

## A novel approach in fish cyber taxonomy

Sudhan C, Moulitharan N

Fisheries College and Research Institute, Thoothukudi, Tamil Nadu, India

### Abstract

A revolution in taxonomic practice is underway that will make taxonomy an even more reliable source of information for ecologists. The process by which information available in or converted into from non-digital sources, standardized electronic format is collated analysed and synthesised into a digital representation of one or more taxon concepts. The end product of which is a combination of outputs in both digital and non-digital format. Cyber taxonomy should not be confused with the practice of creating databases of taxonomic information. Databases are an integral part of the process, but cybertaxonomy refers to a wide range of hardware, software, instrumentation, communication tools, and work practices that collectively allow taxonomists to do their work more efficiently while maintaining the highest levels of excellence.

**Keywords:** fish, cybertaxonomy, hardware, software, instrumentation, communication tools

### Introduction

A revolution in taxonomic practice is underway that will make taxonomy an even more reliable source of information for ecologists. How taxonomic information is created, tested, accessed, thought about, and used, is changing dramatically with the emergence of cyber taxonomy.

### Definitions

A taxonomic work process that involves the use of standardised electronic tools to access information (databases, e-publications) and/or to generate knowledge bases.

The process by which information available in or converted into from non-digital sources, standardized electronic format is collated analysed and synthesised into a digital representation of one or more taxon concepts. The end product of which is a combination of outputs in both digital and non-digital formats.

### Cybertaxonomy

Cyber taxonomy is a contraction of "cyber-enabled taxonomy." It shares the traditional goals of taxonomy: to explore, discover, characterize, name, and classify species; to study their phylogenetic relationships; and to map their geographic distributions and ecological associations. Through cyber infrastructure and the adoption of digital technologies, cyber taxonomy is able to produce results faster and better than ever before (Wheeler 2008, 2010) <sup>[4]</sup>.

Cyber taxonomy should not be confused with the practice of creating databases of taxonomic information. Databases are an integral part of the process, but cybertaxonomy refers to a wide range of hardware, software, instrumentation, communication tools, and work practices that collectively allow taxonomists to do their work more efficiently while maintaining the highest levels of excellence. Consequently, users will find taxonomic information more reliable, comprehensive, and easily accessible. Cyber taxonomy, like traditional taxonomy, is integrative. That is, taxonomists pull together, synthesize, and analyze all available evidence that is informative at the taxonomic level(s) being studied.

Typical data sources include morphological, molecular, fossil, and ontogenetic as well as ethological, physiological, biochemical, and other sources of data, as appropriate. Cyber taxonomy can be visualized as a GIS-like environment with multiple data layers: morphological, distributional, molecular, image, and sound recordings (to name just a few). In addition, there are layers with algorithms and applications to process data in regard to phylogenetic, temporal, spatial, or ecological relations. Users may activate any combination of "layers" to retrieve desired information in a multi-layered "mesh". The possibilities are numerous and diverse: dichotomous or interactive diagnostic keys, checklists of species in particular areas plotted over seasonal occurrences, distribution maps (point data or predicted ranges based on climate and environment), three-dimensional visualizations of phenotypic variation, etc. Just as individual taxonomists have traditionally synthesized all available knowledge into periodically published monographs, international teams of experts collaborating in cyberspace will be able to assemble and maintain, with up-to-the-minute accuracy, all data and information pertaining to the species of a taxon. Even these taxon knowledge bases will be combinable by ecologists seeking to compare and contrast species that co-occur (or might co-occur, given future climate change, for example) in a particular ecosystem.

Taxonomy is the science central to exploring and understanding biodiversity, which exists a greater demand on this science to supply the increasing need of biodiversity knowledge. Cyber taxonomy is an assemblage of electronic taxonomic tools for accelerating species discoveries and application of taxonomic knowledge in biodiversity studies. This new approach utilizes standardized electronic tools to access information and generates knowledge bases, integrating the best of the Information technology revolution into the taxonomic processes, thus expediting the steps in identification and documentation. This is achieved mainly through electronic publications, electronic databases, factsheets and interactive identification keys.

### Taxonomy and Cybertaxonomy

Taxonomy, the science of discovery, naming and classification of living organisms is the basic and most essential tool in biodiversity applications. Taxonomy is more important than ever times of global change. In the light of the recent biodiversity loss and species extinctions, a great demand has emerged for ways to accelerate the discovery, documentation and application of knowledge regarding living organisms. Since, Taxonomy is the science central to exploring and understanding biodiversity, this science is now being pressured to supply the increasing need of biodiversity knowledge.

