ZOOTAXA

1950

Updating the Linnaean Heritage: Names as Tools for Thinking about Animals and Plants

ALESSANDRO MINELLI, LUCIO BONATO & GIUSEPPE FUSCO (EDS)

Magnolia Press
Auckland, New Zealand

ALESSANDRO MINELLI, LUCIO BONATO & GIUSEPPE FUSCO (EDS)
Updating the Linnaean Heritage: Names as Tools for Thinking about Animals and Plants
(*Zootaxa* 1950)

163 pp.; 30 cm.

5 Dec. 2008

ISBN 978-1-86977-297-0 (paperback)

ISBN 978-1-86977-298-7 (Online edition)

FIRST PUBLISHED IN 2008 BY

Magnolia Press

P.O. Box 41-383

Auckland 1346

New Zealand

e-mail: zootaxa@mapress.com

http://www.mapress.com/zootaxa/

ISSN 1175-5326 (Print edition)
ISSN 1175-5334 (Online edition)

Zootaxa 1950: 3–4 (2008)
www.mapress.com/zootaxa/

ISSN 1175-5326 (print edition)
ZOOTAXA
ISSN 1175-5334 (online edition)

Updating the Linnaean Heritage: Names as Tools for Thinking about Animals and Plants

ALESSANDRO MINELLI FLS, LUCIO BONATO & GIUSEPPE FUSCO (EDS)
Department of Biology, University of Padova, Via Ugo Bassi 58B, I 35131 Padova, Italy
Email: alessandro.minelli@unipd.it, lucio.bonato@unipd.it, giuseppe.fusco@unipd.it

Table of contents

Preface

In December 2007, *Zootaxa* celebrated the tercentenary of the birth of Linnaeus with an important collection of 30 articles published as volume 1668 under the title *Linnaeus Tercentenary: Progress in Invertebrate Taxonomy.*

Now, in December 2008, *Zootaxa* publishes another thematic collection, as a contribution to celebrate the 250th anniversary of the publication of the X edition of *Systema Naturae*, the Linnaean work that represents the starting point of current scientific nomenclature in zoology.

The present issue is based on the papers presented at an international meeting held at the University of Padova, Italy, in cooperation with the Linnean Society of London. During the two-day meeting (29–30 May 2008), thirty scientists from seven countries discussed from different perspectives the current role of nomenclature in communicating science. The subject matter was not limited to names for species and higher taxa, but also included anatomical nomenclature and the relationships between names, theory, and generation of questions on development and evolution.

A substantial selection of the papers presented at the meeting is published in this volume of *Zootaxa*, together with three additional articles written by some participants other than invited speakers, also dealing with the same range of topics as those formally presented in Padova.

The editors are most grateful to the Linnean Society of London for its generous support in the preparation and realization of the meeting and to *Zootaxa* for not less generously hosting in its pages a collection from which we hope there will be scope for stimulating further discussion.

Alessandro Minelli FLS, Lucio Bonato and Giuseppe Fusco

Zootaxa 1950: 5–8 (2008)
www.mapress.com/zootaxa/

ISSN 1175-5326 (print edition)
ZOOTAXA
ISSN 1175-5334 (online edition)

Actual usage of biological nomenclature and its implications for data integrators; a national, regional and global perspective*

CHARLES HUSSEY[1], YDE DE JONG[2] & DAVID REMSEN[3]

[1]*Natural History Museum, Cromwell Road, London SW7 5BD UK. E-mail: c.hussey@nhm.ac.uk*
[2]*University of Amsterdam, P.O. Box 94766, NL-1090 GT, Amsterdam, The Netherlands. E-mail: yjong@uva.nl*
[3]*Global Biodiversity Information Facility, Universitetsparken 15, DK-2100 Copenhagen, Denmark. E-mail: dremsen@gbif.org*

* *In:* Minelli, A., Bonato, L. & Fusco, G. (eds) *Updating the Linnaean Heritage: Names as Tools for Thinking about Animals and Plants. Zootaxa*, 1950, 1–163.

Abstract

Biological names play an important role in resource identification and as anchors for all sorts of associated information. This is borne out in ever-expanding online resources but the ways in which names are stored and presented give rise to challenges and pitfalls that can lead to missed or misinterpreted information. These resources must serve a variety of users and keep abreast of changes in nomenclature and systematics. Observations on the use of biological names are presented and some solutions to the challenges are offered.

Key words: Biodiversity, Digital data bases, Nomenclators

Accurate identification of organisms and correct use of biological names is essential in order to apply correct measures in the fields of conservation and to control pest and disease causing organisms. As has been pointed out by Grimaldi and Engel (2005) "*All accumulated information of a species is tied to a scientific name, a name that serves as a link between what has been learned in the past and what we today add to the body of knowledge*". While the veracity of the statement holds true, the nature of taxonomy and nomenclature present significant obstacles to taking advantage of this universal link between a taxon name and the accumulated information.

Efforts to mobilise biodiversity information have now yielded significant online resources, and these are set to grow enormously in the future. The Internet is revolutionising accessibility but also creates its own set of obstacles to discovery and retrieval of information based on taxonomic names. At the time of writing, the Global Biodiversity Information Facility network (http://www.gbif.org) has mobilized nearly 150 million collection and observation records from nearly 3,000 individual datasets. The Biodiversity Heritage Library (http://www.biodiversitylibrary.org) has recently passed the 7 million page mark toward its goal of digitizing an estimated 2–3 million publications relating to species. The National Center for Biotechnology Information (http://www.ncbi.nlm.nih.gov) stores tens of millions of gene sequence relating to more than three hundred thousand taxa. The Biodiversity Information Standards (http://www.tdwg.org) website lists 592 different biodiversity informatics projects, all of which are mobilizing, serving, integrating and exchanging species information. Each of these resources shares a common dependence on taxon names to provide the species context to the associated information.

There are also numerous information resources relating to the compilation and reconciliation of taxon names. The authors have each been engaged in long-term projects that involve collating lists of taxonomic

Accepted by A. Minelli: 24 Oct. 2008; published: 5 Dec. 2008

names from disparate sources and making these accessible through the Internet (National Biodiversity Network Species Dictionary http://www.nhm.ac.uk/nbn, Nature Navigator http://www.nhm.ac.uk/naturenavigator, Fauna Europaea http://www.faunaeur.org, GBIF Electronic Catalogue of Names http://www.gbif.org/prog/ecat, uBio http://www.ubio.org).

In our experience, taxon names present a range of challenges that must be addressed in order to realize their potential as useful data discovery devices. Among the most fundamental are conceptual separations that tend to be blurred and can hinder communications. In particularly there is little appreciation of the difference between taxonomy and nomenclature. This impacts upon concepts as simple as what is meant by the term "correctness" as it relates to taxon names or as fundamental as the term "name" itself. All taxa are referred to by a name but not all names refer to currently recognized taxa. A "correct name" in the taxonomic sense may refer to the valid or accepted name for a taxon, whilst in the nomenclatural sense it may refer to the fitness of the name relative to the codes of nomenclature. The scope of a name can change over time, as the result of improved knowledge, or through differing opinions amongst experts. Strictly speaking, names should be replaced by taxon concepts, which are names linked to explicit usage. This requires a record of the source used in making a taxonomic determination – metadata that is most often missing from a species observation record.

In these days of internet resources, this lack of distinction can lead to confusion, relevant information being missed, or information being incorrectly associated and returned in a search result. Among the challenges are the following:

1. Matching a name entered via a search to a name actually recorded within a data repository. Computers excel at comparing text strings but differences in how names are recorded can result in both false negative and false positive returns.
2. There is, as yet, no comprehensive catalogue or index representing all the taxon names that exist, nor any accurate measure as to the true number of names.
3. It is becoming apparent that copies of some datasets are included in other resources, with or without the agreement of the original data owners. Any errors in the original dataset may therefore be carried through to the derivative resource and may persist even after the original has been corrected.

Bearing in mind that there are around 1.75 million described species (Wilson 2003), it is noteworthy that the number of names assembled by uBio within their *NameBank* currently amount to 11 million distinct name records. This is due primarily to the fact that *NameBank* records distinct verbatim name strings (name+authorship) combinations and reconciles these to a single logical group. The Catalogue of Life (http://www.catalogueoflife.org), on the other hand, starts with quality lists that have been reconciled to single representative name records and therefore, the 2008 edition, representing over 1.1 million species, has a total of 2 million name records. In the United Kingdom, which has around 80,000 species (excluding bacteria and viruses), the NBN Species Dictionary holds 245,000 name strings.

These counts include representational forms of both taxonomically and nomenclaturally valid and invalid names. Many of these will be obsolete names, subsequently made synonyms of current names. In addition, a single name may present a wide range of variability in how it is actually recorded within a dataset. The name may be non code-compliant, have a wrong endings to the species epithet, or be simply misspelled. Variation may also occur in how authorship is represented (such as abbreviations and inconsistent use of diacritical marks). Such variation presents challenges in federated data environments where inconsistency is the rule.

Thus a single taxon name in the more traditional nomenclatural sense may be represented by many small lexical variations of that name, as they have been discovered within biodiversity resources. Nonetheless, the 11 million records within the uBio *Namebank* represent over 4 million distinct taxon names if authorship is no longer a factor. Interestingly there are over 4.8 million distinct name strings, based on this same definition, within the current GBIF indices, which can be reconciled to 3.4 million distinct names. Of these, only 11% are listed in the *Catalogue of Life* and only about 25% are believed to overlap those found within the *NameBank*.

This serves to show the scale of one of the challenges facing biodiversity informatics.

Recording names in use, even where they are erroneous, enables query expansion: provided that effort is put into mapping these names to their accepted forms. Mapping of names can, to a certain extent, be automated: uBio and GBIF, for instance, have developed their *LexMapper* algorithm to handle this. Older names will increasingly need to be tracked, now that specimen collections are being digitised, as well as the historic literature. Homonyms represent an unquantified but significant issue, even within a single biological kingdom, and become even more of an issue when resources span several kingdoms. In the process of digitising *Nomenclator Zoologicus*, 21,000 homonym groups were identified (Remsen et al. 2006). GBIF is developing an All Genus Index (AGI) that should identify all genus-level homonyms (Remsen & Patterson 2007).

Storing, comparing, exchanging and searching for taxonomic names and classification schemes also present challenges. Search portals usually offer the facility to search using the genus or species epithet. However, name strings can contain up to 14 words in the case of plant hybrids. Because database searches rely on string matching, variants in spelling (such as presence or absence of diacritical marks) can lead to missed records – unless such variants have been mapped to accepted forms.

Various data models and exchange standards have been developed over the years (ABCD, Berlin Taxonomic Information Model, Darwin Core, EDIT Common Data Model, Nomencurator, Taxonomic Concept Transfer Schema) to cope with biological names and classifications. It is possible to use these schemas to wrap name data to common formats, even if the underlying database has a unique structure. It should be borne in mind, however, that many of the data providers, particularly those involved in local and national recording schemes, may not be willing or able to use complex systems and, instead, often record and present data using simple spreadsheets or documents.

Biologists look to nomenclators and taxonomic indexing services for help in checking current names and their authorities, which is only possible if synonymies are included. But users will also include conservationists, developers and planners, local and national government, environmental agencies, biological recorders and members of public, who may have different needs. For instance, biological recorders require the inclusion of recording aggregates (an amalgam of species that are difficult to identify in the field) and wish to record against names that they are familiar with. Many users, who are not practising taxonomists, are not concerned with the niceties of nomenclatural and taxonomic rules, such as the use of subgenera and authorities – they just want a reliable name! Some sectors (e.g. birds, butterflies, mammals) routinely use common names. Informal names are also helpful for higher taxonomic groupings. Even biologists will be unfamiliar with names of genera, families and orders outside of their own speciality and it can greatly help if search results assign each scientific name to a familiar higher grouping.

There are numerous initiatives at national level, fewer at regional level and even fewer resources at global level. National coverage, both in terms of expertise and content, is uneven. There are numerous instances where data exists but are yet to be made accessible. Whilst a single checklist can achieve consistency, through being based upon a single taxonomic opinion, when datasets are assembled from multiple sources, these sources may employ different classifications and synonymies. Often, however, there is a preferred classification for a taxonomic group at a national level. The correspondence between vernacular names and scientific names may differ between countries and even the accepted scientific name for a species can vary. Equivalencies can be determined by assigning Globally Unique Identifiers (GUIDs) to taxa. It is, however, important that systems are able to allow for and support different taxonomic opinions. There also needs to be an effective exchange of information between national, regional and global initiatives. In that way, new occurrence records can be fed upwards and changes to nomenclature can be fed back to biological recorders.

What is needed is sustainable, long-term, initiatives that will deliver maintained taxonomic indexes and nomenclators. Whilst it is possible that the Lifewatch (for Europe) and Encylopedia of Life projects (http://www.lifewatch.eu/, http://www.eol.org), together with GBIF, will provide high-level access to data, the challenge is to secure support for the hundreds of individual data contributors. All resources, whether nomenclators or species inventories, should be kept abreast of changes: in order to be able to gauge whether a name is current and also whether a species occurrence is current. This requires continuous effort, and it is not easy to secure funding for this sort of activity. To take things forward, more attention should be given to the mapping of obsolete and malformed names to code-compliant accepted names, to flag the status of names, and to capture vernacular names. Development of a management classification will help ensure that consistent results are returned from searches across distributed datasets. Authorities are necessary to give attribution to a name, but the abbreviated form in which they are presented (in both botany and zoology) does not enable the determination of the underlying bibliographic reference. The increasing availability online of scientific literature should be complemented by a resource that not only links species names to their original description, but does the same for species recombinations (comb. nov.). It is to be hoped that use of GUIDs will become commonplace; with a management system that resolves multiple GUIDs that may get assigned to a single taxon concept. It is important that the provenance of datasets is indicated whenever records are displayed or downloaded. Attribution also provides welcome acknowledgement of the work of data providers, many of whom work on a voluntary basis. Above all, it is at the human level that action is required. Action to promote best practice in the use of names. Also action to mobilise the biological community to assist with error detection and correction, and to both share and consolidate resources, in order that the current duplication of effort may be reduced.

References

Grimaldi, D. & Engel, M.S. (2005) *Evolution of the Insects*. Cambridge University Press, Cambridge, 786 pp.

Remsen, D.P., Norton, C. & Patterson, D.J. (2006) Taxonomic informatics tools for the electronic *Nomenclator Zoologicus*. *Biological Bulletin*, 210, 18–24.

Remsen, D.P. & Patterson, D.J. (2007) *Building an index of all genera: A test case in interchange*. Available from: http://www.tdwg.org/proceedings/article/view/265 (accessed 6 September 2008).

Wilson, E.O. (2003) The encyclopedia of life. *Trends in Ecology and Evolution*, 18, 77–80.

Additional sources

Anonymous (2005) *Taxonomic Concept Transfer Schema*. Available from: http://tdwg.napier.ac.uk/ (accessed 6 September 2008).

Berendsohn, W.G. (page editor) (2007a) *Access to Biological Collection Data (ABCD) Schema*. Available from: http://www.bgbm.org/tdwg/codata/schema/ (accessed 6 September 2008).

Berendsohn, W.G. (2007b) *The Berlin Taxonomic Information Model*. Available from: http://www.bgbm.org/biodivinf/Docs/BGBM-Model/default.htm (accessed 6 September 2008).

Biodiversity Information Standards (2007) *Darwin Core*. Available from: http://www.tdwg.org/activities/darwincore/ (accessed 6 September 2008).

EDIT [European Distributed Institute of Taxonomy] (2008) *Common data model*. Available from: http://dev.e-taxonomy.eu/trac/wiki/CommonDataModel (accessed 6 September 2008).

Ytow, N., Morse, D.R. & Roberts D.McL. (2001) Nomencurator: a nomenclatural history model to handle multiple taxonomic views. *Biological Journal of the Linnean Society*, 73(1), 81–98. See also http://www.nomencurator.org/ (accessed 6 September 2008).

Zootaxa 1950: 9–20 (2008)
www.mapress.com/zootaxa/

ISSN 1175-5326 (print edition)
ZOOTAXA
ISSN 1175-5334 (online edition)

The Linnean foundations of zoological and botanical nomenclature*

OTTO KRAUS

Zoological Institute & Museum, University of Hamburg, Martin-Luther-King-Platz, 20146 Hamburg, Germany.
E-mail: otto.kraus@zoologie.uni-hamburg.de

* *In:* Minelli, A., Bonato, L. & Fusco, G. (eds) *Updating the Linnaean Heritage:Names as Tools for Thinking about Animals and Plants. Zootaxa*, 1950, 1–163.

Table of contents

Abstract

In contrast to primary taxonomic research, nomenclature is a tool to be used in information retrieval and with the aim of ensuring world-wide understanding. Radical changes to its guiding principles would be counterproductive as they would produce disorder in a system of internationally accepted conventions that have developed and matured over decades. Classifications are ordering systems. They are even more utilitarian in function and should not be constantly adapted to potentially ephemeral phylogenies. In principle, the Linnaean foundations of nomenclature and classification survive. Attempts to introduce fundamental changes or even alternatives were proposed a century too late to get enforced; they therefore failed.

Key words: Classification, History of systematics, Linnaean hierarchy, Nomenclature

Accepted by A. Minelli: 24 Oct. 2008; published: 5 Dec. 2008

Zootaxa 1950: 21–38 (2008)
www.mapress.com/zootaxa/

ISSN 1175-5326 (print edition)
ZOOTAXA
ISSN 1175-5334 (online edition)

Zoological vs. botanical nomenclature: a forgotten 'BioCode' experiment from the times of the Strickland Code*

ALESSANDRO MINELLI

[1]*Department of Biology, University of Padova, Via Ugo Bassi 58 B, I 35131 Padova, Italy. E-mail: alessandro.minelli@unipd.it*

* *In:* Minelli, A., Bonato, L. & Fusco, G. (eds) *Updating the Linnaean Heritage: Names as Tools for Thinking about Animals and Plants. Zootaxa*, 1950, 1–163.

Table of contents

Abstract

A code intended to apply to both zoological and botanical nomenclature was drafted in 1842–43 by a commission appointed by the fourth Congress of Italian Scientists on request of the zoologist Carlo Luciano Bonaparte, as a reaction to the recently published Strickland Code (1842). Large excerpts from the latter document and an English translation of documents related to the Italian initiative are presented in appendices.

Key words: Bonaparte (Carlo Luciano), Congresses of Italian scientists, History of nomenclature, Italian zoology

The Strickland Code

Next to Linnaeus' own principia enumerated in his *Philosophia Botanica* (Linnaeus 1751), and more rightly than that work, the earliest document having the nature of a code ruling the scientific nomenclature for living organisms is the so-called Strickland Code (Strickland *et al*. 1842). Some acquaintance with this document is required to get an adequate historical setting for the little known episode in the history of biological nomenclature I present in this article. As Strickland *et al.*'s document is very frequently cited but, arguably, very seldom read, I present ample excerpts from this publication as Appendix 1 below.

Accepted by Z.-Q. Zhang: 24 Oct. 2008; published: 5 Dec. 2008

The divorce between zoological and botanical nomenclature

When Hugh Strickland and his colleagues were working at the document eventually published as a *Series of propositions for rendering the nomenclature of zoology uniform and permanent* (Strickland *et al.* 1842), zoological and botanical nomenclature had already gone along different paths to a sizeable extent.

Some aspects of this divergence were merely linguistic, such as different attitudes in respect to the admissible kinds of names to be employed in scientific literature. For example, following the example of Linnaeus, botanists used extensively generic names derived from persons, while this was quite rare in zoology and even occasionally rejected as legitimate. But other differences between the zoological and the botanical tradition were more technical and eventually became fixed in the internationally adopted Codes, starting with de Candolle (1867) for plant names and the 'Paris Rules' (International Commission on Zoological Nomenclature 1905) for animal names.

In a lucid comparison of a modern version of the zoological and botanical codes (International Commission on Zoological Nomenclature 1985; Voss *et al.* 1983), Jeffrey (1986) identified twelve main differences, including the nature of the link between names and taxa through the adoption of types and the provisions for conservation and rejection of names, that is, the rules limiting the scope of application of the principle of priority. In addition, botanists have name for hybrids as such, while this is excluded from the zoological code. There are also different criteria for what in zoology is the availability of names (the concept is called valid publication in botany). Tautonyms are rejected in botany, but not in zoology (well known examples are *Gorilla gorilla* and *Bufo bufo*); the principle of coordination (cf. Dubois 2008) has wider scope in zoology than in botany, and different in the two codes are the works and dates adopted as starting points for scientific nomenclature, zoology uniformly adopting the tenth edition of *Systema naturae* (Linnaeus 1758), while botanists' main reference is *Species Plantarum* (Linnaeus 1753), but with many exceptions, for selected groups, e.g. mosses, whose nomenclature starts officially with Hedwig (1801). There are also differences in the requirements for orthography, in the treatment of secondary homonyms, and in the use of names adopted for fossil taxa.

By the time the British Association for the Advancement of Science appointed the Strickland committee, another direct consequence of the independence *de facto* of zoological and botanical nomenclature was already evident, that is, the growing number of identical names used for animal and plant genera.

The BioCode

The latter circumstance, anyway, was long ignored in practice and generally proved to be of little consequence until plant and animal names begun to be stored together in large data bases.

But in the long run, the problem of cross-kingdom homonymy was eventually to emerge as one of the main reasons suggesting the importance of harmonization (Ride 1988), if not even a straight unification of zoological and botanical nomenclature. Another and arguably more critical reason behind this move was the problem of regulating nomenclature of the so-called ambiregnal organisms, that is of taxa that have been treated sometimes as protozoans, and thus named according to the rules of zoological nomenclature, sometimes as algae, and thus named in accordance to the botanical nomenclature. This twofold treatment has been often applied to closely related taxa, sometimes even to one and the same taxon. Cross-kingdom homonyms and the nomenclature of ambiregnal organisms have thus prompted efforts at what has become known as the BioCode (Greuter *et al.* 1996, 1998; Hawksworth 1997). This initiative and its eventual fate have been briefly summarised by Greuter (2003) and also by Kraus (2008) in this volume.

A ‘Draft BioCode’ dated 1842–43

A circumstance that has as much as ignored to date is that an effort largely similar to the recent BioCode initiative had been produced exactly at the time the Strickland Code was written. This effort, eventually fated to failure within short, was energetically launched by a prominent figure whose undisputed authority in zoological matters was likely increased by his social status (Stroud 2000). This man, Charles Lucien [=Carlo Luciano] Bonaparte, Prince of Canino, was indeed the son of a brother of Napoleon. Among the zoologists he was renowned for his studies on vertebrates, especially birds. By 1842 Bonaparte had already published the four-volume *American Ornithology* (Bonaparte 1825–33) and a lavishly illustrated monograph of Italian vertebrates (Bonaparte 1832–41). His most lasting contribution to zoology, however, is perhaps his later synopsis of world bird genera (Bonaparte 1850–7).

In the late 1830s, Bonaparte launched a successful series of congresses of the Italian scientists. At that time, Italy was divided into many political units, including parts under foreign rule. The congresses provided good opportunities for people of different Italian states to exchange their views, political as well as scientific: those events thus were instrumental in helping the eventual unification of Italy into a single national state. But Bonaparte’s own strong Italian nationalistic feelings combined with the indisputably international character of his research and his Europe-wide network of acquaintances. Thus, at the fourth meeting of the Italian scientists, held in Padova in 1842, he was able to illustrate the document freshly produced by the Strickland Committee and to offer it in Italian translation (Anon. 1843).

In Bonaparte’s view, the British document was a useful contribution towards a scientific nomenclature less deregulated than this was at the time, but he was not completely happy with the proposal. He was well aware of the differences between zoological and botanical traditions in nomenclature and regarded the latter as much closer to the Linnaean standards. Zoological nomenclature should thus converge towards the current praxis in botany and in order to get this result, he proposed to set up a Commission with the charge of producing a set of rules to be applied to zoological and botanical names alike. The chronicle of this interesting episode in the history of bionomenclature occupies many pages of the proceedings (*Atti*) of the 1842 congress: a long excerpt from that text is given here, in translation, as Appendix 2.

One year later, during the fifth meeting of the Italian scientists, held this time in Lucca, members of the Commission appointed in Padova presented long analytical reports and a draft of code, but against the proposal for a unified code, or its individual articles, were also raised strong objections, some of them by the dean of the Italian botanists Giuseppe Moretti, others, and harsher, by the entomologist Marquis Massimiliano Spinola. A short excerpt from the published chronicle (Anon. 1844) is given below, also in translation, as Appendix 3.

Following the Lucca meeting, the issue of a unified code with rules for the scientific names of organisms disappeared suddenly from the public debate, never to appear again seriously, to the best of my knowledge, before the start of the BioCode initiative.

Divorce, but with advantageous continuing exchange

To be sure, zoological and botanical traditions in nomenclature were often discussed comparatively, but highlighting the difference was not a stimulus to renovate the efforts towards unification. We can suggest that authors, if sensible to the issue, were regarding a single nomenclatural code as an opportunity already lost for ever. Significant, in this respect, are the words of Dall (1877, p. 9): “A serious mistake appears to have been committed at the outset by divorcing Zoological from Botanical nomenclature, as was done by the committee of the British Association. The signal success which has attended the efforts of botanists to unify their nomenclature, when compared to the confusion reigning in some departments of zoology, is sufficient proof of this.”

A few pages later, commenting on the general principles, at that time quite recently expressed by de Candolle (1867) in his *Lois de la nomenclature botanique*, Dall (1877, p. 23) added the following remark: “The manner in which Botany and the different branches of zoology have reached their present state, being far from uniform, and the nature of the organisms treated of being dissimilar, an absolute identity in the application of nomenclature is impracticable even if it were wholly desirable. The fundamental principles, however, and the end to be attained, are the same in both branches of study.”

The difficulties experienced during the 1990s by the BioCode initiative demonstrate that Dall’s judgement was not much off the mark, but his own efforts in comparing zoological and botanical traditions were not without virtue, and this is demonstrated again by the BioCode. Even if a unified code could not be eventually produced, it is certainly as a consequence of the dialogue stimulated by the BioCode initiative that the current edition of the zoological code (International Commission on Zoological Nomenclature 1999) was eventually to include a new Recommendation 1A. stating that “Authors intending to establish new genus-group names are urged to consult the *Index Nominum Genericorum (Plantarum)* and the *Approved List of Bacterial Names* to determine whether identical names have been established under the International Codes of Nomenclature relevant to those lists and, if so, to refrain from publishing identical zoological names.” Existing cross-kingdom homonyms will remain, but their number, at least, should not increase – exactly one of the targets Prince Bonaparte intended to reach by his well intended, though now largely forgotten effort.

Acknowledgements

I am grateful to Dr. Carlo Violani for his help in obtaining copy of the proceedings of the 1842 and 1843 Congresses of the Italian scientists.

References

Anon. (1843) *Atti della Quarta Riunione degli Scienziati Italiani tenuta in Padova nel settembre del MDCCCXLII*. Tipografia del Seminario, Padova.

Anon. (1844) *Atti della Quinta Riunione degli Scienziati Italiani tenuta in Lucca nel settembre del MDCCCXLIII.* Tipografia Giusti, Lucca, 848 pp.

Bonaparte, C.-L. (1825–33) *American Ornithology; or, The Natural History of Birds Inhabiting the United States, not Given by Wilson*. Carey, Lea & Carey, Philadelphia, vi, [ii], 105, [1]; vii, [iii], 95, [1]; [iv], 60; [iv], 142 pp.

Bonaparte, C.-L. (1832–41) *Iconografia della fauna italica per le quattro classi degli animali vertebrati*. Salviucci, Roma, [originally issued in XXX parts with unnumbered pages; actually, 1116 pp. with 182 plates].

Bonaparte, C.-L. (1850–57) *Conspectus generum avium*. E.J. Brill, Lugduni Batavorum, 543+232 pp.

Dall, W.H. (1877) *Nomenclature of Zoology and Botany. A Report to the American Association for the Advancement of Science at the Nashville Meeting, August 31, 1877*. Salem Press, Salem, 56 pp.

De Candolle, A. (1867) *Lois de la nomenclature botanique adoptées par le Congrès International de Botanique tenu à Paris en Août, 1867, suivies d'une deuxième édition de l'introduction historique et du commentaire qui accompagnaient la rédaction préparatoire presentée au Congrès.* Georg, Genève-Bâle & Ballière, Paris, 64 pp.

Dubois, A. (2008) Phylogenetic hypotheses and nomina of taxa. *In*: Minelli, A., Bonato, L. & Fusco, G. (Eds), *Updating the Linnaean Heritage: Names as Tools for Thinking About Animals and Plants*. *Zootaxa*, 1950, 51–86.

Greuter, W. (2003) Biological nomenclature in the electronic era: chances, challenges, risks. In: Legakis, A., Sfenthourakis, S., Polymeni, R. & Thessalou-Legaki, M. (eds.) *The New Panorama of Animal Evolution, Proceedings of the 18th International Congress of Zoology.* Pensoft, Sofia-Moscow, pp. 665–672.

Greuter, W., Hawksworth, D.L., McNeill, J., Mayo, M.A., Minelli, A., Sneath, P.H.A., Tindall, B.J., Trehane, P. & Tubbs, P. (1996) Draft BioCode: the prospective international rules for the scientific names of organisms. *Taxon*, 45, 349–372.

Greuter, W., Hawksworth, D.L., McNeill, J., Mayo, M.A., Minelli, A., Sneath, P.H.A., Tindall, B.J., Trehane, P. & Tubbs, P. (1998) Draft BioCode (1997): the prospective international rules for the scientific names of organisms. *Taxon*, 47, 127–150.

Hawksworth, D.L. (ed.) (1997) *The New Bionomenclature: The BioCode Debate. (Biology International, Special Issue 34).* International Union of Biological Sciences, Paris, 103 pp.

Hedwig, J. (1801) *Species muscorum frondosorum descriptae et tabulis aeneis lxxvii coloratis illustratae - opus posthumum, editum a Friderico Schwaegrichen.* J. Barth, Lipsiae, vi+353 pp.

International Commission on Zoological Nomenclature. 1905. *Règles internationales de la nomenclature zoologique. International Rules of Zoological Nomenclature. Internationale Regeln der zoologischen Nomenklatur.* Rudeval, Paris, 57 pp.

International Commission on Zoological Nomenclature, 1985. *International Code of Zoological Nomenclature. Third Edition adopted by the XXth General Assembly of the International Union of Biological Sciences*. International Trust for Zoological Nomenclature in Association with Brith Museum (Natural History), London, xx+338 pp.

International Commission on Zoological Nomenclature. 1999. *International Code of Zoological Nomenclature. Fourth Edition.* The International Trust for Zoological Nomenclature, London, xxix+306 pp.

Jeffrey, C. (1986) Some differences between the botanical and zoological codes. In: Ride, W.D.L. & Younès, T. (eds.) *Biological Nomenclature Today: A Review of the Present State and Current Issues of Biological Nomenclature of Animals, Plants, Bacteria and Viruses (IUBS Monograph Series, 2).* IRL Press, Oxford, pp. 62–65.

Kraus, O. (2008) The Linnaean foundations of zoological and botanical nomenclature. *In*: Minelli, A., Bonato, L. & Fusco, G. (Eds), *Updating the Linnaean Heritage: Names as Tools for Thinking about Animals and Plants*. *Zootaxa,* 1950, 9–20.

Linnaeus, C. (1751) *Philosophia botanica, in qua explicantur fundamenta botanica cum definitionibus partium, exemplis terminorum, observationibus rariorum, adjectis figuris aeneis.* G. Kiesewetter, Holmiae, 362 pp.

Linnaeus, C. (1753) *Species Plantarum*. Laurentius Salvius, Holmiae,[12]+1200+[31] pp.

Linnaeus C. (1758) *Systema Naturae per regna tria Naturae secundum classes, ordines, genera, species, cum characteribus, differentiis, synonymis, locis. Tomus I. Editio decima, reformata*. Laurentius Salvius, Holmiae, [4]+823+[1] pp.

Ride, W.D.L. (1988) Towards a unified system of biological nomenclature. In: Hawksworth, D.L. (ed.) *Prospects in Systematics (The Systematics Association Special Volume No. 36).* Clarendon Press, Oxford, pp. 332–353.

Strickland, H.E., Henslow, J.S., Phillips, J., Shuckard, W.E., Richardson, J.B., Waterhouse, G.R., Owen, R., Yarrell, W., Jenyns, L., Darwin, C., Broderip, W.J. & Westwood, J.O. (1842). *Report of a Committee appointed "to consider the rules by which the nomenclature of Zoology may be established on a uniform and permanent basis."* John Murray, London, for the British Association for the Advancement of Science. [Also published in 1843 as: Series of propositions for rendering the nomenclature of zoology uniform and permanent, being a report of a Committee for the consideration of the subject appointed by the British Association for the Advancement of Science. *Annals and Magazin of Natural History*, 11, 259–275].

Stroud, P.T. (2000) *The Emperor of Nature: Charles Lucien Bonaparte and his World*. University of Pennsylvania Press, Philadelphia, xvii+371 pp.

Voss, E.G., Burdet, H.M., Chaloner, W.G., Demoulin, V., Hiepko, P., McNeill, J., Meikie, R.D., Nicolson, D.H., Rollins, R.C., Silva, P.C. & Greuter, W. (eds.) (1983) *International Code of Botanical Nomenclature. Adopted by the Thirteenth International Botanical Congress, Sydney, August, 1981 (Regnum Vegetabile 111)*. Bohn, Scheltema & Holkema, Utrecht & Antwerp, xv+472 pp.

Appendix I

Excerpts from Strickland et al. (1842)

SERIES OF PROPOSITIONS FOR RENDERING THE NOMENCLATURE OF ZOOLOGY UNIFORM AND PERMANENT

PREFACE.

[..] The world of science is no longer a monarchy, obedient to the ordinances, however just, of an Aristotle or a Linnæus. She has now assumed the form of a republic, and although this revolution may have increased the vigour and zeal of her followers, yet it has destroyed much of her former order and regularity of government. The latter can only be restored by framing such laws as shall be based in reason and sanctioned by the approval of men of science; and it is to the preparation of these laws that the Zoological Section of the Association have been invited to give their aid. [...]

Among the numerous rules for nomenclature which have been proposed by naturalists, there are many which, though excellent in themselves, it is not now desirable to enforce[1]. The cases in which those rules have been overlooked or departed from, are so numerous and of such long standing, that to carry these regulations into effect would undermine the edifice of zoological nomenclature. But while we do not adopt these propositions as authoritative laws, they may still be consulted with advantage in making such additions to the language of zoology as are required by the progress of the science. By adhering to sound principles of philology, we may avoid errors in future, even when it is too late to remedy the past, and the language of science will thus eventually assume an aspect of more classic purity than it now presents.

Our subject hence divides itself into two parts; the first consisting of *Rules* for the rectification of the present zoological nomenclature, and the second of *Recommendations* for the improvement of zoological nomenclature in future.

PART I.

RULES FOR RECTIFYING THE PRESENT NOMENCLATURE.

[*Limitation of the Plan to Systematic Nomenclature*.]

In proposing a measure for the establishment of a permanent and universal zoological nomenclature, it must be premised that we refer solely to the Latin or systematic language of zoology. We have nothing to do with vernacular appellations. [..]

[*Law of Priority the only effectual and just one*.]

It being admitted on all hands that words are only the conventional signs of ideas, it is evident that language can only attain its end effectually by being permanently established and generally recognized. This consideration ought, it would seem, to have checked those who are continually attempting to subvert the established language of zoology by substituting terms of their own coinage. [..] Now in zoology no one person can subsequently claim an authority equal to that possessed by the person who is the first to define a new genus or describe a new species; and hence it is that the name originally given, even though it may be inferior in point of elegance or expressiveness to those subsequently proposed, ought as a general principle to be permanently retained. To this consideration we ought to add the injustice of erasing the name originally selected by the person to whose labours we owe our first knowledge of the object; and we should reflect how much the permission of such a practice opens a door to obscure pretenders for dragging themselves into notice at the expense of original observers. Neither can an author be permitted to alter a name which he himself has once published, except in accordance with fixed and equitable laws. It is well observed by Decandolle, "L'auteur même qui a le premier établi un nom n'a pas plus qu'un autre le droit de le changer pour simple cause d'impropriété. La priorité en effet est un terme fixe, positif, qui n'admet rien, ni d'arbitraire, ni de partial."

For these reasons, we have no hesitation in adopting as our fundamental maxim, the "law of priority," viz.

§ 1. The name originally given by the founder of a group or the describer of a species should be permanently retained, to the exclusion of all subsequent synonyms (with the exceptions about to be noticed). [..]

---

[1] See especially the admirable code proposed in the 'Philosophia Botanica' of Linnæus. If zoologists had paid more attention to the principles of that code, the present attempt at reform would perhaps have been unnecessary.

[*Not to extend to authors older than Linnæus*.]

As our subject matter is strictly confined to the *binomial system of nomenclature*, or that which indicates species by means of two Latin words, the one generic, the other specific, and as this invaluable method originated solely with Linnæus, it is clear that, as far as species are concerned, we ought not to attempt to carry back the principle of priority beyond the date of the 12th edition of the 'Systema Naturæ.' Previous to that period, naturalists were wont to indicate species not by a *name* comprised in one word, but by a *definition* which occupied a sentence, the extreme verbosity of which method was productive of great inconvenience. It is true that one word sometimes sufficed for the definition of a species, but these rare cases were only binomial by accident and not by principle, and ought not therefore in any instance to supersede the binomial designations imposed by Linnæus.

The same reasons apply also to generic names. Linnæus was the first to attach a definite value to genera, and to give them a systematic character by means of exact definitions; and therefore although the *names* used by previous authors may often be applied with propriety to modern genera, yet in such cases they acquire a new meaning, and should be quoted on the authority of the first person who used them in this secondary sense. It is true, that several of the old authors made occasional approaches to the Linnæan exactness of generic definition, but still these were but partial attempts; and it is certain that if in our rectification of the binomial nomenclature we once trace back our authorities into the obscurity which preceded the epoch of its foundation, we shall find no resting-place or fixed boundary for our researches. [..]

We therefore recommend the adoption of the following proposition:—

§ 2. The binomial nomenclature having originated with Linnæus, the law of priority, in respect of that nomenclature, is not to extend to the writings of antecedent authors. [..]

[*Generic names not to be cancelled in subsequent subdivisions*.] [..]

§ 3. A generic name when once established should never be cancelled in any subsequent subdivision of the group, but retained in a restricted sense for one of the constituent portions.

[*Generic names to be retained for the typical portion of the old genus*.]

When a genus is subdivided into other genera, the original name should be retained for that portion of it which exhibits in the greatest degree its essential characters as at first defined. Authors frequently indicate this by selecting some one species as a fixed point of reference, which they term the "type of the genus." When they omit doing so, it may still in many cases be correctly inferred that the *first* species mentioned on their list, if found accurately to agree with their definition, was regarded by them as the type. A specific name or its synonyms will also often serve to point out the particular species which by implication must be regarded as the original type of a genus. In such cases we are justified in restoring the name of the old genus to its typical signification, even when later authors have done otherwise. We submit therefore that

§ 4. The generic name should always be retained for that portion of the original genus which was considered typical by the author. [..]

[*When no type is indicated, then the original name is to be kept for that subsequent subdivision which* first *received it*.] [..]

§ 5. When the evidence as to the original type of a genus is not perfectly clear and indisputable, then the person who first subdivides the genus may affix the original name to any portion of it at his discretion, and no later author has a right to transfer that name to any other part of the original genus.

[*A later name of the same extent as an earlier to be wholly cancelled*.]

When an author infringes the law of priority by giving a new name to a genus which has been properly defined and named already, the only penalty which can be attached to this act of negligence or injustice, is to expel the name so introduced from the pale of the science. It is not right then in such cases to restrict the meaning of the later name so that it may stand side by side with the earlier one, as has sometimes been done. For instance, the genus *Monaulus*, Vieill. 1816, is a precise equivalent to *Lophophorus*, Tem. 1813, both authors having adopted the same species as their type, and therefore when the latter genus came in the course of time to be divided into two, it was incorrect to give the condemned name *Monaulus* to one of the portions. To state this succinctly,

§ 6. When two authors define and name the same genus, *both making it exactly of the same extent*, the later name should be cancelled *in toto*, and not retained in a modified sense. [..]

This rule admits of the following exception:—

§ 7. Provided however, that if these authors select their respective types from different sections of the genus, and these sections be afterwards raised into genera, then both these names may be retained in a restricted sense for the new genera respectively. [..]

[*A later name equivalent to several earlier ones is to be cancelled*.]

§ 8. If the later name be so defined as to be equal in extent to two or more previously published genera, it must be cancelled *in toto*.

[*A genus compounded of two or more previously proposed genera whose characters are now deemed insufficient, should retain the name of one of them*.] [..]

§ 9. In compounding a genus out of several smaller ones, the earliest of them, if otherwise unobjectionable, should be selected, and its former generic name be extended over the new genus so compounded. [..]

[*A name should be changed when previously applied to another group which still retains it.*]

It being essential to the binomial method to indicate objects in natural history by means of *two words* only, without the aid of any further designation, it follows that a generic name should only have one meaning, in other words, that two genera should never bear the same name. For a similar reason, no two species in the same genus should bear the same name. When these cases occur, the later of the two duplicate names should be cancelled, and a new term, or the earliest synonym, if there be any, substituted. […] It is, we conceive, the bounden duty of an author when naming a new genus, to ascertain by careful search that the name which he proposes to employ has not been previously adopted in other departments of natural history[2]. […] We submit therefore, that

§ 10. A name should be changed which has before been proposed for some other genus in zoology or botany, or for some other species in the same genus, when still retained for such genus or species.

[*A name whose meaning is glaringly false may be changed*.]

Our next proposition has no other claim for adoption than that of being a concession to human infirmity. If such proper names of places as Covent Garden, Lincoln's Inn Fields, Newcastle, Bridgewater, &c., no longer suggest the ideas of gardens, fields, castles, or bridges, but refer the mind with the quickness of thought to the particular localities which they respectively designate, there seems no reason why the proper names used in natural history should not equally perform

---

[2] This laborious and difficult research will in future be greatly facilitated by the very useful work of M. Agassiz, entitled "Nomenclator Zoologicus.

the office of correct indication even when their etymological meaning may be wholly inapplicable to the object which they typify. But we must remember that the language of science has but a limited currency, and hence the words which compose it do not circulate with the same freedom and rapidity as those which belong to every-day life. The attention is consequently liable in scientific studies to be diverted from the contemplation of the thing signified to the etymological meaning of the sign, and hence it is necessary to provide that the latter shall not be such as to propagate actual error. Instances of this kind are indeed very rare, and in some cases, such as that of *Monodon*, *Caprimulgus*, *Paradisea apoda* and *Monoculus*, they have acquired sufficient currency no longer to cause error, and are therefore retained without change. But when we find a Batracian reptile named in violation of its true affinities, *Mastodonsaurus*, a Mexican species termed (through erroneous information of its habitat) *Picus cafer*, or an olive-coloured one *Muscicapa atra*, or when a name is derived from an accidental monstrosity, as in *Picus semirostris* of Linnæus, and *Helix disjuncta* of Turton, we feel justified in cancelling these names, and adopting that synonym which stands next in point of date. At the same time we think it right to remark that this privilege is very liable to abuse, and ought therefore to be applied only to extreme cases and with great caution. With these limitations we may concede that

§ 11. A name may be changed when it implies a false proposition which is likely to propagate important errors.

