



## Overview

# Viva La Revolución! Designing the digital renaissance in zoological taxonomy

David Yeates\*

Schlinger Fellow, Australian National Insect Collection and Taxonomic Research and Information Network (TRIN), CSIRO Entomology, PO Box 1700, Canberra, ACT 2601, Australia.

**Abstract** Web-based, digital taxonomy is developing rapidly, but the *International Code of Zoological Nomenclature* (ICZN) still requires that the primary data element of taxonomy, the description, be published in hard copy. If accepted, a recently proposed amendment to the ICZN allows for electronic publication of taxonomic names, with the names to be registered in a global, publicly accessible, web-based archive called *ZooBank*. This proposal will allow for contemporary models of publication and begin the process of establishing a global register of zoological names. While on the surface this sounds like a reasonably minor step in the pathway to the new taxonomy, the logical implications of this proposal are many and far reaching. For example, this change may lead to further advances so that zoological taxonomy bypasses traditional journal publication entirely, with *ZooBank* or some other electronic web-based outlet operating as the publishing vehicle and names register for the new taxonomy. While electronic publication is an important step in improving taxonomic practice, it is by no means a panacea for the critical shortage of taxonomic expertise in Australia or elsewhere.

**Key words** electronic publication, nomenclature, taxonomy, *ZooBank*.

## INTRODUCTION

Since the 1992 United Nations Earth Summit conference in Rio de Janeiro, Brazil, clear scientific evidence that we are on the verge of a major biodiversity crisis has been steadily mounting. Virtually all components of biodiversity are in steep decline and a large number of populations and species are likely to become extinct this century (Wilson 1985, 1992; Loreau *et al.* 2006). Taxonomy, one of the major scientific disciplines that underpins our knowledge of biodiversity and its management in such a crisis, has also been in decline for many decades (Wilson 1985, 2004). It is now a decade ago that workshops and meetings were held in Darwin and elsewhere that led to the Global Taxonomy Initiative (The Darwin Declaration 1998) after governments of the world realised that the targets and objectives of the Convention on Biological Diversity were critically limited by a lack of taxonomic expertise (Secretariat of the Convention on Biological Diversity 2007).

The magnitude of the problem is simple to describe. The world's 6000 taxonomists currently describe about 17 000 new species per year (State of Observed Species Report 2008), yet there are probably 8 million more species to be described (Wilson 2004). A telling example of the urgent need for progress in taxonomy is that 250 years after the publication of the 10th edition of Linnaeus's *Systema Naturae* we still lack a

catalogue of all scientific names of animals. Australia is home to a megadiverse fauna and a small human population, and the crises in biodiversity and taxonomy is as urgent here as in many developing countries (Lindenmayer 2007; Yeates & Raven 2007; National Taxonomy Forum 2008).

To address these challenges, there have been calls for a wholesale revision of the taxonomic process, especially to take into account contemporary developments in information technology (Godfray 2002; Wheeler *et al.* 2004). A new taxonomy is developing that is web-based, distributed, authoritative, accessible and relevant (Wooley 2006; Wheeler 2008). A number of research groups around the world are developing software and systems to enable the new taxonomy, for example, the *European Distributed Institute of Taxonomy* (<http://e-taxonomy.eu>), the *Creating A Taxonomic E-science* project (<http://www.cate-project.org/index.html>) and the *Taxonomic Research and Information Network* (<http://www.taxonomy.org.au>). These changes are a symptom of the impacts that progress in computing, information and communication technology is having in science and technology generally.

The phenomenal success of the journal *Zootaxa* is testament to the willingness of the taxonomic community to accept changes in the way they do business. The publication model for *Zootaxa* is rapid and 'almost' electronic only. Hard copies of papers are produced to satisfy the paper subscribers and this allows the new names to meet the current requirements of the Code of Zoological Nomenclature. The journal is 'almost' open access: authors can pay \$US20.00 per page to make their

\*david.yeates@csiro.au

works freely available from the *Zootaxa* website to readers who do not have a subscription to the journal. Papers can be published just a few days to a few months after acceptance by the editors. Since the inception of *Zootaxa* in 2001, it has published almost 100 000 pages of taxonomy, including descriptions of over 10 000 new taxa from 5000 authors. The number of pages published has dramatically increased in each year since 2001, with 23 000 pages published in 2007. These statistics show that *Zootaxa* alone now publishes a significant proportion of the world's taxonomic work.