Cyber taxonomy means “Cyber-enabled taxonomy”. It shares the traditional goals of taxonomy – to explore, discover, characterize, name and classify species, to study their phylogenetic relationships and to map their geographic distributions and ecological associations. It uses an assemblage of electronic tools that will aid and equip the science of taxonomy in accelerating species discoveries and application of taxonomic knowledge in recent biodiversity studies. The results of taxonomic studies are ideally suited for dissemination in cyber space. The new internet resources enable and simplify taxonomic work and support filling gaps in biodiversity knowledge.

### Open Access Internet Databases

A number of global scale initiatives make biodiversity data freely available to a diverse user community for a variety of purposes.

1. Encyclopedia of Life (EoL) – <http://www.eol.org>
2. Global biodiversity Information Facility (GBIF) – <http://www.gbif.org>
3. Genbank – <http://www.ncbi.nlm.nih.gov/Genbank>
4. Atlas of Living Australia (ALA) – <http://www.ala.org.au>
5. Species 2000 – <http://www.sp2000.org>
6. ITIS catalogue of Life – <http://www.catalogueoflife.org>
7. Morphbank – <http://www.morphbank.net>
8. Zoobank – <http://www.zoobank.org>
9. Biodiversity Heritage Library (BHL)-<http://www.Biodiversitylibrary.org>
10. Fish base – <http://www.fishbase.org>
11. Open access internet databases generated through cyber taxonomy tools are an integral part of such initiatives which deliver free instant access to biodiversity information.

### Databasing Museum Collections

Data-basing of biological collections in museums increases the utility of biological data that have accumulated over time, which usually stays locked up in museum drawers. This can be achieved using software like SPECIFY, a versatile collection of research data management system with provisions for web interface and distributed generic information retrieval protocols. It can also manage multiple disciplines and have value added advanced options of Geo-referencing by Bio-Geomancer, locality mapping and latitude longitude converters.

Linking the world’s natural history collections into a single virtual database helps to build a better and deeper understanding on various aspects about a tax on. Since specimen labels hold the metadata of specimens, the digitized

images of the original labels allow an instant check of the original collection information at any pointy of time.

### Need For Cyber Taxonomy

1. Conservation, responsible use and trade, and law enforcement, e.g. resource use (fisheries and aquaculture), environmental assessment, and responsible trade mechanisms (e.g. certification, non-detriment findings [Convention on International Trade in Endangered Species of Wild Fauna and Flora – CITES]). Typical end users in this sector are fishery managers, environmental fishery statisticians, stock assessment scientists, law enforcement officers, fishers, fish farmers and fish traders.
2. Education, awareness building, consumer considerations and nonconsumptive uses, e.g. formal education (schools, universities, courses, etc.), aquaria and museums, consumer information, tourism (e.g. diving, sports fishing, and touring). Typical end users in this sector are consumers, environmental non-governmental organizations (NGOs), tourism planners, aquatic tour guides, divers and aquarium staff.
3. Research and development, e.g. scientific assessments and surveys, development of fishing and sampling methods, basic research. Typical end users in this sector (apart from the public or private research funding sources) are bioaquatic and fishery scientists and engineers, ichthyology curators and marine military officers.

### Status of Cyber Taxonomy

When fully implemented, cyber taxonomy will impact nearly every aspect of the creation and use of taxonomic information. Taxonomists require research resources on a scale unlike that of any other life science discipline: primary literature dating back to 1753, thousands of specimens from the full geographic ranges of scores of related species, type specimens to assure nomenclatural stability, and specimens and data from dozens of museums or herbaria in many countries. Cyber taxonomy holds the promise of unprecedented access to such resources, including digital image archives, open-access databases, remotely operable instrumentation, and electronic publishing tools.

Traditional printed sources of taxonomic information are almost always out of date, often by the time of release. Since publication, there may be any number of new species, name changes, distribution records, or natural history observations added, all of which are of great value to the ecologist. Cyber taxonomy will open access to full and current information drawn from a tax on knowledge base that is constantly updated by the taxonomic community.

Cyber taxonomy will increase the arsenal of species identification tools available to the field ecologist, including browser-based field guides, interactive diagnostic keys, automated identifications from photographs, direct access to a specialist, or the collecting of tissue samples for molecular identifications — the latter uniquely useful for associating very dissimilar life stages such as those observed between larval and adult insects.

### Biodiversity and Cyber Taxonomy

Biodiversity sciences have progressed at such a pace that the

taxonomic community has been unable to grow concomitantly to keep up with the influx of biological data. This “taxonomic impediment” has led some to suggest that taxonomy is no longer pertinent and to the development of methodologies that circumvent the taxonomic process. Recent informatics technology allows us to mobilize the 2 most important aphid taxonomy resources: experts and specimens, both distributed globally. “Cyberspecimens,” museum specimens digitally rendered at a resolution sufficient for remote identification, and open “cybertaxonomic” tools will allow the international aphid taxonomic community to carry out large, ambitious, projects.