[*Names not clearly defined may be changed*.]

[..] Two things are necessary before a zoological term can acquire any authority, viz. *definition* and *publication*. Definition properly implies a distinct exposition of essential characters, and in all cases we conceive this to be indispensable, although some authors maintain that a mere enumeration of the component species, or even of a single type, is sufficient to authenticate a genus. To constitute *publication*, nothing short of the insertion of the above particulars *in a printed book* can be held sufficient. [..] Therefore

§ 12. A name which has never been clearly defined in some published work should be changed for the earliest name by which the object shall have been so defined.

[*Specific names, when adopted as generic, must be changed*.]

The necessity for the following rule will be best illustrated by an example. The *Corvus pyrrhocorax*, Linn., was afterwards advanced to a genus under the name of *Pyrrhocorax*. Temminck adopts this generic name, and also retains the old specific one, so that he terms the species *Pyrrhocorax pyrrhocorax*. The inelegance of this method is so great as to demand a change of the specific name, and the species now stands as *Pyrrhocorax alpinus*, Vieill. We propose therefore that

§ 13. A new specific name must be given to a species when its old name has been adopted for a genus which includes that species.

[*Latin orthography to be adhered to*.]

§ 14. In writing zoological names the rules of Latin orthography must be adhered to. [..]

When a name has been erroneously written and its orthography has been afterwards amended, we conceive that the authority of the original author should still be retained for the name, and not that of the person who makes the correction.

PART II.

RECOMMENDATIONS FOR IMPROVING THE NOMENCLATURE IN FUTURE. [..]

[*The best names are Latin or Greek characteristic words*.] [..]

§ A. The *best* zoological names are those which are derived from the Latin or Greek, and express some distinguishing characteristic of the object to which they are applied.

[*Classes of objectionable names*.]

It follows from hence that the following classes of words are more or less objectionable in point of taste, though, in the case of *genera*, it is often necessary to use them, from the impossibility of finding characteristic words which have not before been employed for other genera. We will commence with those which appear the least open to objection, such as

a. Geographical names.—These words being for the most part adjectives can rarely be used for *genera*. As designations of *species* they have been so strongly objected to, that some authors (Wagler, for instance) have gone the length of substituting fresh names wherever they occur; others (*e. g.* Swainson) will only tolerate them where they apply *exclusively*, as *Lepus hibernicus*, *Troglodytes europæus*, &c. We are by no means disposed to go to this length. It is not the less true that the *Hirundo javanica* is a Javanese bird, even though it may occur in other countries also, and though other species of *Hirundo* may occur in Java. The utmost that can be urged against such words is, that they do not tell the *whole truth*. However, as so many authors object to this class of names, it is better to avoid giving them, except where there is reason to believe that the species is chiefly confined to the country whose name it bears.

b. Barbarous names.—Some authors protest strongly against the introduction of exotic words into our Latin nomenclature, others defend the practice with equal warmth. We may remark, first, that the practice is not contrary to classical usage, for the Greeks and Romans did occasionally, though with reluctance, introduce barbarous words in a modified form into their respective languages. Secondly, the preservation of the trivial names which animals bear in their native countries is often of great use to the traveller in aiding him to discover and identify species. We do not therefore consider, if such words have a Latin termination given to them, that the occasional and judicious use of them as scientific terms can be justly objected to.

c. Technical names.—All words expressive of trades and professions have been by some writers excluded from zoology, but without sufficient reason. Words of this class, *when carefully chosen*, often express the peculiar characters and habits of animals in a metaphorical manner, which is highly elegant. We may cite the generic names *Arvicola*, *Lanius*, *Pastor*, *Tyrannus*, *Regulus*, *Mimus*, *Ploceus*, &c., as favourable examples of this class of names.

d. Mythological or historical names.—When these have no perceptible reference or allusion to the characters of the object on which they are conferred, they may be properly regarded as unmeaning and in bad taste. Thus the generic names *Lesbia*, *Leilus*, *Remus*, *Corydon*, *Pasiphae*, have been applied to a Humming bird, a Butterfly, a Beetle, a Parrot, and a Crab respectively, without any perceptible association of ideas. But mythological names may sometimes be used as generic with the same propriety as technical ones, in cases where a direct allusion can be traced between the narrated actions of a personage and the observed habits or structure of an animal. Thus when the name *Progne* is given to a Swallow, *Clotho* to a Spider, *Hydra* to a Polyp, *Athene* to an Owl, *Nestor* to a grey-headed Parrot, &c., a pleasing and beneficial connexion is established between classical literature and physical science.

e. Comparative names.—The objections which have been raised to words of this class are not without foundation. The names, no less than the definitions of objects, should, where practicable, be drawn from positive and self-evident characters, and not from a comparison with other objects, which may be less known to the reader than the one before him. Specific names expressive of comparative size are also to be avoided, as they may be rendered inaccurate by the after-discovery of additional species. The names *Picoides*, *Emberizoides*, *Pseudoluscinia*, *rubeculoides*, *maximus*, *minor*, *minimus*, &c. are examples of this objectionable practice.

f. Generic names compounded from other genera.—These are in some degree open to the same imputation as comparative words; but as they often serve to express the position of a genus as intermediate to, or allied with, two other genera, they may occasionally be used with advantage. Care must be taken not to adopt such compound words as are of too great length, and not to corrupt them in trying to render them shorter. The names *Gallopavo*, *Tetraogallus*, *Gypaetos*, are examples of the appropriate use of compound words.

g. Specific names derived from persons.—So long as these complimentary designations are used with moderation, and are restricted to persons of eminence as scientific zoologists, they may be employed with propriety in cases where expressive or characteristic words are not to be found. But we fully concur with those who censure the practice of naming species after persons of no scientific reputation, as curiosity dealers (e. g. *Caniveti*, *Boissoneauti*), Peruvian priestesses (*Cora*, *Amazilia*), or Hottentots (*Klassi*).

h. Generic names derived from persons.—Words of this class have been very extensively used in botany, and therefore it would have been well to have excluded them wholly from zoology, for the sake of obtaining a *memoria technica* by which the name of a genus would at once tell us to which of the kingdoms of nature it belonged. Some few personal generic names have however crept into zoology, as *Cuvieria*, *Mulleria*, *Rossia*, *Lessonia*, &c., but they are very rare in comparison with those of botany, and it is perhaps desirable not to add to their number.

i. Names of harsh and inelegant pronunciation.—These words are grating to the ear, either from inelegance of form, as *Huhua*, *Yuhina*, *Craxirex*, *Eschscholtzi*, or from too great length, as *chirostrongylostinus*, *Opetiorhynchus*, *brachypodioides*, *Thecodontosaurus*, not to mention the *Enaliolimnosaurus crocodilocephaloides* of a German naturalist. It is needless to enlarge on the advantage of consulting euphony in the construction of our language. As a general rule it may be recommended to avoid introducing words of more than five syllables.

k. Ancient names of animals applied in a wrong sense.—It has been customary, in numerous cases, to apply the names of animals found in classic authors at random to exotic genera or species which were wholly unknown to the ancients. The names *Cebus*, *Callithrix*, *Spiza*, *Kitta*, *Struthus*, are examples. This practice ought by no means to be encouraged. The usual defence for it is, that it is impossible now to identify the species to which the name was anciently applied. But it is certain that if any traveller will take the trouble to collect the vernacular names used by the modern Greeks and Italians for the Vertebrata and Mollusca of southern Europe, the meaning of the ancient names may in most cases be determined with the greatest precision. It has been well remarked that a Cretan fisher-boy is a far better commentator on Aristotle's 'History of Animals' than a British or German scholar. The use however of ancient names, *when correctly applied*, is most desirable, for "in framing scientific terms, the appropriation of old words is preferable to the formation of new ones[3]."

l. Adjective generic names.—The names of genera are, in all cases, essentially substantive, and hence adjective terms cannot be employed for them without doing violence to grammar. The generic names *Hians*, *Criniger*, *Cursorius*, *Nitidula*, &c. are examples of this incorrect usage.

m. Hybrid names.—Compound words, whose component parts are taken from two different languages, are great deformities in nomenclature, and naturalists should be especially guarded not to introduce any more such terms into zoology, which furnishes too many examples of them already. We have them compounded of Greek and Latin, as *Dendrofalco*, *Gymnocorvus*, *Monoculus*, *Arborophila*, *flavigaster*; Greek and French, as *Jacamaralcyon*, *Jacamerops*; and Greek and English, as *Bullockoides*, *Gilbertsocrinites*.

n. Names closely resembling other names already used.—By Rule 10 it was laid down, that when a name is introduced which is *identical* with one previously used, the later one should be changed. Some authors have extended the same principle to cases where the later name, when correctly written, only approaches in form, without wholly coinciding with the earlier. We do not, however, think it advisable to make this law imperative, first, because of the vast extent of our nomenclature, which renders it highly difficult to find a name which shall not bear more or less resemblance in sound to some

---

[3] Whewell, Phil. Ind. Sc. v. i. p. lxvii.

other; and, secondly, because of the impossibility of fixing a limit to the degree of approximation beyond which such a law should cease to operate. We content ourselves, therefore, with putting forth this proposition merely as a recommendation to naturalists, in selecting generic names, to avoid such as too closely approximate words already adopted. So with respect to species, the judicious naturalist will aim at variety of designation, and will not, for example, call a species *virens* or *virescens* in a genus which already possesses a *viridis*.

o. Corrupted words.—In the construction of compound Latin words, there are certain grammatical rules which have been known and acted on for two thousand years, and which a naturalist is bound to acquaint himself with before he tries his skill in coining zoological terms. One of the chief of these rules is, that in compounding words all the radical or essential parts of the constituent members must be retained, and no change made except in the variable terminations. But several generic names have been lately introduced which run counter to this rule, and form most unsightly objects to all who are conversant with the spirit of the Latin language. A name made up of the first half of one word and the last half of another, is as deformed a monster in nomenclature as a Mermaid or a Centaur would be in zoology; yet we find examples in the names *Corcorax* (from *Corvus* and *Pyrrhocorax*), *Cypsnagra* (from *Cypselus* and *Tanagra*), *Merulaxis* (*Merula* and *Synallaxis*), *Loxigilla* (*Loxia* and *Fringilla*), &c. In other cases, where the *commencement* of both the simple words is retained in the compound, a fault is still committed by cutting off too much of the radical and vital portions, as is the case in *Bucorvus* (from *Buceros* and *Corvus*), *Ninox* (*Nisus* and *Noctua*), &c.

p. Nonsense names.—[..] The following are examples: *Viralva*, *Xema*, *Azeca*, *Assiminia*, *Quedius*, *Spisula*. To the same class we may refer *anagrams* of other generic names, as *Dacelo* and *Cedola* of *Alcedo*, *Zapornia* of *Porzana*, &c. Such verbal trifling as this is in very bad taste, and is especially calculated to bring the science into contempt. It finds no precedent in the Augustan age of Latin, but can be compared only to the puerile quibblings of the middle ages. It is contrary to the genius of all languages, which appear never to produce new words by spontaneous generation, but always to derive them from some other source, however distant or obscure. And it is peculiarly annoying to the etymologist, who after seeking in vain through the vast storehouses of human language for the parentage of such words, discovers at last that he has been pursuing an *ignis fatuus*.

q. Names previously cancelled by the operation of § 6.—Some authors consider that when a name has been reduced to a synonym by the operations of the laws of priority, they are then at liberty to apply it at pleasure to any new group which may be in want of a name. We consider, however, that when a word has once been proposed in a given sense, and has afterwards sunk into a synonym, it is far better to lay it aside for ever than to run the risk of making confusion by re-issuing it with a new meaning attached.

r. Specific names raised into generic.—It has sometimes been the practice in subdividing an old genus to give to the lesser genera so formed, the names of their respective typical species. Our Rule 13 authorizes the forming a new specific name in such cases; but we further wish to state our objections to the practice altogether. Considering as we do that the original specific names should as far as possible be held sacred, both on the grounds of justice to their authors and of practical convenience to naturalists, we would strongly dissuade from the *further continuance* of a practice which is gratuitous in itself, and which involves the necessity of altering long-established specific names. [..]

[*Families to end in* idæ, *and Subfamilies in* inæ.] [..]

§ B. It is recommended that the assemblages of genera termed *families* should be uniformly named by adding the termination *idæ* to the name of the earliest known, or most typically characterized genus in them; and that their subdivisions, termed *subfamilies*, should be similarly constructed, with the termination *inæ*.

These words are formed by changing the last syllable of the genitive case into *idæ* or *inæ* as *Strix*, *Strigis*, *Strigidæ*, *Buceros*, *Bucerotis*, *Bucerotidæ*, not *Strixidæ*, *Buceridæ*.

[*Specific names to be written with a small initial*.]

A convenient *memoria technica* may be effected by adopting our next proposition. It has been usual, when the titles of species are derived from proper names, to write them with a capital letter, and hence when the specific name is used alone it is liable to be occasionally mistaken for the title of a genus. But if the titles of *species* were *invariably* written with a *small* initial, and those of *genera* with a *capital*, the eye would at once distinguish the rank of the group referred to, and a possible source of error would be avoided. It should be further remembered that all species are *equal*, and should therefore be written all *alike*. We suggest, then, that

§ C. Specific names should *always* be written with a small initial letter, even when derived from persons or places, and generic names should be always written with a capital.

[*The authority for a species, exclusive of the genus, to be followed by a distinctive expression*.]

The systematic names of zoology being still far from that state of fixity which is the ultimate aim of the science, it is frequently necessary for correct indication to append to them the name of the person on whose authority they have been proposed. When the same person is authority both for the specific and generic name, the case is very simple; but when the specific name of one author is annexed to the generic name of another, some difficulty occurs.

For example, the *Muscicapa crinita* of Linnæus belongs to the modern genus *Tyrannus* of Vieillot; but Swainson was the first to apply the specific name of Linnæus to the generic one of Vieillot. The question now arises, Whose authority is to be quoted for the name *Tyrannus crinitus*? The expression *Tyrannus crinitus*, Lin., would imply what is untrue, for Linnæus did not use the term *Tyrannus*; and *Tyrannus crinitus*, Vieill., is equally incorrect, for Vieillot did not adopt the name *crinitus*. If we call it *Tyrannus crinitus*, Sw., it would imply that Swainson was the first to describe the species, and Linnæus would be robbed of his due credit. If we term it *Tyrannus*, Vieill., *crinitus*, Lin., we use a form which, though expressing the facts correctly, and therefore not without advantage in particular cases where great exactness is required, is yet too lengthy and inconvenient to be used with ease and rapidity. Of the three persons concerned with the construction of a binomial title in the case before us, we conceive that the author who *first* describes and names a species which forms the groundwork of later generalizations, possesses a higher claim to have his name recorded than he who afterwards defines a genus which is found to embrace that species, or who may be the mere accidental means of bringing the generic and specific names into contact. By giving the authority for the *specific* name in preference to all others, the inquirer is referred *directly* to the original description, habitat, &c. of the species, and is at the same time reminded of the date of its discovery; while genera, being less numerous than species, may be carried in the memory, or referred to in systematic works without the necessity of perpetually quoting their authorities. The most simple mode then for ordinary use seems to be to append to the original authority for the species, when not applying to the genus also, some distinctive mark, such as (*sp.*) implying an exclusive reference to the *specific* name, as *Tyrannus crinitus*, Lin. (*sp.*), and to omit this expression when the same authority attaches to both genus and species, as *Ostrea edulis*, Lin.[4] Therefore,

§ D. It is recommended that the authority for a specific name, *when not applying to the generic name also*, should be followed by the distinctive expression (*sp*.).

[*New genera and species to be defined* amply *and* publicly.]

A large proportion of the complicated mass of synonyms which has now become the opprobrium of zoology, has originated either from the slovenly and imperfect manner in which species and groups have been originally defined, or from their definitions having been inserted in obscure local publications which have never obtained an extensive circulation. Therefore, although under § 12, we have conceded that mere insertion in a printed book is sufficient for *publication*, yet we would strongly advise the authors of new groups always to give in the first instance a full and accurate definition of their characters, and to insert the same in such periodical or other works as are likely to obtain an immediate and extensive circulation. To state this briefly,

---

[4] The expression *Tyrannus crinitus* (Lin.) would perhaps be preferable from its greater brevity.

§ E. It is recommended that new genera or species be *amply* defined, and *extensively* circulated in the first instance.

[*The names to be given to subdivisions of genera to agree in gender with the original genus.*]

In order to preserve specific names as far as possible in an unaltered form, whatever may be the changes which the genera to which they are referred may undergo, it is desirable, when it can be done with propriety, to make the new subdivisions of genera agree *in gender* with the old groups from which they are formed. This recommendation does not however authorize the changing the gender or termination of a genus already established. In brief,

§ F. It is recommended that in subdividing an old genus in future, the names given to the subdivisions should agree in gender with that of the original group.

[*Etymologies and types of new genera to be stated.*]

It is obvious that the names of genera would in general be far more carefully constructed, and their definitions would be rendered more exact, if authors would adopt the following suggestion:—

§ G. It is recommended that in defining new genera the etymology of the name should be always stated, and that one species should be invariably selected as a type or standard of reference.

In concluding this outline of a scheme for the rectification of zoological nomenclature, we have only to remark, that almost the whole of the propositions contained in it may be applied with equal correctness to the sister science of botany. We have preferred, however, in this essay to limit our views to zoology, both for the sake of rendering the question less complex, and because we conceive that the botanical nomenclature of the present day stands in much less need of distinct enactment than the zoological. The admirable rules laid down by Linnæus, Smith, Decandolle, and other botanists (to which, no less than to the works of Fabricius, Illiger, Vigors, Swainson, and other zoologists, we have been much indebted in preparing the present document), have always exercised a beneficial influence over their disciples. Hence the language of botany has attained a more perfect and stable condition than that of zoology; and if this attempt at reformation may have the effect of advancing zoological nomenclature beyond its present backward and abnormal state, the wishes of its promoters will be fully attained.

(Signed)
June 27, 1842.

H. E. STRICKLAND.
JOHN PHILLIPS.
JOHN RICHARDSON.
RICHARD OWEN.
LEONARD JENYNS.
W. J. BRODERIP.

J. S. HENSLOW.
W. E. SHUCKARD.
G. R. WATERHOUSE.
W. YARRELL.
C. DARWIN.
J. O. WESTWOOD.

Appendix 2

Excerpts from Anon. (1843), pages 305 through 315, translated by Alessandro Minelli

Proceedings of the Fourth Meeting of Italian Scientists held in Padova, September 1842

Session of 27 September [1842]

The Botanical Section and the Zoological Section are jointly meeting under the Presidency of Prof. [Giuseppe L.] Moretti, to discuss the plan intended to provide these sciences with uniform and lasting nomenclature.

His Excellence the Governor of the Venetian Provinces honours the assembly of his presence.

The minutes of the previous session are approved.

The President invites Prince [Carlo Luciano] Bonaparte to illustrate the plan, as announced.

Prince Bonaparte narrates that the distinct British ornithologist Mr. [Hugh] Strickland first drafted this plan and circulated it among friends including Prince Bonaparte himself, to get their comments and suggestions; that he reformulated it in the light of these friendly comments and subsequently submitted it to the British Association for an evaluation. The latter body delegated the examination of that plan to a Committee that modified it slightly, eventually presenting the document to that Society's meeting held earlier this year in Manchester. The text now newly published as an expression of that Assembly will now be presented by Prince Bonaparte to the Botanical and Zoological Sections meeting together today, on the basis of still unpublished page proofs he has been sent precisely to this aim. However, before starting reading the plan, on behalf of the [Zoological] Section of which he is the President, Mr. Prince kindly asks botanists for help, as their discipline has been more respectful of the rules of nomenclature and has kept strict to the principles established by Linnaeus, just an exceptional minority being those authors who from time to time abandoned that track, while zoologists have quite often abused of nomenclature in the most different ways. Thus the Prince is pleased with the idea first conceived by himself, to look at botany for inspiration – an idea completely overlooked by the British colleagues responsible for this new plan for reforming zoological nomenclature. By unifying in these matters the procedures of botanists and zoologists, he much hopes eventually to provide a good service to the latter. Further, he announces that he will limit himself to present the most fundamental rules of his new plan, and also asks the President to established a Commission with the duty of examining and discussing it. To aptly introduce matters he illustrates the need, widely felt in zoology especially, to reform nomenclature in order to limit the increasing flood of abuse by fixing invariant and indisputable rules. He points out that the hope that these rules will be eventually observed by everybody can only rest on their promulgation by a body of scientists such as a scientific Congress, currently the only authority to be universally obeyed. Then, after fixing the principle that such a reform should only deal with Latin systematic nomenclature, he fixes the principle of priority as its only right and effective guide. In this context he remarks on the purely conventional nature of scientific names, to the exclusion of any current or possible meaning of names as such; and that the only authority to be respected in accepting and conserving a name should be that of the scientist who first used it as the name of an object previously devoid of systematic denomination. On this background he presents the rules as follows.

I. The name originally given by the student who first established a group or a species is to be permanently retained, to the exclusion of any later synonym.

Marquis [Massimiliano] Spinola [entomologist] objects that such a rule would fix the errors of those who created false species or groups, just because they were the first to introduce those names. The Prince replies by providing a more complete explanation of the rule. Marquis Spinola believes that the case when a given group proposed by someone must be suppressed to divide it into a number of smaller groups must be kept outside the domain of application of that rule, but the Prince replies that in that case one of the newly established smaller groups must retain the original denomination previously given to the larger group. Prof. [Filippo] Parlatore [professor of botany in Florence] raises the objection that even [Augustin Pyramus] de Candolle [Swiss botanist] who first established the rule of priority as the most important fundament of nomenclature, nevertheless acknowledged that in some cases exceptions have to be accepted. The Prince agrees that such exceptions must be accepted and anticipates that a paragraph to be discussed later (para. IX) deals precisely with those exceptions. Marquis Spinola remarks that to determine the principle of precedence, simple name priority is not enough, as such an authority has only to be acknowledged to the first student who actually described a given object in scientific terms. On this point the Prince also remarks that a later paragraph (para. XII) is specifically devoted to it. [..]

Continuing reading, the Prince established a limit to this otherwise undefined anteriority, and in agreement to the previously accepted principle to only deal with binomial names used for systematic purposes, as first done by Linnaeus, introduces a second principle, as follows.

II. As binomial nomenclature originated with Linnaeus, the rule of priority should not be extended to pre-Linnaean authors.

Marquis Spinola, while agreeing with the principle of setting a temporal beginning to the botanical and zoological nomenclature, nevertheless remarks that Linnaeus himself did often mix more than one species under one name, especially in the case of animals of the latter classes; as a consequence, it would be advisable to go back in time only to authors more recent than Linnaeus. President Moretti, only speaking for plants, remarks that previous to Linnaeus, [the French botanist Joseph Pitton de] Tournefort had already defined many genera with wonderful exactitude and that Linnaeus himself in the second edition of *Systema Naturae* changed many genus names he had used in the first edition, replacing them with those of Tournefort that have been subsequently retained in the later editions. To the first objection the Prince replies by offering a zoological example, as illustrated in his own written document. This example is about the generic names exquisitely introduced by [the French zoologist Mathurin Jacques] Brisson at the same time as Linnaeus. These names should be retained, but substituting the specific names provided by Linnaeus for the specific phrases, or the first word of these, as provided instead by Brisson. Same way should be treated the analogous examples where with the first word of the specific phrase one can form an accidental binomen: in all these instances, he argues, the Linnaean specific name must be retained even if another author's generic name is adopted. President prof. Moretti gives examples of genera aptly defined by [the Italian botanist Pier Antonio] Micheli and Tournefort prior to Linnaeus; and also reminds that Linnaeus himself in his earlier works employed descriptive phrases to designate individual species, and only by the second edition of *Flora suecica* he begun using binomials. As a consequence, he suggest that 1753, rather than [a specified edition of] *Systema Naturae*, should be accepted as the official starting point of scientific nomenclature. Prince Bonaparte remarks that the same might be said of *Fauna suecica*. On the other hand, the founding father of systematic botany and zoology could well enjoy the privilege of changing the rules. Finally, the Prince expressed the view that it is advisable to accept the universally followed principle to fix *Systema Naturae* as nomenclature's starting point. As for the genera established before him, Linnaeus partly accepted them, partly rejected as invalid, partly finally accepted but changed their name. In this last case it could be possible to favour Linnaeus by accepting, as an exception, his names rather than the older ones.

Subsequently, by distinguishing the different groups of organisms, that is, genera, families, tribes etc. as successively higher levels of generalization, he forms for these groups the third rule.

III. The rule of priority, despite its usefulness as a guide in respect to the names of higher groups, should not be rigorously applied except in the case of genus and species names. [..]

IV. Once established, a generic name should not be cancelled whenever the genus is subsequently split into narrower genera, but retained instead as the name of one of these parts.

V. The original generic name should be reserved to that part of the original genus that the author regarded as typical. [..]

VI. When the original type of a genus is not perfectly clear and unquestionable, the researcher who first subdivides it is free to restrict the original name to one or the other of its sections. His action should not be reversed by any subsequent work.

VII. When two independent authors define and provide with different names the same genus, by giving it the exactly same extension, the younger name must be wholly cancelled.

VIII. If a younger name is introduced to exactly cover the joint extension of two or more previously published genera, that name must be definitely cancelled.

IX. When many smaller genera are combined into one, the oldest of their names must be retained as the name of the whole larger genus thus obtained. [..]
President Prof. Moretti asks about any possibly different treatment for names already in existence vs. those still to be introduced. The Prince explains that the plan he has the honour to present today includes for the two sets of names separate provisions in two distinct articles. [..]

X. A name must be changed if it has already been used as the name of a different genus of either animals and plants, or of another species in the same genus [..]

XIII. A new specific name is to be provided for a species if its older name has been later adopter for a genus including that species.

XIV. The rules of the Latin grammar should be followed in writing zoological and botanical names. [..]

The Prince invites the President to ask the jointly meeting Sections to vote on the general principles of this plan, irrespective of the fact that the same will be subjected to the critical evaluation of the Commission the President will eventually appoint. President Prof. Moretti however disagrees, as the objections raised witness the lack of universal agreement on the plan's general principles, and also because he does not believe that such a vote could be universally binding for all naturalists. On behalf of the zoologists, Prince Bonaparte repeats that he will mainly rely on the evaluation to be expressed by botanists, as more advanced than zoologists in matters of nomenclature. He also expresses his satisfaction as, whilst he expected to find among them the strongest opposition to the basic principle of priority, he found instead that botanists go even further in that direction than his plan itself, discussion being limited to the few exceptions to the rules accepted in his document. Subsequently, President Prof. Moretti fixes the membership of the Commission, following consultation with the Prince about the most suitable names among the zoologists. He thus indicates the following: Marquis Spinola, Cav. [Carlo] Bassi, C[arlo] Porro, [Filippo] de Filippi and Prince Bonaparte [..]. Of botanists, Prof. [Gaetano] Savi, Prof. Parlatore, Prof. [Roberto de] Visiani, Mr. [Vittore Benedetto Antonio] Trevisan, Prof. Moretti as President and Prof. [Giuseppe] Meneghini as the Commission's Secretary. [..]

Signed – President Prof. G. Moretti
Secretaries Prof. G. Meneghini and Prof. F. Parlatore

Appendix 3

Excerpts from Anon. (1844), pages 761 through 763; 792, translated by Alessandro Minelli

Proceedings of the Fifth Meeting of Italian Scientists held in Lucca, September 1843

Session of day 20 September [1843]

[..] The Secretary reads Marquis Spinola's report, who is almost contrary to this project so warmly cherished by many naturalists. President Prince of Canino will later rebut one by one the arguments of the illustrious Entomologist, while limiting himself at the moment to remark that it is easier to destroy a house than to build it; he also expresses his hope that Marquis Spinola will eventually help with the project rather than doubt its success. The next reading is by Mr. Porro, who presents a document written by himself also on behalf of Cav. Bassi and Dr. De Filippi. [..]

This is followed by a discussion, from which an agreement emerges on accepting the XII edition of *Systema Naturae* as the most sensible starting point when establishing priority. [..] The President [..] disagrees on the principle that a name used for an animal genus cannot be used also for a plant genus. [..] De Visiani, Meneghini and Trevisan comment on the point, previously also made by Prince Bonaparte, that botanists, who have not shared the many errors of zoologists in matters of nomenclature, would be ill-advised if they would part from the rules established by Linnaeus. Nevertheless, their document also remarks that "De Candolle has added some useful change to those rules, and others are required by the current state of science and by the recurrence of abuse. Nevertheless, what is most deeply felt by Botanists is the need to rigorously follow the set of rules established by Linnaeus". Furthermore, this Commission regards as insufficient the plan proposed by the British zoologists, while, with suitable modifications, the Linnaean code could be adapted to zoology too. Therefore, they propose:

"1. To examine and whenever required to update the Linnaean rules as pertaining to Botany.

2. To examine with mature judgement the possible application of these same rules to Zoology too. To this aim the work done by British zoologists will provide help, but cannot be taken as the real starting point to reform nomenclature." [..]

Signed – The President Carlo Principe Bonaparte
The Secretaries of the Botany Sections—Dott. L. Masi and L. E. Celi
The Secretary of the Zoology Section—Dott. T. Riboli

Zootaxa 1950: 39–50 (2008)
www.mapress.com/zootaxa/

ISSN 1175-5326 (print edition)
ZOOTAXA
ISSN 1175-5334 (online edition)

ZooBank: Developing a nomenclatural tool for unifying 250 years of biological information*

RICHARD L. PYLE[1] & ELLINOR MICHEL[2]

[1]*Department of Natural Sciences, The Bishop Museum, 1525 Bernice St., Honolulu, Hawaii 96817, USA.*
E-mail: deepreef@bishopmuseum.org
[2]*International Commisssion on Zoological Nomenclature, c/o The Natural History Museum, Cromwell Road, SW7 5BD London, U.K.*

* *In:* Minelli, A., Bonato, L. & Fusco, G. (EDS) *Updating the Linnaean Heritage: Names as Tools for Thinking about Animals and Plants. Zootaxa*, 1950, 1–163.

Table of contents

Accepted by A. Minelli: 24 Oct. 2008; published: 5 Dec. 2008

Abstract

Nomenclature represents the backbone upon which virtually all biological information is organized. However, the practice of zoological nomenclature has changed relatively little since its start in 1758. As modern technology changes the paradigm under which modern scientists exchange information, there is increasing need to capitalize on these same technologies to fortify nomenclature. ZooBank has been proposed as the official registry of names and nomenclatural acts, in zoology, as well as associated published works and their authors, and type specimens. Having a coordinated registry of zoological names, integrated with the existing Code of Zoological Nomenclature, will allow increased efficiency of communication among biologists, and enhanced stability of names. Such a registry would encompass two distinct realms, each with their own set of challenges. Retrospective registration involves the monumental task of aggregating and validating two and a half centuries of existing names, whereas prospective registration must be tightly integrated with the future paradigm in which scientific names are created and managed under new models of publication. The prototype of ZooBank has been hosted at Bishop Museum during its initial development phase. Following the lead of standard-setting bodies in biodiversity informatics, Life Science Identifiers (LSIDs) have been selected for use as the globally unique identifiers for ZooBank registration entries. The first ZooBank LSIDs were issued on January 1st, 2008, and included five new fish species described in a work published that same day, as well as all 4,819 names established in the 10th Edition of Linnaeus' *Systema Naturae*. Three alternate scenarios for implementing mandatory registration in ZooBank have been articulated, each incorporating different degrees of coordination between published works and registration events. A robust discussion involving a broad spectrum of practicing zoological taxonomists is required over the next several years to define the specific implementation aspects of ZooBank.

Key words: Linnaeus, Prospective registration, Registration of names and acts, Retrospective registration, *Systema Naturae*

Introduction

Sense and stability in nomenclature

Stable nomenclature is at the heart of clear and unambiguous communication about biodiversity. Species names provide the most consistent anchor to which all taxonomic, ecological, molecular, conservation, and other biologically relevant data are attached. Legal protection and policy are also linked with names, on the assumption that the groups indicated by the names are consistent through time and among places. Scientific discussion relies on names having unequivocal, context-independent meanings. Medical and veterinary implementation requires communication about unambiguous identifications. Although discovery and delineation of species may receive the emphasis of high-profile press coverage, all taxonomic practice is crucially dependent on a stable nomenclature to provide a steady platform on which to build. The International Commission on Zoological Nomenclature (ICZN) has, for the past 113 years, set the rules by which scientific names for animals are established, as currently set forth in the ICZN Code of Nomenclature ('the Code', 4th Edition, 1999). The ICZN is the single professional organization devoted to ensuring that this work happens in a globally consistent way, providing continuity both for new species discoveries and for the correction of errors and inconsistencies in past works.

Development of a registry for new animal names (prospective registration) and a complete listing of existing names (retrospective registration) has long been a goal for biologists. The stakeholders for a gold-standard registry of animal names are diverse and central to the functioning of many biological sciences and to policy concerned with the living world. They include not only taxonomists, ecologists, and biodiversity informatics specialists but also conservationists, medical and veterinary workers, planners, policy makers, lawyers and even customs enforcers. Their requirements include ready access to a system of unambiguous answers to questions on the availability and validity of animal names that can be retrieved both by ordinary people and machines. The ICZN is meeting this need by developing ZooBank, a web-based registry of animal names

(Polaszek *et al*. 2005a). This will include nomenclatural acts (including new names), publications, authors and information on primary type specimens, and serve as a resolver for LSIDs (Life Science Identifiers), a tool for global communication among bioinformatics projects. It will be both a hub and authority for nomenclatural information.

Taxonomy, nomenclature and typification

Taxonomy and nomenclature are closely allied, but separate and complementary endeavors in developing the language of biodiversity. Discovering and delimiting species is the challenging job of alpha taxonomy; determining relationships and establishing higher taxa is referred to as beta taxonomy. Delimiting both alpha and beta taxa requires using a range of character data to test hypotheses about the inclusiveness of taxon definitions. This can naturally lead to strongly opposing alternative points of view, depending on character selection, method of analysis, and philosophical stance of the taxonomist. Definitions of taxa, from species to genera to higher taxa, can thus change significantly as the iterative process of improving the tests of taxonomic boundaries weighs alternative hypotheses and moves to new conclusions. Although it may be a source of frustration to end-users who simply want defined taxonomic entities, this process of change is a sign of the health of the science of taxonomy. Ultimately, if data accumulation were to saturate and if philosophical perspectives on species definitions were to converge, it is possible that taxonomy would stabilize and reach consensus definitions for taxa (changing only to accommodate ongoing organismal evolution). This situation is not on the horizon.

By contrast, the establishment of scientific names of animals is not a scientific process of testing alternatives; rather, it involves a bibliographic and quasi-legal process of presentation of a name with appropriate supporting documentation in a publication. Although a scientific name is generally established within the context of a published work on taxonomy, its link to actual organisms is through the primary type specimen (or specimens). This process of typification allows the name to be tied to a physical standard (and hence provides an objective basis for identifications), but leaves room for taxonomy to change; different names can be applied to taxa as is appropriate for their new boundaries. Figure 1 presents a tree-based example, in which alternative interpretations by different taxonomists result in different generic groupings, each of which could take a different name depending on the type species of the generic group. The same process could be visualized simply based on variation, with a more inclusive (‘lumping’) perspective requiring one type specimen for a species, thus receiving one name; whereas a more divisive (‘splitting’) perspective requires names derived from several type specimens for the perceived groups. Choosing between available names for types in a group is generally governed by the Principle of Priority, such that name first established should be used for that group (Figure 1). However, even if names are not in current use for a group, if they were originally validly published they are not permanently retired, as they may well be needed in the future. Taxonomic work may split an existing group, because less inclusive taxa are more consistent with data in hand. Having older names ready to apply provides an immediate tool for recovering past information on that taxon. This means that ZooBank must include both names in current use and all past, validly described names.

We want to underscore that the work of nomenclature aims for stability in names, but is completely independent of the process of flexibility in taxonomic interpretation. This philosophy is fundamental to the ICZN’s role, as articulated in the Introduction to the 4th Edition of the ICZN Code which states:

> *There are certain underlying principles upon which the Code is based. These are as follows:*
>
> *(1) The Code refrains from infringing upon taxonomic judgment, which must not be made subject to regulation or restraint.*
>
> *(2) Nomenclature does not determine the inclusiveness or exclusiveness of any taxon, nor the rank to be*

accorded to any assemblage of animals, but rather provides the name that is to be used for a taxon whatever taxonomic limits and rank are given to it.

(3) The device of name-bearing types allows names to be applied to taxa without infringing upon taxonomic judgment. [etc] (ICZN p. xix).

A cartoon graphic for the relationship of the trinity of nomenclature, taxonomy and type specimens is shown in Figure 2.

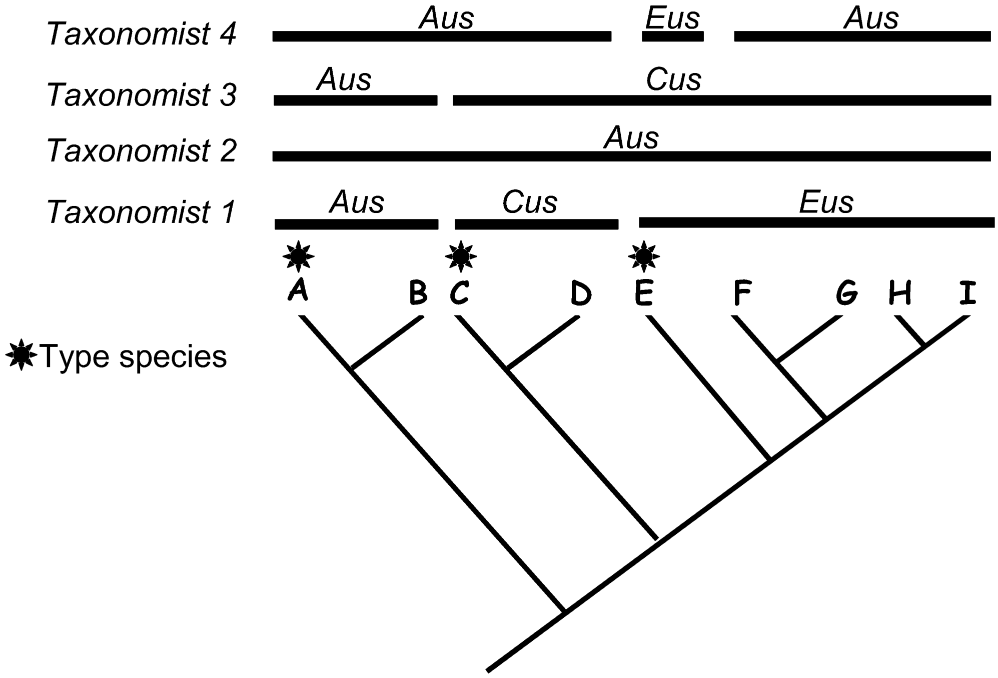

FIGURE 1. Diagrammatic representation of different nomenclatural interpretations for species A through I, mapped to a hypothesized phylogeny. Taxonomist 1 recognizes three new genus names; the genus *Aus* is typified by species A; *Cus* by species C, and *Eus* is by species E. Taxonomist 2 treats them as congeners (the diagram assumes that *Aus* has nomenclatural priority over *Cus*, and *Cus* over *Eus*). Taxonomist 3 recognizes two genera. Taxonomist 4 believes the underlying phylogeny is incorrect, and interprets the breakdown of two genera differently. Typification is necessary to establish which cluster each genus name is associated with, when taxonomic definitions of genera change.

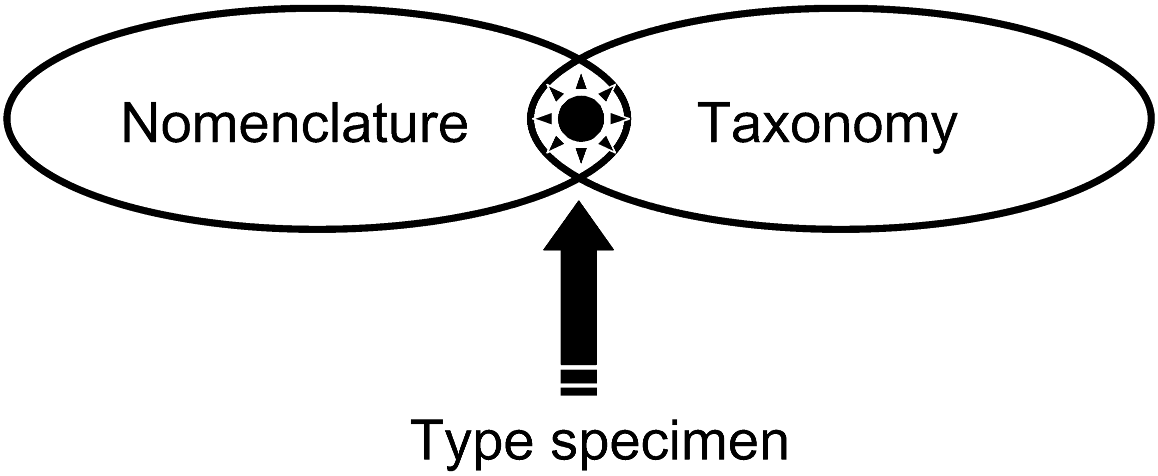

FIGURE 2. Diagrammatic representation of how type specimens represent the intersection of nomenclature and taxonomy.

Why do names need regulation?

The ICZN was established to address a situation in the mid 19th century described as 'increasing chaos in zoological nomenclature' (Melville 1995). Taxonomists were tackling the description and naming of an ever-expanding number of species they encountered from explorations of distant places, in addition to increased recognition of diversity from their home turf. The number of errors, inconsistencies and redundancies in application of names, either in synonymy or homonymy, began to create serious problems in the core objective of the Linnean enterprise of an unambiguous name for each species. Taxonomists working without access to the full literature inadvertently described taxa with the same name as another species in another taxonomic group, creating homonyms. Conversely, different workers in different parts of the world would unknowingly assign different names to what would otherwise be regarded as the same species, thereby creating synonyms.

At its inception, the ICZN acted on behalf of the zoological community at large, and with cooperation among taxonomists despite linguistic, political and taxonomic differences, even through world wars, to achieve a set of stabilizing rules for naming animals. The problems to be addressed then were, at their root, the result of lack of access to published information. The current revolution in information availability means that the information is now increasingly accessible, but in its shear volume the problems of disorganization from inappropriate names are becoming more starkly apparent. The consequences are serious, in that information is inappropriately presented and errors are propagated. Rigorous nomenclature must become dynamically integrated into the tools of cybertaxonomy.

The challenges of prospective and retrospective registration

In an ideal world, ZooBank would include full, verified information on all available names for animals. However, with 16,000–24,000 new additions yearly (N. Robinson, Zoological Record pers. comm., P. Bouchet, pers. comm.) to an estimated 1.7–1.8 million described animal species (Bouchet 2006), the numbers of names to be checked for homonymy and objective synonymy is enormous, so the logistics of populating the registry require strategic approaches. The need for reliable names for biodiversity work is urgent, so the tasks must be partitioned.