The vast majority of species remaining to be described are terrestrial invertebrates, especially insects (Grimaldi & Engel 2005). Any changes to the way that taxonomy is done in future will largely affect the descriptions of new taxa created and published by invertebrate taxonomists. In this sense it is vitally important that Australian insect taxonomists understand the nature, scope and rationale of the proposed changes to the *International Code of Zoological Nomenclature* (ICZN). Availability in the sense of nomenclature means the process of formally making the names available to other scientists through publication. For scientific names to be available in this sense they must fulfill certain requirements that are specified in the *Code*. The proposed changes to the *Code* all relate to the ways that taxonomic names can be made available.

The International Commission on Zoological Nomenclature is an elected representative group of 28 commissioners dedicated to achieving stability and sense in the scientific naming of animals (<http://www.iczn.org/>). From time to time the ICZN produces editions of the ICZN. Zoologists comply with the *Code* on a voluntary basis; however, most journals require that authors comply with the rules of nomenclature in their submitted manuscripts. It is the responsibility of authors, journal editors and reviewers to ensure that manuscripts for publication comply with the *Code*. Compliance is virtually 100%, and the *Code* is a remarkable example of voluntary unity in science.

One notable and relevant exception to the universal application of the ICZN occurs in the taxonomy of Lepidoptera. A large part of the Lepidoptera taxonomic community does not adhere to the ICZN rule (Article 31.2) that the endings of genus and species names should agree in gender. This rule can cause confusion when a species is subsequently transferred to a genus of different gender than its original combination. The transfer necessitates that names proposed as Latin or latinised adjectives or participles in the nominative singular case are given a slight but automatic change in the last few letters of the species name. For example, the Small Copper butterfly was described over 150 years ago as *Thecla aurifer*. Over 100 years ago, it was transferred to the genus *Paralucia*, and is refereed to in the Catalogue of Australian Lepidoptera (Nielsen *et al.* 1996) as *Paralucia aurifer*. However, the two genera are of different gender, and strict application of the *Code* would mean changing the ending of the species name in *Paralucia* to *aurifera*. This can be confusing to computer search algorithms and human users alike because the two slightly different species names refer to exactly the same biological entity. In Lepidoptera, the original spelling of the

species name is used no matter what the gender of the genus if the species is transferred. Although this practice is technically a violation of the *Code*, it has gained wide acceptance among lepidopterists. Given the lack of a classical education for most taxonomists today, and the confusion caused by this rule, perhaps this is an area that needs to be revised in the next edition of the *Code*.

There are other aspects of the rules of nomenclature that will hinder the dawn of the new taxonomy. The four separate *Codes* of nomenclature (there are separate ones for animals, plants including fungi, bacteria and viruses) operate quite independently. In practice, this means that names only compete for priority within their own domain. For example, there are names of animals that are exactly the same as names of plants, but obviously refer to entirely different biological entities. The proteaceous plant genus *Lomatia* Brown 1810 is quite different from the bombyliid fly genus *Lomatia* Meigen 1822, but both are acceptable valid names according to the current *International Codes of Botanical and Zoological Nomenclature*. While a human can decipher the different meanings because of the context in which they are used, it is more difficult for computers to determine that these same names actually mean different things. It is surprising that, within the jurisdiction of each *Code*, a lot of effort goes into avoiding such homonyms, but the issue is entirely ignored between *Codes*. To minimise this problem, it is likely that in the future, new scientific names should compete for priority across all *Codes*.

The *Code of Zoological Nomenclature* has a number of basic principles to foster stability in the names of animals, the two most important being the principle of priority (the first name applied to a taxon is the accepted one), and the principle of types (names are allocated to taxa by means of a type specimen for species or type taxon for genera and families). It is important to recognise that the *Code* regulates the use of names, but does not legislate on the quality of taxonomy. The code is a tool taxonomists use to stabilise the application of names in taxonomy, not the scientific issues associated with the process of taxonomy. For example, there is no requirement for taxonomic works to be published in a peer-reviewed journal. The *Code* simply states that numerous simultaneously produced copies of a taxonomic work must be made available. It is possible to publish taxonomic names without peer review, in a privately published document or journal. The *Code* also does not regulate the quality of taxonomic decisions, the *Code* simply states that the purpose of the work must be to describe taxa and that characters must be given that purport to differentiate new taxa. Whether these characters actually do differentiate taxa efficiently and sufficiently is beyond the scope of the *Code*. Electronic publication of taxonomic names will not increase the incidence of 'rogue' taxonomy because it is quite easy to publish poor quality taxonomy under the current rules using ink and paper or CD-ROMs.