### Delta

DELTA (DEscription Language for TAXonomy) is an advanced computer program developed by CSIRO (Commonwealth Scientific & Industrial Research Organisation, Australia) to handle all kinds of taxonomic data in the most optimal way. It is currently regarded as the state of the art tool in modern taxonomy and used by taxonomists to (re)describe taxa in a standardised format, which makes the information readily available for comparative (phylogenetic) studies, cataloguing of fauna lists and construction of interactive illustrated keys. Processing time and quality control of the data is being optimized by DELTA, as all data have to be treated in an identical manner.

It will provide an overview of current international biodiversity aggregators, such as the Global Biodiversity Information Facility, the Ocean Biogeographic Information System, Tree of Life and GenBank. Attention will be given to the Biodiversity Information Standards (developed by the Taxonomic Databases Working Group) which aim to improve efficient exchange of biodiversity data. Subsequently, encouraged to bring species (or higher level) data of their taxon of interest, which will be used to correctly build a database, construct interactive illustrated keys and generate natural language species descriptions.

### Other Taxonomic Tools Used For Identification

On-site taxonomist Trained taxonomists, preferably with a PhD in systematic biology and postdoctoral experience, are familiar with a large number of species and have specialist competence in a special group (e.g. a family or a fauna). They know about nomenclatural rules and morphometric methods for species identification and have a high awareness of the level of accuracy of their identifications. Moreover, they usually identify species relatively quickly. There may be conceptual differences between individual taxonomists that could lead to limited repeatability of certain identifications, but the accuracy should still be high. Taxonomists are most helpful with fresh or preserved whole specimens.

1. ID-tool: Local (folk) expert: Folk taxonomies are systems of categorization created by non-scientists in order to organize, name, and understand the natural world. Folk taxonomies frequently diverge on some points from the phylogeny established by the scientific study of taxonomy but they also tend to align with scientific classifications on other points.
2. Image recognition system: The identification process is based on the automatic characterization of image visual properties (e.g. colour, texture and shape) using computer vision techniques, i.e. image retrieval and/or classification

approaches that exploit feature vectors and similarity functions.

3. Field guides based on dichotomous keys: A taxonomic key is an ordered sequence of alternative choices, as provided by diagnostic (morphological) characters of organisms, that leads to a reliable identification of an organism or class of organisms. Diagnostic characters used in a key are defined and may be illustrated for clarity. The formal or taxonomic scope of a key is usually restricted to printed material or presented in digital format.
4. Integrated Photo-based Online Fish-Identification System (IPOFIS) exemplifying Interactive Electronic Keys (IEKs): An IPOFIS is a photo-based online fish identification system that integrates three methods: visual inspection, dichotomous keys, and a multiattribute query procedure. Each fish species is represented by multiple colour photographs of different individuals and close-ups of important identification features. The system efficiently organizes and presents these photographs and associated morphometric information in an interactive format that facilitates fast and accurate identification.
5. IPEz (morphometric software): IPEz is an automated, computer- software-based species identification system for marine and freshwater fish species. It uses a large number of morphometric measurements and it is based on machine learning techniques.
6. Other tools: scales, otoliths, bones, Acoustic fish identification, Genetic identification using barcoding, Genetic identification through single nucleotide polymorphisms (SNPs) and other fish base websites or search engines.

### Uses of Taxonomic Information in Ecology

As taxonomists make more species known, it will become possible to improve and better test assumptions and models on species diversity, distributions, co-occurrences, and co varying factors; detect climate change; track species losses or gains; deepen our understanding of interactions among species in ecosystems; improve strategies and priorities for conservation; and predict the expansion or contraction of ranges. Among the uses for taxonomic information in ecology are:

### Species Identifications

Cyber taxonomy will provide a wide range of tools to assist in the immediate and accurate identification of species, the training of field ecologists to identify target taxa, and video-mediated consultation.