The first task for ZooBank is the establishment of a 'Black List' of unavailable names. This would serve as a foundation for on-line quality control that could be applied across projects, for example, by publishers to flag inappropriate usages of unavailable names. This is achievable with the recent digitization of the *Official Lists and Indexes of Names and Works in Zoology* by J.D.D. Smith, which could be implemented (through its 2007 version) within ZooBank.

The creation of a 'White List' of all available names is a much more difficult task, comprised of two parts. **Prospective registration** of new names, as they are published, will require active listing by taxonomists as they describe their thousands of new species each year, in a manner analogous to registering GenBank sequences today. This will rely on large-scale participant buy-in. Publishers are likely to support this, as the LSIDs resulting from ZooBank listing (described below) will allow greater exposure for their publications. The nomenclatural reliability of biodiversity informatics initiatives such as the Global Biodiversity Information Facility (GBIF), Encyclopedia of Life (EoL), and Catalogue of Life (CoL), which aim to present current taxonomic knowledge, will improve through dynamic integration with ZooBank. ZooBank will also increase the scope for linking with type specimen information held in museum databases to wider bioinformatics initiatives.

Working taxonomists appear eager to register names, and the incentives will increase with broader community use. Although the taxonomic community is strongly behind ZooBank, a straw poll at a large conference on the topic (EDIT 2008) indicated that registration mandated by the Code (see below) should wait until the project has sufficient content, exposure and momentum. Ultimately, however, it is envisaged that as electronic (paperless) publication becomes more common, a mandatory registration system will become necessary. This point will be a central issue in discussions for publication of the fifth edition of the Code of Zoological Nomenclature (e.g. http://www.iczn.org/electronic_publication.html).

Retrospective registration of existing names is the greatest challenge for populating ZooBank. Initial assembly of published names can come from historical sources such as Sherborn (1902–1933) and Neave (1939–1996), and various taxon-specific nomenclatural databases. Names will then need to be flagged as to their level of nomenclatural vetting. One suggestion is a coding system. For example, a Gold (or green) flag would indicate the name has been checked to its original published source, a Silver (or yellow) flag indicates it has been checked to a reliable secondary source (e.g. a respected checklist) and a Bronze (or red) flag indicates it has been dumped from an unvetted source. Groups of names of particular interest could then be targeted to be worked-up *in toto*. For example, groups of concern to CITES could be prioritized with targeted funds. It might also be possible, with the completion of the Biodiversity Heritage Library, that 'citizen scientist' initiatives could be enlisted to populate ZooBank. Much of this depends on exactly how the scope of ZooBank will be established (see below), and what protocols for data quality assurance are put into place.

Although the ultimate goal is for ZooBank to act as a definitive source of names, it is the nature of historical work and taxonomic research that new information can cast doubt on past certainties. As a result, ZooBank will need to remain updatable, with allowance for external, expert input. We expect that the development of search tools will increase the possibility that conflicts within the ZooBank database, and also with external sources, could be flagged up for correction. We anticipate that ZooBank will develop into an indispensable tool, allowing unprecedented access to reliable nomenclatural information for biodiversity workers worldwide.

Technical implementation

ZooBank was launched as a functional prototype web site (http://zoobank.org) on January 1st, 2008, coinciding with the 250th anniversary of the official start of Zoological Nomenclature. At its launch the ZooBank registry included 4,819 names established in the 10th Edition of Linnaeus' Systema Naturae (Linnaeus, 1758), as well as five new fish species names established in an article published concurrently with the launch of ZooBank. Since that time, additional nomenclatural acts, published works, authors and type specimens have been both prospectively and retrospectively registered.

Server architecture and software platform

The initial prototype implementation of ZooBank is being developed at the Bishop Museum in Honolulu. The Bishop Museum was selected in part because of its association with the ICZN (former Commissioner and President of the ICZN Neal Evenhuis, and current Commissioner Richard Pyle), and in part because of the existing network facilities and technical support. In particular, the implementation of VMware Virtual Server Architecture allows for multiple distributed and redundant server platforms to be easily established, enabling improved performance and failover support.

In its initial implementation, ZooBank is split across two separate virtual servers, both running the Microsoft Windows 2003 Server operating system. One is a dedicated web server, hosting the user interface and web services (http://zoobank.org), which are being developed using Microsoft ASP.NET, and the VB.NET programming language. The other virtual server hosts the database content, which is implemented with Microsoft SQLServer 2005. The data model is derived from portions of the *Taxonomer* data model (Pyle

2004), with modifications specific to the needs of the ZooBank registry. For its initial implementation, there is only a single instance of the database (with suitable backup); but options for globally-distributed replicated instances are being explored for the long-term.

All of these (and other) technical implementation details are provisional, established only as an initial development environment, and do not necessarily have any bearing on the final specifications for the ZooBank implementation.

Scope of ZooBank registry

At present, four different kinds of data objects may be registered within ZooBank: Nomenclatural Acts, Publications, Authors, and Type Specimens. Each of these classes of data objects has direct implications for nomenclature (in accordance with the Code). Although some of these data objects fall within the overlapping scope of other data initiatives (e.g., the library community for publications and authors, and the natural history collections community for type specimens), their importance to Zoological Nomenclature mandates that ZooBank establish its own registry for these items in order to preserve autonomy, in the event that external data entities do not persist for as long as ZooBank persists. ZooBank was originally conceived as being a registry for zoological names as covered by the Code. However, as has been observed by the biodiversity informatics community, there are as many different notions of a name as there are database systems designed to record them. This is not only true across the different major Codes of nomenclature (Botanical, Bacteriological, Zoological, Viral, and Cultivated Plants) and within the broader biodiversity informatics community, but also within the realm of practicing zoologists. In some cases, names amount to little more than strings of text characters, sometimes inclusive of authorship and/or year, sometimes restricted to the individual name elements themselves. In some interpretations, alternate spellings constitute different names, whereas in other interpretations, such orthographic variations are regarded as alternative representations of the same name. Some database systems are designed to treat a name as only a singular element of a full taxonomic name (e.g. only the species epithet), treating binomials and trinomials more as concatenations of two or three separate names. Others only regard the complete set of name elements in a given combination as a single name.

To avoid ambiguity, the core data object as registered in ZooBank is the Nomenclatural Act. A Nomenclatural Act is a type of Taxon Name Usage instance, which is defined very generally as the usage or treatment of a particular taxon name within some form of documentation (see elaboration of documentation below). In the context of ZooBank, Nomenclatural Acts are those particular usage instances that have some direct or indirect bearing on nomenclatural details, as governed by the Code. The most common types of Nomenclatural Acts are those name-usage instances that constitute the original establishment of new zoological names (i.e., original descriptions) in the family-group, genus-group, and species group (for simplicity and clarity, the registration of such acts are often referred to as as the registration of names; but in fact it is the nomenclatural act establishing the name that is registered). Other Nomenclatural Acts include emendations, lectotypifications, neotypifications, First Reviser actions, and other nomenclatural assertions that have direct bearing on aspects of zoological nomenclature according to the Code. Some have suggested that Nomenclatural Acts may also include particular name usage instances such as species-group names used in combination with a genus-group name other than the original combination. Although such Acts are not directly governed by the Code, they may affect nomenclature indirectly, such as cases involving secondary homonymy. The complete spectrum of taxon name usages that may be registered as Nomenclatural Acts within ZooBank has not been formally established, and is the subject of ongoing discussion.

Inherent to any taxon name usage instance is a documentation instance in which the usage occurred. Such documentation may be interpreted very broadly, but in the context of ZooBank, it is more narrowly limited to works published in accordance with Article 8 of the ICZN Code. Because the Code officially regulates various aspects of published works, they represent the second data object included within the scope of the ZooBank registry. In the context of ZooBank, objects that constitute published works are not limited to traditionally cited units of publications such as journal articles and book, but may also include individual taxon treatments within an article or book. The reason for allowing the inclusion of individual taxon treatments as units of publication within the scope of the ZooBank registry, is to accommodate circumstances where the authorship of the name (= the authorship of the taxonomic treatment of a name, representing a Nomenclatural Act) differs from the authorship of the traditionally-cited parent unit of publication (article, book, etc.).

The third type of object included within the scope of the ZooBank registry is Authors. Although not as extensively governed by the Code as Nomenclatural Acts or Publications, Authors have nevertheless been integral to zoological nomenclature since its inception, and therefore warrant individual registration. In addition to the Authors of registered ZooBank Publications, contributors to the ZooBank registry may also be registered as Authors within ZooBank.

The final object type included within the ZooBank scope is Type Specimens. Of particular importance are primary or name-bearing types, i.e. holotypes, syntypes, lectotypes, and neotypes. Although specimens certainly fall into the domain of natural history museums in terms of data management, they nevertheless play a critical role in nomenclature as they are the physical standard to which the name is tied, and as such fall within the scope of ZooBank. Whether or not secondary (non-name-bearing) types may also be entered into the ZooBank registry has not yet been determined.

ZooBank LSIDs

One of the primary functions of ZooBank is to issue Globally Unique Identifiers (GUIDs) to each registered data object. GUIDs are necessary because taxonomic names, author names, publication citations, and references to type specimens are not unique (e.g. homonymy of names), are subject to inconsistent representations in textual form (e.g. *Homo sapiens* vs. *H. sapiens*; Linnaeus vs. Linn vs. L.; etc.), and are generally fraught with ambiguity. Most trained taxonomists have no difficulty resolving these ambiguities; however, computers are not so adept at making informed inferences and subjective interpretations. They are much more effective when interpreting, linking, and resolving unambiguous GUIDs.

ZooBank follows the lead of the Biodiversity Informatics Standards (formerly the Taxonomic Databases Working Group; TDWG), and the Global Biodiversity Information Facility (GBIF) in adopting Life Science Identifiers (LSIDs) as the type of GUID assigned to registered data objects. LSIDs were originally developed by IBM, and are implemented, maintained and perpetuated primarily by the biodiversity informatics community. LSIDs do not require centralized issuance, and do not directly cost any money to issue. Moreover, there is a growing body of software in development and available for use in implementing and resolving LSIDs.

An LSID has minimally five parts, with an optional sixth part. Each part is delimited by a colon (:) character (Figure 3). The first two parts are always the same for all LSIDs: the lower-case characters urn:lsid. The first part identifies it as a Universal Resource Number (URN), and the second part identifies it as an LSID. The third part is called the Authority Identification, and is usually (but not always) an internet domain name registered to the LSID issuing entity. The authority identification part of all ZooBank-issued LSIDs is zoobank.org. The fourth part is the Namespace Identification, and is used to partition sets of identifiers within a particular authority. In the case of ZooBank, there are four such logical sets, represented in issued LSIDs by the text act (for Nomenclatural Acts), pub (for published works), author (for authors of published works, and for registered users of ZooBank), and specimen (for type specimens). Finally, every LSID must have an Object Identification part. This part must be unique within the Authority + Namespace combination. For ZooBank LSIDs, the object identification is a Universally Unique Identifier (UUID), a standard form of GUID common to many computer applications. There are several reasons why UUIDs were chosen for the object identification part of ZooBank LSIDs (instead of, for example, an integer number or alphanumeric code). The main reason is that UUIDs are themselves globally unique, and thus retain their identity even when stripped of the rest of the LSID parts. UUIDs are not self-resolving (i.e., having only a UUID does not allow you to automatically find out what the UUID represents), but the LSID wrapper structure (particularly the first three parts) allows resolution of the information content associated with LSIDs (in this case, ZooBank registration entities). In the event that LSIDs eventually fall out of favor within the biodiversity informatics community, the UUID itself could be embedded within a different self-resolving GUID protocol (due to its global uniqueness).

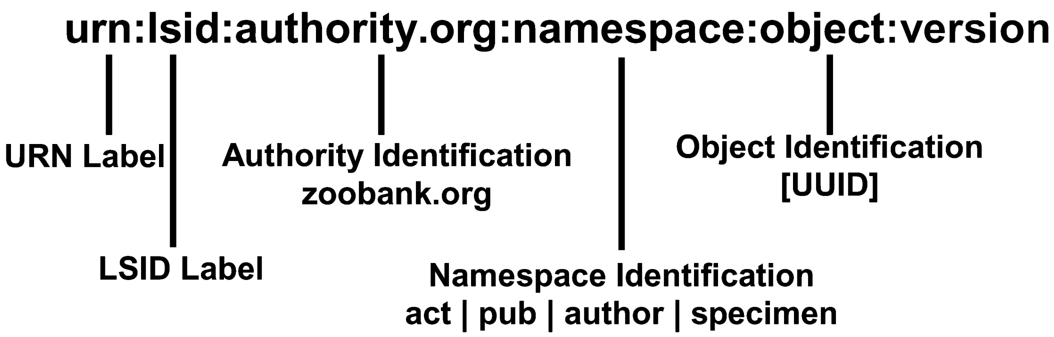

FIGURE 3. Components of an LSID. There does not appear to be a consistent implementation of the version part of the LSID, and it is not incorporated into ZooBank LSIDs.

Examples of existing ZooBank LSIDs include the following (the namespace part is emphasized in bold):

Nomenclatural Act: Original establishment of the species name, *Chromis abyssus*.
urn:lsid:zoobank.org:**act**:8BDC0735-FEA4-4298-83FA-D04F67C3FBEC

Publication: Published work in which *C. abyssus* was established (Pyle *et al*., 2008).
urn:lsid:zoobank.org:**pub**:68376390-7809-46FF-9EC4-1371B4AAD0FF

Author: First author of this published work (Richard L. Pyle).
urn:lsid:zoobank.org:**author**:8C466CBE-3F7D-4DC9-8CBD-26DD3F57E212

Type Specimen: Holotype of *C. abyssus* (BPBM 40861).
urn:lsid:zoobank.org:**specimen**:FDE70A5C-59C3-407B-B9A6-5A9A2DA14BD1

It should be emphasized that LSIDs (like UUIDs and all other GUIDs) are intended for use by computers, not by humans. LSIDs (again, like other GUIDs) are designed to operate behind the scenes, out of view of human eyes. They allow easy and unambiguous identification when linking electronic resources together, and it is in this context that they should be optimized.

Scenarios for mandatory registration

There has been a great deal of discussion concerning the implications of enforcing mandatory registration through amendments to or a new edition of the Code. In their announcement of ZooBank, Polaszek *et al*. (2005a) indicated that their intention was to make ZooBank a mandatory requirement (governed by the ICZN Code) for future names and nomenclatural acts. In a follow-up technical article, Polaszek *et al*. (2005b) outlined two proposed scenarios for how mandatory registration might be implemented in the Code, and a third scenario was proposed by Doug Yanega in a series of messages posted to the ZooBank email discussion list (Yanega, pers. comm.). A more cohesive description of these three alternative scenarios for mandatory registration (including diagrammatic flow charts) is included in Polaszek *et al.* (2008), and a brief summary is provided below.

It is important first to clarify the definition of three terms, which, for the purposes of this article, are as follows:

Registration: The process of entering a complete record in the ZooBank registry.

Publication: ICZN-compliant published works, as defined in Chapter 3 (Arts. 7–9) of the 4th Edition of the ICZN Code.

Availability: A nomenclatural act (such as a scientific name applied to an animal taxon) that meets the criteria of availability set forth in the Code.

Scenario #1: (Publication+Registration)=Availability

The first scenario posits that the act of registration would simply be added to the existing requirements of the Code, such that in order to be available under the Code, a name or nomenclatural act would need to be *both* published in accordance with existing Code rules, *and* separately registered in ZooBank. Registration could take place either before or after publication. If registration occurs before or within two years of publication, the date of availability is the publication date; but if registration is completed more than two years after publication, the date of availability is registration date (except in certain extenuating circumstances, as evaluated by the ICZN.)

Advantages of this scenario are that it would require a relatively small change to existing taxonomic practice, and it maintains implicit quality control via traditional publication venues. Moreover, many perceive this scenario as being the most likely to gain broad acceptance by the taxonomic community.

One disadvantage to this scenario is that it would require a somewhat complex procedure involving asynchronous publication and registration events, arbitrary time periods affecting date of availability, and petitions to the Commission in certain special circumstances. In particular, the temporal decoupling of publication and registration events establishes a somewhat ambiguous gray zone after publication and before registration when names & acts are assumed to be available, even though they are technically not available until registered. Also, this scenario still suffers from all the complexities and ambiguities associated with traditional paper publication entangled with nomenclatural availability. Finally, it may also require an increase in active role of ICZN staff (with associated costs) to process registration requests and verify Code compliance for issuance of GUIDs and exposing registration details to the public.

Scenario #2: Registration=Availability

In this scenario, the process of registration itself would be all that is required for availability of new names and nomenclatural acts. Prior or subsequent publication through traditional venues is encouraged, but would not be integral to nomenclatural availability.

The main advantage of this scenario is that the legalities of nomenclatural availability under the ICZN Code and the science of taxonomy are disentangled from each other. This philosophy is fundamental to the ICZN's role, as articulated in the Introduction to the 4th Edition of the ICZN Code (as quoted previously).

Other advantages of this scenario include the elimination of ambiguity concerning dates of availability, the rendering of existing complexities of nomenclatural availability of published works as moot, and the minor increases in the active role of ICZN staff.

The main disadvantage of this scenario is that it would represent a fundamental change to the way names and nomenclatural acts are established (i.e. altogether eliminating publication process as part of the requirements for availability). There is concern by many that by implication, taxonomists would lose their primary benchmark for establishing professional status (i.e. their CVs would have fewer publications listed). Even if taxonomists followed through with proper taxonomic descriptions in published form, there is concern that journals might no longer publish taxonomic descriptions if the articles no longer carry the prestige of establishing new names and acts in accordance with ICZN rules. Moreover, although there are no existing requirements in the Code for peer-review or any other form of explicit quality control (for the taxonomy associated with the nomenclatural acts), the existing requirements for publication result in a de facto standard of peer-review and quality control. This would potentially be lost if nomenclatural acts were dissociated from the richer context of taxonomic work that often is included as part of published nomenclature. Finally, there is some concern that if the process of conferring availability of names under the Code were so simplified, lazy taxonomists might never get around to publishing the full description after the name is registered, potentially creating many names without robust taxonomic definitions. Even worse, bad taxonomists (and non-taxonomists) might abuse the system by registering hundreds of bogus and unneeded names, perhaps for unscrupulous reasons (e.g. selling names for money).

Scenario #3: Publication=Registration=Availability

In this scenario, the ZooBank web site would host a full-blown, edited, peer-reviewed online journal (like *ZooTaxa* or *Zookeys*) in which ***all*** names and nomenclatural acts must be published. In this scenario, the science of taxonomy becomes an explicit part of the nomenclatural process (by Code rules). Submitted manuscripts would be open to non-anonymous review by any interested or concerned taxonomist.

There are many potential advantages to this scenario. For example, all taxonomic publications would appear in a single venue (as is now done for bacteria), instead of scattered across thousands of journals. There would no longer be a potential for one author to steal another's work by trying to submit a plagiarized work to a journal that has a faster turnaround time. All manuscripts would be examined by a large contingent of reviewers, instead of just a handful, greatly improving the reviews as well as democratizing the process. These reviews would be public (instead of anonymous), so personal grudges or biases of the reviewers would be exposed to scrutiny by the whole community. Moreover, a dedicated nomenclatural journal of this sort would mean that the review criteria would explicitly address all necessary aspects of code-compliance and proper nomenclature. Indeed, this scenario would enjoy all of the other advantages of an online review process (fast, iterative, open to bidirectional feedback, etc.), and, perhaps most importantly, would not be subject to any copyright restrictions.

Equally significant are the potential disadvantages to this scenario. Foremost, it would represent a ***major*** and ***fundamental*** change to the way taxonomy is done, both in terms of legalities of nomenclature and for the science of taxonomy. The legalities of nomenclatural availability and the subjective science of taxonomy would, for the first time, be formally coupled under Code rules. Although the open review process proposed under this scenario is appealing, many taxonomic groups do not have many (or even any) experts who would serve as reviewers, and thus submitted manuscripts may never receive appropriate peer review (although this is no less true in the current publication paradigm). Such a system would impose a huge burden on the taxonomic community to provide peer reviews to 16,000–24,000 new names each year (again, in theory this would be no different from the current paradigm). It has also been pointed out that under this scenario, existing journals that depend on taxonomic descriptions and nomenclatural acts to fill their pages and maintain a subscriber base may be driven out of business. Also, the criteria for determining how, when and by whom a submitted manuscript should be deemed accepted will always be a subjective and contentious issue.

These are by no means the only possible scenarios for implementing mandatory registration in ZooBank. Many other possibilities exist, including various aspects of these three scenarios, as well as other factors not accounted by them. Much careful discussion and consideration will be required before a working scenario can be crafted, and the associated technical infrastructure developed. It is of vital importance that this discussion be open to the broadest possible array of practicing taxonomists (not all of whom are able to participate in online discussion forums).

What is clear, however, is that the basic notion of online registration for nomenclatural acts (and associated publications) is generally desired by the majority of practicing taxonomists who have participated in discussions so far. As with so many aspects of science (and in particular issues concerning the Code) ‘the devil is in the details’. Nevertheless, scientific names for animals are every bit as relevant and important to a wide variety of different fields in biology and medicine, as they were two hundred and fifty years ago in the time of Carl Linnaeus.

Acknowledgements

We wish to thank Sandro Minelli, Jon Todd, Ken Johnson, Andrew Polaszek, Yde de Jong, Charles Hussey, and the many ICZN Commissioners and other interested taxonomists who have contributed many of the ideas included herein.

References

Bouchet, P. (2006) The magnitude of marine biodiversity. *In*: Duarte, C. (Ed.) (2006). *The exploration of marine biodiversity: Scientific and technological challenges*. Fundacin BBVA, Bilbao, pp. 31–62.

EDIT (European Distributed Institute of Taxonomy) (2008). *Future Trends of Taxonomy - Annual General Meeting, Carvoeiro, Portugal, January, 2008.* EDIT, Paris.

Linnaeus, C. (1758) *Systema Naturae per regna tria naturae, secundum classes, ordines, genera, species, cum characteribus, differentiis, synonymis, locis.* Tomus I. Editio decima, reformata. Laurentus Salvius, Holmiae.

Melville, R.V. (1995) *Towards stability in the names of animals: A history of the International Commission on Zoological Nomenclature 1895–1995*. International Trust for Zoological Nomenclature, London.

Neave, S.A. (1939–1996) *Nomenclator Zoologicus*. Vols I–X. Zoological Society of London, London (available at http://uio.mbl.edu/NomenclatorZoologicus/browse.html)

Polaszek, A., Agosti, D., Alonso-Zarazaga, M., Beccaloni, G., de Place Bjrn, P., Bouchet, P., Brothers, D.J., Earl of Cranbrook, Evenhuis, N., Godfray, H.C.J., Johnson, N.F., Krell, F.-T., Lipscomb, D., Lyal, C.H.C., Mace, G.M., Mawatari, S., Miller, S.E., Minelli, A., Morris, S., Ng, P.K.L., Patterson, D.J., Pyle, R.L., Robinson, N., Rogo, L., Taverne, J., Thompson, F.C., van Tol, J., Wheeler, Q.D. & Wilson, E.O. (2005a) Commentary: A universal register for animal names. *Nature*, 437, 477. (http://www.iczn.org/Nature_Commentary.pdf)

Polaszek, A., Alonso-Zarazaga, M., Bouchet, P., Brothers, D.J., Evenhuis, N., Krell, F.-T., Lyal, C.H.C., Minelli, A., Pyle, R.L., Robinson, N.J., Thompson, F.C. & van Tol, J. (2005b) ZooBank: the open-access register for zoological taxonomy: Technical Discussion Paper. *Bulletin of Zoological Nomenclature*, 62, 210–220. (http://www.iczn.org/ZooBank_Paper.htm)

Polaszek, A., Pyle, R. & Yanega, D. (2008) Animal names for all: ICZN, ZooBank, and the New Taxonomy. *In*: Wheeler, Q.D. (Ed.). *The New Taxonomy*. CRC Press, Boca Raton, pp. 129–142.

Pyle, R.L. (2004) Taxonomer: A relational data model for managing information relevant to taxonomic research. *Phyloinformatics*. 1, 1–54. (http://systbio.org/files/phyloinformatics/1.pdf)

Pyle, R.L., Earle, J.L. & Greene, B.D. (2008) Five new species of the damselfish genus *Chromis* (Perciformes: Labroidei: Pomacentridae) from deep coral reefs in the tropical western Pacific. *Zootaxa*. 1671, 3–31. (http://www.mapress.com/zootaxa/2008/f/zt01671p031.pdf)

Sherborn, C.D. (1902–1933) *Index Animalium; sive, Index nominum quae ab A.D. MDCCLVIII generibus et speciebus animalium imposita sunt*. Trustees of the British Museum. London (available at http://www.sil.si.edu/digitalcollections/indexanimalium/)

Zootaxa 1950: 51–86 (2008)
www.mapress.com/zootaxa/

ISSN 1175-5326 (print edition)
ZOOTAXA
ISSN 1175-5334 (online edition)

Phylogenetic hypotheses, taxa and nomina in zoology*

ALAIN DUBOIS

Muséum national d'Histoire naturelle, Systématique & Evolution, USM 602 Taxonomie & Collections, Reptiles & Amphibiens, CP 30, 25 rue Cuvier, 75005 Paris, France. E-mail: adubois@mnhn.fr

In: Minelli, A., Bonato, L. & Fusco, G. (eds) *Updating the Linnaean Heritage: Names as Tools for Thinking about Animals and Plants. Zootaxa*, 1950, 1–163.

Table of contents

Accepted by A. Minelli: 24 Oct. 2008; published: 5 Dec.2008

Abstract

Taxonomic paradigms have changed several times during the history of taxonomy, yet a single nomenclatural system, so-called Linnaean, has remained in force all along. It is theory-free regarding taxonomy as it relies on ostensional allocation of nomina to taxa, rather than on intensional definitions of nomina (e.g., "phylogenetic definitions"). Nomina are not descriptions, definitions or theories but simple labels designating taxa. Both for theoretical and practical reasons, this system should be maintained for the allocation and validity of nomina under a cladistic taxonomic paradigm. Whereas taxa can be cladistically defined by *apognoses* or *cladognoses*, nomina should remain attached to taxa through onomatophores, combined in some cases with a Principle of Coordination. Under such a system, the allocation of nomina to taxa is automatic, unambiguous and universal, and nomenclature does not infringe upon taxonomic freedom. However, to avoid misunderstandings and to solve some current problems, the current *Code* of zoological nomenclature should be improved in several respects. The distinction should be made clear between *taxonomic categories*, which have biological definitions, and *nomenclatural ranks*, which do not, as they give only a position in a nomenclatural hierarchy: if used consistently under a cladistic paradigm, they simply allow to express hypotheses about successive branchings and sister-taxa relationships. Taxa referred to a given rank in different groups cannot therefore be considered equivalent by any biological criterion. The nomenclatural rules should cover the whole taxonomic hierarchy, which is currently not the case in zoology. The recent strong increase in the number of higher taxa which results from cladistic analyses may quickly lead to chaos and problems in communication if the nomina of these taxa continue to be based on personal tastes and opinions. There is an urgent need for the zoological *Code* to cover these nomina with automatic and stringent rules leaving no place to subjective interpretation. Just like for those currently covered by the *Code*, the status of these nomina should be established in their first publication (*nomenclatural founder effect*). The *Code* should be protected against alternative nomenclatural systems by rejecting as unavailable all nomina and nomenclatural acts published without respecting the basic Linnaean system of nomenclatural hierarchy of ranks.

Key words: Allocation of nomina to taxa, Apognoses, Cladistic hypotheses, Cladognoses, *Code*, Definitions of taxa, Diagnoses, Equivalence between taxa, Hypotheses, Linnaean nomenclatural hierarchy, Monosemy, Nomenclatural founder effect, Nomenclatural parsimony, Nomenclatural ranks, Nomenclature, Nomina, Onomatophores, Polysemy, Principle of Coordination, Redundancy, Taxa, Taxonomic categories, Taxonomy

Printing conventions

In the text and tables below, species-series and genus-series nomina are printed, as usual, in lower case *italics*, whereas nomina of higher-ranked taxa are printed in small capitals, with the following distinction: family-series nomina are in *ITALICS*, whereas class-series nomina are in **BOLD**. In this paper, "the *Code*" designates the edition currently in force of the *International Code of Zoological Nomenclature* (Anonymous 1999) and "ICZN" the International Commission on Zoological Nomenclature.

A preliminary statement

At the beginning of the "century of extinction" (Dubois 2003), the science of biology is facing a new paradigm, which results from the combination of two different facts: the taxonomic impediment and the biodiversity crisis. This statement is summarized in the following sentences: "In face of the biodiversity crisis, the need for urgency could be no greater. (...) The grand biological challenge of our age is to create a legacy of knowledge for a planet that is soon to be biologically decimated." (Wheeler *et al.* 2004: 285). This well-known statement will not be discussed further here but these ideas will be kept in mind throughout the discussion below.

Taxonomy and nomenclature

Science is the study of reality, not reality itself. It provides methodologies, concepts, theories, models and hypotheses. Theories and hypotheses are accepted and used as valid as long as they have not been refuted. This does not mean that they are "true" or that they reflect exactly the reality.

Taxonomy is the *science* of classification of organisms. It recognizes classificatory units, the *taxa* (singular *taxon*). Within any given classification or *ergotaxonomy* (Dubois 2005c), taxa may be defined according to a taxonomic paradigm, i.e., a theory of biological classification.

Nomenclature is a *technique* allowing to name the taxa. It makes use of Latin or Latin-like "scientific names" or *nomina* (singular *nomen*) (see Dubois 2000b), that allow finding the taxonomic information they refer to. A nomen is just a label—not a description, a diagnosis, a definition, a coordinate, a model or a theory. A nomen may be either *defined* by *intension* (e.g., "all black animals") or *extension* (e.g., a list of black animals), or simply *attached* to a taxon by *ostension* (e.g., pointing to a particular black animal).

A nomen is not a taxon. A taxon may be defined without being named: for example, it may simply be described, diagnosed or defined, or it may be designated by a code or a number (numericlatures). A nomen may be created without designating a taxon (*nomen nudum*). Several distinct nomina may designate the same taxon (*synonymy*): this may result from objective or nomenclatural synonymy, or *isonymy* (Dubois 2000b), or from subjective or taxonomic synonymy, or *doxisonymy* (Dubois, 2000b). Two identical nomina may designate different taxa: as we will see, this may result (1) from *homonymy*, i.e., identity or similarity between *different* nomina, or (2) from *eponymy*, a situation resulting from the partially *polysemic* nature of the current nomenclatural systems relying on a *Principle of Coordination*, according to which the same nomen designates several coordinated taxa, e.g., a genus and one of its subgenera, or (3) from changes in the ergotaxonomies used by different authors and therefore in the intensions and extensions of taxa.

The aim of taxonomy is to provide a scientific classification of living organisms. Like all scientific disciplines, taxonomy relies on scientific theories or paradigms. Several "schools" of taxonomy or *taxonomic paradigms* have been in force during the long history of biology since the middle of the XVIIIth century. They differ mostly by their concepts of taxa. This is a well-known and long-discussed matter (e.g., Mayr 1982), that needs only to be very briefly reminded here.

Under an *essentialistic* or typological taxonomic paradigm, taxa were viewed as corresponding to an essence (Platonic *eidos*), fixed forever by their creator and unliable to change or evolve: the duty of taxonomists was then understood as to discover the "design" of God and to translate it into a classification. Under a *phenetic* taxonomic paradigm, the classification was based on overall similarity, and the role of taxonomists was seen mostly as developing reliable methods for measuring this similarity. Both these approaches, as well as others not mentioned here, are now largely obsolete, although the last one may have to be revived in the future, at least for the taxonomy of organisms in which lateral gene transfer is an important evolutionary phenomenon and concerns a large part of the genomes (Doolittle 1999). Two main taxonomic paradigms are currently in force in zoology and botany. Under a *cladistic* taxonomic paradigm, the classification is based on cladistic relationships, and the role of taxonomists is seen mostly as developing reliable methods for inferring these relationships and reconstructing the "tree of life": the only taxa recognized are groups considered *holophyletic* (Ashlock 1971) or "monophyletic" sensu Hennig (1950, 1966), i.e., including an ancestor and all its descendants. Finally, under an *evolutionary* or synthetic paradigm, classification is understood as aiming at reflecting the patterns of evolution, considered not only as a series of cladogenetic events, but also of anagenesis and adaptation: therefore two kinds of taxa can be recognized, either holophyletic or *paraphyletic* (Hennig 1950, 1966), both categories of taxa that are *homophyletic* (Dubois 1986, 1988b) or "monophyletic" sensu Haeckel (1866), i.e., non-polyphyletic. These current taxonomic paradigms cannot be considered as "the final word" in the history of taxonomy, the "ultimate taxonomic paradigm", especially as they do not account for phenomena like lateral gene transfer and reticulate evolution, and do not take into account the complex phenomena that are involved in the relationships between genetics and development (Evo-Devo), so that there is no *a priori* reason for rejecting the possibility that they might be replaced in the future by another paradigm or several (Dubois 2005c).

In contrast, nomenclature is not a science but a technique, a *tool* at the service of taxonomy. In order to play properly this role, to remain universal and to follow the changes occurring in taxonomic paradigms while keeping a high robustness in the nomina of taxa, nomenclature should not be linked to a scientific theory of classification, but should depend on a set of stringent, universal and stable rules, i.e., on a *Code*, that can be used under any taxonomic paradigm.

Nomenclatural rules should therefore be *theory-free* regarding taxonomy (Dubois 2007a). This tool should be as neutral as possible, in order to respect what the *Code* calls "the freedom of taxonomic thought and action". This is similar to grammatical rules relative to language or literature: they do not tell us what to say or write, but how to do it, and they are universal and stringent for proper communication. The rules should be devised in such a way that they can be used by all taxonomists whatever their opinions on taxonomy, and that they apply to nomina created within any given taxonomic paradigm, even after their transfer into another paradigm.

In some recent publications, the term "clade" has been used to designate some kinds of taxa, i.e., groups that are considered holophyletic under a cladistic hypothesis of relationships. In the scientific field of phylogenetic research, hypotheses can be built about relationships between organisms. Clades no doubt have existed and exist in the real world, but the "clades" resulting from cladistic analysis are not the "real clades" of the real world: they are theories, hypotheses, that can be refuted. Several terms have been proposed to designate such hypotheses (see Dubois 2006b: 826), among which the clearest one is that of "cladon" (Mayr 1995). Cladons are taxa of a particular kind, recognized under the tree of life (cladistic) paradigm. Cladons, just like any other kind of taxa, can be *defined*, but not *discovered*. Below, they will simply be designated as "taxa", not "clades", as I do not think that taxonomy deals with clades, but with hypotheses about clades.

Categories and ranks

Most scientific classifications of living organisms so far have used hierarchical nomenclatural systems with successive ranks from the highest to the lowest. Thus, in zoology, the partial nomenclatural hierarchy regulated by the *Code* recognizes at least 11 ranks: superfamily, family, subfamily, tribe, subtribe, genus, subgenus, species group, species, subspecies group and subspecies. Additional ranks between superfamily and subtribe can be used also if necessary, but not between subtribe and subspecies.

In zoology, the nomenclatural hierarchy has been arbitrarily divided into five *nominal-series* (Dubois 2000b). Each nominal-series includes several *ranks*. Several recent authors made a confusion between rank and nominal-series, so some clarifications may be useful.

Three nominal-series ("groups of names") are recognized by the *Code*, which regulates the use of their nomina. The *family-series* includes nomina of the ranks superfamily (ending in *–OIDEA*), family (ending in *–IDAE*), subfamily (ending in *–INAE*), tribe (ending in *–INI*), and subtribe (ending in *–INA*), and possible additional ranks without fixed endings. Bour & Dubois (1985, 1986) and Dubois (2006a) proposed standard endings for some of these ranks. The *genus-series* includes only two ranks allowed by the *Code*, genus and subgenus (see Dubois 2006b,d, 2007b). The *species-series* includes four ranks recognized by the *Code*, species-group (as "aggregate of species"), species, subspecies-group (as "aggregate of subspecies") and subspecies, no additional ranks being allowed by the *Code* (see Dubois 2006b).

Besides, two nominal-series are not recognized and regulated by the *Code*. The *class-series* (Dubois 2000b) includes all nomina of taxa of the highest ranks: order, class, phylum, reign, etc. These nomina are currently not regulated by the *Code*, but this would be very useful to avoid the progressive instauration of a chaos in higher nomenclature with the increase in the number of higher taxa that follow the multiplication of molecular phylogenies. Detailed rules have been proposed to incorporate these nomina into the *Code* (Dubois 2006a). The *variety-series* (Dubois 2005a,c) includes the nomina of all taxa of the lowest ranks, such as variety, natio, form, etc. Incorporating these nomina into the *Code* would also be useful, in order to allow naming entities at low levels of analysis, e.g. for phylogeographic studies or for conservation biology (Dubois 2006b).

Nominal-series play a crucial role in the functioning of zoological nomenclature, much more important than ranks by themselves. This is misunderstood by all those who qualify the nomenclatural system of the zoological *Code* as a "rank-based nomenclature". The two main characteristics of this nomenclatural system, which regulate the valid nomen of any taxon within the frame of any taxonomic arrangement, are the allocation of nomina to taxa through *ostension* using *onomatophores* (Simpson 1940; Dubois & Ohler 1997; Dubois 2005c), and the recognition of three distinct *nominal-series* among which a *Principle of Coordination* is in force (Dubois 2005c). Ranks by themselves play no role in the establishment of the valid nomen of a taxon. It would be more appropriate to call this system "onomatophore-based nomenclature" (Dubois 2005c), "ostensional nomenclature", or "ostensional eponymic nomenclature", but such terms would be pedant and cumbersome, so it seems better to keep the traditional term "Linnaean nomenclature", although this nomenclatural system is quite different from that used by Linnaeus himself (Moore 2003; Dubois 2005c, 2006c).

The basic function of nominal-series is to allow nomenclatural parsimony, as any nomen given to a taxon *within* a nominal-series is available also for other taxa in the same nominal-series. This is made possible in this system because all nomina in the same nominal-series interact concerning: (1) coordination and eponymy; (2) synonymy; (3) homonymy; (4) priority. Among all nomina that may potentially apply to any given taxon of a nominal-series in a given ergotaxonomy, the valid nomen is usually established by priority of publication.

On the other hand, nomina do not interfere *between* nominal-series for eponymy, synonymy, homonymy and priority. The only interactions between nomina of different nominal-series are: (1) through the use of some nomina in a lower series as onomatophores for nomina of an upper series: "type-species" for nominal genera or subgenera, and "type-genera" for nominal family-series taxa; (2) in a few very special cases (Art. 32, 33, 35, 39 and 40 of the *Code*) which concern only family-series nomina (Dubois 2008c).

Are nomenclatural ranks useful or harmful?

The usefulness of nomenclatural ranks has been challenged recently by some authors who support unranked nomenclatural systems recognizing only "taxa" (or sometimes "clades") which are not referred to ranks. In fact, few of these authors really follow a fully unranked nomenclatural system, which would require abandoning also the ranks species and genus. Most of them in fact adopt a "partially ranked" nomenclatural system, using Linnaean nomenclature for taxa at the ranks species and genera, and sometimes also families, superfamilies, subfamilies and tribes, but using the term "taxon" for all other taxa above, between or below the latter in their hierarchy. In particular, as will be discussed in more detail below, some authors use *pseudoranked* nomenclatural systems (Dubois 2007a), in which they refer some taxa to formal ranks (genus, subfamily, family) but without respecting the hierarchical arrangement of taxa in which sister-taxa are afforded the same rank.

Discussing the usefulness of ranks will require exploring two distinct questions: (1) the problem of arbitrariness of ranks and equivalence of taxa referred to the same rank; (2) the hierarchical organisation of taxonomic information.

The equivalence of taxa referred to the same taxonomic category

In the scientific literature dealing with biodiversity, in various fields including evolution, palaeogeography, biogeography, ecology, conservation biology, etc., there is a widespread use of ranks for taxonomic comparisons between taxa, faunae, periods, etc. Such works rely for example on numbers of genera, families, orders or classes to compare faunae in different regions or at different epochs. This would suggest that evolutionary patterns, taxonomic richness and diversity, etc., can be inferred from nomenclatural patterns, as such calculations rely on the nomenclatural ranks afforded to taxa. Are such comparisons warranted? They would be so only if taxa at the same rank in different groups were "equivalent", at least by some criteria.

In this context, equivalence requires common criteria (Schaefer 1976). Such criteria can be used either to *establish* taxa that are equivalent in some respect, or to *measure* the equivalence between taxa previously established using other criteria. Various criteria have been used for this purpose (Dubois 1988b), e.g., among others, quantitative metataxonomic criteria, phenetic criteria relying on characters, relational criteria or absolute age of taxa. Let us consider some of these approaches more closely.

The metataxonomic criterion of Van Valen (1973) relies on the number of taxa at different ranks in different taxonomies. Some taxonomies appear "well-balanced" (e.g., **CHONDRICHTHYES**), others "oversplit" (e.g., **AMPHIBIA**) and others "overlumped" (e.g., **AVES**), but these disparities can have various causes, from different taxonomic practices in different zoological groups to genuine differences in their evolutionary patterns. Therefore, this criterion can hardly be used to *standardize* the use of ranks in zoology, but it can provide interesting information when *comparing* classifications (Dubois 1988a-b).

Phenetic criteria relying on *characters* (obtained from morphological, molecular, karyological, ethological, bioacoustical or other data) allow estimates of various "distances" between taxa (Dubois 1988b) and provide measurements of variability and dispersion within various taxa. Such criteria can also be used to *compare* the taxonomies of these different groups, but hardly to *standardize* the use of ranks in zoology. Within a given group, it is possible to use some ecological or behavioural characters to homogenize the use of ranks: for example, in the **AMPHIBIA**, it has been suggested that holophyletic taxa with different reproductive modes be afforded the rank of genus, except when this contradicts the crossability criterion discussed below (Dubois 1988b, 2004b). However, such criteria cannot be generalized to the whole taxonomic hierarchy, as they can be used only for closely related taxa sharing homologous characters (Schaefer 1976).