A recent attempt (Fisher & Smith 2008) to publish some new species names in the electronic journal *PLoS ONE* caused heated discussion among ICZN commissioners and taxonomists alike. While the descriptions in the article were excel-

lent, and included extensive illustrations, distribution maps, a molecular phylogeny and embedded links to web resources (Global Unique Identifiers, GUIDs), the current *Code* explicitly disallows electronic publication. The new species names coined in this article should not be used by other scientists because the journal does not comply with the current *Code* of nomenclature. Since the publication of Fisher and Smith (2008), paper and pdf copies of the article were deposited in six, major publicly accessible libraries and the scientific names were thereby made available from this paper version.

This episode catalysed the International Commission on Zoological Nomenclature to propose changes to the *Code* that allow a radical change to the taxonomic process (ICZN 2008). In short, the proposed rule changes will allow for species to be named in publications appearing in electronic form only, provided that these names are registered in a central web-accessible database, *ZooBank*. These kinds of changes have been considered by the taxonomic community for some time, and the botanical community expressly rejected mandatory registration in 2006. Why are these proposed changes to zoology so radical? Why are they so important?

### THE ICZN PROPOSAL

In brief, the International Commission on Zoological Nomenclature proposed the following three amendments to the *Code of Zoological Nomenclature* in October 2008 ([http://www.iczn.org/electronic\\_publication.html](http://www.iczn.org/electronic_publication.html)). After a 12-month period of scrutiny and comment from the zoological community, the Commission will reconsider the amendments and vote upon them. In the case of a positive vote, the amendments will come into effect during 2010.

- 1 That names of new species, genera and families can be validly published in an electronic-only format. There is no need for a paper issue to make the names available as long as a mechanism can be found to archive these digital data reliably.
- 2 That names published in electronic-only form must be registered in a specific on-line database *ZooBank* (ICZN 2005, <http://www.ZooBank.org/>). A registration number provided by *ZooBank* must be quoted for each name published in electronic-only form. The registration information must include the internet address of the archiving organisation. Names published in paper only or paper and electronic form do not have to be registered.
- 3 The publication of descriptions of species, genera and family names on CD-ROM or DVD will not be allowed under the *Code* after 2009.

### HOW WILL THESE CHANGES AFFECT THE STATUS QUO?

There is currently no central list, repository or clearing house for new taxonomic names. New species and other taxa are

described in hundreds of journals and other published works throughout the world, including this one. Many of these publications are difficult to obtain, and not available in electronic form. Some of these publications, and hence the names they contain, can be found by searching *ISI Web of Knowledge* (<http://apps.isiknowledge.com/>) but many cannot. A large proportion of taxonomic journals are not indexed by the *Web of Knowledge*. Thompson Reuters attempts to gather and report on all new taxonomic names in *Zoological Record*, and while this is an extremely valuable online service, some names are overlooked or misspelled. Hence it is almost impossible (short of searching through hundreds of journals) for a taxonomist to provide an authoritative, real-time list of the new species of insects described each year. Initiatives such as the *Catalogue of Life* rely on other 'aggregators' such as *Zoological Record*, Species 2000, ITIS and individual scientists to provide names. For Diptera, for example, Chris Thompson and colleagues have spent a large portion of their careers developing a global list of Diptera names (<http://www.diptera.org>).

The growing range of initiatives (*The Encyclopedia of Life* and the *Atlas of Living Australia* being three examples) in biodiversity informatics that need access to authoritative lists of current taxonomic names need to be better served than the current model of *post hoc* aggregation. The ICZN hopes that *ZooBank* will serve as such an authoritative archive and clearing house, initially for taxonomic names published electronically, but eventually for all taxonomic names. These proposed amendments start the process.

The critical issue of concern for the ICZN, if electronic publication is introduced, is the ability of these electronic media to be accessed in the future. Archiving systems for electronic media are developing (e.g. Portico and LOCKSS), and some standards are available (ISO standard 14721: 2003 for Open Archive Information System), but it is fair to say that a winning technology has not yet emerged. The great advantage of paper publication has been its quality as an archiving medium. Even the paper-based texts of the earliest works of zoological nomenclature are easily accessed now without any technological tools.

### HAS THE ICZN GONE FAR ENOUGH?

The proposed changes to the *Code* require that three pieces of information must be included in *ZooBank*:

- 1 The name itself, and the author/s of it.
- 2 The relevant Internet address and bibliographic information so that others can find the work in which the name was proposed.
- 3 For species names, the collection that contains the type specimen/s (and for genus group names, the type species and similarly for family group names, the type genus).

