sector provides a clear indication whether the listing within a genus or higher taxon has been reviewed by a taxonomic contributor.

6. Continuous and current curation

Species lists should be maintained and updated as new taxonomic treatments are published or errors are detected.

- **Goal:** Each sector has the ability to make timely updates in response to newly published names, taxonomic revisions, and feedback from partners and users.
- **Minimum:** The taxonomic community for each sector has the ability to make well-managed updates at regu-

lar intervals and to maintain a log of unsatisfied update requests, possibly with the assistance of COL.

7. Stable delivery

Every sector should be maintained in a database or repository that increases the probability of its long-term survival.

- **Goal:** Each sector is hosted by an institution that can provide stable long-term support for the dataset and work with COL to implement best practices in data management (stable record identifiers, full metadata, standards compliance, community support).
- **Minimum:** Each sector is maintained and regularly versioned in a structured data format supported by COL, and COL maintains consistency in record identifiers between versions.

8. Community-managed

Species lists should have a governance system that fosters open collaboration between all relevant taxonomists with an interest in contributing.

- **Goal:** Each sector is maintained by a team of representative taxonomic experts who collaborate to maintain a dataset that is broadly supported by the wider taxonomic community as an appropriate consensus view.
- **Minimum:** Each sector has an appropriate and publicly shared governance system that enables equitable involvement of all relevant taxonomists and associated experts, including quality control and dispute resolution processes.

9. Licensing

As fundamental information resources for science and society, species lists should be shared under terms that guarantee free and simple reuse.

- **Goal:** Each sector is made available under a CC0 or CC BY license.
- **Minimum:** Each sector is made available under a CC0 or CC BY license.

10. Metadata

Users need to understand the source, origins, methods, scope, and restrictions associated with species lists.

- **Goal:** Each sector is provided to COL with all metadata elements required to showcase the dataset, to explain its scope and provenance, to support good citation, and to enable others to provide feedback or assistance.
- **Minimum:** Same as goal—this requirement is foundational.

11. Acknowledgment

Taxonomic research is an important activity, and species lists have a role in raising its visibility and a responsibility to credit those who contribute to their construction.

- **Goal:** For each taxonomic sector, taxonomic contributors and preferably also developers and data managers should be identified and acknowledged through structured metadata in forms that support accurate citation and acknowledgments for each data record.
- **Minimum:** For each taxonomic sector, metadata includes clear identification of all taxonomic contributors to the dataset.

12. Global representation

Taxonomy is a global science with global use, so species lists should incorporate the perspectives of taxonomists working in every region.

- **Goal:** The taxonomic community for each sector includes representatives from most regions where the taxonomic group occurs and is studied.
- **Minimum:** The taxonomic community for each sector offers access for taxonomists from any part of the world to contribute and influence taxonomic decisions.

13. Monitoring and reporting

The taxonomic community responsible for each sector reports annually on the state of its list.

- **Goal:** Each taxonomic community reports annually on the state of its list and its list governance, following a simple protocol.
- **Minimum:** The taxonomic community for each sector has a baseline assessment of its list quality and the processes used to derive it.

Call to action

Building COL has been a lengthy effort and work continues to deliver the vision of the founding partners. However, the centrality of this information to all biological sciences and the urgency of delivering the information necessary for a sustainable future on Earth make it imperative that we deliver a resource that is effectively complete and readily maintained and corrected over time.

In 2018, GBIF convened the second Global Biodiversity Informatics Conference (GBIC2) in Copenhagen. This led to a call for all stakeholders with an interest in biodiversity—for science, conservation, sustainability, and human welfare—to collaborate around a Call to Action (GBIF, 2018) for an alliance for biodiversity knowledge. COL was identified early as a key example of a collaborative effort that depends on and demonstrates the goals and vision for this alliance.

In this spirit, and recognising that we all have much still to improve and that success will be enhanced from greater openness and collaboration, we urge all taxonomists to work with us to build the best possible checklist of the world's species, and all governments, non-government organisations, and other parties that depend on biodiversity knowledge to support and build on this activity. Recent developments in the tools and processes used within COL will ensure that efforts to improve the COL Checklist will contribute cumulatively to delivering a constantly improving resource important to countless users.