Relational taxonomic criteria or *relacters* (Dubois 2004b), based on real interactions between organisms (not on comparisons of characters by scientists), such as the mixiological criterion at the species level (Mayr 1940, 1942, 1963; Dubois 2008d,f) or the crossability criterion at genus level (Dubois 1981, 1988b, 2004b) require certain precautions for proper use, but these are not always respected. For example, according to the "biological species concept"of Mayr (1940, 1942), the mixiological criterion at the species level states that whenever two entities freely exchange genes *in nature*, i.e., when an unbiased bidirectional introgressive gene flow exists between them in a contact zone, these two entities are part of the same species taxonomic unit. But the fact that, in captivity or in artificial conditions, individuals of two entities are able to give birth to living offspring, is not by itself sufficient to consider them conspecific, as suggested by some (e.g., Samadi & Barberousse 2006), as various factors, e.g. behavioural, can impede gene flow between them in nature: many cases are known of species fully separated in nature although they are genetically compatible. In contrast, at generic level, the crossability criterion simply states that whenever two species are liable, in natural *or artificial* conditions, to produce viable hybrids, these two species should be referred to the same generic taxonomic unit, but the reverse is not true: intersterile species may well be congeneric (for details, see Dubois 1988b). When used carefully and rigorously (i.e., respecting these precautions), relacters are very helpful for a standardization of taxonomy at and between the ranks species and genus in some zoological groups. However, such criteria cannot be generalized to the whole taxonomic hierarchy, as they cannot apply to ranks above genus and to organisms that are not bisexual or that have peculiar meioses (Dubois 2008d, 2008f), not to mention the fact that they cannot be applied to fossils.

The proposal to standardize the use of ranks over the whole zoology through the absolute age of taxa was first made by Hennig (1936, 1950, 1966), subsequently abandoned and recently revived by Avise & Johns (1999) and Avise & Mitchell (2007). Thanks to the recent improvement of molecular dating methods of of cladogeneses, this criterion is very promising and might be implemented in the future, but for the time being this is difficult, because of three different kinds of problems. First, this approach cannot be of generalized use today because of missing information for many groups of organisms. This problem will progressively be solved as data are accumulated. Second, this approach applies only to organisms living synchronically, e.g., nowadays, as otherwise all fossil taxa would have to be given higher ranks simply because they lived long ago! The datation of fossil groups for which no representatives exist in the current fauna is today more difficult, but progress in dating methods for these groups can be expected. As molecular and palaeontological data increase, it will be possible to estimate the absolute age of all major taxa *at any period* of the earth's history, and therefore to use this method for rank assignation of taxa, but these ranks will be valid only for comparisons of synchronic taxa (living at the same period). Thirdly, implementing such a change in the allocation of ranks to taxa would pose strong problems regarding "taxonomic tradition", as well illustrated by Avise & Johns (1999): in the cases of the cichlid fishes of lake Victoria, of anthropoid primates and of fruit flies of the genus *Drosophila*, any time-scale standardization in the ranks given to taxa would result in changing drastically the ranks traditionally given to taxa in at least two of these three groups. Therefore, such a drastic change would be impossible to carry out through the individual action of some zoologists, and could be so only through a collective action of the international community in one or several large international meetings (Dubois 2007a), the organization of which may take a few decades, if it ever occurs.

In conclusion, for the time being, there exists no method for a *general* standardization of the "meaning" of ranks over the whole of zoology and palaeontology. The "meaning" of the rank family or genus is by no way equivalent in flatworms, beetles and birds. Therefore, any comparison between faunas or taxonomies using the ranks of taxa as a criterion (e.g., quantitative comparisons based on numbers of taxa at some ranks) is unwarranted and misleading (Minelli 2000). This statement was one of the main reasons why several recent authors rejected the use of ranks in taxonomy. But is this reason valid? It would be so only if nomenclatural ranks were viewed as identical with taxonomic categories, an opinion that is shared by many but that is questionable. Dubois (2005c, 2007a) proposed to recognize a basic distinction between these two concepts, stating that one refers to taxonomy and the other one to nomenclature.

The criteria of equivalence between taxa briefly reviewed above are of two kinds: biological and chronological. Biological criteria are all of limited use for equivalence, as they can be used only at low taxonomic levels (species and genus), and are not relevant in various situations. Chronological criteria are potentially general but face three problems (missing data, applicability only for synchronic taxa and taxonomic tradition) that preclude their implementation over the whole of zoology for the time being. This is true, but, as discussed below, the use of such criteria in some situations can however be informative as it allows to obtain useful information regarding the patterns of evolution. Sets of taxa defined by such criteria can be designated as *taxonomic categories*. Taxonomic categories are categories of taxa that share some common features and are equivalent by some taxonomic criterion. They do not provide information on cladogenetic relationships, but this information can be provided by *nomenclatural ranks*. On the other hand, nomenclatural ranks are nomenclatural tools which only provide information on the detailed hierarchical structure of a taxonomic hierarchy, but no information on the evolutionary peculiarities of the taxa in this hierarchy. To make this unusual distinction fully clear, a few words must be said about taxonomic categories as here defined.

Taxonomic categories

The interest in taxonomic categories, particularly in "evolutionary categories", was high in the middle of the XXth century, at the time of the "evolutionary synthesis" (Mayr 1982). In contrast with the current situation, the focus of interest of evolutionary biologists then was not so much in the cladistic relationships, but mostly in "micro-evolutionary processes", such as speciation, hybridization, adaptation, etc. Evolutionary taxonomic categories were developed then because they were considered a useful tool for evolutionary biologists: referring a taxon to one of these categories (Rassenkreis, Artenkreis, superspecies, semispecies, subspecies, etc.) is a simple and brief way to provide some information, for example on the "stage" in the process of speciation reached by various allopatric entities in an archipelago. This is part of the special domain of taxonomy that deals with species and other low-level taxa, which has been designated as *microtaxonomy* (Mayr & Ashlock 1980) or *eidonomy* (Dubois 2008d). In strong contrast with nomenclatural ranks, such categories are indeed *defined* by biological and/or chronological (i.e., evolutionary) criteria, and all taxa referred to one of these categories are indeed equivalent by these criteria. Many such taxonomic categories, most of which are at the nomenclatural rank of species, just above or just below, have been proposed, mostly in the first half of the XXth century. Several good reviews and syntheses about these eidonomic categories, their definitions and usefulness, with examples of evolutionary situations referred to by these concepts and terms, are available (Bernardi 1956, 1957, 1980; Haffer 1986). More recently, other eidonomic categories (klepton, klonon, kyon, etc.) were established to accommodate bisexual, unisexual or asexual entities that have special reproductive modes, often with particular gametogeneses with special meioses, metameioses or ameioses (Dubois & Günther 1982; Dubois 1991, 2008d,f; Bogart *et al.* 2007).

Although most of these taxonomic categories are situated around the species level, this approach can be useful also at the genus level, as shown by the proposal to use data from interspecific hybridization as a relacter to delimit genera (Dubois 1988b). Combined with the requirement to recognize as taxa only groups that appear holophyletic with the data available, and with the use of other criteria such as the reproductive mode, this criterion allows to greatly enhance the objectivity, repeatability and equivalence of generic taxa, as illustrated in the **AMPHIBIA** (Dubois 1987, 2004b), where genera recognized on the basis of cladistic data alone (e.g., Frost *et al.* 2006) are in no respect equivalent or even comparable, and may be considered much less useful to many biologists.

This evolutionary or synthetic approach to taxonomy is certainly not "fashionable" today, where few taxonomists work at population level, consider adaptation as an important taxonomic criterion or work on hybrid zones or artificial hybridization to obtain information that they will use in their taxonomic work. However, the idea that cladistic relationships between organisms is the only information interesting for taxonomists is a very strange one indeed, and it seems reasonable to think that times will come when young taxonomists become again interested in these matters. When this occurs, they will certainly take advantage of the rich literature of the last century on these questions, and of the evolutionary taxonomic categories created by the authors of this period.

As explained above, these categories are based on biological or chronological criteria, but not on cladistic data. They can be used in some cases to make taxa in different groups "equivalent" in some respect, but this has no generality over the whole animal kingdom as these criteria are irrelevant and non-usable in many cases. Thus, the mixiological criterion as used at species level cannot be used in allopatry or allochrony, or between species that are intersterile, whereas in contrast, knowing that two entities prove interfertile in captivity or under artificial fertilization does not tell us whether gene flow occurs between their populations when they get in contact in the field: the only way to have the answer to this question is through field work and observation of the contact zone. Such categories are often more meaningful to field naturalists than to laboratory workers.

The confusion between taxonomic categories as here defined and nomenclatural ranks has long created difficulties in theoretical discussions among taxonomists and even among evolutionary biologists. They stem in many cases from the use of the same term to designate different concepts. This is particularly true of the term *species*, which has been used in many senses in biology, and even in two fully different senses in systematics, to designate either a taxonomic category or a nomenclatural rank.

The many "species concepts" that are repeatedly discussed by theoreticians of systematics apply to the *taxon* species: they are taxonomic concepts. In contrast, the use of the *rank* species to designate a taxon is a matter of nomenclature. In order to avoid the confusion between the two distinct uses of the term "species", it appears useful, either to create two new terms, or to restrict the use of this term to one of these two situations. Being more parsimonious, the second solution was retained by Dubois (2007a, 2008d,f) who proposed to restrict the use of the term "species" to the nomenclatural rank. As a matter of fact, whatever the taxonomic paradigm they use, or even if they use none (which is often the case!), all biologists designate the organisms they study under a binominal Latin nomen, their species nomen, like *Drosophila melanogaster* or *Homo sapiens*. Such nomina, which carry by themselves no information on the criteria used to build the classification, are those which are found not only in all scientific publications, but also in all official texts and lists in force in commerce, customs, laws, conservation biology, etc. It seems therefore better to restrict the use of the term "species" to this nomenclatural acceptation of the term. For the taxonomic concept designating a unit in a classification, Dubois (2007a) suggested to use a similar term but ending in *–on*, just like the term "taxon": *specion*. As there are several distinct concepts of specion, they can be given different designations, such as *mayron* for the "biological species concept", *simpson* for the "evolutionary species concept", *klepton* for a taxon accommodating entities reproducing by "hybridogenesis" or gynogenesis, or *klonon* for those reproducing by parthenogenesis (for more details, see Dubois 2008d,f). These terms, and many others that could be coined in a similar way, designate different kinds of specions, i.e., taxonomic categories defined according to different criteria. They are therefore *alternative* categories, which have no hierarchical relationships between them. Nomenclaturally, all these taxonomic categories include taxa which are referred to the same nomenclatural rank, that of *species*.

The same distinction can be made for taxa referred to nomenclatural ranks above the species, e.g., *genion* for a taxonomic category that can be recognized on the basis of biological criteria like crossability for taxa referred to the rank genus (Dubois 2007a, 2008d). However, this process soon reaches its limits when one moves upwards in the nomenclatural hierarchy, because, for reasons discussed above, there are no common biological criteria allowing to define a taxonomic category like family, order, class or reign. The chronological criterion of the absolute age of taxa could allow defining such categories, but the time is not yet ripe for this, as we have seen.

Nomenclatural hierarchy: the Principle of Coordination

A hierarchical presentation of biological classifications has been used long ago, even before the works of Linnaeus, which however are striking as showing a very consistent use of such a system (Dubois 2007c). A hierarchical presentation is indeed much more informative than a non-hierarchical one (Knox 1998), just like a database or even a simple index where all items are listed alphabetically are less useful for finding an information than hierarchical ones (Dubois 2007a). This practical aspect of taxonomic hierarchies, as a device for storing, tracking and retrieving information is an important one, that should not be ignored or despised (Cracraft 1974; Mayr 1982; Ashlock 1984; Benton 2000). Besides and quite independently, although initially they were viewed as expressing the *scala naturae*, i.e., a scale of value and importance of organisms credited to God, it so happens that hierarchical taxonomies are particularly efficient for presenting cladistic information. Under the "tree of life" paradigm, diversification of organisms can be reduced to a succession of cladogenetic events. This is a simplified vision of evolution, which ignores the importance of anagenetic change within lineages as well as all phenomena of reticulate evolution, speciation through hybridization and lateral gene transfer, but this simplification provides an efficient tool for building taxonomies. Under this paradigm, each cladogenetic event in the "tree" can be expressed by a new hierarchical level in a classification. A hypothesis of cladistic relationships can be expressed in different ways, such as a tree, a list with successive indentations from the margin, or a hierarchical arrangement of taxa successively included in each other. In this case, inclusion of a taxon G in another one F (such as a genus in a family) expresses the fact that the cladogenesis which gave birth to G was posterior to that which produced F. In terms of logic, successive inclusion of taxa is strictly equivalent to a tree with successive branchings. This is for example the case of figures 2 and 3 in Dubois (2006a).

This taxonomic hierarchical representation of phylogeny can be expressed nomenclaturally, and this is the role of ranks. Although ranks were not used for this purpose in the early days of taxonomy, it turned out that they can play this role very well. However, to use the nomenclatural hierarchy as a reflection of the structure of a cladogram or a phylogenetic tree requires a few assumptions. It seems that misunderstanding these assumptions played a role in the recent rejection of ranks by some taxonomists.

The first important assumption is that sister taxa must always be referred to the same nomenclatural rank (Raikow 1985; Sibley & Ahlquist 1990): they are therefore *parordinate* (Dubois 2006b: 827). Second, any taxon is *subordinate* to a *single* upper taxon, which must be referred to the *just upper rank*. It may be *superordinate* to two or more taxa of *just lower rank*. In such a system, the relations between all taxa that are connected by superordination, parordination or subordination are *relations of coordination*. In the absence of such relations between them, two taxa may be described as being in a relation of *alienordination* (from the Latin *alienus*, "foreign", and *ordo*, "order")[1]. Thus, in the recent **AMPHIBIA**, according to the cladistic relationships currently agreed upon by most authors (e.g., Frost *et al.* 2006), and according to the higher nomenclature of Dubois (2004a, 2005d), the taxon **BATRACHIA** is the sister-taxon of the **GYMNOPHIONA**: they are parordinate taxa that must be given the same rank, in this case that of superorder. Both are subordinate to the subclass **NEOBATRACHI**, and the superorder **BATRACHIA** is superordinate to the orders **ANURA** and **URODELA**. The latter are alienordinate to any other taxon that is not directly related to them by coordination, e.g., the **GYMNOPHIONA**.

Under the *Code*, to be nomenclaturally available, any nomen must be published following a set of stringent rules (Dubois 2005c, 2008e). This includes the need to refer this nomen to one of the three nominal-series recognized by the *Code*: family-, genus- or species-series. As ranks and nominal-series have no biological meaning by themselves, this assignation is arbitrary, mostly guided by tradition and consensus. This poses (or should pose) no problem to taxonomists, who know (or should know) that ranks are meaningless, and it should be their duty to explain this to laymen in order to avoid them to believe the contrary (Dubois 2006c).

The *Principle of Coordination* is a major rule of the *Code*, which states that, within a nominal-series, among all the parordinate taxa that are subordinate to the same superordinate taxon, one, called in the *Code* the "nominotypical taxon", must bear the same nomen (with the same nomenclatural author and date) as this superordinate taxon. The nomen of the subordinate taxon is identical in spelling to that of the superordinate taxon in the species-series (subspecies *temporaria* of the species *Rana temporaria*) and in the genus-series (subgenus *Rana* of the genus *Rana*), but must be emended to indicate the rank in the family-series (subfamily *RANINAE* of the family *RANIDAE*). These different terms are not different *nomina*, as they keep the same author, date and onomatophore, but are different "avatars", or *morphonyms* (Dubois 2000b), of the same nomen. The terminology provided by the *Code* is not precise enough as it does not allow to point to the status of all these morphonyms of the same nomen created by the Principle of Coordination, so Dubois (2006b) proposed to call *epinym* the morphonym designating the superordinate taxon, and *hyponym* the morphonym designating the subordinate taxon. Epinym and hyponym are two *eponyms* of a single nomen, and the relation between

---

[1] This term is here substituted to the term *xenordination* (Dubois 2006b: 827) which was ill-formed, being a combination of Greek and Latin roots.

eponyms can also be called *relation of eponymy*. Among the nomina available for all the subordinate taxa of a taxon, the hyponym is established by the Principle of Priority. As for the nomen of the parordinate taxon, it may be called the *getonym* of the latter (from the Greek *geiton*, “neighbour”, and *onoma*, “name”), and the relation between getonyms a *relation of getonymy.* The relation of coordination, which involves not only eponyms but also getonyms at all ranks, is more comprehensive than the relations of eponymy and getonymy, i.e., it is a combination of both. Nomina that are in a relation of alienordination are *telonyms* (from the Greek *tele*, “far away”, and *onoma*, “name”).

Because the zoological nomenclatural hierarchy is divided into five successive “slices”, the nominal-series, the relation of coordination does not cover the whole nomenclatural hierarchy but is limited to subsets of the latter. Thus, in a given ergotaxonomy, one such subset may be composed of a superfamily, its subordinate families, their subordinate subfamilies, tribes and subtribes, but it stops upwards when one reaches the lowest rank of the class-series (e.g., suborder) and downwards when one reaches the rank genus. The set of nomina which are involved in the relation of coordination in this case, from superfamily to subtribe, may be called a *coordinate nomenclatural set*. It corresponds to a set of nomina which, in a given ergotaxonomy, designate taxa that are either superordinate, parordinate or subordinate to each other, but it excludes all those which are alienordinate.

Monosemic and polysemic nomenclatural systems

The existence of the Principle of Coordination in the *Code* results in this nomenclatural system being partly polysemic. In grammar and linguistics, *monosemy* applies to a situation where one word has only one meaning, whereas in *polysemy* one word has several meanings.

Polysemy is very widespread in all “natural languages”. The same word may have several meanings, but the proper meaning is usually easy to identify because of the context. However, in science, in order to make communication more precise and to avoid possible confusions, most “technical” words have only one meaning. This is not the case in zoological nomenclature, because of the Principle of Coordination: several coordinate taxa bear the same eponym. Is this justified? This feature of the *Code* has recently been criticized, so what are the pros and cons of eponymy in nomenclature?

The major argument against eponymy in nomenclature is that it may be a source of ambiguity, particularly for non-taxonomists. Thus, Hillis (2006) suggested that “Google users” may be confused, when they search for a nomen, e.g., *Rana*, to receive replies that concern both the genus *Rana* and its hyponymous subgenus *Rana*, as they do not understand that the same nomen can designate two distinct taxa. This is certainly true. The question is whether a scientific discipline must be directed by laymen, be they “Google users” or governmental or other “experts”, or by the scientists involved in the discipline itself. Until now, science has been mostly directed by scientists, and, to take just one example, the systems of designation of atoms and of molecules have been decided by chemists: no “user” will challenge their decision to designate the atom of copper by Cu and the carbon dioxide molecule by CO_2.

Until now, the partially polysemic nature of zoological nomenclature does not seem to have been criticized by zoologists themselves, as they have apparently been able to distinguish easily between “genus *Rana*” and “subgenus *Rana*”. However, specialists should not remain deaf to the comments from non-specialists. Times are no doubt changing regarding access to scientific information, which is now much wider through the web than it has ever been, and this request for removing the possible ambiguity of some nomina should not be ignored, but discussed.

The main argument in favour of partial eponymy in nomenclature is *nomenclatural parsimony* (Dubois 2006b), i.e., the need of fewer nomina for the same number of taxa. For example, in the case of a superfamily A that contains families, subfamilies, tribes and subtribes, the same nomen A may designate 5 taxa, several other nomina 4 taxa, others 3 and 2 taxa. Altogether, over the whole of zoological taxonomy, this allows to spare the creation of a large proportion of nomina (a proportion which would be interesting to estimate from real ergotaxonomies).

Sparing the creation of nomina allows to make nomenclatural databases simpler and "lighter". It is furthermore fully justified in view of the fact that taxonomies are constantly changing and improving, as more data are obtained and as taxonomic methods and concepts evolve, so that many nomina once created disappear as synonyms. When this occurs to an eponym (a nomen that is used as valid at different ranks in a taxonomy), only one nomen is concerned by this synonymisation, whereas, if each taxon had been given a different nomen, several nomina would have been created uselessly and would have to be stored in synonymies. Thanks to the existence of eponymy, it is often possible to establish a new taxon without having to create a nomen, as nomina once considered synonyms can be resurrected from synonymy. In contrast, the suppression of eponymy in zoological nomenclature might possibly be another inducement for some zoologists to create nomina, just to "attach their names" to them, a real problem in biological nomenclature (Dubois 2008b; Evenhuis 2008).

Despite these advantages of polysemy, the question raised by Hillis (2006) needs consideration. It is clear that, in the coming decades, more and more non-specialists will have access to taxonomic databases and data, and that some of them will be confused by polysemy. Would it then be a good idea to remove eponymy from zoological nomenclature? Contrary to what is stated by some, who consider that eponymy is inherent to Linnaean nomenclature, this would be fully possible, by changing only one rule of the *Code*, i.e., by replacing the Principle of Coordination by another system, as will be discussed below. Before discussing this however, we need to come back to the distinction between *definitions of taxa* and *definitions of nomina*. This distinction is not made by some supporters of "phylogenetic nomenclature", and this confusion is basic for many endless discussions between the latter and partisans of keeping the Linnaean nomenclatural system.

A few final words of caution must be added here regarding the meaning of the term *eponymy*. The situation it describes can be, and has been, confused with two other situations regarding biological nomenclature. *Eponymy* is the situation where the *same nomen* (same author, date and onomatophore) is used *in the same ergotaxonomy* as the *valid nomen* for *several distinct, coordinate taxa*. In contrast, *homonymy* is the situation where *different nomina* (generally with different authors, dates and onomatophores, with a few exceptions, when the same author used the same nomen for naming two different nominal taxa) are *nomenclaturally available*—which results in one of them, usually the junior one, being rejected as *invalid*. Finally, a third situation results from the fact that zoological nomina under the *Code* are not *defined* by closed intension or extension, but *attached* to taxa by ostension (Stuessy 2000, 2001; Keller *et al.* 2003). This results in the same nomen being liable to designate quite different taxa in different ergotaxonomies, the only requirement being that these taxa must include the onomatophore of this nomen. The reasons why this is highly preferable to a system of closed intension or extension were explained in detail elsewhere (Dubois 2005a, 2006c, 2007a): if a nomen corresponded to a strict, unchangeable definition and/or content of the taxon, a new nomen would have to be coined every time a subordinate taxon or even a specimen is added to the taxon or removed from it, so that there would be no continuity in the use of nomina and no simple way to understand the taxonomic history of a group, as is now possible through "synonymies" or more exactly *logonymies* (see Dubois 2000b). The situation here described, where the *same nomen* applies to *different taxa*, but *in different ergotaxonomies*, is neither homonymy nor eponymy, and its clear distinction from the latter two requires a special designation. For this situation, I propose the term *astatonymy* (from the Greek *astatos*, "unstable", and *onoma*, "name"). This situation is extremely common in zoology, by far more than the situation where the nomen has always designated exactly the same taxon since its creation, which may be called *menonymy* (from the Greek *meno*, "I stay, I am stable", and *onoma*, "name").

Definitions of taxa

Definitions of taxa are a matter of taxonomy, not of nomenclature. Different taxonomic "schools" use different kinds of definitions of taxa. Nowadays, no taxonomic school claims to be "Linnaean", i.e., to use "Linnaean" definitions of taxa. There exist no such things as "ICZN-taxa" (Joyce *et al.* 2004) because the *Code* does not provide any guideline for defining taxa, being theory-free regarding taxonomy. In current taxonomy, only two kinds of definitions of taxa are widely used: phenetic definitions or *diagnoses*; and cladistic or "phylogenetic" definitions, or *cladognoses* (Dubois 2007a: 43).

Diagnoses are definitions of taxa which are not associated with a cladistic hypothesis. They are based on "character states" or *signifers* (Ashlock 1985) that are considered to be shared by all members of the taxon and absent in all non-members.

Cladognoses are definitions of taxa that are associated with a cladistic hypothesis. They may be based either on characters or on relations.

Cladognoses may be based on signifers (1) that are considered to be shared by all members of the taxon and absent in all non-members, and (2) that are regarded, on the basis of a cladistic analysis and hypothesis, to be autapomorphic for the taxon. Such cladognoses have received the long and cumbersome designation of "apomorphy-based definitions" (de Queiroz & Gauthier 1990), but may be called more shortly *apognoses* (Dubois 1997). Alternatively, cladognoses may be based directly on the hypothesized cladistic relationships between taxa. Such cladognoses, which received no general designation by de Queiroz & Gauthier (1990) and their followers, can be called *coinognoses* (from the Greek *koinos*, "common, kindred", and *gignosko*, "I know"). They are of two kinds: "node-based definitions" (de Queiroz & Gauthier 1990) or more briefly *rhizognoses* (from the Greek *rhiza*, "root", and *gignosko*, "I know"), and "branch-based definitions" (de Queiroz & Gauthier 1990) or more shortly *caulognoses* (from the Greek *kaulos*, "stalk", and *gignosko*, "I know"). Although de Queiroz & Gauthier (1990) stated that these definitions apply to *nomina*, they in fact apply to *taxa*, as they are based on statements about the organisms and are a matter of taxonomy (Dubois 2005c).

Allocation of nomina to taxa

Three systems have been used in zoological nomenclature to allocate nomina to taxa: extensional definition, intensional definition and ostensional allocation. They were described in some detail elsewhere (Dubois 2006c, 2007a) and only a summary will be presented below.

(1) *Extensional definitions of nomina* provide a statement about the extension or circumscription of the taxon designated by the nomen. This may be a list of the members of the taxon (inclusive extension) or of its non-members (exclusive extension). This system is theory-free regarding taxonomy. There are two distinct ways of defining nomina by this system.

(a) *Closed extension*, which allows no modification of the extension of the taxon (addition or removal of members), is always monosemic. This system is favoured by some computer specialists. It would be appropriate for a taxonomy including a very low number of units or taxa, all already known or close to it (such as the planets of the solar system or anthropoid primates). As this is not the case in zoology, where millions of species await discovery, description and naming, this system is much too rigid to be appropriate in zoological nomenclature.

(b) *Open extension* is more flexible, as it allows changes in the content of the taxon without having to change its nomen. The situation is complex if *removal* of members is allowed, as this often results in a modification of the original intension of the taxon and can lead in some cases to a significant drift from the original taxon, which in fact amounts to a change in the intension of the nomen and therefore to the creation of a new nomen (a junior homonym of the original one): this system is inappropriate in zoological taxonomy. But if only *addition* of members to the taxon is allowed, the original intension may be kept, which allows to avoid such a drift. This latter system does not provide a strict definition of the taxon to which the nomen may apply, as no limits are assigned to the taxon. Pushed to its extreme, this system is equivalent to that of ostension presented below.

(2) *Intensional definitions of nomina* provide a statement of some of the *characters*, *properties* or *relations* (e.g., as inferred from cladistic analysis) considered to be shared by all members of the taxon. Such a system is always monosemic and tied to a taxonomic paradigm. Three major kinds of such definitions have been in use in zoological taxonomy, the last two being still in use by different authors nowadays. In the case of *essentialistic* intension, the nomen was supposed to express the Platonic *essence* of the taxon. In *phenetic* intension, the nomen is bound to a *diagnosis* of the taxon (a list of its differential characters). In *cladistic* intension (used e.g. in "phylogenetic nomenclatures"), the nomen is bound to a *cladognosis* of the taxon, which can be either a list of its autapomorphic characters (*apognosis*) or a statement of its hypothesized cladistic relationships (*coinognosis*). Intensional definitions of nomina are inappropriate in zoological taxonomy for not being universal, being linked to a theory of taxonomy. Furthermore, in some of the proposals of "phylogenetic nomenclatures" published in the recent years, this system is partly subjective and circular within language, for relying only on verbal definitions of taxa without the compulsory use of reference specimens.

(3) In *ostensional allocation of nomina to taxa*, nomina are not *defined* but *pointed to* by an *onomatophore* (nomen-bearing device), that may be either a *specimen* or a *nomen* (which itself in the end refers to a specimen through a species-series nomen). This system provides neither characters, properties or relations for the taxon, nor a list of its members or non-members, nor its limits: all this is left to taxonomy and is not part of the nomenclatural process. This system is *objective* as based on specimens. It can be either monosemic or polysemic.

Interestingly, this system of ostension is not proper to biological nomenclature. A similar system is rather frequently used in geography. Many administrative divisions in many countries are named by reference to the name of their major city or of a river that flows, at least in part, within the province or district. Just like in biological nomenclature, naming a province "Padova" only tells us that the city of Padova is included in the province, but not the limits of the latter. Just like in biological nomenclature also, in some countries, several progressively comprehensive administrative units included in one another (district, province, etc.) may bear the same name, that of their main city. Just like in biological nomenclature also, the extension of administrative divisions and their boundaries are liable to change without entailing a change in their name. This may even cause problems to zoologists when trying to locate an ancient collection locality for specimens: this is the case of the frog species *Rana maritima* Risso, 1827, which had been stated to have been collected at Napoléon's time near the sea in the French département of the "Alpes Maritimes", an administrative division that straddled the current French département of the "Alpes-Maritimes" and the Italian regione of Liguria (Dubois & Ohler 1995). Countries usually keep their names although their extension and boundaries may change: just compare the successive maps of "France" during the last five centuries… In the case of geography however, the reference to an onomatophore is not always strict and stable, as even the capital of a country may shift from a city to another, and there existed for several centuries a "Roman Empire" that did not include the city of Roma! However, in many cases, the nomination of administrative divisions relies on a system close to that of onomatophores in biological nomenclature.

Five different systems of ostensional allocation of nomina to taxa have been described in details by Dubois (2006b), which are only summarized here.

(a) The system of *onomatophores with a Principle of Coordination* is in force in the three nominal-series recognized by the *Code*. This partially polysemic system uses inclusive ostension with eponymy. The valid nomen of any ergotaxon is the oldest among all the nomina created for protaxa whose onomatophores are included in the ergotaxon.

(b) A system of *onomatophores with absolute ranks* would tie each nomen to a rank, thus suppressing the nominal-series. In such a fully monosemic system, the valid nomen of each ergotaxon would be the oldest one for a taxon of this rank including an onomatophore referred to this taxon. This system is the only one that would deserve the designation of "rank-based nomenclature" which is used improperly by some to designate Linnaean nomenclature. It is not to be recommended, as it would put inappropriate emphasis on ranking and appear to support the idea that ranks have a meaning by themselves, other than expressing sister-taxa relationships and hierarchical relationships between taxa. Furthermore, it would result in many nomina having to be often abandoned, during the frequent process of modification of taxonomic hierarchies with upgrading or downgrading of taxa without modification of their intension, extension and nomen.

(c) A system of *progressive additivity of onomatophores* can be thought of in order to allow nomenclature to be monosemic without having to tie nomina to ranks. Within a nominal-series, the onomatophores of two lower taxa could be added to make the onomatophore of an upper, more inclusive, taxon. This system could be appropriate if all the nomenclature of organisms was created at once, simultaneously, but it is not so within a nomenclatural system that has been incremented progressively over decades, for reasons explained in details elsewhere (Dubois 2006c: 21–24).

(d) The system of *indissoluble set of onomatophores with inclusive ostension* is a monosemic system without eponymy relying on special onomatophores composed of one or several specimens or taxa indissolubly linked together. Unlike in the *Code* for genus-series and species-series nomina, such onomatophores cannot be modified by restriction to one specimen or taxon among several originally included. In this system, a nomen applies to the *least inclusive* taxon including entirely its onomatophore, and the valid nomen of any taxon is the oldest one meeting this requirement. This system is one possible solution for obtaining a monosemic nomenclatural system, i.e., for getting rid of eponymy, but it allows mostly to name the least inclusive taxa in a taxonomy. When additional taxa are discovered that should be included in the taxon according to its original intension, the original nomen cannot be kept for the more inclusive taxon including them (Dubois 2006c: 25). To solve this problem, a last system of allocation of nomina to taxa has to be considered.

(e) The system of *indissoluble set of onomatophores and onomatostases with bidirectional ostension*, first proposed by Dubois (2004a, 2005b, 2005e, 2006a) for class-series nomenclature, is a monosemic system without eponymy relying on indissoluble onomatophores and *onomatostases*. The latter are specimens or taxa originally and expressly excluded from the taxon for which the nomen was coined. In this system, a nomen applies, within an ergotaxonomy, to the most inclusive taxon including all its onomatophore and excluding all its onomatostase, and the valid nomen of any taxon is the oldest one meeting this requirement.

When proposing rules for class-series nomenclature in zoology, Dubois (2004a, 2005b,e, 2006a) suggested to use a combination of the two latter nomenclatural systems for the establishment of the valid nomen of a taxon in the nominal-series. This suggestion was based on the idea that, as higher zoological nomenclature has never been regulated by the *Code* until now, implementing such rules nowadays should be done carefully, without disrupting the existing nomina for higher zoological taxa. Another proposal of rules for these nomina (Alonso-Zarazaga 2005) fails to solve this problem, as it would result in considerable changes in higher zoological nomenclature. This latter proposal was based on the idea that the Principle of Coordination should be extended to this nomenclature. As this rule has never been in force in the class-series and as it results in a partly polysemic nomenclatural system which poses problems of ambiguity, it would be a bad idea to follow this suggestion. The rules proposed by Dubois allow to keep the long-established nomina for higher taxa and to respect monosemy for the nomina of such taxa, and should be preferred on that account (Dubois 2006c).

The question is now: should this example be followed further by shifting the whole of zoological nomenclature to a monosemic system?

Should zoological nomenclature shift to a fully monosemic system?

We now have the elements to discuss the possible shift of zoological nomenclature from a partly polysemic system to a fully monosemic one. As we have seen, this would require only to change a single rule of the *Code*, i.e., to replace the Principle of Coordination by one of the monosemic nomenclatural systems discussed above. The two most appropriate ones for this change are *indissoluble onomatophore with inclusive ostension* and *indissoluble onomatophore and onomatostase with bidirectional ostension*. This shift would pose no theoretical problem, but several practical ones.

It would require the creation of thousands or probably hundreds of thousands of nomina, to replace the epinyms and hyponyms made invalid by this change of rule. For example, if a genus *Rana* contains several subgenera, in the current system one of them has to bear also the nomen *Rana*, which should then be replaced by another nomen, in order for all subgenera to have nomina distinct from that of the genus, as in the nomenclature of Hillis & Wilcox (2005), which is invalid under the *Code* (Dubois 2006b,d, 2007b).

Let us consider a superfamily including taxa of four family-series ranks from family to subtribe as evoked above. In the case of an unbalanced taxonomy with 9 taxa, where each family-series taxon contains only two subordinate taxa, one of which does not contain subordinate family-series taxa, the *Code* requires 5 nomina to name these 9 taxa, but removing eponymy would require the creation of 4 nomina (tables 1–2), therefore 80 % additional nomina . With the same hierarchy but a balanced, completely resolved, taxonomy (without polytomies), 16 nomina instead of 31 are needed under the *Code* for 31 taxa (tables 3–4), so that transfer to a monosemic nomenclature would require 94 % additional nomina. The number of additional nomina needed increases with the resolution of the tree, and then with the number of ranks. In all cases however, the nomenclature of all the taxa of a *coordinate nomenclatural set* in any given nominal-series requires far less nomina in a partially polysemic nomenclatural system. In the case of a fully resolved tree and of recognition of a new taxon at each dichotomy, transfer from a polysemic nomenclature with n nomina to a monosemic one requires (n–1) additional nomina, or, to put the same thing differently, the complete nomenclature of n taxa under the Principle of Coordination requires (n+1)/2 nomina, against n nomina under any monosemic system.

The total number of eponyms used in the current zoological nomenclature is unknown, but certainly high, especially in the family-series, as subgenera and subspecies, which are frequently used in some zoological groups, especially the best studied ones, are almost ignored in other groups, mostly for reasons of tradition. Therefore, although removing polysemy from the *Code* would be theoretically possible, it would be a very heavy, time- and money-consuming endeavour. Furthermore, this would be a very specialized work, which should be carried out by well-trained, professional taxonomists. Would this be a good choice during the century of extinctions, when taxonomists are not numerous enough and lack professional positions, student grants, funding for field and laboratory work, and for collection management? Asking the question provides the answer: although in an ideal world it might be considered desirable to shift from a partially polysemic nomenclatural system to a fully monosemic one, this is not an urgent matter today and this should be postponed until we have highly improved our knowledge of the living species of the planet (see the introductory preliminary statement above). Besides, transferring millions of data from a system to another would be bound to entail loss of information, and likely to introduce errors. This would be inappropriate at all times. The implementation of such a shift would appear to be justified once we have a much more exhaustive inventory of the living organisms of the planet, a complete database with all necessary information on all zoological taxa and nomina, and a fully automatic method for the replacement of invalid epinyms and hyponyms by existing nomina or by new nomina created especially for this purpose. Until this is the case, it is necessary to keep the current nomenclatural system in zoology, within the frame of which millions of nomina and nomenclatural acts have already been stored.

TABLE 1. The family-series nomenclature of a hypothetical unbalanced zoological taxonomy, according to the *Code*'s nomenclatural system based on the Principle of Coordination. Nomina of taxa are symbolized by letters and numbers: those in **bold** indicate eponyms, i.e., nomina that are used as valid for several taxa at different ranks. In this system, 5 nomina are sufficient to designate the 9 taxa recognized.

- Superfamilia **A01**
  - Familia **A01**
    - Subfamilia **A01**
    - Subfamilia A03
      - Tribus **A01**
      - Tribus A04
        - Subtribus **A01**
        - Subtribus A05
  - Familia A02

TABLE 2. The family-series nomenclature of the same hypothetical unbalanced zoological taxonomy as in table 1, according to a hypothetical nomenclatural system based on bidirectional ostension (see Dubois 2007a). Nomina of taxa are symbolized by letters and numbers. This system does not recognize eponyms. The four nomina in *italics* are nomina the creation of which is required by the suppression of eponymy. All other taxa keep the same nomina as in Table 1. In this system, 9 nomina are required to designate the 9 taxa recognized.

- Superfamilia A01
  - Familia *B01*
    - Subfamilia *B02*
    - Subfamilia A03
      - Tribus *B03*
      - Tribus A04
        - Subtribus *B04*
        - Subtribus A05
  - Familia A02

TABLE 3. The family-series nomenclature of a hypothetical fully balanced zoological taxonomy, according to the *Code*'s nomenclatural system based on the Principle of Coordination. Nomina of taxa are symbolized by letters and numbers: those in **bold** indicate eponyms, i.e., nomina that are used as valid for several taxa at different ranks. In this system, 16 nomina are sufficient to designate the 31 taxa recognized.

- Superfamilia **A01**
  - Familia **A01**
    - Subfamilia **A01**
      - Tribus **A01**
        - Subtribus **A01**
        - Subtribus A09
      - Tribus **A05**
        - Subtribus **A05**
        - Subtribus A10
    - Subfamilia **A03**
      - Tribus **A03**
        - Subtribus **A03**
        - Subtribus A11
      - Tribus **A06**
        - Subtribus **A06**
        - Subtribus A12
  - Familia **A02**

- Subfamilia **A02**
  - Tribus **A02**
    - Subtribus **A02**
    - Subtribus A13
  - Tribus **A07**
    - Subtribus **A07**
    - Subtribus A14
- Subfamilia **A04**
  - Tribus **A04**
    - Subtribus **A04**
    - Subtribus A15
  - Tribus **A08**
    - Subtribus **A08**
    - Subtribus A16

TABLE 4. The family-series nomenclature of a hypothetical fully balanced zoological taxonomy, according to a hypothetical nomenclatural system based on bidirectional ostension (see Dubois 2007a). Nomina of taxa are symbolized by letters and numbers. This system does not recognize eponyms. The fifteen nomina in *italics* are nomina the creation of which is required by the suppression of eponymy. In this system, 31 nomina are required to designate the 31 taxa recognized.

- Superfamilia A01
  - Familia *B01*
    - Subfamilia *B02*
      - Tribus *B04*
        - Subtribus *B08*
        - Subtribus A09
      - Tribus A05
        - Subtribus *B09*
        - Subtribus A10
    - Subfamilia A03
      - Tribus *B05*
        - Subtribus *B10*
        - Subtribus A11
      - Tribus A06
        - Subtribus *B11*
        - Subtribus A12
  - Familia A02
    - Subfamilia *B03*
      - Tribus *B06*
        - Subtribus *B12*
        - Subtribus A13
      - Tribus A07
        - Subtribus *B13*
        - Subtribus A14
    - Subfamilia A04
      - Tribus *B07*
        - Subtribus *B14*
        - Subtribus A15
      - Tribus A08
        - Subtribus *B15*
        - Subtribus A16

The problem of redundancy

Another criticism that has been raised against the use of ranks in zoological nomenclature is the problem of redundancy of taxa. In order to express nomenclaturally a taxonomy based on a cladistic hypothesis, two kinds of taxa require to be named (Dubois 2007a: 48–49): (1) all taxa including subordinate taxa referred to the same coordinate nomenclatural set; (2) all taxa which are parordinate to the latter, even if they do not include subordinate taxa of the same coordinate nomenclatural set. If the latter are not named, because this would be "redundant" for example with the nomen of the taxon of the next lower nominal-series, this results in having sister-taxa with different ranks, a nomenclature which does not convey any information about the hierarchical structure of taxonomy and therefore on cladistic relationships. Such *pseudoranked* nomenclatures (Dubois 2007a: 34) exist, and will be discussed below.

Following a long tradition in zootaxonomy, Dubois (2006a: 217, 2007a: 50) and Kuntner & Agnarsson (2006) further suggested that, in all ergotaxonomies, a third kind of taxa should always be recognized and named, namely taxa belonging to the seven primary key ranks regnum, phylum, classis, ordo, familia, genus and species. The purpose of this proposal is that, once they have been studied, at least superficially, all organisms of the earth should be referred, sometimes provisionally, to a taxon of these seven ranks. Unlike in the two preceding cases, this is not meant at always expressing cladistic relationships (although in most cases it also plays this role). For example, a taxon T may be referred to the rank classis to express its parordination to another classis including many subordinate taxa at various ranks, but the class T may include only one species: in this case, naming also an order, a family and a genus for this species is indeed redundant, as it does not carry additional cladistic information. However, naming these taxa is important to comply with an important function of biological classifications, besides its "explanatory" one about evolution: that of providing a universal system of storage and retrieval of information (Cracraft 1974; Mayr 1982, 1997; Ashlock 1984; Benton 2000; Dubois 2005c). The existence of these seven "compulsory" ranks for all organisms would greatly facilitate the building of taxonomic and nomenclatural databases and the search for this information in the latter. In a database like the *Zoological Record*, a single hierarchy using only, but always, these seven ranks can be used for all organisms of the earth. In all cases of redundancy of nomina within a nominal-series (e.g., a single order in a class), Dubois (2004a, 2006a: 203) proposed to implement a modified Principle of Coordination, using the same eponym for both taxa, although in this case the hyponym has no getonym: this avoids the useless creation of a new nomen for the redundant subordinate taxon.

Dubois (2006a) provided a review of all ranks that have been used by zoologists in their taxonomic hierarchies in the past, and proposed a standardisation of this system for the whole of zoology, with 209 ranks including 9 primary key ranks (e.g., family), 10 secondary key ranks (e.g., phalanx) and 10 subsidiary ranks (e.g., subfamily) for all key ranks. This system should be largely sufficient to cover all the needs for future taxonomies. As a matter of fact, no real ergotaxonomy recognizes taxonomically all the nodes of a tree, as this would result in much too cumbersome and useless classifications. Despite this high number of potential ranks, which are sometimes quite useful as illustrated below in table 10, most ergotaxonomies only use about 10 to 20 ranks (see e.g. Minelli 1991). Among these, the 7 primary key ranks listed above should always be present. In order not to upset the tradition, the best known nomina, such as **AMPHIBIA** or **AVES**, should as far as possible be allocated to primary, not secondary or subsidiary ranks (Dubois 2006a: 224, 2007a: 50). This can usually be realized easily, provided some attention and care are given to this question, as ranks are fully arbitrary and can be fixed by arbitrary decision.