If the ICZN's proposed changes are accepted, a search of *ZooBank* will return names, references to the original sources of these names, and type repositories for all zoological

species-group names published electronically after 2010. This will be a great advance for taxonomists and non-taxonomists. For taxonomists, the *ZooBank* interface can provide automatic checking for *Code* compliance, such as fixing genders (comments on the Lepidoptera-style abandonment of this need, notwithstanding) and stems and checking for homonymy. For non-taxonomists it will provide authoritative access to all taxonomic names, correctly spelled and *Code* compliant. *ZooBank* or something like it is a critical component of the digital web-based revolution in taxonomy. Initially, names published on paper or in both paper and electronic form don't need to be added to *ZooBank*, but the ICZN strongly encourages authors to do so.

The proposed process requires authors of electronically published taxonomy do two different but connected things to make names available: (1) publish them in an electronic journal that has an acceptable archiving standard, and (2) register the names in *ZooBank*. If only one of these things is completed, the name is not available (publication + registration = availability, Polaszek *et al.* 2008). This creates the possibility of a 'grey area', if for example, the name is registered but not yet published to the appropriate archiving standard, and vice versa. In practice, authors will be given provisional registration numbers from *ZooBank* to include in publications, and when the electronic publication appears, authors will amend the *ZooBank* reference to include the full bibliographic record. The date of publication will be the date when the electronic publication appeared, not the date of registration. If it is found that a registered name has not been published in an appropriately archived electronic journal, the case is referred to the ICZN for a decision

### JOURNALS: ARE THEY PART OF THE PROBLEM, OR PART OF THE SOLUTION?

What is the role of journals and their publishers in taxonomy? Management of the peer review process is as important for taxonomy as for any other field of science, and this is managed by journals and their publishers. Journals also have an important role in maintaining the quality of words and images published for the scientific record. Journals themselves also provide a permanent archive of their contents, and the ICZN requires that this now be formally standardised for electronically published taxonomic journals. The editorial policies of journals are important, so that editors and reviewers can ensure that zoological authors comply with the *Code*.

The obvious solution that would clear up the 'grey area' caused by the two-step process described above is to make *ZooBank* the archiving system (publication = registration = availability, Polaszek *et al.* 2008) for the taxonomic descriptions themselves, but there are a number of obstacles in the way. Authors currently assign copyright for their work to a journal publisher, so they may be in breach of copyright by adding their description to *ZooBank*. A simple solution to this problem would be for journal publishers such as Wiley Blackwell to

agree that standard copyright law should not apply to taxonomic works – their content is the factual record of observation for the public record.

Current copyright legislation aside, it is easy to imagine a possible future process whereby *ZooBank* registration itself confers publication and availability. *ZooBank* or an associated journal would need to take on the functions that journals now do for taxonomy. It would become a dedicated nomenclatural and taxonomic outlet, managing the peer review and archiving function for taxonomy. This is already the case in bacterial taxonomy, where only taxa appearing in the *International Journal of Systematic and Evolutionary Microbiology* are made available under the Code of Bacterial Nomenclature.

### WHAT IS IN IT FOR TAXONOMISTS?

Registration of a zoological name in *ZooBank* will be free, and available via a web dialogue box. It will, however, take taxonomists a few more minutes of additional time for each taxonomic name in addition to the current process. Modern taxonomists only describe a few hundred taxa in their lifetime (Evenhuis & Thompson 2004), so this will hardly be much of a burden over a professional career. The fact that the registration process will bring with it *Code* compliance is a definite advantage as most working taxonomists do not have an intimate knowledge of the *Code* or of the complexities of Latin and Greek grammar. As *ZooBank* grows it will become a resource that taxonomists can use to find names and other information connected to them. Lastly, the work of taxonomists will be more easily found and will be more accessible and visible via *ZooBank*, and this must be an advantage.

It is important to see the current proposed amendments to the *Code* as an important first step in the new taxonomy. A number of recommendations included in the proposed amendments will increase the value of *ZooBank* to the scientific community. Recommendations are 'guides to best practice' but not enforceable as the rules are. The ICZN intends to recommend that taxonomic works published electronically are also published simultaneously in paper, and that all new names (electronic or not) are entered into *ZooBank*. *ZooBank* can only become a more valuable resource for taxonomists and others if other information such as descriptions, distributions and type localities are also entered into the *ZooBank* register.