COL seeks the involvement of all who work in the field of taxonomy to ensure the completeness of the checklist, to verify the quality of the data records, and to work with colleagues to develop the best possible consensus perspectives and to document areas of uncertainty and disagreement.

COL urges taxonomic institutions to encourage and support the work of taxonomists in producing or contributing to GSDs as part of the role associated with their positions, with an adequate allocation of time and appropriate recognition for these efforts, even when not associated with traditional publications.

COL seeks the engagement of all who work to conserve, manage, and use biodiversity to review the current COL Checklist and to engage with the COL community to ensure that it delivers its information and products in a form that meets the needs of all parties.

COL also urges all parties to recognise the importance of taxonomic knowledge for humankind to interact in positive ways with the natural world and encourages all funding bodies to support the work of taxonomists in delivering this knowledge openly and freely for us all.

Acknowledgements This paper is one of a series published by the IUBS Working Group on the Governance of Taxonomic Lists. We are grateful for funding and support from the International Union of Biological Sciences and its Executive Officer Nathalie Fomproix, administration of the grant by Roanne Ramsey (Charles Darwin University), and formatting by Jeremy Garnett of Top End Editing. The paper also celebrates 20 years since the foundation of Catalogue of Life (COL) as a collaboration. This paper has benefited greatly from discussions with all members of both the Working Group and the COL Global Team, including those not listed as co-authors.

Funding Financial support was received from the International Union for Biological Sciences and Charles Darwin University.

Availability of data and material Data sharing is not applicable to this article as no datasets were generated or analysed during the current study.

Declarations

Conflict of interest The authors declare no competing interests.