Pseudoranked nomenclatures

Although they have until now been considered by most taxonomists as far less informative than the Linnaean nomenclatural system (as clearly shown by the fact that the overwhelming majority of ergotaxonomies published nowadays follow the latter), unranked nomenclatural systems are theoretically justified and follow an internal logic. This is not the case however of nomenclatures that may be known as *pseudoranked* (Dubois 2007a: 34). Such nomenclatures have been used in recent years without any theoretical justifications by a few authors who remain, so to speak, half-way between Linnaean and unranked nomenclatures: they claim that ranks are useless or harmful, and they use unranked nomina (just designated as “taxa”) for the most comprehensive taxa of their taxonomies, but they still use nomina referred to formal ranks for families, genera and species, and also sometimes for superfamilies, subfamilies and tribes. It is difficult to understand this intermediate attitude, except as a way to avoid seing “their” taxa named validly under the *Code* by others, as has happened already in a few amusing cases (e.g., Pleijel 1999, Muona 2006). Having two strings to their bows will allow such authors to see “their” new nomina survive at any rate in the future, whatever nomenclatural system ultimately wins the “war of nomenclature”.

However, for such nomina to be available and valid under the *Code*, they should follow strictly the rules of the latter, which is not always the case. Several examples of similar situations are available in recent publications dealing with the **AMPHIBIA**. The cases of the nomenclatures of the salamander genus *Eurycea* by Hillis *et al.* (2001) and of the frog genus *Rana* by Hillis & Wilcox (2005) were discussed in detail elsewhere (Dubois 2006b,d, 2007b) and need not be so again here. Let us consider here three other recent works, those of Vieites *et al.* (2007), Frost *et al.* (2006) and Grant *et al.* (2007), which are good illustrations of this problem.

The nomenclature of the salamander family *PLETHODONTIDAE* recently proposed by Vieites *et al.* (2007: Online Supporting Information), shown here in table 5, is not acceptable under the *Code*, for two distinct reasons. First, it includes a rank (supergenus) which is not recognized by these rules. This rank, which would belong in the genus-series, would indeed be useful in zoological nomenclature (Dubois 2006b), but until the *Code* is modified to allow for the recognition of more than two ranks in this nominal-series, this rank should not be used, at least in a formal nomenclature following the *Code*. This nomenclature is also invalid under the *Code* for not respecting the Principle of Coordination. It is not possible to recognize a tribe *SPELERPINI* in the subfamily *HEMIDACTYLIINAE* without also recognizing at least one other tribe, the *HEMIDACTYLIINI*. A third, related, problem with this nomenclature, although this does not by itself make it invalid under the current *Code*, is that it is based on a partially resolved tree with polytomies where parordinate taxa are given different ranks. Let us just consider the four taxa immediately subordinate to the subfamily *HEMIDACTYLIINAE*. The nomenclature proposed fails to convey any cladistic information, as it is impossible to know, from the nomina of these taxa alone (i.e., without seeing a tree or a complete taxonomy of this family), that the genera *Batrachoseps* and *Hemidactylium*, the “supergenus” *Bolitoglossa* and the tribe *SPELERPINI* are the four members of an unresolved polytomy.

TABLE 5. The supraspecific nomenclature of the family *PLETHODONTIDAE* proposed by Vieites *et al.* (2007: Online Supporting Information). This nomenclature is not valid under the *Code*, which does not recognize a rank “supergenus”, and which requires to follow the Principle of Coordination: if a tribe *SPELERPINI* is recognized in the *HEMIDACTYLIINAE*, the *Code* requires to recognize at least one other tribe, the *HEMIDACTYLIINI*. The use of ranks in this nomenclature is non-informative, as parordinate taxa are not afforded the same rank. In the table below, within each rank, taxa are presented in alphabetical order of their nomina, unlike in Vieites *et al.* (2007), who apparently followed a “phylogenetic” order (although these hypothesized cladistic relationships are not expressed in their nomenclature).

- Familia *PLETHODONTIDAE*
  - Subfamilia *HEMIDACTYLIINAE*
    - Genus *Batrachoseps*
    - Supergenus *Bolitoglossa*
      - Genus *Bolitoglossa*
      - Genus *Bradytriton*

- Genus *Chiropterotriton*
- Genus *Cryptotriton*
- Genus *Dendrotriton*
- Genus *Ixalotriton*
- Genus *Lineatriton*
- Genus *Nototriton*
- Genus *Oedipina*
- Genus *Parvimolge*
- Genus *Pseudoeurycea*
- Genus *Thorius*

Genus *Hemidactylium*
Tribus *SPELERPINI*
    Genus *Eurycea*
    Genus *Gyrinophilus*
    Genus *Pseudotriton*
    Genus *Stereochilus*
Subfamilia *PLETHODONTINAE*
    Genus *Aneides*
    Supergenus *Desmognathus*
        Genus *Desmognathus*
        Genus *Phaeognathus*
    Genus *Ensatina*
    Supergenus *Hydromantes*
        Genus *Atylodes*
        Genus *Hydromantes*
        Genus *Speleomantes*
    Genus *Karsenia*
    Genus *Plethodon*

TABLE 6. A possible supraspecific nomenclature of the family *PLETHODONTIDAE* following the taxonomy proposed by Vieites *et al.* (2007: Online Supporting Information) but respecting the *Code* and affording the same rank to parordinate taxa. Within each rank, taxa are presented in alphabetical order of their nomina. Nomina are followed by the date of their nomenclatural creation but not their authors, for reasons explained by Dubois (2008b). Tribal nomina between quotation marks are informal nomina without availability in zoological nomenclature. They are mentioned here just to show what the nomenclature of this family could be if the erection of these tribes was judged useful by specialists of this group. If it were the case, these nomina should be formally published with a diagnosis and a statement of intention of creating a new nomen, as, for the time being, no available nomina exist to name these tribes.

Familia *PLETHODONTIDAE* 1850
    Subfamilia *HEMIDACTYLIINAE* 1856
        Tribus "*BATRACHOSEPINI*"
            Genus *Batrachoseps* 1841
        Tribus *BOLITOGLOSSINI* 1856
            Genus *Bolitoglossa* 1854
            Genus *Bradytriton* 1983
            Genus *Chiropterotriton* 1944
            Genus *Cryptotriton* 2000
            Genus *Dendrotriton* 1983
            Genus *Ixalotriton* 1989
            Genus *Lineatriton* 1950
            Genus *Nototriton* 1983
            Genus *Oedipina* 1868
            Genus *Parvimolge* 1944

    - Genus *Pseudoeurycea* 1944
    - Genus *Thorius* 1869
  - Tribus *HEMIDACTYLIINI* 1856
    - Genus *Hemidactylium* 1838
  - Tribus *SPELERPINI* 1859
    - Genus *Eurycea* 1822
    - Genus *Gyrinophilus* 1869
    - Genus *Pseudotriton* 1838
    - Genus *Stereochilus* 1869
- Subfamilia *PLETHODONTINAE* 1850
  - Tribus "*ANEIDINI*"
    - Genus *Aneides* 1849
  - Tribus *DESMOGNATHINI* 1850
    - Genus *Desmognathus* 1850
    - Genus *Phaeognathus* 1961
  - Tribus *ENSATININI* 1850
    - Genus *Ensatina* 1850
  - Tribus "*HYDROMANTINI*"
    - Genus *Atylodes* 1868
    - Genus *Hydromantes* 1848
    - Genus *Speleomantes* 1984
  - Tribus "*KARSENIINI*"
    - Genus *Karsenia* 2005
  - Tribus *PLETHODONTINI* 1850
    - Genus *Plethodon* 1838

TABLE 7. A possible supraspecific nomenclature of the family *PLETHODONTIDAE* following the taxonomy proposed by Vieites *et al.* (2007: Online Supporting Information) but respecting the *Code* and affording the same rank to parordinate taxa. Within each rank, taxa are presented in alphabetical order of their nomina. Nomina are followed by the date of their nomenclatural creation but not their authors, for reasons explained by Dubois (2008b). In this taxonomy, no tribes are recognized in the subfamily *PLETHODONTINAE*, which results in recognizing subgenera in the genera *Desmognathus* and *Hydromantes*.

- Familia *PLETHODONTIDAE* 1850
  - Subfamilia *HEMIDACTYLIINAE* 1856
    - Tribus "*BATRACHOSEPINI*"
      - Genus *Batrachoseps* 1841
    - Tribus *BOLITOGLOSSINI* 1856
      - Genus *Bolitoglossa* 1854
      - Genus *Bradytriton* 1983
      - Genus *Chiropterotriton* 1944
      - Genus *Cryptotriton* 2000
      - Genus *Dendrotriton* 1983
      - Genus *Ixalotriton* 1989
      - Genus *Lineatriton* 1950
      - Genus *Nototriton* 1983
      - Genus *Oedipina* 1868
      - Genus *Parvimolge* 1944
      - Genus *Pseudoeurycea* 1944
      - Genus *Thorius* 1869
    - Tribus *HEMIDACTYLIINI* 1856
      - Genus *Hemidactylium* 1838
    - Tribus *SPELERPINI* 1859

      - Genus *Eurycea* 1822
      - Genus *Gyrinophilus* 1869
      - Genus *Pseudotriton* 1838
      - Genus *Stereochilus* 1869
  - Subfamilia *Plethodontinae* 1850
    - Genus *Aneides* 1849
    - Genus *Desmognathus* 1850
      - Subgenus *Desmognathus* 1850
      - Subgenus *Phaeognathus* 1961
    - Genus *Ensatina* 1850
    - Genus *Hydromantes* 1848
      - Subgenus *Atylodes* 1868
      - Subgenus *Hydromantes* 1848
      - Subgenus *Speleomantes* 1984
    - Genus *Karsenia* 2005
    - Genus *Plethodon* 1838

TABLE 8. A possible supraspecific nomenclature of the family *Plethodontidae* following the tree of figure 1 of Vieites *et al.* (2007: Online Supporting Information). This nomenclature respects the *Code* and affords the same rank to parordinate taxa. Within each rank, taxa are presented in alphabetical order of their nomina. Nomina are followed by the date of their nomenclatural creation but not their authors, for reasons explained by Dubois (2008b). Family-series nomina between quotation marks are informal nomina without availability in zoological nomenclature. They are mentioned here just to show what the nomenclature of this family could be if the erection of these taxa was judged useful by specialists of this group. If it were the case, these nomina should be formally published with a diagnosis and a statement of intention of creating a new nomen, as, for the time being, no available nomina exist to name these taxa. For the subtribus "*Pseudotrinonina*", the nomen *Mycetoglossina* Bonaparte, 1850 would have been available, if it had not been "suppressed" (invalidated) by ICZN (Anonymous 1997). The family-series ranks below subtribe and their endings follow the proposals of Dubois (2006a). If the genera *Atylodes* and *Speleomantes* were to be downgraded to the rank of subgenera of a single genus, the latter ought to bear the nomen *Speleomantes* for the reasons given by Crochet (2007).

- Familia *Plethodontidae* 1850
  - Subfamilia *Hemidactyliinae* 1856
    - Tribus *Hemidactyliini* 1856
      - Subtribus *Bolitoglossina* 1856
        - Infratribus "*Batrachosepita*"
          - Genus *Batrachoseps* 1841
        - Infratribus *Bolitoglossita* 1856
          - Genus *Bolitoglossa* 1854
          - Genus *Bradytriton* 1983
          - Genus *Chiropterotriton* 1944
          - Genus *Cryptotriton* 2000
          - Genus *Dendrotriton* 1983
          - Genus *Ixalotriton* 1989
          - Genus *Lineatriton* 1950
          - Genus *Nototriton* 1983
          - Genus *Oedipina* 1868
          - Genus *Parvimolge* 1944
          - Genus *Pseudoerycea* 1944
          - Genus *Thorius* 1869
      - Subtribus *Hemidactyliina* 1856
        - Genus *Hemidactylium* 1838
    - Tribus *Spelerpini* 1859
      - Subtribus "*Pseudotrinonina*"

      - Infratribus “*GYRINOPHILITA*”
        - Genus *Gyrinophilus* 1869
      - Infratribus “*PSEUDOTRITONITA*”
        - Genus *Pseudotriton* 1838
        - Genus *Stereochilus* 1869
    - Subtribus *SPELERPINA* 1859
      - Genus *Eurycea* 1822
- Subfamilia *PLETHODONTINAE* 1850
  - Tribus “*HYDROMANTINI*”
    - Subtribus “*HYDROMANTINA*”
      - Genus *Hydromantes* 1848
      - Genus *Speleomantes* 1984
        - Subgenus *Atylodes* 1868
        - Subgenus *Speleomantes* 1984
    - Subtribus “*KARSENIINA*”
      - Genus *Karsenia* 2005
  - Tribus *PLETHODONTINI* 1850
    - Subtribus *DESMOGNATHINA* 1850
      - Infratribus *DESMOGNATHITA* 1850
        - Clanus “*ANEIDITOI*”
          - Genus *Aneides* 1849
        - Clanus *DESMOGNATHITOI* 1850
          - Genus *Desmognathus* 1850
          - Genus *Phaeognathus* 1961
      - Infratribus *ENSATINITA* 1850
        - Genus *Ensatina* 1850
    - Subtribus *PLETHODONTINA* 1850
      - Genus *Plethodon* 1838

There would be not one, but several ways to reconcile the ergotaxonomy adopted by Vieites *et al.* (2007) with a nomenclature respecting the *Code* and providing cladistic information through the use of ranks. This uncertainty is not problematic, as it is due to the fact that ranks are just arbitrary tools that carry by themselves no information on the taxa, their characters, their divergence or other non-cladistic data, but only information on the hierarchical structure of the ergotaxonomy and hence on the cladistic hypothesis adopted. Tables 6 and 7 present two possible nomenclatures for this family following the ergotaxonomy of Vieites *et al.* (2007). In the nomenclature of table 6, all genera afforded the rank genus by these authors are maintained at this rank, but then this compels to recognize several additional tribes to respect the taxonomic structure adopted and the rank equivalence between parordinates. Table 7 presents the same taxonomy with a different nomenclature, where the first rank subordinate to subfamily in the *PLETHODONTINAE* is genus, which obliges to downgrade five “genera” to the rank subgenus. Both taxonomies of tables 6 and 7 are *Code*-compliant and carry exactly the same cladistic information.

Interestingly, in their presentation of their taxonomy of the subfamily *HEMIDACTYLIINAE*, Vieites *et al.* (2007) did not follow the alphabetical order of the nomina of the taxa. They presented the taxa under the following succession: *Hemidactylium*, *Batrachoseps*, *Bolitoglossa* and *SPELERPINI*. Although they did not explain their reason for doing so, this is probably because they meant to express cladistic relationships through this succession. This is actually a common way of listing taxa in zoological monographs, revisions, etc. As a matter of fact, in the tree of their figure 1, their *SPELERPINI* appear as the sister-group of a group including the other three taxa, and, among the latter, their genus *Hemidactylium* appears as the sister-group of the group composed of their genus *Batrachoseps* and their supergenus *Bolitoglossa*. However, this expression of hypothesized cladistic relationships through the order of presentation of taxa in a list or table is highly confusing and ambiguous (it is not even quite clear if this was indeed the intention of the authors!). The only way to express clearly a cladistic hypothesis into a taxonomy using formal nomenclatural ranks like genus, tribe or subfamily is through the use of different ranks for superordinate and subordinate taxa, and the same rank for parordinate taxa.

Any taxonomist who uses a tree as a working hypothesis for building a taxonomy has two possibilities. The first one is to accept the whole tree as valid and to express all nodes through recognizing a new taxon for each of them, following a hierarchy of nomenclatural ranks for successive branchings. The other possibility is to recognize taxonomically and nomenclaturally only some nodes, leaving some unresolved polytomies. This may be due to doubts on the validity of the tree (e.g., because of medium or low values of Bayesian, bootstrap or other indices), or to a decision to use only a few ranks in order to have a simpler and "lighter" taxonomy and nomenclature. This choice is entirely in the hands of the author, and does not depend in any way on nomenclatural rules. However, once a choice has been made between these possibilities, the resulting taxonomy and nomenclature should not be ambiguous.

In their taxonomy of the *PLETHODONTIDAE*, Vieites *et al.* (2007) decided not to recognize taxonomically all the nodes of their tree, which is their full right. But, then, this taxonomy leaves some polytomies unresolved and this should be reflected in the nomenclature, as is the case e.g. in the figures 6 and 7 here. Another possibility would have been to recognize more taxa and ranks, in order to follow more closely the cladistic hypothesis of their figure 1. One possible way of doing so is shown here in figure 8. This is not meant to support this ergotaxonomy here, but to show that it is always possible to build a meaningful taxonomy and nomenclature on the basis of a given cladistic hypothesis. Many more unknown species of *PLETHODONTIDAE* remain to be discovered, and much more information remains to be obtained from the salamanders of this group, so that the taxonomy of the latter we have in 2008 is certainly not the "final word" on this question and further changes may be expected in the future: this is the normal way in which taxonomy evolves and there is nothing worrying about this. But, at any time during the taxonomic history of a group, taxonomists should care about producing well built and clear ergotaxonomies, with rigorous nomenclatures, providing unambiguous information on the cladistic hypotheses they accept as valid at this given stage of research.

A last interesting comment regarding this nomenclature is that, although it does not follow a strictly ranked taxonomy, it uses ranks, and in a way that suggests that ranks are credited with a "value" and a "meaning" by themselves—although the fact that this belief is unwarranted is precisely the main reason given by most authors for not using ranks! Some genera appear as parordinate to clusters of genera, so that, in this nomenclature, the rank genus does not convey any cladistic message. But then, different nomenclatural ranks are afforded to the two parordinate clusters of genera recognized in the subfamily *HEMIDACTYLIINAE*: "supergenus" and tribe. No explanation for this discrepancy is provided. As in both cases there is only one rank between subfamily and genus, one would expect to see these two generic clusters afforded the same rank, either "supergenus" or tribe. Why isn't it the case? One possible explanation could be that reinstating the nomen *Atylodes* as a valid generic nomen may have been used as a transitional step in order to provide this nomen with some "usage", making it possible later to synonymize the nomen *Speleomantes* with it and to nullify the nomenclatural correction of Crochet (2007) to the invalid nomenclature of Wake *et al.* (2005). If this is not the case, then the only possible reason that can be thought of to explain this unbalanced nomenclature is that the *SPELERPINI* were considered as "more divergent" from the other genera of the subfamily than the "supergenus *Hydromantes*". This would suggest that ranks, which were not used to convey cladistic information, can convey phenetic information on "the importance of divergence" between taxa, therefore reinstating a belief in ranks having a "meaning" *by themselves*, distinct from that of providing information on the hierarchical structure of the taxonomy and therefore of the tree used as a basis for the latter. Therefore, this pseudoranked nomenclature falls fully into the main criticism raised by several recent authors against the use of ranks: considering the latter somewhat "equivalent" from one group to an other.

In fact, the same tendency can be observed in several other recent pseudoranked nomenclatures. Another good example is provided by the nomenclature used by Frost *et al.* (2006) in their ergotaxonomy of the **AMPHIBIA**. As discussed elsewhere already (Dubois 2007a: 34), this nomenclature is pseudoranked because, in many occasions, different ranks are afforded to parordinate taxa. It is difficult to understand the rationale for allocating ranks to taxa in this nomenclature. At first look, one could think that it a simple matter of progressive incrementation of levels above genus. Taxa could simply be given successive ranks when going upwards in the hierarchy: subfamily, family, superfamily, then various levels of unranked "taxa". This is not the case, however, because the hierarchy of ranks used above genus in this nomenclature is different from one group to another. Although all genera in this ergotaxonomy are referred to families, the ranks subfamily and superfamily are not used consistently: in some cases, genera are directly referred to a family and the latter to a "taxon", whereas in other cases the ranks subfamily or superfamily, or both, are interpolated. Here again, the only possible explanation seems to be that subfamilies are viewed as "less divergent" than families, and superfamilies "more divergent" than families, hence giving credit to the idea that ranks are meaningful by themselves and somewhat "equivalent" by some mysterious criteria.

Such problems are present throughout the ergotaxonomy of Frost *et al.* (2006). To save space, it will be enough to illustrate them with an extract of this classification, dealing with their "taxon **HYLOIDES**" (table 9). The purpose here, as in the case studied above, is not to challenge or even discuss their taxonomy, but to examine the nomenclature used to express this taxonomy. Table 9 shows clearly that in many cases parordinate taxa are given different ranks, and are often even referred to different nominal-series. In some cases, this nomenclature also does not eliminate completely the so-called "redundant taxa": the families *HEMIPHRACTIDAE* and *THOROPIDAE* and the subfamilies *PELODRYADINAE*, *TELMATOBIINAE* and *ALLOPHRYNINAE* are redundant with their unique genera (other similar cases exist in other parts of their ergotaxonomy that are not examined here). Therefore this argument is not valid to explain why they did not follow a Linnaean ranked taxonomy, at least for all taxa below the class-series. This nomenclature is as poorly informative as that of the preceding example. Here also, there would be many ways of expressing nomenclaturally the cladistic relationships on which the ergotaxonomy is based, and table 10 provides only one of them. In this table, the choice was made, following the example of the ergotaxonomy of **AMPHIBIA** of Dubois (2005d), to use only family-series nomina above the rank genus. This does not mean that, in the end, class-series nomina will not have to be used for the higher nomenclature of **AMPHIBIA**, but time is clearly not ripe for this. Frost *et al.* (2006) ignored some existing class-series nomina and therefore proposed new nomina for taxa that had already nomina (see e.g. Dubois & Ohler 2008). They did not follow rigorous or even clear rules for the allocation of nomina to higher taxa (using sometimes original extension or intension, sometimes "tradition", sometimes opinions and tastes) and therefore had a lax concept of synonymy for class-series nomina, etc. Consequently, several of their nomina would have to be changed if their taxa had to be named in the nominal class-series. Developing this here would be beyond the scope of the present work, and, for more simplicity, only family-series nomina are used in the suprageneric nomenclature of table 10. This nomenclature is fully informative by itself about the cladistic hypothesis retained by Frost *et al.* (2006) to build their taxonomy. It makes use of 12 distinct ranks in the family-series, and even more could be used if needed. Because of the Principle of Coordination, many of these nomina are eponyms and are used under different morphonyms at different ranks. These different spellings allow to distinguish between them and to identify immediately the getonyms by their endings. Such a nomenclature is much more informative about the status of nomina and taxa than that of the pseudoranked nomenclature used by Frost *et al.* (2006).

TABLE 9. The suprageneric nomenclature of the "taxon **HYLOIDES**" proposed by Frost *et al.* (2006). Note that, according to the nomenclatural rules proposed by Dubois (2004a, 2005b,e, 2006a,c), the nomen used for the latter taxon is invalid, being a junior homonym of several other class-series nomina (see Kuhn 1967); the same is true of the nomen of their "taxon **RANOIDES**". In the table below, within each rank, taxa are presented in alphabetical order of their nomina, unlike in Frost *et al.* (2006), who apparently followed a "phylogenetic" order (not expressed in their nomenclature). Nomina are followed by the date of their nomenclatural creation but not their authors, for reasons explained by Dubois (2008b). Shortly after publication of this nomenclature, several parts of it were modified by Grant *et al.* (2006).

- Taxon **HYLOIDES** 2006
  - Taxon **NOTOGEANURA** 2006
    - Taxon **AUSTRALOBATRACHIA** 2006
      - Familia *BATRACHOPHRYNIDAE* 1875
        - 3 genera
      - Superfamilia *MYOBATRACHOIDEA* 1850
        - Familia *LIMNODYNASTIDAE* 1971
          - 8 genera
        - Familia *MYOBATRACHIDAE* 1850
          - 13 genera
    - Taxon **NOBLEOBATRACHIA** 2006
      - Familia *HEMIPHRACTIDAE* 1862
        - 1 genus
      - Taxon **MERIDIANURA** 2006
        - Familia *BRACHYCEPHALIDAE* 1858
          - 15 genera
        - Taxon **CLADOPHRYNIA** 2006
          - Familia *CRYPTOBRANCHIDAE* 2006
            - 2 genera
          - Taxon **TINCTANURA** 2006
            - Familia *AMPHIGNATHODONTIDAE* 1882
              - 2 genera
            - Taxon **ATHESPHATANURA** 2006
              - Familia *HYLIDAE* 1815
                - Subfamilia *HYLINAE* 1815
                  - 38 genera
                - Subfamilia *PELODRYADINAE* 1858
                  - 1 genus
                - Subfamilia *PHYLLOMEDUSINAE* 1858
                  - 7 genera
              - Taxon **LEPTODACTYLIFORMES** 2006
                - Taxon **CHTHONOBATRACHIA** 2006
                  - Familia *CERATOPHRYIDAE* 1838
                    - Subfamilia *CERATOPHRYINAE* 1838
                      - 6 genera
                    - Subfamilia *TELMATOBIINAE* 1843
                      - 1 genus
                  - Taxon **HESTICOBATRACHIA** 2006
                    - Taxon **AGASTOROPHRYNIA** 2006
                      - Familia *BUFONIDAE* 1825
                        - 48 genera
                      - Superfamilia *DENDROBATOIDEA* 1850
                        - Familia *DENDROBATIDAE* 1850
                          - 11 genera
                        - Familia *THOROPIDAE* 2006
                          - 1 genus

- Familia *Cycloramphidae* 1850
  - Subfamilia *Cycloramphinae* 1850
    - 11 genera
  - Subfamilia *Hylodinae* 1858
    - 3 genera

- Taxon **Diphyabatrachia** 2006
  - Familia *Centrolenidae* 1951
    - Subfamilia *Allophryninae* 1978
      - 1 genus
    - Subfamilia *Centroleninae* 1951
      - 3 genera
  - Familia *Leptodactylidae* 1838
    - 11 genera

- Familia *Sooglossidae* 1931
  - 2 genera

TABLE 10. A possible suprageneric nomenclature of the "taxon **Hyloides**" of Frost *et al.* (2006) using only family-series nomina and following the rules of the *Code*. In the table below, within each rank, taxa are presented in the same order as in table 9, to facilitate comparisons. Nomina are followed by the date of their nomenclatural creation but not their authors, for reasons explained by Dubois (2008b).

- Anofamilia *Hylaidai* 1815
  - Hyperfamilia *Hylaidia* 1815
    - Epifamilia *Myobatrachoidia* 1850
      - Superfamilia *Batrachophrynoidea* 1875
        - Familia *Batrachophrynidae* 1875
          - 3 genera
      - Superfamilia *Myobatrachoidea* 1850
        - Familia *Limnodynastidae* 1971
          - 8 genera
        - Familia *Myobatrachidae* 1850
          - 13 genera
    - Epifamilia *Hyloidia* 1815
      - Superfamilia *Hemiphractoidea* 1862
        - Familia *Hemiphractidae* 1862
          - 1 genus
      - Superfamilia *Hyloidia* 1815
        - Familia *Brachycephalidae* 1858
          - 15 genera
        - Familia *Hylidae* 1815
          - Subfamilia *Cryptobranchinae* 2006
            - 2 genera
          - Subfamilia *Hylinae* 1815
            - Infrafamilia *Amphignathodontinei* 1882
              - 2 genera
            - Infrafamilia *Hylinei* 1815
              - Tribus *Hylini* 1815
                - Subtribus *Hylina* 1815
                  - 38 genera
                - Subtribus *Pelodryadina* 1858
                  - 1 genus
                - Subtribus *Phyllomedusina* 1858
                  - 7 genera

- Tribus *BUFONINI* 1825
  - Subtribus *BUFONINA* 1825
    - Infratribus *CERATOPHRYITA* 1838
      - Clanus *CERATOPHRYITOI* 1838
        - 6 genera
      - Clanus *TELMATOBIITOI* 1843
        - 1 genus
    - Infratribus *BUFONITA* 1825
      - Clanus *BUFONITOI* 1825
        - Subclanus *BUFONILOI* 1825
          - 48 genera
        - Subclanus *DENDROBATILOI* 1850
          - Infraclanus *DENDROBATISOI* 1850
            - 11 genera
          - Infraclanus *THOROPISOI* 2006
            - 1 genus
      - Clanus *CYCLORAMPHITOI* 1850
        - Subclanus *CYCLORAMPHILOI* 1850
          - 11 genera
        - Subclanus *HYLODILOI* 1858
          - 3 genera
  - Subtribus *LEPTODACTYLINA* 1838
    - Infratribus *CENTROLENITA* 1951
      - Clanus *ALLOPHRYNITOI* 1978
        - 1 genus
      - Clanus *CENTROLENITOI* 1951
        - 3 genera
    - Infratribus *LEPTODACTYLITA* 1838
      - 11 genera
- Hyperfamilia *SOOGLOSSAIDIA* 1931
  - Familia *SOOGLOSSIDAE* 1931
    - 2 genera

Interestingly, only five months after publication of the work by Frost *et al.* (2006), the same research team published another work (Grant *et al.* 2006) which modified significantly the cladistic hypothesis and taxonomy of the "taxon **ATHESPHATANURA**" defined in the first work and presented here in table 9. As a result, a class-series nomen published on 15 March 2006 (**DIPHYABATRACHIA**) was invalidated on 15 August 2006 by the same authors, and several new class-series nomina were created. No doubt further changes can be expected in this ergotaxonomy in the coming years, as more species and genes, from well-identified vouchers, are analysed, as more morphological and other characters are taken into account, etc. (see e.g. Wiens 2007). This strongly supports the statement above that naming all these class-series taxa in these works was largely premature. Using only family-series nomina, like in table 10 here and in Dubois (2005d), provides a temporary, but non-ambiguous, nomenclature, during a period of intense research on the cladistic relationships among **AMPHIBIA**. Family-series nomina follow the Principle of Coordination, so they need far fewer nomina than class-series nomina to express the same relationships: whereas the suprageneric nomenclature of table 9 uses 34 nomina, that of table 10 uses only 21 distinct nomina, including 7 eponyms, for exactly the same taxonomy, hence sparing 38 % of the nomina. In the future, whenever this ergotaxonomy is changed, epinyms and hyponyms can easily be abandoned, without having to store these morphonyms in synonymies and online nomenclatural databases. In contrast, all the nomina created by Frost *et al.* (2006), some of which were considered valid by their own authors only for five months, will have to be stored permanently in such databases. When the present period of intense research is over, and when cladistic relationships among **AMPHIBIA** are more consensual and stable, it will be time to provide a class-series nomenclature for these animals.

A nomenclature like that shown in table 10 is complex, as it uses many ranks and similar nomina with different endings. Such a nomenclature is not meant to be used daily by taxonomists, and is of little interest to non-taxonomists. When dealing with a species, in most cases the latter only need to know its generic and specific nomen, and sometimes its familial allocation. In this respect, the nomenclature of table 10 is highly informative, as it places many taxa, previously referred to several families, into a single family *HYLIDAE*. The latter is by no way equivalent to the traditional family *HYLIDAE* in the batrachological literature, but is a very large group that encompasses several of the former families of anurans. For examples, the two families *HYLIDAE* and *BUFONIDAE* had been recognized as taxa of the same rank in all amphibian classifications since Gray (1825), which suggests that they had been considered more or less "equivalent" in phenetic terms, but, according to Frost *et al.*'s (2006) taxonomy, these two taxa now appear in table 10 as just two subsets at different ranks (tribe *HYLINI* and subclan *BUFONILOI*) of a much larger unit, along with various other groups. Keeping the same rank family for both, as done by Frost *et al.* (2006) themselves (table 9), obscures this message. Of course, the rank family in the nomenclature of table 10 is fully arbitrary, and this rank could well be moved upwards or downwards in the scale, but then all other ranks would have to follow and the relations of subordination and parordination between taxa would remain the same. At any rate, if the cladistic relationships and the taxonomy presented by Frost *et al.* (2006) are accepted as valid, keeping the traditional *HYLIDAE* and the *BUFONIDAE* as two families is as misleading and illogical as keeping the **REPTILIA** and the **AVES** as two classes: in any taxonomy supposed to reflect cladistic relationships, these latter two taxa can be retained, but at different ranks, e.g., classis for the former and phalanx for the latter (Dubois 2006a: 193).

Discussion: from phylogenetic taxonomy to hierarchical nomenclature

Reasons have been given above and elsewhere (Knox 1998; Dubois 2005c, 2007a) for preferring a hierarchical taxonomic and nomenclatural system to a non-hierarchical one. Nominal-series and ranks provide an efficient way to express nomenclaturally the hierarchical organisation of the taxonomy. In Linnaean nomenclature, the way nomenclature is used to express the taxonomy is regulated by precise and stringent rules, which leave no space for "interpretations", "opinions" or "tastes". These rules require to follow three steps, the three "storeys of the nomenclatural house" (Dubois 2005a–c,e), to establish the valid nomen of a taxon under a given ergotaxonomy. *Availability* of nomina under the *Code* requires following some simple but stringent rules. *Allocation* of a nomen to a taxon is made by (1) its *onomatophore* in all cases and (2) the *Principle of Coordination* in all cases where the ergotaxonomy uses more than one rank in the *coordinate nomenclatural set* considered. Finally, the *validity* of a nomen to designate a given taxon is usually established automatically by the Principle of Priority, and only in a few cases by recourse to "usage" or by a vote of ICZN.

Once again, nomenclature is not taxonomy. Whereas taxonomy defines taxa, nomenclature does not: it only allocates and validates nomina for the designation of a given taxon in a given ergotaxonomy. The same nomen can have different meanings in different taxonomies (*astatonymy*) or at different ranks in the same taxonomy (*eponymy*). To understand a cladistic hypothesis from a taxonomy, much more important than the nomina themselves are the relations between nomina as shown by the ranks: the relations of eponymy and getonymy are highly informative regarding phylogeny.

An ergotaxonomy without a nomenclature cannot be communicated easily, and a nomenclature makes sense only by reference to an ergotaxonomy. As different taxonomic paradigms are used by different authors, the relation between an ergotaxonomy and a nomenclature has to rely on objective criteria (the onomatophores) and on criteria that rely on the internal structure of the nomenclatural hierarchy (relations of eponymy, getonymy and telonymy).

The combination of a taxonomic paradigm with the partly polysemic ostensional system of the *Code* provides a hierarchical nomenclature that reflects through its ranks the structure of the tree accepted as taxonomic hypothesis for the group, at a given stage of research. Within a given coordinate nomenclatural set, three kinds of taxa require to be named, the first two in order to express nomenclaturally the cladistic hypothesis used to build the ergotaxonomy, and the third one in order to have a fully efficient nomenclature for the storage and retrieval of taxonomic data: (1) all taxa including subordinate taxa referred to the same coordinate nomenclatural set; (2) all taxa which are parordinate to the latter, even if they do not include subordinate taxa of the same coordinate nomenclatural set; (3) if they were not named already for the first two reasons, all taxa that correspond to one or several of the seven primary key ranks regnum, phylum, classis, ordo, familia, genus and species. These three different reasons for naming taxa in a nomenclature devised in order to reflect a cladistic hypothesis are illustrated in figure 1 in Dubois (2007a: 49).

The taxonomic information and concepts on which an ergotaxonomy is based can be made clear by definitions that are given of the *taxa*, not of the *nomina*. Under a *phenetic* taxonomic paradigm, a taxon may be defined by a *diagnosis*, whereas a nomen is allocated to this taxon using its onomatophore and the Principle of Coordination. Under a *cladistic* paradigm, a taxon may be defined by a *cladognosis* (i.e., either an *apognosis* or a *caulognosis*, or both), but there is no difference in the way a nomen is allocated to this taxon: this is still through its onomatophore and the Principle of Coordination. The same applies to an ergotaxonomy following an *evolutionary* paradigm, i.e. recognizing homophyletic but not necessarily holophyletic taxa: a taxon may be defined by a combination of cladognosis and diagnosis, but its nomen remains attached to its taxon through onomatophore and coordination. Dubois (2007a: 60–68) provided the example of an ergotaxonomy in which taxa are defined both by cladognoses and diagnoses, whereas nomina are attached to taxa through the rules of the *Code*.

As we have seen, the hierarchical organisation of taxonomic information through nomenclatural ranks is very informative and useful. It is therefore strange that, unlike in botany, nomenclatural rules in zoology only apply to a portion of the nomenclatural hierarchy, from superfamily to subspecies. There exists in fact no theoretical or practical reason for the *Code* to limit the number of potential ranks, either outside the three nominal-series it recognizes, or also within these series (Dubois 2006b). This prevents universal, non-ambiguous and automatic nomenclature for all taxa that may be recognized by a taxonomist, and these limitations put the *Code* in a situation of inferiority relative to alternative nomenclatural proposals such as those, entitled “phylogenetic nomenclatures”, that have been published in the last two decades.

This problem exists at three levels in the nomenclatural hierarchy: (1) above the rank superfamily, nomenclatural rules for the class-series should be integrated into the *Code*; if these rules are well devised, this would be possible without threatening the tradition of use for well-known nomina of higher taxa (Dubois 2004a, 2005b,e, 2006a,c); (2) for lower nomenclature below the rank subspecies, nomenclatural rules for the variety-series should also be integrated into the *Code*, especially for use in phylogeographic analysis and for conservation biology purposes (Dubois 2006b); (3) within the three nominal-series recognized by the *Code*, no limitation should exist in the number of potential ranks that can be used by taxonomists, as is the case currently (above superfamily and below subtribe in the family-series, and above, between and below the two and four ranks recognized by the *Code*, respectively in the genus- and species-series); this would be useful for example to name taxa at several infrageneric and supraspecific ranks (Dubois 2006b,d, 2007b). Additionally, it would be useful to implement in the *Code* a few rules or, at least, recommendations, regarding the different kinds of ranks (primary and secondary key ranks, subsidiary ranks), and also the endings of the nomina at all ranks in the family-series. Detailed suggestions in these respects have been offered (Dubois 2006a).

If ICZN decides to incorporate these new nominal-series and ranks into the *Code*, it will be important to care for implementing rules that respect a basic feature of the current rules of the *Code*: the *nomenclatural founder effect* (Dubois 2005e). This means that the nomenclatural status of a nomen is fixed once and for all in the original publication where this nomen is created. This applies to its author, date, nominal-series of allocation, spelling and, above all, onomatophore. As the latter is the tool that allows objective, automatic, stable and universal allocation of the nomen to a taxon or several taxa (in polysemic nomenclatural systems), it is crucial that it cannot be changed in subsequent publications—even by its original author!

Nomenclatural rules in all nominal-series should be strict in not allowing any change in the onomatophore (or onomatophore and onomatostase in bidirectional ostensional systems) of a nomen, because if this were allowed it would open the Pandora's box of "emendations" and "redefinitions" of nomina, which may be a virtually endless process, as illustrated in the recent literature (e.g., Laurin & Anderson 2004 and references therein; Frost *et al.* 2006; Vidal & Hedges 2005; Martin & Benton 2008). In our special historical period, the century of extinctions, taxonomists have certainly more urgent matters to deal with than permanently "redefining" nomina that have been published one or two centuries ago! Allocation of nomina to taxa should be automatic and not liable to open discussions. Nomina should remain permanently attached to their original onomatophore (and onomatostase if relevant). If a nomen, given the rules of allocation of nomina to taxa, can be used for a taxon recognized in a recent ergotaxonomy, it must be kept. If it does not, a new nomen must be coined for the taxon, but an existing nomen should not be "redefined"!

Two exceptions only should be accepted to the nomenclatural founder effect: (1) whenever the original publication left some ambiguity regarding the onomatophore (e.g., for a species nomen, several "syntypes" belonging to different biological species, or, for a genus nomen, several "originally included species", now referred to different genera, without designation among them of a "type-species"), this ambiguity should be clarified by a *first-reviser action*—which then plays the role of nomenclatural founder effect; (2) whenever, because of the original onomatophore, strict application of the rules results in threatening a *sozonym* (Dubois 2005b–c), i.e., a *genuinely very well-known* nomen, used in many publications *outside the specialized field of systematics*, then ICZN should be entitled to use its Plenary Power to replace the original onomatophore by another one in order to solve this problem. In all other cases, no change in onomatophore should be allowed because otherwise nomenclatural chaos will soon be in order, as is currently the case in higher zoological nomenclature (Dubois 2005c, 2006a).

A final, but important, suggestion, may be offered here regarding the necessary changes that should be brought to the *Code* to avoid this almost bicentennial set of rules to be threatened by alternative nomenclatural systems (Dubois 2008a,e). One of the first step that was taken by the founders of this nomenclatural system (Strickland *et al.* 1843) was to draw a line of delimitation between authors following these rules, and those not following them. This step was the rejection outside the nomenclatural system of all publications that did not follow the Linnaean system of binominal nomenclature for species. This rule is still in vigour today, and allows rejecting as unavailable all nomina proposed under other rules (e.g. Muona 2006). The same could, and should in my opinion, be done today regarding publications that do not follow the Linnaean principle of a hierarchy of ranks. This would require some modifications in the *Code*. First of all, the *Code* should give clear *definitions of nomenclatural ranks*, contrasting them with taxonomic categories, and it should provide guidelines for their use in zoological nomenclature (Dubois 2007a: 54). It should make compulsory: (1) that in any nomenclature following the *Code*, two parordinate taxa be always given the same nomenclatural rank (a genus cannot be parordinate to a family), even if they do not include the same number of lower-ranked taxa; and (2) that a subordinate taxon cannot share the same rank as a taxon subordinate to it, or vice versa (a species cannot be included in a species, a genus in a genus, a family in a family). Such statements should be presented as *stringent rules*, not mere "recommendations".

I even suggest that the *Code* should go still one step further, by stating that any new nomen or nomenclatural act provided in a publication where this basic Linnaean use of nomenclatural hierarchy is not followed should be rejected as unavailable in zoological nomenclature—just like works that do not follow a binominal species nomenclature are rejected as unavailable under the *Code*. Such an act of *self-protection* of the *Code* against alternative nomenclatural systems and, what is perhaps worse, about half-way systems that try to be compatible with several non-miscible nomenclatural systems, would greatly clarify the matter by requiring from all taxonomists to choose which system they decide to follow. Nobody is obliged to adhere to the *Code*, but it should be clear to all readers of a taxonomic work whether its author adheres to it or not, and hence if its nomina or nomenclatural acts should be taken into account by taxonomists adhering to the *Code*. Such a clarification would certainly be very beneficial to the universality and efficiency of zoological nomenclature, at a time when zoologists need to concentrate their efforts on describing the vanishing biodiversity of our planet, not on producing new nomenclatural systems every two weeks.

Conclusion

Nomenclatural rules must be completely disconnected from taxonomic paradigms, i.e., they must be theory-free regarding taxonomy. A given nomenclature makes sense only within the frame of a given ergotaxonomic scheme. The request for "nomenclatural stability" sometimes presented to taxonomists amounts in fact to an anti-scientific plea for "taxonomic stability", i.e., for ignorance (Gaffney 1979; Dominguez & Wheeler 1997; Benton 2000). As long as taxonomy remains a living scientific discipline, taxonomic schemes will be permanently changing. At least, because of the taxonomic impediment, this will be so for many decades yet, and the request for "taxonomic stability" is a weapon against the discipline of taxonomy that should not be accepted by taxonomists or by other biologists (Dubois 1998, 2000a).