*ZooBank* itself, or electronically enabled taxonomy are not panaceas for the crisis in taxonomy and biodiversity (Thiele & Yeates 2002; de Carvalho *et al.* 2007): they are simply more efficient ways to publish one of the results of taxonomic research in a web-enabled world. To counter the crisis in taxonomy, we need both an increase in the taxonomic effort, and an increase in taxonomic throughput per unit effort during the taxonomic process. The publication step is very rarely the rate-limiting one in the taxonomic process. The limiting steps are often associated with curating research material to an appropriate state for critical taxonomic observations, and determining the identity of taxa in a study group that have

already been given names. In practice, this usually requires extensive personal study of old type specimens held in museums overseas.

## ACKNOWLEDGEMENTS

This document benefitted from comments and suggestions made by Bruce Halliday (CSIRO Entomology), Donald Hobern (Director, Atlas of living Australia) and Thomas Pape (Zoological Museum, Copenhagen). Ted Edwards kindly orientated me to the nomenclatural situation in Lepidoptera.

## REFERENCES

- de Carvalho MR, Bockmann FA, Amorim DS *et al.* 2007. Taxonomic impediment or impediment to taxonomy? A commentary on systematics and the cybertaxonomic-automation paradigm. *Evolutionary Biology* **34**, 140–143.
- Evenhuis NL & Thompson FC. 2004. Bibliography of and new taxa described by D. Elmo Hardy (1936–2001). *Bishop Museum Bulletin of the in Entomology* **12**, 179–222.
- Fisher BL & Smith MA. 2008. A revision of Malagasy Species of *Anochetus* Mayr and *Odontomachus* Latreille (Hymenoptera: Formicidae). *PLoS ONE* **3**, e1787.
- Godfray HCJ. 2002. Challenges for taxonomy. *Nature* **417**, 17–19.
- Grimaldi D & Engel M. 2005. *Evolution of the Insects*. Cambridge University Press, Cambridge, UK.
- ICZN. 2005. A universal register for animal names. *Nature* **437**, 477.
- ICZN. 2008. Proposed amendment of the International Code of Zoological Nomenclature to expand and refine methods of publication. *Zootaxa* **1908**, 57–67.
- Lindenmayer D. 2007. *On Borrowed Time: Australia's Environmental Crisis and What We Must Do about it*. Penguin Books and CSIRO Publishing, Camberwell, Victoria, Australia.
- Loreau MA, Oteng-Yeboah MTK, Arroyo D *et al.* 2006. Diversity without representation. *Nature* **442**, 245–246.
- National Taxonomy Forum. 2008. *Proceedings of the National Taxonomy Forum*. Federation of Australian Scientific and Technological Societies and The Australian Biological Resources Study.
- Nielsen ES, Edwards ED & Rangsai TV. 1996. *Checklist of the Lepidoptera of Australia*. Monographs on Australian Lepidoptera Vol. 4. CSIRO Publishing, Collingwood, Victoria, Australia.
- Polaszek A, Pyle R & Yanega D. 2008. Animal names for all: ICZN, zoobank and the new taxonomy. In: *The New Taxonomy* (ed. QD Wheeler), pp. 129–141. Taylor and Francis, Boca Raton, FL, USA.
- Secretariat of the Convention on Biological Diversity. 2007. *Guide to the Global Taxonomy Initiative*. CBD Technical Series No. 30. p. 195.
- State of Observed Species Report. 2008. [Accessed 26 Nov 2008.] Available from URL: <http://species.asu.edu/files/sos.pdf>
- The Darwin Declaration. 1998. Report of a workshop 'Removing the Taxonomic Impediment' Darwin, Australia, 3–5 February 1998, as endorsed by The Conference of the Parties to the Convention on Biological Diversity (COP CBD).
- Thiele K & Yeates DK. 2002. Tension arises from duality at the heart of taxonomy. *Nature* **419**, 337.
- Wheeler QD, ed. 2008. *The New Taxonomy*. Taylor and Francis, Boca Raton, FL, USA.
- Wheeler QD, Raven PH & Wilson EO. 2004. Taxonomy: impediment or expedient? *Science* **303**, 285.
- Wilson EO. 1985. The biological diversity crisis: a challenge to science. *Issues in Science and Technology* (Fall), 20–29.
- Wilson EO. 1992. *The Diversity of Life*. Harvard University Press, Boston, USA.
- Wilson EO. 2004. Taxonomy as a fundamental discipline. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences* **359**, 739.
- Wooley JB. 2006. Name registration: one fewer impediment to taxonomy. *Bulleting of Zoological Nomenclature* **63**, 98–101.
- Yeates DK & Raven PH. 2007. Australian biodiversity: threats for the present, opportunities for the future. *Australian Journal of Entomology* **46**, 177–187.

Accepted for publication 30 May 2009.