References

- Biodiversity Heritage Library (BHL). (2021). *Biodiversity heritage library*. BHL. <https://biodiversitylibrary.org/>
- Bisby, F., Shimura, J., Ruggiero, M., Edwards, J., & Haeuser, C. (2002). Taxonomy, at the click of a mouse. *Nature*, *418*, 367. <https://doi.org/10.1038/418367a>
- BOLD Systems. (2021). *Barcode of Life Data System*. BOLD Systems. <http://www.boldsystems.org/>
- Borsch, T., Berendsohn, W., Dalcin, E., Delmas, M., Demissew, S., Elliott, A., Fritsch, P., Fuchs, A., Geltman, D., Güner, A., Haevermans, T., Knapp, S., le Roux, M. M., Loizeau, P.-A., Miller, C., Miller, J., Miller, J. T., Palese, R., Paton, A., & Zamora, N. (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. <https://doi.org/10.1002/tax.12373>
- CABI. (2021). *CABI databases*. CABI. <http://www.speciesfungorum.org/>
- Catalogue of Life (COL). (2021). *Catalogue of Life*. COL. <https://www.catalogueoflife.org/>
- Chapman, A. D. (2009). *Numbers of living species in Australia and the world* (2nd edition). Department of Agriculture, Water and the Environment, Australian Government.
- Coleman, C. O. (2015). Taxonomy in times of the taxonomic impediment – Examples from the community of experts on Amphipod Crustaceans. *Journal of Crustacean Biology*, *35*(6), 729–740. <https://doi.org/10.1163/1937240X-00002381>
- Committee on Data (CODATA). (2021). *CODATA*. CODATA. <https://codata.org/>
- Convention on Biological Diversity (CBD). (2021). *Convention on biological diversity*. CBD. <https://cbd.int/>
- Convention on International Trade in Endangered Species of Wild Fauna and Flora (CITES). (2021). *CITES*. CITES. <https://cites.org/>
- Convention on the Conservation of Migratory Species of Wild Animals (CMS). (2021). *Convention on the conservation of migratory species of wild animals*. CMS. <https://cms.int/>
- Cresswell, I., & Bridgewater, P. (2000). The global taxonomy initiative - quo vadis? *Biology International: The news magazine of the International Union of Biological Sciences*, *38*, 12–16.
- Encyclopedia of Life (2019). *Global access to knowledge about life on Earth*. Encyclopedia of Life. <https://eol.org/>
- Fisher, R., Knowlton, N., Brainard, R. E., & Caley, M. J. (2011). Differences among major taxa in the extent of ecological knowledge across four major ecosystems. *PLoS One*, *6*(11), e26556. <https://doi.org/10.1371/journal.pone.0026556>
- Fricke, R., Eschmeyer, W. N. & Van der Laan, R. (Eds.). (2021). Eschmeyer's catalog of fishes. California Academy of Sciences. <http://researcharchive.calacademy.org/research/ichthyology/catalog/fishcatmain.asp>
- Froese, R. and Pauly, D. (Eds.). (2021). *FishBase* (version 02/2021). Fishbase. www.fishbase.org
- Garnett, S. T., & Christidis, L. (2017). Taxonomy anarchy hampers conservation. *Nature*, *546*, 25–27. <https://doi.org/10.1038/546025a>
- Garnett, S. T., Christidis, L., Conix, S., Costello, M. J., Zachos, F. E., Bánki, O. S., Bao, Y., Barik, S. K., Buckeridge, J. S., Hobern, D., Lien, A., Montgomery, N., Nikolaeva, S., Pyle, R. L., Thomson, S. A., van Dijk, P. P., Whalen, A., Zhang, Z.-Q., & Thiele, K. R. (2020). Principles for creating a single authoritative list of the world's species. *PLoS Biology*, *18*(7), e3000736. <https://doi.org/10.1371/journal.pbio.3000736>
- Global Biodiversity Information Facility (GBIF). (2014a). *New approaches to data licensing and endorsement*. GBIF <https://www.gbif.org/news/82363/new-approaches-to-data-licensing-and-endorsement>
- GBIF. (2014b). *Summary of responses to GBIF consultation: Licensing of data within the Global Biodiversity Information Facility*. GBIF.
- GBIF. (2018). *An alliance for biodiversity knowledge*. GBIF. <https://www.biodiversityinformatics.org/>
- GBIF. (2021). *Free and open access to biodiversity data*. Global Biodiversity Information Facility. <https://gbif.org/>
- Guala, G. F. (2016). The importance of species name synonyms in literature searches. *PLoS One*, *11*(9), e0162648. <https://doi.org/10.1371/journal.pone.0162648>
- Hawksworth, D. L. & Kalin-Arroyo, M. T. (1995). Magnitude and distribution of biodiversity. In V. Heywood (Ed.), *Global biodiversity assessment* (pp. 107–191). Cambridge University Press.
- Index Fungorum. (2021). *Index Fungorum*. Index Fungorum. <http://www.indexfungorum.org/>
- iNaturalist. (2021). *iNaturalist*. iNaturalist. <https://inaturalist.org/>
- Integrated Taxonomic Information System (ITIS). (2021). *Integrated taxonomic information system*. ITIS. <https://itis.gov/>
- International Union of Biological Sciences (IUBS). (2021). *International union of biological sciences*. IUBS. www.iubs.org
- International Committee on Taxonomy of Viruses (ICTV). (2021). *International committee on taxonomy of viruses*. ICTV. <https://talk.ictvonline.org/>
- International Commission on Zoological Nomenclature (ICZN). (2019). *Rotifer parts of the LAN*. ICZN. <https://www.iczn.org/list-of-available-names/rotifer-lan/>
- ICZN (2021). *Welcome to the ICZN*. ICZN. <https://www.iczn.org/>
- ICZN (n.d.). *List of available names*. ICZN. <https://www.iczn.org/list-of-available-names/>
- Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services (IPBES). (2021). *IPBES*. IPBES. <https://ipbes.net/>
- International Plant Names Index (IPNI). (2021). *International plant names index (IPNI)*. IPNI. <https://ipni.org/>
- International Union of Microbiological Societies (IUMS). (2021). *IUMS - International union of microbiological societies*. IUMS. <https://www.iums.org/>
- Linnaeus, C., (1753), *Species plantarum*. <https://www.biodiversitylibrary.org/page/358012>
- Linnaeus, C., (1758), *Systema Naturae* (10th ed.). <https://www.biodiversitylibrary.org/page/726886>