The current nomenclatural rules of the *Code* allow taxonomists to do their work well, and particularly to express cladistic hypotheses clearly and unambiguously under the form of hierarchical nomenclatures using onomatophores, nominal-series, ranks and eponymy. These rules should not be drastically modified, but they can and should greatly be improved, especially in order to draw a clear line between users of these rules and non-users.

Working taxonomists are a small and "endangered" group of scientists, and they bear on their shoulders the historical responsibility to try their best to discover as many as possible of the living species of our planet before they get extinct just before our eyes, in the indifference of most of our contemporaries. The main urgency for taxonomy is not to implement brand new nomenclatural rules, but to improve the existing ones in order to facilitate this work of inventory. The use of a clear methodology, with clear concepts expressed under a clear, precise, unambiguous terminology, for taxonomy and nomenclature, are liable to help them in this difficult task.

References

Anonymous [International Commission on Zoological Nomenclature] (1997) Opinion 1873. Hemidactyliini Hallowell, 1856 (Amphibia, Caudata): conserved. *Bulletin of zoological Nomenclature*, 54, 140–141.

Anonymous [International Commission on Zoological Nomenclature] (1999) *International code of zoological nomenclature*. Fourth edition. London, International Trust for zoological Nomenclature, xxix + 306 pp.

Alonso-Zarazaga, M.A. (2005) Nomenclature of higher taxa: a new approach. *Bulletin of zoological Nomenclature*, 62, 189–199.

Ashlock, P.D. (1971) Monophyly and associated terms. *Systematic Zoology*, 20, 63–69.

Ashlock, P.D. (1984) Monophyly: its meaning and importance. *In*: Duncan, T. & Stuessy, T. F. (Eds.), *Cladistics: perspectives on the reconstruction of evolutionary history*, Columbia University Press, New York, pp. 39–46.

Ashlock, P.D. (1985) A revision of the *Bergidea* group: a problem in classification and biogeography (Hemiptera-Heteroptera: Lygaeidae). *Journal of the Kansas entomological Society*, "1984", 57 (4), 675–688.

Avise, J.C. & Johns, G.C. (1999) Proposal for a standardized temporal scheme of biological classification for extant species. *Proceeding of the national Academy of Sciences of the USA*, 96, 7358–7363.

Avise, J. C. & Mitchell, D. (2007) Time to standardize taxonomies. *Systematic Biology*, 56, 130–133.

Benton, M.J. (2000) Stems, nodes, crown clades, and rank-free lists: is Linnaeus dead? *Biological Reviews*, 75, 633–648.

Bernardi, G. (1956) Contribution à l'étude des catégories taxonomiques. I. Avant-propos, nomenclature et définitions. *Bulletin de la Société entomologique de France*, 61, 194–200.

Bernardi, G. (1957) Contribution à l'étude des catégories taxonomiques. II. Les Règles internationales de la nomenclature zoologique et la notation des catégories taxonomiques. *Bulletin de la Société entomologique de France*, 62, 224–250.

Bernardi, G. (1980) Les catégories taxonomiques de la systématique évolutive. *In*: Bocquet, C., Génermont, J. & Lamotte, M. (Eds.), *Les problèmes de l'espèce dans le règne animal*, Vol. 3, *Mémoires de la Société zoologique de France*, 40, 373–425.

Bogart, J. P., Bi, K., Fu, J., Noble, D.W.A. & Niedzwiecki, J. (2007) Unisexual salamanders (genus *Ambystoma*) present a new reproductive mode for eucaryotes. *Genome*, 50, 119–136.

Bour, R. & Dubois, A. (1985) Nomenclature ordinale et familiale des Tortues (Reptilia). *Studia geologica salmanticensia*, vol. especial 1, *Studia palaeocheloniologica*, I, Ediciones Universidad, Salamanca, 77–86.

Bour, R. & Dubois, A. (1986) Nomenclature ordinale et familiale des Tortues (Reptilia). Note complémentaire. *Bulletin mensuel de la Société linnéenne de Lyon*, 55, 87–90.

Cracraft, J. (1974) Phylogenetic models and classification. *Systematic Zoology*, 23, 71–90.

Crochet, P.-A. (2007) Nomenclature of European plethodontid salamanders : *Speleomantes* Dubois, 1984 has precedence over *Atylodes* Gistel, 1868. *Amphibia-Reptilia*, 28, 170–172.

De Queiroz, K. & Gauthier, J. (1990) Phylogeny as a central principle in taxonomy: phylogenetic definitions of taxon names. *Systematic Zoology*, 39, 307–322.

Dominguez, E. & Wheeler, Q. D. (1997) Taxonomic stability is ignorance. *Cladistics*, 13, 367–372.

Doolittle, W.F. (1999) Phylogenetic classification and the universal tree. *Science*, 284, 2124–2128.

Dubois, A. (1981) Hybridation interspécifique et notion du genre en zoologie. *Comptes rendus de l'Académie des Sciences*, (3), 292 (A), 201–203.

Dubois, A. (1986) A propos de l'emploi controversé du terme “monophylétique”: nouvelles propositions. *Bulletin mensuel de la Société linnéenne de Lyon*, 55, 248–254.

Dubois, A. (1987) Miscellanea taxinomica batrachologica (I). *Alytes*, “1986”, 5 (1–2), 7–95.

Dubois, A. (1988a) Some comments on the genus concept in zoology. *Monitore zoologico italiano*, (n. s.), 22, 27–44.

Dubois, A. (1988b) The genus in zoology: a contribution to the theory of evolutionary systematics. *Mémoires du Muséum national d'Histoire naturelle*, (A), 140, 1–123.

Dubois, A. (1991) Nomenclature of parthenogenetic, gynogenetic and “hybridogenetic” vertebrate taxons: new proposals. *Alytes*, 8, 61–74.

Dubois, A. (1997) An evolutionary biologist's view on the science of biology. *Alytes*, 15 (3), 133–136.

Dubois, A. (1998) List of European species of amphibians and reptiles: will we soon be reaching “stability”? *Amphibia-Reptilia*, 19 (1), 1–28.

Dubois, A. (2000a) Nomenclature zoologique: quelques problèmes et besoins actuels. *Bulletin de la Société française de Systématique*, juin 2000, 24, 13–23.

Dubois, A. (2000b) Synonymies and related lists in zoology: general proposals, with examples in herpetology. *Dumerilia*, 4 (2), 33–98.

Dubois, A. (2003) The relationships between taxonomy and conservation biology in the century of extinctions. *Comptes rendus Biologies*, 326 (suppl. 1), S9–S21.

Dubois, A. (2004a) The higher nomenclature of recent amphibians. *Alytes*, 22 (1–2), 1–14.

Dubois, A. (2004b) Developmental pathway, speciation and supraspecific taxonomy in amphibians. 2. Developmental pathway, hybridizability and generic taxonomy. *Alytes*, 22 (1–2), 38–52.

Dubois, A. (2005a) Les règles de la nomenclature familiale en zoologie. *In*: Dubois, A., Poncy, O., Malécot, V. & Léger, N. (Eds.), *Comment nommer les taxons de rang supérieur en zoologie et en botanique?*, *Biosystema*, 23, 17–40.

Dubois, A. (2005b) Propositions pour l'incorporation des nomina de taxons de rang supérieur dans le *Code international de nomenclature zoologique*. *In*: Dubois, A., Poncy, O., Malécot, V. & Léger, N. (Eds.), *Comment nommer les taxons de rang supérieur en zoologie et en botanique?*, *Biosystema*, 23, 73–96.

Dubois, A. (2005c) Proposed Rules for the incorporation of nomina of higher-ranked zoological taxa in the *International Code of Zoological Nomenclature*. 1. Some general questions, concepts and terms of biological nomenclature. *Zoosystema*, 27 (2), 365–426.

Dubois, A. (2005d) *Amphibia Mundi*. 1.1. An ergotaxonomy of recent amphibians. *Alytes*, 23 (1–2), 1–24.

Dubois, A. (2005e) Proposals for the incorporation of nomina of higher-ranked taxa into the Code. *Bulletin of zoological Nomenclature*, 62 (4), 200–209.

Dubois, A. (2006a) Proposed Rules for the incorporation of nomina of higher-ranked zoological taxa in the *International Code of Zoological Nomenclature*. 2. The proposed Rules and their rationale. *Zoosystema*, 28 (1), 165–258.

Dubois, A. (2006b) New proposals for naming lower-ranked taxa within the frame of the *International Code of Zoological Nomenclature*. *Comptes rendus Biologies*, 329 (10), 823–840.

Dubois, A. (2006c) Incorporation of nomina of higher-ranked taxa into the *International Code of Zoological Nomenclature*: some basic questions. *Zootaxa*, 1337, 1–37.

Dubois, A. (2006d) Naming taxa from cladograms: a cautionary tale. *Molecular Phylogenetics & Evolution*, 42, 317–

330.

Dubois, A. (2007a) Phylogeny, taxonomy and nomenclature: the problem of taxonomic categories and of nomenclatural ranks. *Zootaxa*, 1519, 27–68.

Dubois, A. (2007b) Naming taxa from cladograms: some confusions, misleading statements, and necessary clarifications. *Cladistics*, 23, 390–402.

Dubois, A. (2007c) Nomina zoologica linnaeana. *In*: Zhang, Z.-Q. & Shear, W.A. (Eds.), *Linnaeus tercentenary: progress in invertebrate taxonomy*, *Zootaxa*, 1668, 81–106.

Dubois, A. (2008a) Zoological nomenclature: some urgent needs and problems. *In*: *Future trends in taxonomy*, EDIT Symposium, Carvoeiro (Portugal), 21-23 January 2008, 15–18.

Dubois, A. (2008b) A partial but radical solution to the problem of nomenclatural taxonomic inflation and synonymy load. *Biological Journal of the Linnean Society*, 93, 857–863.

Dubois, A. (2008c) A family-series nomen based on a junior synonym is not thereby made invalid: the case of Adamansini Huber, 2007 (Cyprinodontiformes, Nothobranchiidae). *Zootaxa*, 1757, 66–68.

Dubois, A. (2008d) Drôles d’espèces. Hybridation, perturbations de la méiose et spéciation dans le règne animal: quelques points délicats de terminologie, d’éidonomie et de nomenclature. *In*: Prat, D., Raynal-Roques, A. & Roguenant, A. (Eds.), *Peut-on classer le vivant? Linné et la systématique aujourd’hui*, Belin, Paris, 169–202.

Dubois, A. (2008e) Le *Code international de nomenclature zoologique*: présentation, philosophie, règles majeures, problèmes actuels. *In*: Prat, D., Raynal-Roques, A. & Roguenant, A. (Eds.), *Peut-on classer le vivant? Linné et la systématique aujourd’hui*, Belin, Paris, 355–402.

Dubois, A. (2008f) Qu’est-ce qu’une espèce animale? *Annales de la Société entomologique de France*, in press.

Dubois, A. & Günther, R. (1982) Klepton and synklepton: two new evolutionary systematics categories in zoology. *Zoologische Jahrbücher: Systematik*, 109, 290–305.

Dubois, A. & Ohler, A. (1995) Catalogue of names of frogs of the subgenus *Pelophylax* (Amphibia, Anura, genus *Rana*): a few additions and corrections. *Zoologica Poloniae*, “1994”, 39 (3-4), 205–208.

Dubois, A. & Ohler, A. (1997) Early scientific names of Amphibia Anura. I. Introduction. *Bulletin du Muséum national d’Histoire naturelle*, (4), 18 (3-4), 297–320.

Dubois, A. & Ohler, A. (2008) Nomina Amphibiorum. 1. The status of the amphibian nomina created by Merrem (1820) and Ritgen (1828), with comments on the nomenclatural standards of integrative taxonomy. *Zootaxa*, accepted pending modifications.

Evenhuis, N.L. (2008) The “mihi itch”—a brief history. *Zootaxa*, 1890, 59–68.

Frost, D.R., Grant, T., Faivovich, J., Bazin, R.H., Haas, A., Haddad, C.F.B., de Sá, R.O., Channing, A., Wilkinson, M., Donnellan, S.C., Raxworthy, C.J., Campbell, J.A., Blotto, B.L., Moler, P., Drewes, R.C., Nussbaum, R.A., Lynch, J. D., Green, D.M. & Wheeler, W.C. (2006) The amphibian tree of life. *Bulletin of the American Museum of Natural History*, 297, 1–370.

Gaffney, E.S. (1979) An introduction to the logic of phylogeny reconstruction. *In*: Cracraft, J. & Eldredge, N. (Eds.), *Phylogenetic analysis and paleontology*, Columbia University Press, New York, pp. 79–111.

Grant, T., Frost, D.R., Caldwell, J.P., Gagliardo, R., Haddad, C.F.B., Kok, P.J.R., Means, D.B., Noonan, B.P., Schargel, W.E. & Wheeler, W.C. (2006) Phylogenetic Systematics of dart-poison frogs and their relatives (Amphibia: Athesphatanura: Dendrobatidae). *Bulletin of the American Museum of natural History*, 299, 1–262.

Gray, J. E. (1825) A synopsis of the genera of Reptiles and Amphibia, with a description of some new species. *Annals of Philosophy*, (2), 10, 193–217.

Haeckel, E. (1866) *Generelle Morphologie der Organismen*. Zweiter Band. *Allgemeine Entwickelunggsgeschichte der Organismen*. Georg Kramer, Berlin, xlc + 462 pp, 8 pl.

Haffer, J. (1986) Superspecies and species limits in vertebrates. *Zeitschrift für zoologische Systematik- und Evolutionsforschung*, 24, 169–190.

Hennig, W. (1936) Beziehungen zwischen geographischer Verbreitung und systematischer Gliederung bei einigen Dipterenfamilien: ein Beitrag zum Problem der Gliederung systematischer Kategorien höherer Ordnung. *Zoologischer Anzeiger*, 116, 161–175.

Hennig, W. (1950) *Grundzüge einer Theorie der phylogenetischen Systematik*. Deutscher Zentralverlag, Berlin, vii + 370 pp.

Hennig, W. (1966) *Phylogenetic systematics*. Univ. Illinois Press, Urbana, Chicago & London, vii + 263 pp.

Hillis, D.M. (2006) Constraints in naming parts of the tree of life. *Molecular Phylogenetics & Evolution*, 42, 331–338.

Hillis, D.M., Chamberlain, D.A., Wilcox, T.P. & Chippindale, P.T. (2001) A new species of subterranean blind salamander (Plethodontidae: Hemidactyliini: *Eurycea*: *Typhlomolge*) from Austin, Texas, and a systematic revision of central Texas paedomorphic salamanders. *Herpetologica*, 57, 266–280.

Hillis, D.M. & Wilcox, T.P. (2005) Phylogeny of the New World true frogs (*Rana*). *Molecular Phylogenetics & Evolution*, 34, 299–314.

Joyce, W.G., Parham, J.F. & Gauthier, J.A. (2004) Developing a protocol for the conversion of rank-based taxon names to phylogenetically defined clade names, as exemplified by turtles. *Journal of Paleontology*, 78, 989–1013.

Keller, R.A., Boyd, R.N. & Wheeler, Q.D. (2003) The illogical basis of phylogenetic nomenclature. *The botanical Review*, 69, 93–110.

Knox, E., (1998) The use of hierarchies as organizational models in systematics. *Biological Journal of the Linnean Society*, 63, 1–49.

Kuhn, O. (1967) *Amphibien und Reptilien*. Gustav Fischer, Stuttgart, vii + 124 pp.

Kuntner, M. & Agnarsson, I. (2006) Are the Linnean and phylogenetic nomenclatural systems combinable? Recommendations for biological nomenclature. *Systematic Biology*, 55, 774–784.

Laurin, M. & Anderson, J.S. (2004) Meaning of the name Tetrapoda in the scientific literature: an exchange. *Systematic Biology*, 53(1), 68–80.

Martin, J. E. & Benton, M. J. (2008) Crown clades in vertebrate nomenclature: correcting the definition of Crocodylia. *Systematic Biology*, 57, 173–181.

Mayr, E. (1940) Speciation phenomena in birds. *American Naturalist*, 74, 249–278.

Mayr, E. (1942) *Systematics and the origin of species from the viewpoint of a zoologist.* Columbia University Press, New York, xiv + 334 pp.

Mayr, E. (1963) *Animal species and evolution.* Harvard University Press, Cambridge, xiv + 797 pp.

Mayr, E. (1982) *The growth of biological thought*. Belknap Press, Cambridge, Mass. & London, xiii + 974 pp.

Mayr, E. (1995) Systems of ordering data. *Biology & Philosophy*, 10, 419–434.

Mayr, E. (1997) *This is biology. The science of the living world.* Belknap Press, Cambridge, Mass. & London, xvii + 327 pp.

Mayr, E. & Ashlock, P.D. (1991) *Principles of systematic zoology.* Second edition. McGraw-Hill, New York, xx + 475 pp.

Minelli, A. (1991) *Biological systematics: the state of the art.* Chapman & Hall, London, xvi + 387 pp.

Minelli, A. (2000) The ranks and the names of species and higher taxa, or a dangerous inertia of the language of natural history. *In*: Ghiselin, M. T. & Leviton, A. E. (Eds.), *Cultures and institutions of natural history: essays in the history and philosophy of sciences*, California Academy of Sciences, San Francisco, 339–351.

Moore, G. (2003) Should taxon names be explicitly defined? *The botanical Review*, 69, 2–21.

Muona, J. (2006) To be named or not to be named—*Heteropodarke pleijeli* sp. n. (Annelida, Polychaeta). *Cladistics*, 22 (6), 634–635.

Pleijel, F. (1999) Phylogenetic taxonomy, a farewell to species, and a revision of *Heteropodarke* (Hesionidae, Polychaeta, Annelida). *Systematic Biology*, 48, 755–789.

Raikow, R.J. (1985) Problems in avian classification. *In*: Johnston, R. F. (Ed.), *Current Ornithology*, Vol. 2, Plenum, New York, 187–212.

Risso, A. (1827) *Histoire naturelle des principales productions de l'Europe méridionale et particulièrement de celles des environs de Nice et des Alpes Maritimes.* Levrault, Paris, xvi + 480 pp.

Samadi, S. & Barberousse, A. (2006) The tree, the network, and the species. *Biological Journal of the Linnean Society*, 89, 509–521.

Schaefer, W.W. (1976) The reality of the higher taxonomic categories. *Zeitschrift für zoologischer Systematik und Evolutions-Forschung*, 14, 1–10.

Sibley, C.G. & Ahlquist, J.E. (1990) *Phylogeny and classification of birds. A study in molecular evolution*. Yale University Press, New Haven & London, xxiii + 976 pp.

Simpson, G. . (1940) Types in modern taxonomy. *American Journal of Science*, 238, 413–431.

Strickland, H.E., Henslow, J.S., Phillips, J., Shuckard, W.E., Richardson, J., Waterhouse, G.R., Owen, R., Yarrell, W., Jenyns, L., Darwin, C., Broderip, W.J. & Westwood, J.O. (1843) Series of propositions for rendering the nomenclature of zoology uniform and permanent, being the Report of a Committee for the consideration of the subject appointed by the British Association for the Advancement of Science. *Annals & Magazine of natural History*, (1), 11, 259–275.

Stuessy, T.F. (2000) Taxon names are *not* defined. *Taxon*, 49, 231–233.

Stuessy, T.F. (2001) Taxon names are *still* not defined. *Taxon*, 50, 185–186.

Van Valen, L. (1973) Are categories in different phyla comparable? *Taxon*, 22, 333–373.

Vidal, N. & Hedges, S.B. (2005) The phylogeny of squamate reptiles (lizards, snakes, and amphisbaenians) inferred from nine nuclear protein-coding genes. *Comptes rendus Biologies*, 328, 1000–1008.

Vieites, D.R., Min, M.-S. & Wake, D. . (2007) Rapid diversification and dispersal during periods of global warming by plethodontid salamanders. *Proceedings of the national Academy of Sciences of the USA*, 104 (50), 19903–19907.

Wake, D.B., Salvador A. & Alonso-Zarazaga, M.-A. (2005) Taxonomy of the plethodontid salamander genus *Hydromantes*. *Amphibia-Reptilia*, 26, 543–548.

Wheeler, Q.D., Raven, P.H. & Wilson, E.O. (2004) Taxonomy: impediment or expedient? *Science*, 303, 285.

Wiens, J.J. (2007) Review of "The amphibian tree of life". *Quarterly Review of Biology*, 82, 55–56.

Zootaxa 1950: 87–95 (2008)
www.mapress.com/zootaxa/

ISSN 1175-5326 (print edition)
ZOOTAXA
ISSN 1175-5334 (online edition)

Anatomical nomenclature: homology, standardization and datasets*

GREGORY D. EDGECOMBE

Department of Palaeontology, Natural History Museum, Cromwell Road, London SW7 5BD, U.K.
E-mail: g.edgecombe@nhm.ac.uk

* *In:* Minelli, A., Bonato, L. & Fusco, G. (eds) *Updating the Linnaean Heritage: Names as Tools for Thinking about Animals and Plants. Zootaxa*, 1950, 1–163.

Table of contents

Abstract

Strictly homology-based character names have the benefit of a consistent, evolutionary basis but must overcome practical problems in terms of the function that names serve as tools for communication. Character names should be fixed at the level of primary (rather than secondary) homology in order to maintain nomenclatural stability between competing phylogenies and to allow characters to potentially re-optimize with the addition of data. Inconsistent rules determine the priority of names for anatomical structures, in marked contrast to the stability and clarity provided by Codes for taxonomic nomenclature. Standardized anatomical nomenclature is amenable to a web-based, ontology-driven framework. Imagery and associated metadata linked to phylogenetic datasets facilitate character documentation, nomenclatural stability, and repeatability without requiring a formal process of typification.

Key words: Primary homology, Priority, Secondary homology, Standardized nomenclature, Typification

Introduction

Biologists (indeed, all humans) name anatomical structures for innumerable reasons, all of them ultimately hinging on communicating some idea about identity. My focus herein will be on the way we name anatomical structures for a particular endeavor in comparative biology, the construction of datasets that we use as tools for inferring evolutionary history.

Accepted by A. Minelli: 24 Oct. 2008; published: 5 Dec. 2008

Zootaxa 1950: 96–102 (2008)
www.mapress.com/zootaxa/

ISSN 1175-5326 (print edition)
ZOOTAXA
ISSN 1175-5334 (online edition)

Morphological nomenclature, between patterns and processes: segments and segmentation as a paradigmatic case*

GIUSEPPE FUSCO

Department of Biology, University of Padova, via Ugo Bassi 58/B, I-35131 Padova, Italy. E-mail: giuseppe.fusco@unipd.it

* *In*: Minelli, A., Bonato, L. & Fusco, G. (eds) *Updating the Linnaean Heritage: Names as Tools for Thinking about Animals and Plants*. Zootaxa, 1950, 1–163.

Table of contents

Abstract

The words we use for describing biological systems and their transformations through development and evolution can recurrently perform as ‘conceptual traps’, i.e. as representations that limit the possibilities of improving our understanding of the very processes they are called to describe. The main focus of this contribution is on the paradigmatic case of segmentation. Limits and drawbacks of the concept of ‘segment’ are critically discussed. Its value as a descriptive unit does not entitle it as a sensible unit for other uses, as for instance investigating the evolution of the developmental process of segmentation.

Key words: Arthropods, Development, Evo-devo, Evolution, Parasegments, Re-segmentation, Segmental mismatch

Introduction

Developmental biology and evolutionary biology are sciences of change through time. Formulating and testing hypotheses within these two disciplines, or at their interface (the so called evolutionary developmental biology, or evo-devo), needs sound names. Names are essential to describe organism features, both (anatomical) patterns and (physiological) processes, but also to describe the changes of these features along ontogeny and across generations, the two processes that produce developmental and evolutionary patterns, respectively. Thus, although the question of names is not a distinctive aspect of this area of biology, it is nonetheless deeply entangled with scientific investigation and communication.

Accepted by A. Minelli: 24 Oct. 2008; published: 5 Dec. 2008

Some issues related to the use of names in morphology are well known and universally acknowledged by specialists. One is that the use of the same name for two structures, or two features, in two different organisms is easily taken as an implicit declaration of homology (see Edgecombe 2008). Attempts to avoid this over-interpretation generates an over-proliferation of morphological and anatomical terminology. However, this is generally accounted for in systematics, for instance in compiling data matrices for phylogenic reconstruction, although paradoxes remain, such as that of the segmentally patterned animals (e.g., kinorhynchs, rotifers, cestodes) that are usually qualified as non-segmented (Minelli & Fusco 2004). Another issue is that scientific progress is not just accumulation of new knowledge, with its specific new nomenclature. There is a continuous revision and refinement of what we already know, where the original meaning of words tends to be stretched, or the words are variably qualified through adjectives, prefixes and suffixes, to fit the new state of knowledge. This process is often a source of terminological confusion. Think for instance of the changing attitudes in respect to the concept of homology, starting with its meaning in pre-Darwinian times, to go throughout the new synthesis, the advent of cladistics, and the so called genomics era (Minelli 2003). Names are undoubtedly in dynamic relationship with knowledge, and continuous vigilance is thus in order.

However, this is just the tip of the iceberg. Names and name usage can exhibit more sly drawbacks. The words we use for describing biological systems and their transformations risk continuously to perform as 'conceptual traps', i.e. as concepts that limit our capability to ask and address sensible questions. There are many examples in descriptive morphology ('tagma', 'body axis', 'body plan'), in development ('embryo', 'larva', 'adult'), and in evolution ('heterochrony', 'novelty', 'adaptation') (see Minelli and Fusco 1995, Minelli *et al*. 2006).

The main focus of this article is on segmentation and its evolution, both as morphological pattern and developmental process, as it represents a paradigmatic case of a 'false friend' in biology. The term 'segmentation' is used to describe both a morphological feature (a form of body symmetry) and the developmental process that generates it. The two concepts are obviously related, but in some way they are independent from each other, and can be analyzed separately.

Segmentation pattern: from structures to axis to structures

Segmentation is a form of body symmetry, in particular it is a type of translational symmetry. A segmental pattern can be defined as the serial occurrence of homologous structures along an axis of the body, for instance along the main body axis.

The scolopender is a good example of a segmented animal, and its body architecture can easily be described on the basis of its symmetry. An introductory textbook description might start as follows: "The body of a scolopender consists of a head and a trunk. The first trunk segment bears a pair of poisonous maxillipedes; this is usually followed, depending on the species, by 21 or 23 leg-bearing segments, followed in turn by a limbless genito-anal region." Onto this 'segmental frame' it is easy to add further anatomical details, to specify, for instance, that "dorsal sclerites of leg-bearing segments II, IV, VI, IX, XI, XIII, XV, XVII, XIX are relatively shorter with respect to the contiguous segments", or that "spiracles, the openings of the tracheal system, are usually borne on pleural sclerites limited to segments III, V, VII, X, XII, XIV, XVI, XVIII, XX".

The value of such a descriptive framework will not be questioned. But let's reflect a moment on its implications. More or less consciously, the periodic pattern is thus intended to apply to the whole trunk, rather than to a specified series of serially homologous structures, such as legs, sclerites, or spiracles. The scolopender is qualified as a segmented animal because its body is comprised of a certain number of segments, rather than because its body presents one or more series of segmental (repetitive) structures. When the segmental pattern is attributed to the whole body axis, rather than to a specific set of repetitive structures along it, the idea of a body 'comprised of' a certain number of 'body-blocks' or 'modules' will inescapably result (Budd 2001).

This unnecessary, apparently harmless, conceptual (or semantic) shift, from 'a segmented animal is one with segmental structures' to 'a segmented animal is made of segments', conveys a number of drawbacks.

For instance, descriptive complications arise when different serial structures along the same axis show discordant serial arrangements. Similar occurrences are traditionally referred to as cases of 'segmental mismatch', a label that with a body-block concept of segment is clearly an oxymoron, because if the body is comprised of segments, there are no distinct segmental series that can either match or not. Segmental mismatch is not rare, nor is it taxonomically confined to a single clade. Many myriapods exhibit segmental mismatch: in symphylans and craterostigmomorph centipedes there are more tergites than leg pairs, whereas in scutigeromorph centipedes and tetramerocerate pauropods the number of tergites is smaller than the number of leg pairs (review in Fusco 2005). In millipedes, for the most part of the trunk, there are two leg pairs for each tergal plate, but the first three or four tergal plates correspond to one leg pair each. Still more, in some taxa the number of dorsal sclerites is not predictive of the number of leg pairs (and vice-versa), as there is only a statistical correlation between the two numbers (Enghoff *et al*. 1993).

Adopting a body-block concept of segment, the evolution of segmentation reduces to a limited set of evolutionary options. Beyond changing the number of segments, variation is just produced by fusion and/or splitting of pre-existing segmental structures. Thus, even in case of segmental mismatch, it seems sensible to ask which is the exact correspondence between different segmental series. For instance, in the case of millipedes, on the basis of what seemed to be the obvious correspondence between dorsal and ventral segmental structures in those species, as the julids, where these elements form a continuous ring, the same correspondence was extended to non ring-forming species. But, descriptive embryology of the pill millipede *Glomeris marginata* suggests a different correlation between dorsal and ventral segmental units (review in Janssen *et al*. 2006). More recently, studying the expression of some segmentation genes, in *Glomeris* again, it finally turned out that dorsal and ventral serial structures are independently established in the embryo, and that the antero-posterior boundaries of the prospective dorsal sclerites do not correlate with either the antero-posterior boundaries of the anlagen of ventral or dorsal structures (Janssen *et al*. 2004). Expression patterns of segmentation genes in the prospective ventral and dorsal tissues are different as well (Janssen *et al*. 2008).

The evolution of segmental mismatch can occur through pathways of change that are other than those resulting from the abstract logic of fusion or splitting of pre-existing segmental structures. Beyond providing developmental genetic basis for dorso-ventral mismatch, Janssen *et al.*'s studies (2004, 2008) showed that, for millipedes at least, the body-block segment is at best a descriptive or a functional body unit (Minelli 2004). But its value as a descriptive unit does not entitle it as a sensible unit for other uses, as for instance to investigate the evolution of the developmental process of segmentation.

Segmentation and re-segmentation

A pattern that presents translational symmetry can be easily described as the periodic occurrence of a repeating motif of a given length (period). However this does not imply that this motif is unique. The number of possible motifs in a periodic pattern is indeed infinite (Fig. 1). The arbitrary choice of which one is the most appropriate for a given task is either a question of convenience or a mere convention. In a linear periodic pattern, a motif can be characterized by anterior and posterior boundaries, but these boundaries, as such, are not inherent elements of the pattern.

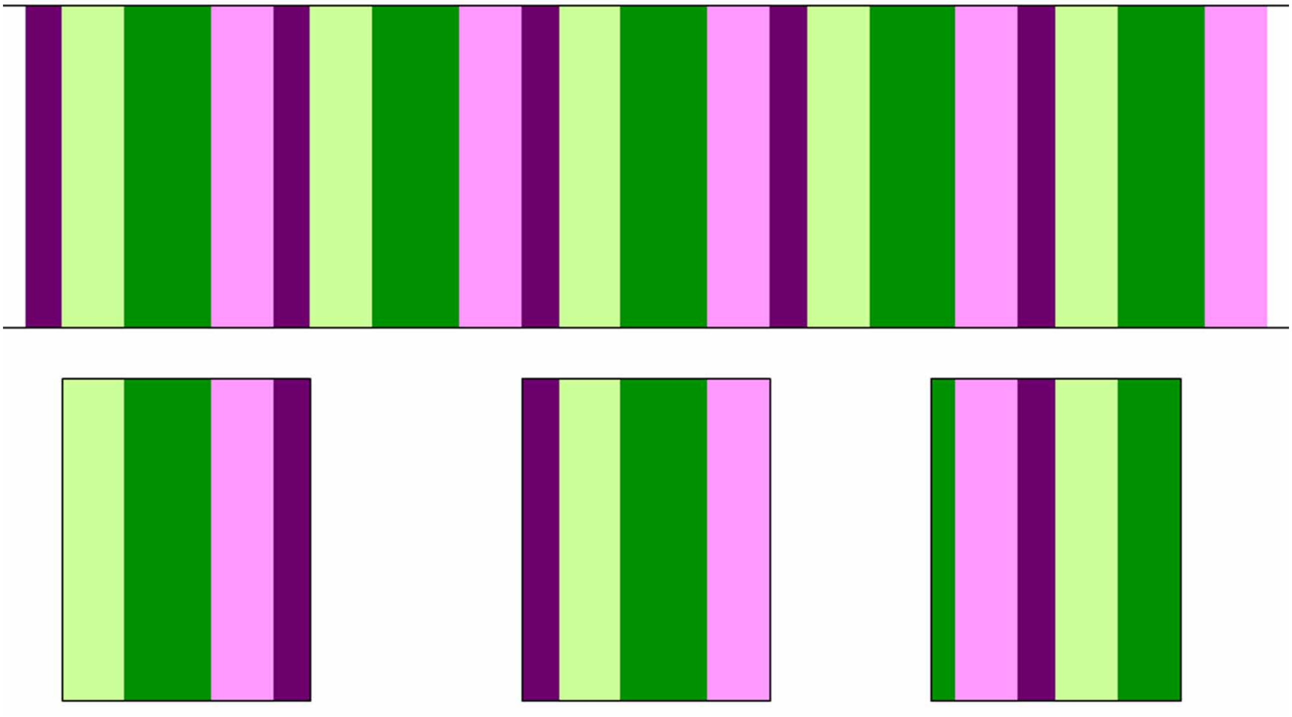

FIGURE 1. A periodic pattern with three of the infinite repeating motifs that can be used for describing it.

These almost trivial geometrical considerations are evidently not as obvious in studying animal segmental patterns. In standard descriptions of gene expression patterns, it is common practice to use a reference system based on segmental units delimited by segmental boundaries, as if these could be established univocally. However, while anatomical elements (e.g. sclerites), cell domains (e.g. compartments, i.e. cell populations of (poly)clonal origin), or gene expression domains, can have objective boundaries, segments, that is the units (sections) into which the axis is subdivided, have only the boundaries given by their definition, for the simple geometrical reasons mentioned above.

Developmental genetic studies, in *Drosophila* first (Martinez-Arias & Lawrence 1985), and later in other arthropods (review in Damen 2007), showed that in the early embryo the functional segmental unit is different from the functional segmental unit in the later embryonic, larval and adult stages. In the early embryo, segment-polarity genes (like *engrailed* or *wingless*) define parasegmental boundaries. These are limits of clonal restriction and often limits of expression domain for later regulatory genes, like the *Hox* genes. Parasegments are out of phase with respect to the later segments, as defined on the basis of trunk articulation. For some authors (e.g., Lawrence 1992), only parasegments are real entities, i.e. "units of internal description" of the embryo, whereas segments, being mere anatomical outcomes, as developmental units "may only exist in the mind of scientists". Other authors (e.g., Deutsch 2004) consider the passage from the parasegment to the segment reference system as an important ontogenetic shift in segmental organization. This has sometimes been called re-segmentation (e.g., Prud'homme *et al*. 2003), in analogy to a similar process described for vertebrate somitogenesis (e.g., Saga & Takeda 2001). My opinion is that neither parasegments nor segments are necessary developmental units. Paraphrasing Lawrence (1992), what might be only in the mind of the scientists is the embryo's need for 'units of internal description'. If the initiation/localization of a morphogenetic process producing a segmental structure (e.g., the insect leg) needs a pre-existing specific periodic signal (e.g., *wingless* expression), this does not entail that there is also a need for an early subdivision of a body domain (e.g., the thorax) into a number of abstract 'units of internal description', either parasegments or segments.

Re-segmentation is not a process per se. It comes to life only in virtue of the arbitrary definitions of two different (and out of phase) descriptive units. Simply, different morphogenetic processes, in early and later development, in different ectodermal and mesodermal tissues, exploit the same framework of positional information provided by the expression of a pool of segmentation genes. It is true that in *Drosophila* early embryogenesis, during germ-band retraction, parasegmental grooves disappear and the segmental ones start to form, but this does not require a shift in the 'internal reference systems'.

In the arthropods where it has been investigated, re-segmentation does not involve in the same way all the segmental structures of the body, thus the 'segmental organization' of the adult exoskeleton can be associated to the 'parasegmental organization' of its nervous system (Deutsch 2004). This differential 'frame shift' has been interpreted as an adaptation for permitting movement and for improving motor control in animals with an articulated exoskeleton, as arthropods are. Although along the main body axis there are indeed many segmental structures (sclerites, myomers, neuromers) reciprocally out of frame (for very good functional reasons), this does not require the presence of two (or more) segmental 'reference systems'. For segmental structures with the same period, one will suffice. Pattern formation needs positional information, but 'units of internal description' are not indispensable. In *Glomeris* it is sensible to distinguish between dorsal and ventral *segmentation*, or dorsal and ventral segmental patterns, but there is no need to define dorsal and ventral *segments* (Janssen *et al.* 2008).

Similarly to the concept of segment as a body block, the idea of the segment (or parasegment) as a necessary developmental unit is a descriptive heredity that constraints the way in which we can conceive the evolution of development. In this case, it imposes to the embryo a developmental logic based on repeated units, apparently necessary to obtain the whole segmental pattern. Evolution of developmental pathways hardly feel this constrictions, which only emerge from the geometry of our descriptions.

Segmentation process: producing segments or not

There are good reasons for not taking segments too literarily, and these reasons can be found in the way in which segmentation develops and evolves. This is certainly a semantic issue, but it is not a 'mere semantic issue', since the meaning associated with these words conditions investigation and understanding of real biological process, both in development and evolution.

To understand the evolution of these developmental pathways we should keep segmental patterns (that sometimes can be easily described as a series of segments) well distinct from segmentation processes (that can produce segments or not). In absence of direct observational data, segments should be considered just as epiphenomenal units (Fusco 2005). The overall aspect of a segmented animal depends on the level of concordance between different segmental series of structures. When many structures occur with the same periodicity, the effect of a body comprised of a series of segments is obtained. To some extent, this could be the case of the scolopender. On the contrary, when different structures are patterned on the basis of distinct discordant regulative signals, the result is a body affected by segmental mismatch, as in the case of the millipede.

The diversity of arthropod segmental patterns fits suitably with segmentation evolving through 'segment-free' pathways of change.

Conclusions

If morphological nomenclature hides insidious 'conceptual traps' that can bias the investigation of biological patterns and processes, is there a general method to avoid them? The problem, evidently, is one of identifying such fallacies. Clues of the presence of a conceptual trap might be a logically inconsistent nomenclature or an ineffectually complex terminology. For instance, descriptive systems based on the depiction of a 'general rule' (‘the trunk of an arthropod is made of a series of segments’) followed by a list of ‘exceptions’ (‘not all segments are exactly the same’, ‘not all structures present a segmental arrangement’, or ‘not all structures present the same segmental arrangement’) are very common. This practice, beyond being logically unsatisfactory and leading to an uncontrolled proliferation of morphological nomenclature, tends to hide the fact that we are probably missing some relevant aspect of the pattern (or process) under consideration.

Without the intention of advancing a radical solution (if there is one) to this weakness of morphological nomenclature, here is a short list of simple rules of thumb that can perhaps help avoiding at least the most macroscopic fallacies.

i. Try to disregard, as much as possible, the burden of non-declared meanings and implications that the usage of a given term has produced and accumulated along its history. Segmentation is not an ‘all or nothing’ condition. Arthropods, annelids and vertebrates are segmented just up to a point. And, there are many non-arthropod, non-annelid, non-vertebrate animals that are segmented animals, up to a point.

ii. Acknowledge and accept the non-resolvable approximate nature of morphological nomenclature. It always depends on the level of description. Any periodic pattern can disappear at an enough close scrutiny. Think of the actual variation of any biometric variable along the main axis (for instance, in an arthropod, a cuticle parameter varying from sclerite to arthrodial membrane), it is never periodic in the strict sense with which this word is used in mathematics. Periodicity emerges only at a convenient level of description (for instance just considering the alternation of sclerites (0) and arthrodial membranes (1): 0101010...). Segmentation exists just up to a point.

iii. Keep always in mind the distinction between the level of description from the level of reality. A model M can be defined as a representation of the system S that can be usefully employed for answering questions on S. Any model has its range of application, beyond that it is only an inadequate representation. As a general rule, a descriptive model cannot be used as the starting point for addressing questions of development and evolution of the modelled system. Segments are elements of a representation.

In short, before usage, ask what that name is for.

Acknowledgments

L. Bonato, C. Brena, W. Damen, J. Deutsch and A. Minelli provided insightful comments on a previous version of this article.

References

Budd, G.E. (2001) Why are arthropods segmented? *Evolution & Development*, 3, 332–342.

Damen, W.G.M. (2007) Evolutionary conservation and divergence of the segmentation process in arthropods. *Developmental Dynamics*, 236, 1379–1391.

Deutsch, J.S. (2004) Segments and parasegments in arthropods: a functional perspective. *BioEssays*, 26, 1117–1125.

Edgecombe, G.D. (2008) Anatomical nomenclature: homology, standardization and datasets. *In*: Minelli, A., Bonato, L. & Fusco, G. (Eds), *Updating the Linnaean Heritage: Names as tools for thinking about animals and plants*. *Zootaxa*, 1950, 87–95.

Enghoff, H., Dohle, W. & Blower, J.G. (1993) Anamorphosis in millipedes (Diplopoda) — the present state of knowledge and phylogenetic considerations. *Zoological Journal of the Linnean Society*, 109, 103–234.

Fusco, G. (2005) Trunk segment numbers and sequential segmentation in myriapods. *Evolution & Development*, 7, 608–617.

Janssen, R., Budd, G.E., Damen, W.G.M. & Prpic, N.-M. (2008) Evidence for Wg-independent tergite border formation in the millipede *Glomeris marginata*. *Development Genes and Evolution*, 218, 361–370.

Janssen, R., Prpic, N.-M. & Damen, W.G.M. (2004) Gene expression suggests decoupled dorsal and ventral segmentation in the millipede *Glomeris marginata* (Myriapoda: Diplopoda). *Developmental Biology*, 268, 89–104.

Janssen, R., Prpic, N.-M., & Damen, W.G.M. (2006) A review of the correlation of tergites, sternites, and leg pairs in diplopods. *Frontiers in Zoology*, 3, 2.

Lawrence, P.A. (1992) *The making of a fly*. Blackwell, Oxford, 240 pp.

Martinez-Arias A & Lawrence, P.A. (1985) Parasegments and compartments in the *Drosophila* embryo. *Nature*, 313, 639–642.

Minelli, A. (2003). *The development of animal form*. Cambridge University Press, Cambridge, 342 pp.

Minelli, A. (2004) Bits and pieces. *Science*, 306, 1693–1694.

Minelli, A., Brena, C., Deflorian, G., Maruzzo, D. & Fusco, G. (2006) From embryo to adult — beyond the conventional periodization of arthropod development. *Development Genes and Evolution*, 216, 373–383.