1. Checklists: Cybertaxonomy will make checklists available for particular localities and ecosystems that are complete and up to the minute.
2. Names: Scientific names are the unique identifiers used to store and retrieve what we know in databases and publications. Cybertaxonomy will allow taxonomists to improve the reliability, stability, and informativeness of names and to retrieve all relevant data and citations recorded under older names.
3. Phylogenetic Classification: By viewing species within their phylogenetic context, ecologists can make predictions about their contributions to ecological functions. Closely related species frequently have similar

- food sources, reproductive strategies, physiologies, and behaviors.
4. Geographic Distributions: Cybertaxonomy is mobilizing the information content of the natural history collections of the world, who's estimated three billion specimens provide a wealth of information about the distributions of species, including irreplaceable historical records.
  5. Virtual Ecological Assemblages: For taxonomic research, specimens are curated according to their phylogenetic relationships. With the tools of cybertaxonomy, the ecologist can virtually reassemble all the specimens, regardless of taxon, collected in any one place at the same time or over a sequence of times.
  6. Conservation-relevant Information: Cybertaxonomy will facilitate access to information about the status, abundance, and rarity of species, the species-richness of particular localities, and other data relevant to conservation evaluation and reserve designs.
  7. Morphologically Structured Information: As image archives grow, incorporating both digitized publications and images of specimens, it will be possible to harvest and analyze such visual information to understand phenotypic variation in relation to environmental conditions, population structure, morphoclines, and other factors.
  8. Online Access to Museum Specimens: Telemicroscopy has been used by pathologists and histologists for decades and will soon be able to network global collections such that actual specimens in addition to stored images may be accessed, manipulated, compared, measured, and studied in real time.

### Conclusion

Cyber taxonomy is not only changing the way that taxonomists work but also the ways in which ecologists can access and make use of taxonomic data, information, and knowledge. As new ways to harvest, structure, and analyze taxonomic and related natural history information emerge, ecologists will be able to understand complex ecosystems in greater detail detect and monitor environmental change more precisely, and more effectively achieve goals for sustainable ecological services.

The new generation electronic publication tools act as an efficient interface bridging the gap between the specialist taxonomic community and a wider public through accelerated easy and effective dissemination of taxonomic knowledge. This will enable the conventional taxonomist to communicate more efficiently like never before on even minute organisms whose knowledge remained largely confined only among the expert circle. Knowledge thus disseminated can initiate numerous studies leading towards development of novel and ecologically significant insights on community structure and environmental influences, which ultimately enriches our knowledge on the living world.

### References

1. Gaston KJ, O'Neill MA. Automated species identification: why not? *Philosophical Transactions of the Royal Society B*. 2004; 359:655-667.
2. Polaszek, A. A universal register for animal names. *Nature* 2005; 437:477.
3. Wheeler QD, ed. *The New Taxonomy*. The Systematics Association Special Volumes Series 76. Boca Raton, FL: CRC Press, 2008.
4. Wheeler QD. What would NASA do? Mission-critical infrastructure for species exploration. *Systematics and Biodiversity*. 2010; 8:11-15.
5. Wilson EO. *The encyclopedia of life*. Trends in Ecology and Evolution. 2003; 18:77-80.
6. Winterton SL. Revision of the stiletto fly genus *Neodialineura* Mann (Diptera: Threvidae): an empirical example of cybertaxonomy. *Zootaxa*. 2009; 2157:1-33.
7. <http://cybertaxonomy.eu/blog/2007/04/11/what-is-cybertaxonomy.html>
8. Helfman GS, Collette BB, Facey DE. *The Diversity of Fishes*. Blackwell Science. 1997, 528.
9. Lagler KF, Bardach JE, Miller RR. *Ichthyology*. John Wiley & sons. 1962, 545.
10. Michael SW. *Reef Fishes*. Volume 1. A Guide to Their Identification, Behaviour, and Captive Care. Microcosm. 1998, 624.
11. Allen GR. *Marine Fishes of Tropical Australia and South-east Asia*. Western Australian Museum. 1997, 292.
12. Kuitert RH. *Coastal Fishes of South-Eastern Australia*. Crawford House Press. 1993, 437.
13. Kuitert RH. *Guide to Sea Fishes of Australia*. New Holland. 1996, 433.
14. Paxton JR, Hoese DF, Allen GR, Hanley JE. *Zoological Catalogue of Australia Vol.7 Pisces Petromyzontidae to Carangidae*. Canberra: Australian Biological Resources Survey. 1989; 2(12):1-665.
15. Randall JE, Allen GR, Steene RC. *Fishes of the Great Barrier Reef and Coral Sea*. Crawford House Press. 1997, 557.
16. <http://onlinelibrary.wiley.com/doi/10.1111/1744-7917.12088/abstract>
17. <http://taxonomytraining.eu/content/basics-taxonomy-describing-illustrating-and-communicating-biodiversity#overlay-context=content/modern-taxonomy-2014-2015%3Fdestination%3Dnode/233>
18. <http://www.fao.org/docrep/019/i3354e/i3354e.pdf>