- Lücking, R. (2020). Three challenges to contemporaneous taxonomy from a lichen-mycological perspective. *Megataxa* 1(1), 78–103. <https://doi.org/10.11646/megataxa.1.1.16>
- McClure, C. J. W., Lepage, D., Dunn, L., Anderson, D. L., Schulwitz, S. E., Camacho, L., Robinson, B. W., Christidis, L., Schulenberg, T. S., Iliff, M. J., Rasmussen, P. C., & Johnson, J. (2020). Towards reconciliation of the four world bird lists: Hotspots of disagreement in taxonomy of raptors. *Proc. R. Soc. B.*, 287, 2020068320200683. <https://doi.org/10.1098/rspb.2020.0683>
- Microbiology Society. (2021). *International journal of systematic and evolutionary microbiology*. Microbiology Society. <https://www.microbiologyresearch.org/content/journal/ijsem>
- National Center for Biotechnology Information (NCBI). (2021). *GenBank Overview*. NCBI. <https://www.ncbi.nlm.nih.gov/genbank/>
- Observation International. (2021). *Observation international*. Observation International. <http://www.observation.org/>
- Open Tree of Life (2021). *Opentree*. <https://tree.opentreeoflife.org/>
- Palomares, M. L. D. & Pauly, D. (Eds.). (2021). *SeaLifeBase* (version 04/2021). SeaLifeBase. <https://www.sealifebase.org>
- Pensoft. (2021). *Science Publisher & Technology Provider*. Pensoft. <https://pensoft.net/>
- Plazi. (2017). *Welcome*. Plazi. <http://plazi.org/about/about-plazi/>
- Pyle, R. L., Barik, S. K., Christidis, L., Conix, S., Costello, M. J., van Dijk, P. P., Garnett, S. T., Hobern, D., Kirk, P. M., Lien, A. M., Orrell, T. M., Remsen, D., Thomson, S. A., Wambiji, N., Zachos, F. E., Zhang, Z-Q., Thiele, K. R. (2021). Towards a global list of accepted species V. The devil is in the details. *Organisms Diversity & Evolution* in press.
- Ruggiero, M. A., Gordon, D. P., Orrell, T. M., Bailly, N., Bourgoin, T., Brusca, R. C., Cavalier-Smith, T., Guiry, M. D., & Kirk, P. M. (2015). A higher level classification of all living organisms. *PLoS ONE*, 10(4), e0119248. <https://doi.org/10.1371/journal.pone.0119248>
- Sneath, P. (2005). The preparation of the Approved Lists of Bacterial Names. *Int J Syst Evol Microbiol*, 55(6), 2247–2249. <https://doi.org/10.1099/ijms.0.64137-0>
- Species 2000. (2019). *Home*. Species 2000. <https://species2000.org/>
- Species File Group. (2013). *Current species files*. Species File Group. http://help.speciesfile.org/index.php/Current_Species_Files
- Taxonomic Databases Working Group (TDWG). (2021). *Biodiversity information standards (TDWG)*. TDWG. <https://tdwg.org/>
- UNITE. (2020). *rDNA ITS based identification of Eukaryotes and their communication via DOIs*. UNITE Community. <https://unite.ut.ee/>
- Whittaker, R. H. (1972). Evolution and measurement of species diversity. *Taxon*, 21(2/3), 213–251. <https://doi.org/10.2307/1218190>
- World Register of Marine Species (WoRMS). (2021a). *An authoritative classification and catalogue of marine names*. WoRMS. <http://www.worldfloraonline.org/>
- WoRMS. (2021b). *Species databases*. WoRMS. <https://www.marinespecies.org/subregisters.php/>
- Zhang, Z. Q. (2013) Phylum Arthropoda. *Zootaxa*, 3703, 17–26. <https://doi.org/10.11646/zootaxa.3703.1.6>
- ZooBank (2021). *The official registry of zoological nomenclature*. ZooBank. <http://zoobank.org/>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.