Minelli, A. & Fusco, G. (1995) Body segmentation and segment differentiation: the scope for heterochronic change. *In*: McNamara, K.J. (ed.), *Evolutionary change and heterochrony*. Wiley & Sons, Chichester, pp. 49–63.

Minelli, A. & Fusco, G. (2004) Evo-devo perspectives on segmentation: model organisms, and beyond. *Tends in Ecology and Evolution*, 19, 423–429.

Prud'homme, B., de Rosa, R., Arendt, D., Julien, J.-F., Pajaziti, R., Dorresteijn, A., Adoutte, A., Wittbrodt, J. & Balavoine, G. (2003) Arthropod-like expression patterns of *engrailed* and *wingless* in the annelid *Platynereis dumerilii* suggest a role in segment formation. *Current Biology*, 13, 1876–1881.

Saga, Y. & Takeda, H. (2001) The making of the somite: molecular events in vertebrate segmentation. *Nature Reviews Genetics*, 2, 835–845.

Zootaxa 1950: 103–122 (2008)
www.mapress.com/zootaxa/

ISSN 1175-5326 (print edition)
ZOOTAXA
ISSN 1175-5334 (online edition)

Plant structure ontology: How should we label plant structures with doubtful or mixed identities? *

BRUCE K. KIRCHOFF[1], EVELIN PFEIFER[2] & ROLF RUTISHAUSER[2]

[1]*Department of Biology, University of North Carolina at Greensboro, Greensboro, USA. E-mail: kirchoff@uncg.edu*
[2]*Institute of Systematic Botany, University of Zurich, Switzerland. E-mail: rutishau@systbot.uzh.ch*

* *In:* Minelli, A., Bonato, L. & Fusco, G. (eds) *Updating the Linnaean Heritage: Names as Tools for Thinking about Animals and Plants. Zootaxa*, 1950, 1–163.

Table of contents

Abstract

This paper discusses problems with labelling plant structures in the context of attempts to create a unified Plant Structure Ontology. Special attention is given to structures with mixed, or doubtful identities that are difficult or even impossible to label with a single term. In various vascular plants (and some groups of animals) the structural categories for the description of forms are less distinct than is often supposed. Thus, there are morphological misfits that do not fit exactly into one or the other category and to which it is difficult, or even impossible, to apply a categorical name. After presenting three case studies of intermediate organs and organs whose identity is in doubt, we review five approaches to categorizing plant organs, and evaluate the potential of each to serve as a general reference system for gene annotations. The five approaches are (1) standardized vocabularies, (2) labels based on developmental genetics, (3) continuum morphology, (4) process morphology, (5) character cladograms. While all of these approaches have important domains of applicability, we conclude that process morphology is the one most suited to gene annotation.

Key words: Character cladogram, Continuum, Evo-devo, Flower, Flowering plants, Gene ontology, Intermediate organs, Leaf, Morphological nomenclature, Organ identity, Plant ontology, Plant structure ontology, Root

Accepted by A. Minelli: 24 Oct. 2008; published: 5 Dec. 2008

Zootaxa 1950: 123–152 (2008)
www.mapress.com/zootaxa/

ISSN 1175-5326 (print edition)
ZOOTAXA
ISSN 1175-5334 (online edition)

Learning from Linnaeus: towards developing the foundation for a general structure concept for morphology*

LARS VOGT

Systematik und Evolution der Tiere, Institut für Biologie, FU Berlin, Königin Luise Str. 1–3, D-14195 Berlin, Germany.
E-mail: lars.vogt@zoosyst-berlin.de

* *In:* Minelli, A., Bonato, L. & Fusco, G. (eds) *Updating the Linnaean Heritage: Names as Tools for Thinking about Animals and Plants. Zootaxa,* 1950, 1–163.

Table of contents

Accepted by A. Minelli: 24 Oct. 2008; published: 5 Dec. 2008

Abstract

Morphology has fundamental problems regarding aperspectival objectivity of its data—morphological terminology is often based on homology assumptions, lacks standardization, and has problems with comparability, reproducibility, and transparency. This is astonishing given that with his sexual system Linnaeus had already established a high degree of aperspectival objectivity in morphology that unfortunately has been lost subsequently. In the first part of the article a brief introduction to the history of classification is given that provides an answer to the question why morphology only initially has been gripped by the general trend towards objectification that started in the seventeenth century. The conceptual shortcomings of Aristotle's concept of essences and its link to the definition of species and taxa in natural philosophy play an important part in this development. The only solution to the problem of essences was to link it to the evolutionary concept of homology, which explains why morphological terminology today often rests on homology assumptions. By taking a closer look at Linnaeus' sexual system, basic principles for developing a general structure concept for morphology are discussed, which would provide the conceptual basis for establishing a high degree of aperspectival objectivity for morphological data. The article concludes with discussing the role of data bases and ontologies for developing a data standard in morphology. A brief introduction to the basic principles of Resource Description Framework (RDF) ontologies is given. A morphological ontology has high potential for establishing a general morphological structure concept if it is developed on grounds of the following principles: morphological terms and concepts must be defined taxon-independently, homology-free, preferably purely anatomically, and if functionally only by clearly indicating the trait's active participation in a specific biological process.

Key words: Aperspectival objectivity, Bio-ontology, Essentialism, Morphological data, Linguistic problem of morphology, RDF, Standardization

Introduction

Morphology represents a set of methods and techniques for producing data about anatomical and organizational facts of organisms. As such, it does not represent a theory or an explanatory hypothesis. When it comes to preparing morphological descriptions, morphology is all about the textual representation, documentation, and comparison of structural diversity and patterns of structural equivalences between organisms and their traits, thereby being only *assisted* by various imaging techniques for the empirical substantiation of these descriptions. Therefore, morphological terminology and language assume a central methodological role in morphology. Only if the language and terminology used in morphological descriptions are capable of reliably transporting the relevant information in an unambiguous way and independent of individual morphologists, and only if they enable the comparison of morphological data across a broad taxonomic range, will morphology meet the high degree of comparability and communicability of data that is being increasingly demanded in the age of a growing importance of data bases in biology.

Unfortunately, morphology lacks standardization and common acceptance of morphological terms and lacks a formalized method of recording and documenting morphological descriptions (Vogt *et al. submitted*). Thus, morphology has fundamental problems with its terminology. As a consequence, morphological terminology and morphological descriptions vary from author to author, the meaning of morphological terms often changes through time, and the applicability of morphological terms is often restricted to a specific taxonomic group and cannot be easily adapted to other groups. In scientific research practice, this non-standardization of morphological terminology and the diversity in quality, organization, and style of morphological descriptions frequently lead to divergent descriptions of equivalent traits or to identically described morphological traits that are in fact not identical (see linguistic problem of morphology, Vogt *et al. submitted*; see also Ramírez *et al.* 2007).

These linguistic ambiguities pose fundamental problems for comparative morphological studies, being the source for repeated misunderstandings among morphologists, undermining the possibility to reliably communicate morphological data. Reliable communication of data, however, represents a necessary prerequisite for the division of labor not only among morphologists conducting comparative studies over a broad taxonomic range, but also for all kinds of co-operations in which morphologists are involved or morphological data are analyzed. Thus, it seems that morphology is hard pushed these days to prove that its standards of objectivity, comparability, and communicability still hold up to non-morphological biological data, as for instance DNA sequence data. Considering these fundamental problems, it is not surprising that some biologists even claim that morphology has already lost its traditionally prominent role in phylogenetics (see e.g., Scotland *et al.* 2003), since comparability of data represents a *sine qua non* of phylogenetic research practice.

The interaction between phylogenetics on the one hand, or more traditionally biological taxonomy and classification, and morphology on the other hand represents a liaison with many different facets and a continuous story of mutual interference. This is not surprising since taxonomy and classification have been one of the initial fields of application of morphology, in which traditionally it always had been very strong. As a consequence, much of morphological terminology and methodology has been strongly influenced by the needs and requirements of generating classifications.

In the following I will provide a brief introduction to the history of classification and its impact on the development of morphological terminology and methodology, including conceptions of naturalness and the epistemic status and role of observation and empirical investigation. This historical excursion is intended to give a historical explanation for the question why objectification did not catch on in morphology, while in many other biological disciplines objectification has advanced to a level that established a high degree of transparency, reproducibility, communicability, and inter-subjective consensus regarding empirical data. This question becomes even more interesting when considering that, initially, morphology has been gripped by the general trend towards objectification that started in the seventeenth century. In the second part of the article I will show that there is a lot to learn from Linnaeus' approach to classification in terms of increasing objectification in morphology. Based on Linnaeus's sexual system, I develop the basics of a general structure concept for morphology and argue that it takes in a key role in the context of objectification of morphology. I conclude the article with a brief introduction to biological data bases and standardized controlled vocabularies (i.e., bio-ontologies) and how they can serve as a basis for establishing a general structure concept in morphology and its broad dissemination.

The historical burden of essentialism

Aristotle and essentialism

In order to be able to identify regularities in biology and to generalize about the biology of organisms, a concept for abstracting individual organisms into types (i.e., classes, kinds, families) is required. From a logical point of view any given set of organisms shares an infinite amount of equivalent properties—in other words, any set of organisms could be conceptualized as some sort of kind. Thus, unfortunately, possessing same traits does not necessarily imply ontological equivalence of the respective organisms. As a consequence, generalizations in biology would become impossible as long as biologists would not manage to differentiate between 'real' kinds and artificial kinds; but how to recognize and define 'real' kinds?

This problem was known to ancient Greek scholars, who recognized the necessity to develop a concept for differentiating between essential and accidental properties. Aristotle defines 'real' kinds according to their *essential* properties and classifies them according to the method of logical division of *per genus et differentiam* (Fristrup 2001). The English word 'essence' comes from Latin *essentia* (from *esse*, 'to be'), which represents a translation of Aristotle's ancient Greek phrase *to ti ēn einai* (i.e., 'what it is for a thing to be'), denoting a thing's essence.

An Aristotelian essence represents the attribute or set of attributes that make an entity what it fundamentally is. Without its essence, the entity would lose its identity. Essential properties are real physical properties of the 'nucleus' (i.e., substance) of a thing. For any specific kind of entity, there is a set of essential properties, all of which any entity of that kind *must* have. As a consequence, if two objects share the same essence, they can be considered to be truly *identical* with respect to this aspect of their 'nucleus' and therefore can be classified as instances of the same 'real' kind. That is the reason why, according to Aristotle, essences are fundamentally linked to definitions of different *kinds* of entities.

Aristotelian definitions are hierarchically organized, resulting in a hierarchy of classes and their subclasses (Fig. 1). The defining attributes of a class are inherited downstream to its subclasses (i.e., downward propagation). Thus, if a given entity is an instance of a specific class, it is necessarily also an instance of all those classes of which this class is a sub-class. This hierarchy represents a taxonomy (i.e., *taxonomy* in a broad sense) of more and more specialized concepts, which implies a hierarchical organization of terms (i.e., taxonomic inclusion, Bittner *et al*. 2004).

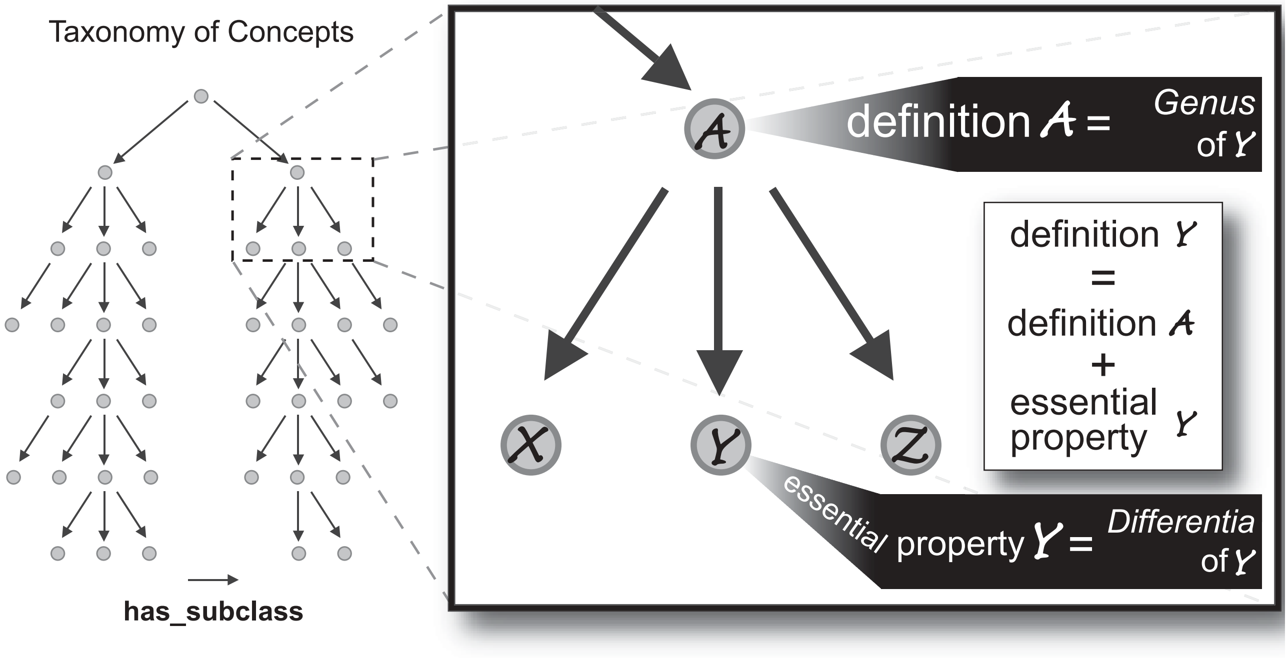

FIGURE 1. Aristotelian definitions: The definition of the 'parent' kind is inherited by all of its 'first child generation' kinds and forms the *Genus* part of their definitions. These represent properties that an instance of a 'child' kind *necessarily* has to possess. On the other hand, the essential property of a 'child' kind represents the distinguishing property that is in combination with the *Genus* part *sufficient* for the recognition of an instance of that kind, since all and only individuals that are instances of this kind do possess this property. This forms the *Differentia* part of the kind's definition. *Genus* and *Differentia* together represent the essence of the kind and at the same time its definition.

When defining a new kind of entity, Aristotle refers to this conceptual relationship of hierarchical specializations. In his definitions, Aristotle distinguishes between what he calls 'genus' and 'differentia'. 'Genus' represents attributes referring to essential properties of all of the respective 'parent' kinds, which an instance *necessarily* has to possess for membership to the 'child' kind. However, these properties are not sufficient for membership recognition. 'Differentia', on the other hand, represent attributes referring to the essential properties of the 'child' kind. 'Differentia' are required to distinguish the kind to be defined from all other kinds of the same hierarchical level. If no other entity, but the instances of the kind to be defined, possesses a certain property and if *all* instances of that kind possess it without any exception, that property qualifies as 'differentia'—the essential property of the kind. Only the combination of 'genus' and 'differentia' is *sufficient* for recognition of membership to the kind. Thus, on *any* hierarchical level of a classification based on Aristotelian definitions the essential properties of all 'parent' kinds provide the 'genus' part of the definition, and the essential properties of the 'child' kind its 'differentia' part; and 'genus' and 'differentia' together represent the essence of the kind (Fig. 1).

Naturalness, essences, and the epistemic role of observation

To Aristotle, observation and accurate description of biological individuals (i.e., single specimens) obtain a central epistemic function in natural philosophy. Following Aristotle, the study of individual natural phenomena is a substantial prerequisite for a philosophical representation of the natural world. Thus, it is not surprising that Aristotle bases his definitions on observation and the study of specimens. His terminology is clearly grounded in observation. According to Aristotle, essences are real physical properties that can be, in principle, discovered through observation.

However, whether a given morphological trait of an organism represents an essential property of its corresponding kind is not directly testable and had to be inferred through comparison and empirical investigations. Aristotle's concept of essences lacked clearly and unambiguously applicable recognition criteria, with the consequence that statements about essences always remain hypothetical and cannot be sufficiently validated in principle.

Aristotle's essentialism had a major influence on all subsequent classificatory attempts in biology and represents a paradigm concept for most biologists to follow. After Aristotle, every biologist who wanted to define a taxonomic group or a morphological trait had to deal in one way or another with Aristotle's concept of essence. Thereby it underwent several major alterations, all of which also significantly influenced morphological methodology and terminology. It is no overstatement to say that what follows in the history of biological classification can be characterized as an enduring reaching out for Aristotelian essences, which at its turn strongly influenced the epistemic status of morphology and the conceptualization of morphological data.

The influence of medieval hermetism

While during the time of Aristotle empirical investigations take on an important role within sciences, this fundamentally changes at the latest in the Roman Empire of the second century. Culturally it is a melting pot of most diverse peoples and languages, a mixture of different ideas and ideologies, in which, officially, all different kinds of religions and deities are more or less tolerated (Eco 1988), but in which the young Christian community was significantly growing and constantly gaining influence. In this time the concept of truth, as it was delivered from the Greek rationalistic tradition, experiences a large crisis. On the search for a *single* truth within the multiplicity of most different religions and cultures, hermetism gains importance and influence.

Many ideas, most of which were spiritually related, influenced the development of hermetic thinking. For instance the Jewish-Christian idea of the existence of a universal language that all peoples spoke until the building of the tower of Babel, when God decided to give the workers different languages to prevent them from finishing their work. This universal language was believed to be closest to the language of Paradise, in which, since it was the language of God, only the truth and nothing but the truth could be spoken. The hermetic idea was that every language carries pieces of this old universal language of Babel, with Hebrew being closest to it, followed by ancient Greek. Therefore, hermetists believed that manuscripts in different languages, preferably Hebrew and ancient Greek, all carry traces of the language of Babel and, thus, hidden pieces of truth that only had to be revealed and discovered by studying and comparing the texts. Thus, in hope that each book holds a spark of truth and that all books confirm each other someway somehow, hermetists focused in their search for truth exclusively on the content of books. Thereby, the principle of *tertium non* *datur* (i.e., excluded middle principle: a statement can only be true or false—either A=B or A≠B, a third possibility being considered to be impossible) is invalidated. Consequently, various things are regarded to be simultaneously true, even if they contradict each other, and empirical research loses its importance (Eco 1988).

If books, however, tell the truth although they contradict each other, each of their words is to be interpreted as an allusion, as an allegory. The universe is interpreted as a network of relations, in which each thing reflects and means all other things—a universe of universal sympathy, which a human being will only comprehend through a web of allegories.

Obviously hermetists were seeking the truth that is lying *behind* the objects and that *transcends* observation and description. In doing so, they did not use language as Aristotle did—words and things were not conceived as necessarily separate concepts and categories, but were believed to belong essentially together, connecting humanity, nature, and the divine with one another, with the ultimate goal to restore the paradisiacal union between them. This hermetic philosophy renders the inference of empirical evidence, with which one could eliminate or refute hypotheses, impossible.

As a consequence and in reference to the Christian influence on hermetism, the medieval notion of truth was more complex and multilayered than ours today: to a literal truth, a complex structured, transferred moral-spiritual truth was associated. On the one hand, this drastically constrained the possibilities of interpreting empirical phenomena and the number of potential explanations, since they were not allowed to contradict the religiously shaped world view of people of this time. On the other hand, they were tremendously extended due to the idea of a potentially infinite repertoire of possible mechanisms of godly intervention.

From the second century until the early Middle Ages, it is characteristic for biological classification that one does not differentiate between observation, document, and fable, which represents an obvious influence of medieval hermetism. Signs are understood as intrinsic properties of their denotations—they are thought to be essentially connected. Thus, following the ideas of hermetism, one assumes things possess at the most visible point of their surface *signatures* or markers, which are to indicate what is substantial (i.e., essential) with them. The signature assigns the meaning to the organism. Heart-shaped leaves of a plant, for instance, must possess a certain effect on the human heart due to their similarity, or they are at least related to it through some cosmic connection. Another example is the belief in mercury salve helping to treat syphilis because mercury is signed by the planet Mercury, which at its turn signs the market place where syphilis has been usually contracted.

Paradoxographies and early encyclopaedists

The classical natural miracles and marvelous creatures of the paradoxographies, to which the Blemmyes (headless creatures with their mouth on their belly), the Sciapodes (who like to lie in the shade of their single large foot) and the Cyclopes (one-eyed creatures) belong, represent an ancient literary genre about which even Aristotle wrote and whose meaning and role is still uncertain (the Greek texts may have served as collections of samples for rhetoricians; Daston & Park 1998). Since paradoxographies were considered to represent essential components of any encyclopedia they have been copied into each medieval encyclopedia and were presented therein as knowledge to be true without any critique—obviously, people of the Middle Ages actually believed in the existence of the fabulous creatures of the paradoxographies.

Starting from third and up to the fourteenth century, the typical compilers of historical and topographic encyclopedias rarely left their own hometown, but, instead, supported their work solely with the content of older encyclopedias. In doing so they knew to tell tales from real and legendary countries inhabited by all kinds of fabulous animals and strange creatures, including those from the paradoxographies (Daston & Park 1998). Thereby, the influence of the Christian world view also had a strong impact. The lion, for instance, was believed to possess the habit to obliterate traces with his tail so that no hunter can take up his track. This proposition was believed to be true since the lion was considered to be a symbol of Christ who, for his part, erased the sins of mankind. In this way Christian hermetism provided a rational basis that gave a certain internal coherence to this proposition. The same applies to Phoenix, a bird-like creature that burns every 500 years on an altar and three days later resurges again from his own ashes, because in the Middle Ages the Phoenix represents the symbol of the redeemer (Eco 1988).

At the end of the twelfth century some scholars and naturalists start to be critically opposed towards the ancient knowledge. Gerald of Wales (lat. Giraldus Cambrensis) for instance stresses the importance of personal experience of flora and fauna. He is one of the first to criticize the mistakes of antiquity and to doubt the existence of the fabulous creatures described in the paradoxographies. For the first time a medieval topography does not exist as pure literary compilation—Cambrensis' work on the history and topography of Ireland, written around 1185, is enriched with personal observations and criticism of Bede, Solinus, and others.

This development has also been influenced by the fact that with the eleventh century the time of the extensive oriental expeditions and voyages begins. Europe opens its cultural borders and ends its isolation as a consequence of, for instance, the opening of the commercial routes and the Mongolian peace. Missionaries, ambassadors, mercantilists, and researchers start to travel distant countries.

Nevertheless, to many authors it still applies: what is to be found in exotic countries is written in the ancient books and the Bible. Therefore, it does not surprise that at the beginning of the fourteenth century illustrations were still added to the report of Marco Polo's journey to China. Those illustrations were added by third persons and showed creatures like the Blemmyes, the Sciapodes, and the Cyclopes, which, following the ancient writings, were to be expected in India, although they had not been described by Marco Polo (Eco 1988). This impressively points out the enormous authority that the ancient writings had in this period of time and the negligible value that was given to observation and personal experience.

The influence of medieval hermetism extended even into the seventeenth century. In books such as his *History of Serpents and Dragons* (Aldrovandi 1648), the influential Ulisse Aldrovandi, professor in Bologna and owner of the largest and most famous collection of naturalia in Europe during the seventeenth century, still mentions for every creature and organism listed the recommended method to catch it, its allegorical use, on which coat of arms it is to be found, known legends and narrations, as well as the best way to serve it with sauce. All this information is listed in addition to and mixed within pieces of knowledge about its biology, following no specific categorical order. Thus, even well into the late Renaissance, organisms are still understood hermetically within the *entire* semantic net that connects them with the world, thereby *not* distinguishing words and things as necessarily separate concepts and categories.

Naturalness, essences, and the epistemic role of observation

With the rise of hermetism, the epistemic role of observation fundamentally changes. This has a major impact on biological classification. While Aristotle gives observation a central epistemological function, his early medieval commentators do not follow him therein—to them all observation that is based on single occurrences is accidental. Observation and empirical comparison no longer serve as a basis for classifications. Definitions of kinds of biological entities are no longer given in terms of physical properties only. Hermetic essences are physical properties that have to bear spiritual meaning. Morphological terminology is not anymore grounded only in observation. What is considered to be real is not necessarily observable but must be referable to the Bible or ancient texts.

Therefore, the medieval philosophers do not regard it as their task to observe natural phenomena and to discover explanations for them. They rather dedicate their time to the study of the universal truths, which they believe to exist as a network of relations of analogy, all dependent on an underlying, singular big metaphysical cause. They were convinced that the universal truths can only be discovered by studying the books or by receiving them from their teachers, which they then only had to improve and refine. This method is also known as *doctrina* because it depends on the passing on of knowledge by instructors (*doctores*).

Emancipation from the ancient heritage

With the invasion of the Iberian Peninsula by Islamic Moors in 711 and the following establishment of the emirate of the Ommiades of Córdoba in 756, the cultural exchange between Arab and European cultures dramatically increased. In the Islamic cultural centers (Baghdad, Damascus, Cairo, Mecca, Samarqand) Oriental and Hellenic knowledge was merged by the Arabian scholars and provided the foundation of an Islamic science, which reached its climax in the ninth and tenth century. Among Arabian Moors who occupied Spain there were scholars who earned their living by translating into Latin ancient texts that had previously been translated into Arabic. This contributed considerably to the intellectual Renaissance in Europe.

A major consequence of this influence is to be seen in the introduction of new opinions about the order of nature, in which nature was seen no longer as the direct expression of divine arrangements, but was thought to be subordinated via internal orders formed by causal chains. With Adelardus and other scholars of that time, the idea of an autonomous natural order was brought into the scientific discourse. This marks an important break with medieval hermetism, as it implies the assumption that the structure of nature and the universe must exist *independent* of humankind and human culture. This represents a precondition for the establishment of *ontological objectivity* (see e.g., Daston & Galison 1992).

The discovery of the 'New World' and the collections of pharmacists and physicians

A significant influence, not only culturally and economically, but also regarding scientific methodology was provided by the discovery of the '*New World*' of America by Christopher Columbus in the year 1492. An immense quantity of unknown species and exotic naturalia are made accessible for study. Girolamo Cardano, lawyer, mathematician, professor of medicine in Pavia and Bologna, and owner of a collection of naturalia, refers to naturalia from America to impressively point out the gaps of ancient knowledge (Cardano 1557). Insights of this kind lead to the consequence that from this point the authority of the Greek and Roman authors experienced a collapse.

Giovanni Battista Olivi is one of the first who studies the naturalia of the New World for their own sake (Olivi 1584). Others will follow him. Gradually, something like the discipline of *natural history* develops, emancipating itself initially only slowly from medicine. In natural history, natural objects are investigated independently from their possible therapeutic applications. With the emergence of the discipline of natural history the great time of biological classifications begins.

Natural history is understood to represent the task of moulding language as to represent things – natural history cares about an unambiguous *designation of the visible*. Regarding content and meaning, natural history is basically a classificatory discipline that deals with the description and classification of plants, animals, and minerals (Kanz 2002). Later, by Francis Bacon, this new discipline will be given an important role within his reformed system of natural philosophy, the 'new philosophy'.

The encyclopedic work of Jan Jonston (1657) represents an impressive example for the corresponding change in thinking, in which a majority of *formerly* essential entries are now missing, which would have been listed by, for instance, Ulisse Aldrovandi. Jonston limits his description of organisms to known biological data as well as the possibilities of the utilization of the organism for human purposes — well-known legends and the like are not thought to belong to the organism in principle and are therefore ignored. Henceforth, as a consequence of the advent of the discipline of natural history, observation and its documentation are separated from fables, but are from then on understood as representational types in their own right. This represents a very important step towards the establishment of *ontological objectivity* (Daston & Galison 1992) in biology.

Bacon's empirical facts

By the growing popularity of scientific collections and the increasingly effective publishing of scientific essays due to the new book printing techniques, contacts among scientists increase within Europe, crossing cultural and national borders. As a result, an active scientific exchange develops. This is the ground on which, during the seventeenth century, a new epistemology of empirical *facts* develops together with a community of empirical researchers.

Francis Bacon was one of the leading figures of this ‘new philosophy’. He had the idea that nature has just to be interpreted correctly, that nature tells its own tale and scientists only have to learn how to listen to her. Nature would reveal herself as soon as scientists manage to rule out their intervention (Galison 1998). This is based on the idea that the ultimate structure of reality is independent of humankind and human culture (*ontological objectivity sensu* Daston & Galison 1992).

In trying to do so, Bacon conceives natural history on the ground of his own epistemological system, in which natural history takes in an epistemological key function (Spedding *et al.* 1857–1874). According to Bacon’s *Novum organum* (Bacon 1620), natural history must serve as source for facts and as the empirical foundation for natural philosophy. Facts represent a new category of scientific experience, detached from explanations, illustrations, or conclusions. Following Bacon, scientists should put aside their terms and should start to deal with facts.

Bacon’s empirical facts are grounded in observation. However, observation is always spatio-temporally fixed and therefore represents a localized event. Observation not only depends on the conditions present at a specific location but also on a given individual with all its idiosyncrasies. Thus, the important question that Bacon had to answer was, how necessarily subjective experience can be transformed into objective empirical facts. Which conditions have to be met to turn a ‘view from somewhere’ (Porter 1992) into a ‘view from nowhere’ (Nagel 1986)?

According to Bacon, the process from the study of particulars to the identification of facts is considered to be burdensome, since the senses have to cross the corset of a strict method of order, processing, and evaluation to be immune against deceptions. Bacon develops boards, so-called *praerogative instances*—boards of agreement, differences, gradations, and repulsion of properties in nature. They are to serve as methodological aids to the discovery of the laws of natural properties and their fundamental forms. The praerogative instances provide the methodological core to Bacon’s conception of natural history, which in its turn was conceived to be free of any theory, and to represent one of the first attempts to establish *aperspectival objectivity* in science.

Aperspectival objectivity (see Daston 1992, 1998; Daston & Galison 1992; *procedural objectivity sensu* Heintz 2000; for a critique of the claim for perspective-independent objectivity in science see, e.g., Kukla 2006) is about communicability of scientific results and claims that something is more objective than something else if it relies less on the specific individual who generated the results, their social position and character.

As a consequence of Bacon’s conception of natural history, observation and its documentation in form of descriptions are *methodologically sharply separated* from the conclusions drawn from them, including all explanatory hypotheses and theories. This represents a major methodological improvement. The focus on facts and their consistent distinction from conclusions should also prove to be favorable for the just emerging scientific societies and their meetings, since one can usually talk about facts more objectively and without personal arguments than it is possible with theories. Bacon’s ‘natural histories’ subsequently developed into the ‘*facts*’ of natural philosophy of the late seventeenth century, which in their turn represent the attempt to generate pure descriptions that should be free of any theory or conclusion and which represent the precursor of our modern notion of empirical data.

During the seventeenth century more and more scientists and philosophers became aware of the subjectivity of perception. Locke (see also Galilei, reprinted in Drake 1957), for instance, distinguishes between primary qualities such as shape, size, distance, solidity, and volume that, according to him, exist in the external world in the same way humans perceive them, and secondary qualities such as color, taste, texture, smell, and sound that, following him, do not exist in things themselves but depend on the perceiver’s senses (Locke 1689/1979). Following this distinction, Bacon’s natural history had to focus on primary and not on secondary qualities.

Naturalness, essences, and the epistemic role of observation

During the seventeenth century more and more scientists and philosophers became aware of a) the subjectivity of perception and b) the requirement of natural philosophy to be grounded on perception. As a consequence, observation and empirical comparison regain their epistemic role. But scientists are at the same time suspicious of singular observations of individual specimens and apply rigorous methods and techniques. They believe that to discipline observation requires the distinction of different observational categories in order to minimize its subjectivity (e.g., by distinguishing primary and secondary qualities). Consequently, essences are considered to exist *independent* of human beings (ontological objectivity) but require specific methodological procedures for their identification and documentation (aperspectival objectivity).

The age of classification

Bacon also had a major influence on biological classification. Jungius, a student of Cesalpino, started to develop criteria for generating something like Baconian empirical facts for morphology. He applied a philosophically derived rationale to the observation of plants. Jungius believed that well defined terms, just like numbers in mathematics, represent stable and objective values. Based on this assumption, he defined clear and unambiguously applicable botanical terms that he stripped off of all untested physiological interpretations, thereby clearly separating observation from conclusion (Jungius 1678; von Sachs 1875; Beck 1969). Jungius' goal was to develop a *primary language* (i.e., a purely descriptive one), with which observation and its documentation through descriptions should become independent of the respective individual observer and universally communicable. By standardizing morphological terminology and descriptions of morphological traits, Jungius attempted to establish *aperspectival objectivity* (Daston & Galison 1992) within botanical morphology and classification.

The standardization of experimental and linguistic practices becomes more and more popular during the nineteenth century, thereby stressing the importance of *communicability* of scientific results among scientists. Objective knowledge comes to be defined as communicable knowledge and requires scientists to standardize their methods of measurement and communication. Jungius thereby follows the most important strategy for linguistic standardization: quantification and formalization—to use formulae, numbers, and graphs whenever possible (Heintz 2000).

One of the first to adopt Jungius' terminology for the purpose of biological classification was Ray. Thereby Ray followed Cesalpino (1583) and Tournefort (1694) in focusing on the properties of flowers and used only distinct and exactly definable properties of morphological traits for classification (Ray 1703).

Cesalpino, Jungius, Ray, and Tournefort prepared the ground for Linnaeus and his new and very successful approach to biological classification. Linnaeus reduced and limited observation to a few categories only, so that one is left not only with what is analyzable in the somewhat confusing opulence of representations but also with what anybody can recognize and identify and, thus, with what can receive a name that everyone understands (Linnaeus 1735, 1751). With his sexual system (i.e., *Clavis systematis sexualis*; for more detail see below), Linnaeus proposed a very pragmatic way to make biological classification a less subjective procedure, at least for botany. Linnaeus' sexual system provides a theory for *taxonomic character* that takes in the function of a *secondary language* (i.e., analytical/explanatory), with which classification should become independent of the individual taxonomist.

Biological classification before Linnaeus is characterized by a plurality of contradictory approaches (Ereshefsky 1997). Actually, Heywood (1985) concludes that before Linnaeus biological taxonomy must have been a rather chaotic discipline, which stands out by its miscommunications and misunderstandings. The tremendous success and the broad acceptance of the Linnaean system are probably due to his rather pragmatic choice of criteria for the conceptualization of classification (for his contribution to the theory of biology see Müller-Wille 1999). Linnaeus' approach provided comparatively clear and simple rules for the construction of classifications, which also included rules for the denomination of species and taxa. This significantly increased the possibility of communication among taxonomists (Ereshefsky 1997; Stevens 1997).

Naturalness, essences, and the epistemic role of observation

The ‘new philosophy’ of Bacon, Descartes, and Locke, had a major influence on taxonomists. In order to satisfy the claim of *transparency* and *reproducibility* that accompanied this new style of doing empirical research, taxonomists strived for making classification a less subjective procedure by relying on mathematics and logic. They limited themselves to the study of distinct, clearly and unambiguously definable morphological traits, with which they attempted to establish a terminological standard in morphology-based classification. This also improved possibilities to *communicate* one’s findings and discuss and agree upon possible classificatory relationships (i.e., aperspectival objectivity). As a consequence, essences are understood to represent real physical properties of traits that are observable, unambiguously describable, and that can be identified independently from a particular taxonomist.

Causal reasoning and the role of function in classification

Development of the laboratory method

Medieval natural philosophers strive for *scientia* (e.g., theology and theoretical medicine), which, according to the definition of Aristotle, represents safe knowledge based on classical syllogisms and, thus, ultimately on postulation and deduction. In contrast to *scientia* stands *artificium*, which is associated with handicraft and which represents another form of researching that is based on experience and the development of laboratory techniques and instruments. Instead of the safe knowledge of *scientia*, *artificium* was considered to only be able to produce reliable opinions, as can be found in practical medicine and agriculture.

Due to the increasing urbanization and the increasing trade at the end of the fourteenth century, the urban bourgeoisie flourishes and a solvent commercial, crafts, as well as intellectual urban elite emerges. These elites let the market for professional medical supply boom. As a consequence of this development and due to its increased social importance, the field of medicine that is concerned with the diagnosis, the description, and the treatment of individual diseases, receives an enhanced position. Noble patrons promote for instance the research on spas, i.e. medical springs. Giovanni Dondi, physician and professor of medicine, is engaged in the observation and description of thermal springs in the proximity of its hometown Padua (Dondi 1372–1374). Giovanni is not alone as Ugolino da Montecatini (1471), Michele Savonarola (1448–1449), and many others of that time were busy with the study of thermal springs as well. During their research they quickly become aware of the fact that the classical methods of *scientia*, which reduce their reasoning to deduction from first principles and the study of literature, are not suitable for the study of thermal springs. Thus, only with the aid of empirical experience and the application of the methods of *artificium* could successful research be accomplished in this field, leading to the development of methods and techniques for experimental exploration and measurement of more complex observable relations. Thus, it is in the tradition of *artificium* that methods of experimental exploration and measurement developed, which represent the core of the emerging new *laboratory method*.

Influenced by the principles that Francis Bacon introduced in his *Novum Organum* (Bacon 1620), the ‘Invisible College’ was founded in the seventeenth century—a precursor of the Royal Society of London. Its members devoted themselves to Bacon’s ‘new philosophy’ and were dedicated to acquire knowledge through experimental investigation (Gingrich 2004). Among its members was Robert Boyle. Assisted by Robert Hook, Boyle went through numerous modifications and changes in design and construction of his air pump, finally leading in 1659 to his ‘Pneumatic Engine’, with which he began to run a series of experiments. Through this experimentation Boyle discovered that the volume of gas varies inversely to its pressure. This, by now famous law resulted from his extensive experimental work with the air pump (Boyle 1660). Boyle’s experimental approach marks a change in paradigm in science, representing a consequence of what is often referred to as the scientific revolution. It is paradigmatic for the enthronement of empiricism and experimentation as primary instruments for gaining knowledge (Shapin & Schaffer 1985; Shapin 1996).

In the beginning of experimentation, experimental results had to be validated and authenticated by trustworthy witnesses. At that time, trustworthiness of a witness was primarily defined in terms of their social position—only the unbiased judgment of a *gentleman* can witness experimental results. Scientific societies take in a central position in the establishment of this *social objectivity* (Heintz 2000). With an increase of standardization of experimental methods and techniques, with the development of instruments for data production, which replace the scientist as an observer, and with the (international) standardization of measurement procedures as well as measuring units, *gentlemen*-science is suppressed and social objectivity is replaced by mechanical (and aperspectival) objectivity (see Heintz 2000). *Mechanical objectivity* (Daston & Galison 1992; *methodical objectivity sensu* Heintz 2000) requires the ruling out of all individual and subjective influences of body and mind and forbids judgment and interpretation in documentation of observation and reports on it (Daston & Galison 1992).

Causal reasoning

Isaac Newton formulates the first of the 'rules for the operation of comprehension', a call for economy of thinking in natural philosophy, using only as many causes for the explanation of natural phenomena as necessary (Newton 1687). This methodological principle, also called parsimony, which goes back as far as to Aristotle, but is commonly attributed to William Ockham ('Ockham's razor'), enjoys at the end of the seventeenth century increasing popularity. Parsimony provides the methodological restriction for the choice of the best explanation which is necessarily required for empirical research. With it, functionality moves into the focus of naturalists—the simplicity of nature and the economy of their instruments are connected with the sobriety of purposes (Daston & Park 1998). The conception of the regularity of causes is associated with that of the regularity of effects. Also among the anatomists of this time, the coupling of morphology/anatomy and functionality plays a prominent role. To them, function begins to represent the most important issue. Thus, already in 1718, the Parisian anatomist Jean Mery thinks of the 'machine of the human body' (Mery 1718).

The role of functional morphology in classification

In the distinct and usually very reliably and easily preservable forms of plants, natural history finds an ideal object for its research (additionally, by mailing seeds, specimens could by traded between botanical gardens). This contributes substantially to the boom of botany, resulting in the formation of many botanical chairs at the universities during this time. The actual practice of classification, however, proves that not every trait provides differentiating properties and can serve as a taxonomic character—not every property represents an essential property. The scientific task of taxonomists therefore consists in finding suitable traits in order to receive accurate names for the objects to be classified. This brings up some difficulties. One has to ask oneself what is 'suitable'—a problem that results from the concept of essence lacking clear recognition criteria. As a consequence, a multiplicity of different classifications are conceivable, each of which is based on different sets of traits. Michel Adanson (1763), for instance, came up with 65 different classification systems this way.

Classification and taxonomy prior to causal reasoning arranged the knowledge about organisms according to the possibility of representing them within a system of names. As a consequence, many different classifications are possible and their scientific value is evaluated on grounds of rather pragmatic criteria (e.g., Stevens 1997). Classification experiences a major change in paradigm with the development of the idea of a *hierarchical* natural order that can be discovered by taxonomists through causal reasoning, in which function takes on a central role.

At the beginning of the nineteenth century the term and concept of *organization* develops, which refers to the internal physique and physiology of an organism. The term and concept of *function* receives specific attention, since it relates individual morphological traits to the entire organism and, thus, to other morphological traits. The morphological traits together with their functional relations establish the organization of the organism. On the basis of the concept of organization, a trait can be evaluated in reference to its functions and their importance to the organism: important traits provide functions that are essential for the survival of the organism, whereas less important traits do not. Thus, by defining an internal law (i.e., the function), organization mediates between *morphological trait* (i.e., primary language, purely descriptive) and *taxonomic character* (i.e., secondary language, analytical/explanatory). Function is understood to permit a certain trait to adopt the value of a taxonomic character. As a consequence, classification receives a completely new conception. With the criterion of function classification becomes a natural system. The hierarchical order of classification is neither merely given by God nor only dependent on the cognitive constraints and requirements of humans anymore. Function determines the hierarchy within the system.

Cuvier, the developer of the modern discipline of comparative anatomy (Coleman 1964), assigns sets of organs to a specific *function* and tries to reveal similarities by comparing them (Cuvier 1800, 1817, 1825)—although Linnaeus justified the epistemic role of his sexual system as providing the key traits for classification on the function of the sexual organs for reproduction, thereby following the tradition of Cesalpino, this reference to function is owed to an Aristotelian concept of 'being' (see, e.g., Larson 1967) rather than a modern causal account of function. According to Cuvier, function is not assigned to anatomy anymore, but instead receives primacy—which was the matter of debate of the famous argument between Cuvier and Geoffroy Saint-Hilaire in 1830 (see e.g., Appel 1987), in which Geoffroy Saint-Hilaire argued that the body plan of an organism constrains how organ functions are manifested (i.e., form determines function), whereas Cuvier argued that function determines how organs are designed (i.e., function determines form). According to Cuvier, any similarities between organisms are due to common functions. During his comparative studies of the internal organization of organisms Cuvier noticed that individual organs can no longer be conceived without the other organs and that all other organs would have to change as soon as one of them changes (Cuvier 1800; principle of functional correlation, Russell 1916). Thus, Cuvier introduced the idea of organic integration into biological thinking, which is fundamental for our modern conception and understanding of the organism (Fristrup 2001).

Furthermore, Cuvier is convinced of being able to recognize an internal hierarchy of morphological traits, with some traits possessing a greater functional importance for the organism than others (principle of subordination of characters; e.g., Coleman 1964; Farber 1976; Eigen 1997). On the basis of the most important function he tries to find the most important type of taxonomic character. This type of character, in its turn, has to serve as foundation for the higher ranked taxa in a classification. In doing so he initially dedicates himself to the investigation of blood circuits, then to digestion, and later to nervous systems. On the basis of having identified four types of nervous systems, Cuvier classifies animals into four distinct basic classes, the '*embranchements*' (i.e., morphological types—for a discussion of Cuvier's type concept see Eigen 1997): Articulata (i.e., arthropods and segmented worms), Mollusca (i.e., all other soft-bodied bilaterally symmetrical invertebrates), Radiata (i.e., cnidarians and echinoderms), and Vertebrata.

Naturalness, essences, and the epistemic role of observation

Many different and mutually contradicting classifications were proposed for animals on the basis of Linnaeus' method of classification. But also regarding the classification of higher ranked taxa in botany Linnaeus' sexual system failed to provide a consistent solution, as he himself had to admit (Linnaeus 1751; see also Larson 1967). This was unsatisfactory and required the introduction of additional criteria to classification. With the advent of causal reasoning and physiology in biology, the investigation of functional relationships within an organism became more and more important. This influenced not only classification but also morphology. As a consequence, essences were understood to represent real physical properties of traits that serve *functions* which are very important for the *survival* of the organism.

The problem with essences and its unsatisfactory solution

This brief introduction to the history of biological classification illustrates the important role that the concept of essence had in biological classification and morphology. With his concept of essence, Aristotle provided a means to rationally distinguish between artificial and 'real' kinds, thereby allowing for abstraction and generalizations over the overwhelming diversity of biological beings. Thus, it is not surprising that Aristotle's essentialism played a very important role in biological classification ever since and that it significantly influenced morphology too. Unfortunately, already when introduced by Aristotle himself, the concept of essence was somehow ill-conceived and could not be satisfactorily clarified. This might be due to the fact that the concept of essence, which refers to specific properties of a given kind of things, has always been linked to the concepts of species and taxa, to which essences are supposed to provide diagnostic criteria. Considering the problems that biologists had and still have with agreeing upon a sound concept of species and taxa (for an overview see, e.g., Claridge *et al.* 1997, especially the contribution of Mayden therein; Wilson 1999, especially the contribution of Boyd therein; Pleijel & Rouse 2000; Wheeler & Meier 2000; Pigliucci 2003), the enduring ambiguity that accompanies the concept of essence is not very surprising.

As for the concept of species and taxa, the concept of essence brings about two problems. First, the *ontological* problem of what exactly is an essence (i.e., its theoretical definition), and second, the *epistemological* problem of how to recognize and identify essential properties (i.e., its recognition criteria). Aristotle's conception was unclear regarding what exactly essences are ontologically, except that they were considered to be real physical properties. In order, in theory at least, to be able to distinguish essential from accidental properties, Aristotle had to assume the existence of an *invisible* 'nucleus', the 'substance', that every thing possesses and that bears essential properties. Other than that, essential properties could only be discovered through observation and comparison, guided by the classificatory method of logical division *per genus et differentiam*. Unfortunately, this procedure is ambiguous and can result in many different, mutually contradictory classifications. This does not necessarily pose any problems, as long as one does not assume the existence of a natural order, which would allow only one classification to be true and all others to be false.

The emergence of Christian hermetism did not really help to clarify the concept of essence, either. Quite contrary, on the basis of understanding fables as another type of property possessed by an organism, sharing the same epistemic status as biological properties, medieval hermetists modified the definition of essence to bear spiritual meaning. The invisible 'nucleus' was not merely the bearer of real physical properties anymore, but also of relationships of similarity to spiritually meaningful things and characters from the Bible or from ancient Greek texts. Since its cultural connotations are understood as essentially belonging to an organism in the same way as its biological properties, reality *cannot* be understood as existing independent of humankind and human culture (i.e., ontological *subjectivity*). Since the Bible and ancient Greek texts are considered to provide the main source for gaining knowledge, observation and empirical investigation were not sufficient anymore to identify essences, and the applicability of the concept of essences was further hampered.

The increasing exchange between Christian European and Islamic Arab culture marks the turning point in European history of science. Arab scholars introduced the idea of a natural order that exists independently from humankind and cultural knowledge. As a consequence, names and fables have to be separated from biological knowledge, because they are now considered to belong to different ontological categories. This marks the beginning of the establishment of ontological objectivity in Western science and philosophy and a break with one of the central paradigms of medieval hermetism.

The scientific revolution, with its claim of transparency and reproducibility, and Linnaeus' new approach to taxonomy, with its focus on communicability, marks a significant step forward towards clarification of the concept of essences. By identifying the problem of subjectivity of individual observations and the necessity to clearly separate observation from explanation, methods for data production and documentation became more important. From then on, essences were considered to be real physical properties that are observable, unambiguously describable and that can be identified independent from individual morphologists. As a consequence, morphological terminology became more standardized and morphologists became aware of the critical role that language plays in objective data representation. The age of classification, with such protagonists as Jungius and Linnaeus, marks the high time of aperspectival objectivity in morphological terminology. When comparing their standards with current standards of morphological data conceptualization and documentation, the question immediately arises why the standards that they developed have not been reached ever since?

Unfortunately, in the long term, the trend of objectification of morphological terminology did not yield the expected success in biological classification, at least not in zoology. Cuvier, who regarded Linnaeus as the greatest genius in biological classification (Eigen 1997), realized that Linnaeus' sexual system fails to provide the foundation for a classification that unambiguously represents the existing natural order. As a consequence, and along with the hype that accompanied the emergence of causal reasoning, taxonomists were eager to further modify the concept of essence in classification, understanding essences to represent real physical properties of traits that serve functions that are very important for the survival of an organism. This can be interpreted as an improvement of the theoretical definition of the concept of essence, since it replaces the spiritual context of medieval hermetism and adds with function a component that is experimentally accessible to the idea of a 'nucleus'. However, comparative anatomical studies reveal that similar traits can have very different functions and equivalent functions can be fulfilled by morphologically diverse traits, suggesting that the relation between form and function is very flexible. Moreover, it is not clear how one can determine the importance of a function of a trait. In other words, recognition criteria for essential properties were still very ambiguous and, therefore, the application of the concept of essence still very problematic.

However, the idea of a hierarchical natural order of organisms allowed taxonomists to differentiate between accidental and essential properties on a *heuristic* basis *without* requiring reference to the functions of the respective traits: With the increase of comparative anatomical studies the idea develops that sameness relations between morphological traits can be differentiated into two different categories. On the one hand, there are those properties that appear to be accidentally equivalent. These properties, which are called *analogies*, occur isolated. Their sameness can be traced back to an equivalence of form and function (Rieppel 1993). On the other hand there are those properties that appear to be 'truly' identical. These properties, which are called *affinities* (Strickland 1840a, 1840b) or *homologies* (Owen 1843), can be distinguished from analogies by their occurrence in reciprocally corroborating aggregates. Starting point for this distinction was the idea that a hierarchical natural order of organisms would have to stand out because of the natural affinity of the corresponding organs (Whewell 1840). In other words, characteristic for affinities/homologies is that classifications based on different affinities/homologies tend to confirm each other by congruence.

Thus, by assuming a hierarchical natural order and, thus, the existence of real correlates for species and taxa, which in its turn provides a means to reasonably organize organisms into classes, an independent criterion for testing whether a trait represents an Aristotelian essence became available. A putative essential property can be tested against sets of other putative essential properties in terms of congruence. Obviously, for the first time in the history of the concept of essences, something like a (heuristic) recognition criterion is available. However, a satisfactory explanation for the existence of affinities/homologies was still lacking and, thus, the ontological status of essences (i.e., their theoretical definition) was still unclear. Nevertheless, the concept of affinity/homology was commonly accepted among nineteenth century comparative anatomists (Panchen 1999).

In the light of the theory of evolution, the concepts of affinity/homology and analogy experienced considerable modifications (for overviews see Hall 1994; Bock & Cardew 1999; Rieppel 1993). As a consequence, nowadays we understand (morphological) homologies as traits that share equivalent properties with one another due to common ancestry, whereas homoplasies represent traits that share equivalent properties due to other reasons, but not common ancestry (Lankester 1870). For the first time ever, with the theory of evolution the basis for an unambiguous, historically grounded theoretical definition of essence becomes available. From then on, homologous traits take in the role of essences in classification: whenever similarity between organisms is interpreted to be based on homologous traits, these traits are considered to be truly identical, providing the grounds for concluding identity of the trait bearing organisms, which at its turn establishes the identity of the corresponding species or taxon. Thus, the concept of homology provides a solution to the problem of essences, with which biologists have struggled for such a long time.

Morphological terminology and homology—the downside of the solution

Morphological methodology has been strongly influenced by the concept of essences. Thus, it is not surprising that we nowadays have to deal with a multiplicity of morphological terms that imply homology. The influence of essentialism on morphological thinking is so strong that even today most morphologists cannot imagine a morphological terminology free of homology assumptions. However, homology transcends the perceptually given by providing an explanation for the sameness of traits. If descriptions of morphological traits are based on homology assumptions, they depend on particular phylogeny hypotheses, which in their turn provide historical explanations for the perceived distribution pattern of sameness and differences of traits. As a consequence, much of morphological terminology is phylogeny-sensitive, requiring a change in terminology whenever a currently preferred phylogenetic hypothesis is replaced with another one due to new data, leading to a continuous change of meaning in many morphological terms. If morphological terminology is phylogeny-sensitive, morphological data are conceptually not clearly separated from conclusions.

Unfortunately, terminological standardization has been further hindered by lack of communication between morphological specialists of different taxa. As a consequence, morphological terminology has developed and grown independently within different taxonomic communities in the past and still does so today, with the effect that morphologists assign different terms to equivalent morphological traits or the same terms to different traits. This leads to a major problem regarding communicability of morphological data across large taxonomic ranges. Especially with respect to the comparative method, it represents a fundamental problem that morphology has to face. Unfortunately, Jungius and Linnaeus were the last biologists to successfully attempt to develop a general standard for morphological terminology. It seems as if morphology does not strive for the major achievements of the scientific revolution anymore: establishing a high degree of *aperspectival objectivity.*

What can we learn from Linnaeus

First attempts to establish a basic degree of mechanical and aperspectival objectivity in morphology can be seen in Locke's (1689/1979) distinction of primary (subject independent) and secondary (subject dependent) qualities in classification. In order to exclude some of the subjectivity that is necessarily connected to individual observations, taxonomists like Jungius, Tournefort, and Ray already excluded most secondary qualities from classificatory considerations. Furthermore, by focusing primarily on fructification traits, they also assumed that specific traits are more suitable for generating a consistent classification than others. As a consequence, the relevant area of matter for classification has been confined and restricted: to Jungius, Tournefort, and Ray not all morphological empirical phenomena were relevant to classification anymore. Distinguishing relevant from irrelevant phenomena, however, requires an epistemological criterion that goes beyond Locke's differentiation. This is where the *concept of structure* comes into play.

Linnaeus' morphological structure concept

Linnaeus' tremendous success can be traced back to four aspects of classification, to which Linnaeus made significant contributions—at least, when considering them in combination:

1) *Defining taxa* on the basis of Aristotelian definitions. Linnaeus defined taxa on the basis of five predi-

cates, which are derived from Aristotle's definition by *genus* and *differentia* (see e.g., Ereshefsky 1997; for a criticism that Linnaeus followed Aristotelian essentialism see Winsor 2003, 2006a, 2006b; for a reply see Stamos 2005):

Definition	A *statement* about necessary traits (i.e., essence)
Genus	Genus part of an Aristotelian definition, inherited from its parent taxon
Differentia	Distinguishing part of the definition
Property	The necessary traits as such (i.e., the taxon's essence)
Accidents	Typical traits that are not essential

2) *Classification* by Aristotelian logical division *per genus et differentiam*. Linnaeus offered clear and simple rules for constructing classifications. For pragmatic and other reasons, he introduced new ranks and rules for naming genera and species (Stevens 1997; Larson 1967). Linnaeus was also the first to propose a classification with a *strictly* encaptic hierarchy of non-overlapping classes, a Linnaean hierarchy, thereby significantly contributing to the theoretical advancement of biology (see Müller-Wille 1999).

3) Linnaeus' confinement to the sexual system—his *taxonomic characters*. Linnaeus owed a lot to the work of Cesalpino, Jungius, Ray, and Tournefort, who formalized morphological descriptions in botany and who already used fructification characters for botanical classification. Linnaeus was convinced that, due to pragmatic reasons (but see also Larson 1967 for his Aristotelian reasons), the various traits of a plant's sexual organs are best suited for botanical classification. Linnaeus considered them to be easy to work with, being most complex organs that incorporate many characters (31: calyx with 7 parts, corolla with 2, stamen with 3, pistil with 3, pericarp with 8, seed with 4, and receptacle with 4 parts; Atran 1990), which can be described precisely (Ereshefsky 1997). Thus, he used them as '*Property*' for defining botanical taxa.

4) Linnaeus' *taxonomic facts*—a botanical ***structure concept*.** Linnaeus described each of the 31 fructification traits according to four categories (Linnaeus's defining attributes; Linnaeus 1735, 1751; see also Larson 1967), each of which is based on a single perceptual judgment: 1) the quantity of observed elements (i.e., *numerus*), 2) their basic geometrical form (i.e., *figura*), 3) their relative size (i.e., *proportio*), and 4) their spatiotemporal distribution (i.e., *situs*). Applied to all fructification traits of a plant, one receives 31 descriptions, each of which consists of four 'values'. As a consequence, the description of fructification traits became parametrized. This not only established terminological standardization, but also a standardization of description which established a degree of aperspectival objectivity that was formerly not known to morphology and that has not been reached ever since.

The combination of these four aspects of Linnaeus' method of classification allowed plant taxonomists of his time to arrive at similar conclusions and to unambiguously communicate their morphological findings (his method failed, however, on the level of orders and classes, as Linnaeus himself had to admit; see Larson 1967).

Linnaeus *taxonomic facts* are obtained as a result of the concision and reduction of perception to four *categories* that exclusively refer to Locke's primary qualities. On the basis of observation, comparison, and *perceptual judgment*, morphologists decide which 'value' a given trait adopts. One could also say that each category poses a question that can be answered in reference to morphological investigations. A category is only applicable, and thus a trait only describable, if morphologists can unambiguously assign a specific 'value' to it. In other words, organizing morphological descriptions on the basis of these categories forces morphologists to make clear *perceptual judgments* in reference to criteria that demand mathematization or formalization, and thus always a standardization of statements about traits and their properties. The 'values' that a trait obtains in a description should be independent of the individual morphologist—the ideal would be that different morphologists assign the same 'values' to a given trait. The respective 'values' of a trait thus represent 'facts' about the trait. In combination, the four values or variables—one for each category—describe what represents the *morphological structure* of a trait.

The idea of restricting morphological descriptions to a predefined set of categories and their corresponding value-spaces represents an ingenious way to deal with the overwhelming diversity of morphological traits. The restriction and abstraction of the phenomenal field to only those phenomena that can be grasped by the structure concept not only *translates* morphological diversity into standardized and analytically accessible bits of information, but also establishes a high degree of communicability and comparability of morphological data and, therefore, a high degree of *aperspectival objectivity.*

Developing a general structure concept for morphology

What is 'structure'?

Structure, in general, can be understood as a fundamental notion covering the observation, recognition, dependencies, and stability of patterns and relationships of objects and processes. The concept of structure is as old as Western philosophy and science and provides an indispensable foundation of nearly every mode of inquiry and discovery in science, philosophy, and art (Pullan 2000). The term 'structure' evokes connotations of organization, connection, orientation, framework, and others—but it is *order,* which is most central to the concept of structure (Pullan 2000).

The set of relations between different parts and aspects of a given complex whole determines the latter's structure. Structure represents a way to conceive properties and relations of a complex whole, and without some notion of structure it would be very difficult for anybody to develop a conceptualization of something.

While a general notion of structure provides a general concept for structuring the overwhelming diversity of the phenomenal world, thereby mediating between phenomena (as representations of sense impressions) and their corresponding concepts (as representations of real objects and processes), when dealing with the world we live in, a potentially infinite plurality of specialized structure concepts have necessarily to be developed (Pullan 2000).

Basic principles for developing a structure concept

A structure concept is developed with a specific practical purpose in mind: it should facilitate in generating data of a specific type and quality that are relevant for a specific scientific discipline, research program, or investigation. In order to successfully develop such a proper structure concept, it is inevitable to understand the characteristics of high quality data for the given scientific question and theoretical framework.

Probably the most basic characteristic of data, commonly shared by most if not by all fields of empirical research, is that it documents some sort of observational experience, conducted by either a human being or by instruments and machines. However, scientifically relevant observational experiences usually involve information about properties and relations of real objects and their behaviour. Thus, on a very basic level, data represent descriptions of properties, relations, and behaviour of specific types of objects and processes. As such, they represent descriptions which are existence statements that do not only go beyond the necessarily private phenomenal world of an observer's experience, but also beyond descriptions of particular phenomena. Instead, these descriptions represent hypotheses about the existence of entities and their properties, which are based on *observational judgments*. In other words, these descriptions provide answers to questions regarding the entity's properties and relations, such as for instance what shape does the entity have; what is adjacent to it; whether it is continuous with some other entity; what is its temperature; how does it react to exposure to light.

A structure concept should provide a method for standardizing and formalizing such descriptions. It should consist of perceptual categories, which pose questions that can be answered in reference to empirical investigation, observation, and measurement. Ideally, the structure concept is formalized to a degree that it restricts the observer in what is allowed as an answer for each question posed by the structure concept. In other words, the structure concept should provide a set of empirical questions (i.e., categories) and with each question a set of ‘values’ (i.e., a ‘value-space’) that are allowed as an answer. Each ‘value-space’ is determined by a range of allowed numerical values (e.g., natural numbers), Boolean values (i.e., ‘*YES*’ or ‘*NO*’), or by a limited set of defined terms (i.e., a controlled vocabulary). Linnaeus’ sexual system, whose application was restricted to sexual organs of plants, came close to being such a formalized structure concept.

A clear and unambiguous structure concept should furthermore provide criteria for distinguishing relevant from irrelevant information. This includes discounting all information that does not meet previously specified and commonly accepted criteria for objectivity (i.e., ontological, aperspectival, and mechanist objectivity) as well as differentiating between phenomena that refer to real entities that are relevant to the ongoing investigation from those that are irrelevant. As a consequence, a specific structure concept necessarily always depends on the theoretical and methodological framework of a given investigation and is therefore always context-dependent. In other words, for different scientific purposes and different domains of matter, different structure concepts have to be developed and applied.

Foundations of a morphological structure concept

In order to develop the foundations for a general morphological structure concept, some questions have to be addressed first. The first question to be answered is what morphological data represents. As I have argued above, descriptions in form of existential statements grounded and substantiated in observation and experimentation represent empirical data. In the context of phylogenetics, many morphologists consider phylogenetic characters and character matrices to represent morphological data. Taking Bacon’s separation of empirical facts from scientific conclusions (i.e., explanatory hypotheses and theories) as a paradigm of scientific objectivity, however, phylogenetic characters cannot represent morphological data in the strict sense, since they incorporate homology hypotheses (i.e., putative character and character state homologies; Brower & Schawaroch 1996; see also Freudenstein 2005) and are therefore explanatory and not purely descriptive. Unfortunately, the documentation of morphological facts as discrete characters and character states becomes more and more popular among biologists, especially in the context of morphological data bases, and seems to become a standard for summarizing comparative morphological data (e.g., Ramírez *et al*. 2007).

Images of morphological traits, just like morphological character matrices, do not represent morphological data in the strict sense, either. An image cannot represent data since, as long as no description accompanies the image, its perception remains stuck in the necessarily subjective private phenomenal realm, which is to a large degree open to personal interpretation. Thus, only morphological descriptions qualify as morphological data in the strict sense.

The second question to be answered is what properties morphological descriptions should have in order to meet standard criteria for mechanical and aperspectival objectivity in morphology. First and foremost, morphological descriptions require a highly formalized and standardized morphological terminology. However, many morphological terms presuppose homology of traits. If the correct application of morphological terminology requires individual morphological traits to be homologous, one would have to know the phylogeny of the trait-bearing organisms before one could give traits a common name and describe them, since homology relations between traits can only be decided upon reference to a phylogeny. The phylogeny, in its turn, can only be reconstructed on the basis of data about distribution patterns of similar morphological traits, which, however, can only be documented and analyzed using morphological concepts and terminology in the first place. In other words, the problem is that if morphological terminology rests on homology assumptions, we cannot produce morphological data without knowing the homology relations beforehand, which, in its turn requires knowledge about the underlying phylogeny that we can only obtain on the basis of morphological data. Obviously, resting morphological terminology on homology assumptions inevitably leads to circular reasoning in phylogenetics. Therefore, in order to avoid circularity, it is essential that all morphological concepts that are used for morphological descriptions be defined without reference to homology relations.

Unfortunately, the notion of basing morphological terminology on homology assumptions represents the currently prevalent practice in morphology. Although it obviously violates Bacon's claim of separation of empirical facts from scientific conclusions, which was previously recognized as an epistemological hallmark regarding transparency and reproducibility of modern sciences, it nevertheless became commonly accepted. As I have argued above, it is most likely that this practice resulted from the impact of both the theory of evolution on biological thinking in general and the constraints of essentialist thinking of morphologists in particular.

In order to re-establish the high degree of aperspectival objectivity in morphology that Linnaeus reached for fructification traits, and in order to expand it to the entire structural diversity of morphological traits, it is inevitable that morphological terminology must be freed of all homology assumptions. Thus, in order to establish a high degree of comparability of morphological data, morphological terms should only represent structural kind terms, which are purely descriptive and free of evolutionary or other explanatory connotations. Furthermore, for allowing comparisons over broad taxonomic ranges, the applicability of morphological terminology should be taxon-independent in principle.

What is the structure of a morphological trait?

Considering the aforementioned criteria, the structure of a morphological trait consists of a set of properties corresponding to the trait and their particular values. Thereby, ideally, the list of possible properties and their definitions are provided by a general morphological structure concept. In order to describe the morphological trait, the morphologist only has to ask the corresponding question about each possible property in the list and study the particular morphological trait for an answer. Ideally, an answer, which takes in the form of a 'value', is chosen from a defined and controlled vocabulary or from a defined interval of numbers that refers to the specific property. As a consequence, morphological data would consist of pairs of property-value descriptions, referenced to a particular morphological trait and based on a general morphological structure concept, which in its turn provides the definitions and meanings to the terms (i.e., possible properties and their possible 'values') used in the descriptions. In other words, the structure of a morphological trait is a standardized list of all of its intrinsic and properties that are describable and relevant to a given scientific research program.

Data bases, ontologies, and data standards

The role of data bases in biology

In life sciences, the rate at which new data, especially molecular data, are generated increases exponentially, and this continuous increase requires the development of tools for easy sifting through and analyzing of large amounts of data (Brazma 2001). This is one of the reasons why data bases become more and more popular in life sciences. Some data bases, such as Pubmed, Ensembl or the UCSC Genome Browser, have already become essential resources, which are being used by many scientists on a daily basis (Stein 2003).

Besides many general data bases for molecular data, a lot of specialized data bases have been developed that are restricted to data from a specific model organism (e.g., FlyBase for *Drosophila*, flybase.bio.indiana.edu; Arabidopsis Information Resource for *Arabidopsis thaliana*, www.arabidopsis.org). Other data bases are devoted to a specific taxonomic group (e.g., Antbase, antbase.org; Fishbase, www.fishbase.org; AmphibiaWeb, amphibiaweb.org).

With their technical possibilities, including the convenient management of all kinds of different information, such as images and other media files, the mapping of for instance collection sites of specimens on global maps and satellite images, the possibility to link all sorts of entries with one another, such as information of a specimen in a morphological data base to its corresponding information in a data base of the museum where it is permanently deposited, biological data bases have the potential to significantly contribute to an increase of transparency and reproducibility of biological data and thus to an increase in objectivity of biological data in general (Vogt *in press*).

Data bases can thus provide a valuable resource for enabling detailed documentation of all relevant information regarding the generation of all kinds of particular empirical data. Thereby, every data base has to define *what* information can be uploaded by *whom* in which *format*. As a consequence, each data base develops its own standardized way of storing and presenting data, which requires the development or the adoption of a corresponding structure concept. Thus, it is not surprising that already today some data bases take on an important role in biological research practice, with the effect of significantly increasing the degree of mechanical and aperspectival objectivity within biology. Terminological problems, such as the lack of standards of gene names and spellings (Brazma 2001; Stein 2003), caused, for instance, fundamental problems with comparability of molecular data, turning the initial purpose of the development of molecular data bases upside down. This forced molecular data base developers to put a lot of effort into the development of defined and controlled vocabularies, in order to deal with these problems. As a consequence, the comparability of molecular data within data bases has significantly increased, with new and better standards of data documentation and representation becoming commonly accepted.

Morphological data bases

Within the last decade, some interesting morphological data bases became available. MorphBank (morphbank.net) is an open web repository of images for the documentation of specimens and vouchers for sharing research results in taxonomy, morphometrics, morphology, and phylogenetics. Another project, MorphoBank (http://morphobank.geongrid.org), is a GenBank-like repository for storing digital images (Pennisi 2003). It catalogues images and allows the labeling of structures on the images and the display of editable phylogenetic matrices, which are linked to images within the data base. A different project, Digital Morphology (DigiMorph, http://www.digimorph.org), is an archive of digital morphological images and 3D models.

Unfortunately, none of the aforementioned morphological data bases stores and documents morphological descriptions and, thus, morphological data in the strict sense. Instead, they focus on providing convenient tools for management of images, specimen information, and homology hypotheses in the form of character matrices. Thus, it is not surprising that none of these data bases provide a defined and formalized, taxon-independent, and homology-free morphological terminology for preparing morphological descriptions.

The Morphological Descriptions Data Base (MorphDBase, http://www.morphdbase.de) *attempts* to provide a platform for uploading different types of phenotypic information including all kinds of media files *and* morphological descriptions. These descriptions will be based on a morphological ontology (i.e., *MorphOntology*, http://www.morphdbase.de; for more information on ontologies see following paragraph), which is currently being developed and will be available in the near future.

Ontologies for standardizing structure concepts

Some biological data bases use *ontologies* (not to be mistaken with *Ontology* in philosophy, which is the study of ‘being’ or ‘existence’), which provide a defined and controlled vocabulary. An ontology consists of a vocabulary of terms with their corresponding concepts and some specifications of their meaning that are used to describe a certain reality. The concepts of an ontology are described both by their meaning and their relationship to each other (see also Bard 2003; Bard & Rhee 2004). An ontology is a formal way of representing knowledge of a particular scientific field through concepts and represents, as such, a data standard (Wang *et al.* 2005). It is based on a set of formal rules and assertions that describe the relationships between the concepts in a computer parsable form.

The Gene Ontology (GO; Gene Ontology Consortium 2006) represents a well-established ontology and probably the most commonly known within biology. GO provides a standardized, controlled vocabulary for genome annotation systems, cataloguing information about the structural and cellular location of gene products, about the processes to which these products contribute, and the functions that they fulfill (Stevens *et al.*

2000; Bard 2003). Hitherto, many data bases that manage molecular data have incorporated the GO annotation sets, such as for instance the Saccharomyces Genome Database (SGD, http://www.yeastgenome.org), FlyBase (http://flybase.bio.indiana.edu), Mouse Genome Informatics (MGI, http://www.informatics.jax.org), Arabidopsis Information Resource (TAIR, http://www.arabidopsis.org), and other genome centers, such as for instance the National Center for Biotechnology Information (NCBI, http://www.ncbi.nlm.nih.gov) (Blake 2004). Unfortunately, regarding their applicability, most bio-ontologies available today are restricted to one specific model organism, with GO representing a rare exception.

An introduction to resource description framework (RDF) ontologies

An ontology has to be highly standardized and formalized in order to be applicable with description logics and utilizable for many different software applications. The Resource Description Framework (RDF, http://www.w3.org/RDF) has become the most accepted general method for modeling knowledge. RDF is a (meta-) data model and not a specific description language for metadata—it is data describing *all* kinds of web resources. In order to serialize (i.e., make it computer-parsable) RDF it requires syntax. Typically, RDF uses a defined XML syntax (Beckett 2004) or N3 (Berners-Lee 2005) and the semantics via reference to RDF Schema Language (RDFS) (Brickley 2004) or Ontology Web Language (OWL) (McGuinness & van Harmelen 2004). RDFS and OWL represent languages that are based upon RDF and offer support for machine processing and inferences (Wang *et al.* 2005).

In RDF, relationships between resources are described by connecting one resource to another through a relation, resulting in a RDF triple: '*Resource_X* **relation** *Resource_Y*'. A resource is anything that is identifiable by a uniform resource identifier (URI; e.g., a web address) reference (Manola & Miller 2004). By convention, the resource to the left of the relation is called '*Subject*', while the resource to the right is called '*Object*', and the relation '**property**' (in the remainder of this article, every '*Subject*' and '*Object*' will be written in italics while every '**property**' will be in bold font), resulting in the typical RDF triple formalism of '*Subject* **property** *Object*'. The '*Subject*' represents the object that is being described, the '**property**' specifies the relationship or property type between '*Subject*' and '*Object*', and the '*Object*' specifies the value of the property and is either another resource (i.e., a URI) or a literal string (i.e., a sequence of letters or numbers that is only stored by the computer without applying semantics to it, as for instance comments and numbers).

Each RDF triple can be modeled as a graph comprising two nodes connected by a directed arc (Fig. 2). A collection of such RDF graphs can jointly form a directed labeled graph (DLG) (Fig. 3). Such a DLG in its turn can, in theory, model most domain knowledge (Wang *et al.* 2005) and is a useful tool for analysis using graphs logics. A collection of RDF triples or graphs can be used to represent an ontology.

FIGURE 2. A RDF triple modeled as a directed labeled graph (DLG). *Subject* and *Object* represent the nodes and the '**property'** the edge that connects the nodes.

Defining concepts in RDF

Within an ontology, concepts are defined by a set of RDF triples. Ideally, all concepts are defined on the basis of Aristotelian definitions—*per genus et differentiam* (it is, however, possible to define a concept only on the basis of the 'genus' part and a specification of the concept of which it represents a specialized sub-concept, without explicitly specifying the 'differentia' part of its definition). As a consequence, specialized concepts inherit all defining triples of their more general 'parent' concepts.

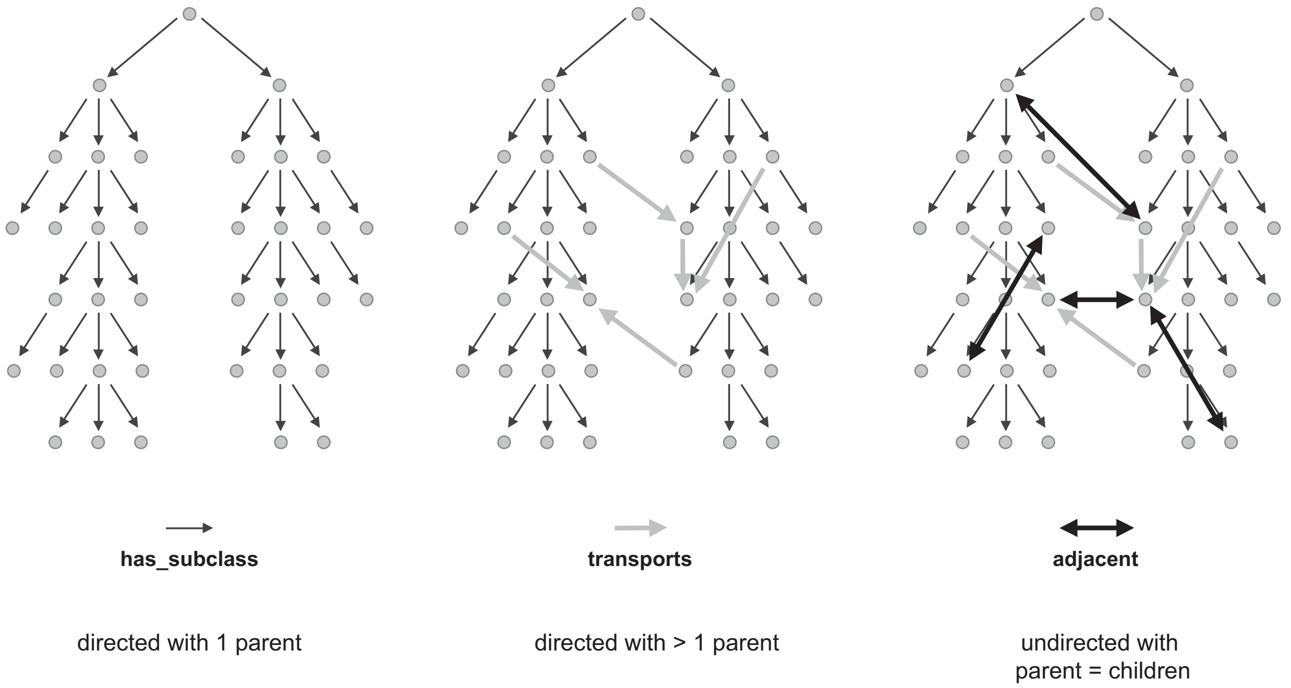

FIGURE 3. Different types of graphs. **a)** A unidirectional rule that allows only a single parent (e.g., '**has_subclass**', which is the inverse property to '**is_a'**). It can be modeled as a simple directed graph representing a tree. **b)** A unidirectional rule that allows for more than one parent (e.g., '**transports'**) can be modeled as a directed acyclic graph, in which the graph itself can be traversed in several ways, with more than one path linking two nodes. **c)** A bidirectional rule that imposes no directional constraints (e.g., '**adjacent**'), resulting in an undirected graph.

For example, a polarized junctioned cell could be defined as a junctioned cell (i.e., *genus*) that has an apico-basal orientation (i.e., *differentia*). Since it represents a specialized junctioned cell, it *necessarily* also has to possess the defining properties of junctioned cells (i.e., a junctioned cell is a cell that has an intermolecular bond with at least one cell-junction of another cell); and since junctioned cells represent a special kind of cell, a polarized junctioned cell would *necessarily* also have to possess all defining properties of cells (e.g., having as its parts a cell membrane and at least one organelle). This definition can be visualized as a graph (Fig. 4a). By organizing different property types into general categories such as topological properties versus functional properties, and by color coding them, one can also easily visually differentiate between different aspects of a definition, as well as differentially navigate through the network of relationships that exist between different concepts of an ontology by only focusing on the properties of interest and blinding out those that are not of interest.

Since most terms and concepts in an ontology should be defined through Aristotelian definitions, terms and concepts are related to one another in a network of different '**property**' relations, with a hierarchical taxonomy of class-subclass relations (i.e., '**is_a**') as a backbone, which at its turn results in a *taxonomy* of more and more specialized concepts, implying a hierarchical organization of terms (i.e. taxonomic inclusion, Bittner *et al.* 2004).

The concepts of an ontology represent classes of defined terms and their inter-relationships and should *not* contain empirical data (i.e., instances) in principle. However, statements about individual objects or individual processes can be linked as instances of concepts to the ontology. This can be done through the '**instance_of**' property. If within a data base empirical data are linked to an ontology in such a way, one receives what is called a *knowledge base* (Stevens *et al.* 2000).

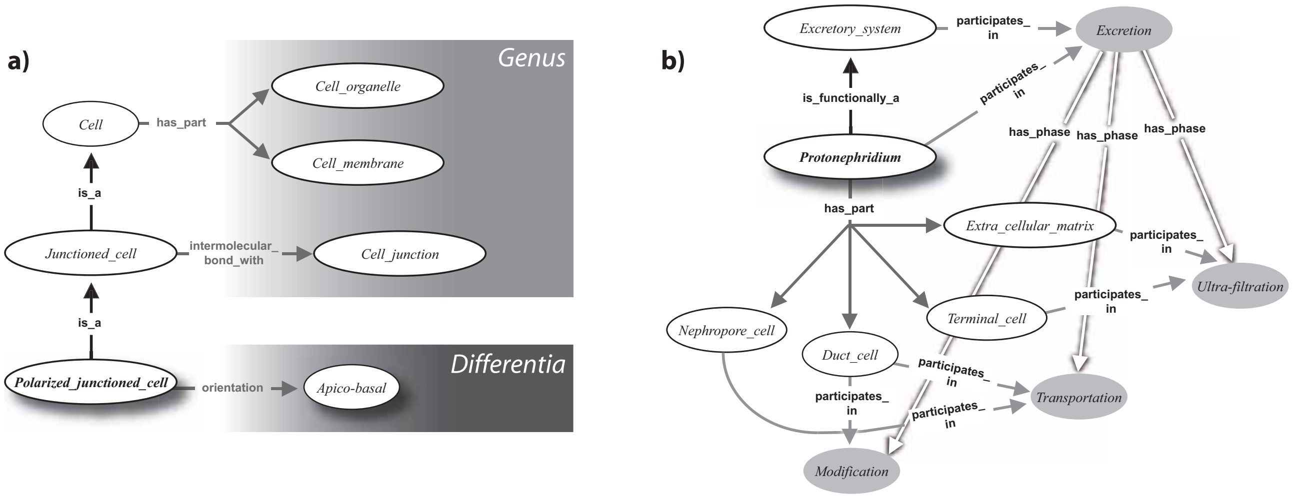

FIGURE 4. Example of two definitions of concepts expressed in RDF: **a)** the definition of 'polarized junctioned cell' (for more information see text); **b)** the definition of 'protonephridium'—a protonoephridium always consists of a nephropore cell, a duct cell, a terminal cell, and extracellular matrix. However, some structures cannot be defined satisfactorily without reference to dispositions of being able to actively participate in specific biological processes. This is also the case for protonephridium, which participates in the process of excretion. This process can be partitioned into different phases, which at their turn have different parts of the protonephridium as their participants.

Defining relations in RDF

The relations of an ontology play a very important role since they carry all the semantic content. Thus, all types of relations specified in an ontology must be carefully defined. In addition to providing free text definitions for each property of an ontology, one can define them according to their logical properties.

The '**is_a**' property, for instance, which stands for the class–subclass relationship between a specialized concept and its more general 'parent' concept, is *transitive* (if 'A_1 **is_a** A_2' AND 'A_2 **is_a** A_3', then 'A_1 **is_a** A_3'), *reflexive* ('A_1 **is_a** A_1'), and *antisymmetric* (if 'A_1 **is_a** A_2' AND 'A_2 **is_a** A_1', then A_1 and A_2 are identical), but it is not symmetric; whereas '**adjacent**', when applied as a relation between instances, is only symmetrical (if 'A_1 **adjacent** A_2', then 'A_2 **adjacent** A_1'). The '**instance_of**' property, on the other hand, is neither transitive, nor reflexive, nor symmetric, nor antisymmetric.

Further logical properties can be specified for each property of an ontology. One can define the concepts or types of literal strings (e.g., numerical values, specific intervals, Boolean values, or free text) that are allowed to be used as a possible '*Subject*' (called the *domain* of the property) and those that are allowed as a possible '*Object*' (called the *range* of the property) in a RDF triple together with this specific property. The property '**actively_participates_in**', for instance, specifies a material object that participates in a process. Thus, the domain of '**actively_participates_in**' has to be restricted to material objects only and its range to processes only. The specification of the domain and the range of each property that is defined in an ontology thereby not only constraints its applicability, but can also be utilized for enforcing, at least to some degree, logical coherence within sets of RDF triples, as for instance triples about particular relations of individual objects representing instances of concepts of the ontology.

Using an ontology for inferences

The possible applications of traditional dictionaries and glossaries, which only represent indexed sets of terms and definitions, are by far outclassed by those of an ontology. By applying descriptions logics, one can utilize the logical properties of relations defined in an ontology in order to make inferences. For instance, given the information that my left arm *A* is part of my body *X*, my left hand *B* is part of my arm *A*, and my left index finger *C* is part of my left hand *B*, appropriate software tools can infer that, due to the *transitivity* of the parthood relation, not only my left arm *A* is part of my body *X*, but also my left hand *B* and my left index finger *C* (if ‘*Arm*$_A$ **part_of** *Specimen*$_X$’ AND ‘*Hand*$_B$ **part_of** *Arm*$_A$’ AND ‘*Finger*$_C$ **part_of** *Hand*$_B$’, then ‘*Hand*$_B$ **part_of** *Specimen*$_X$’ AND ‘*Finger*$_C$ **part_of** *Specimen*$_X$’).

While this application seems to be trivial at first sight, it turns out to be invaluable when it comes to searching for relevant information within very large data bases. So for instance, when annotating the content of images using an ontology, one could annotate the information that the image depicting a complete organism also depicts its parts (e.g., its head and thorax) and all further subparts, simply by annotating that the image depicts an instance of a specific body organization. This would be enough in case this body organization is defined within the ontology as necessarily possessing a head and thorax as its parts. As a consequence, when searching for thorax within the data base, all images depicting this body organization could be retrieved as well, thereby guaranteeing that all images showing heads will be found.

Ontology as a structure concept

A structure concept requires the standardization and formalization of a specialized terminology that is required for making scientific descriptions (i.e., empirical data). Fortunately, to provide such a specialized terminology is exactly one of the key purposes of scientific ontologies. Each ‘**property’** of an ontology that refers to properties and relations of the things and processes to be described can be understood as a particular question that the structure concept poses to the scientist in reference to this given thing or process, just like Linnaeus’ categories of his sexual system (e.g., what **shape** does the entity to be described have; what is **adjacent** to it; whether it is ‘**continuous_with’** some other entity; what is its **temperature**; how does it **react** upon exposure to light). The thing or process to be described is represented by the ‘*Subject*’ in a RDF triple. The ‘*Object*’ of an RDF triple, on the other hand represents the answer to this question and specifies a specific value for the trait to be described (see Fig. 5). Actually, Linnaeus’s definitions of plants can be easily translated into RDF statements (see Fig. 6).

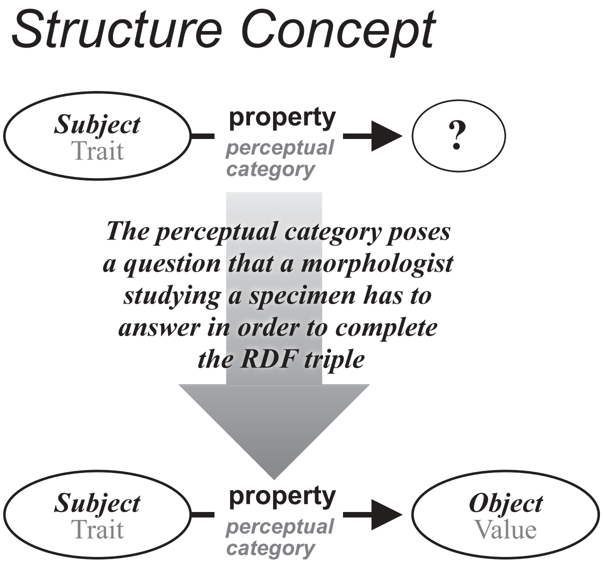

FIGURE 5. Implementation of the structure concept in RDF ontology: The trait to be described is represented by the ‘*Subject*’ of a RDF triple. The ‘**property**’ represents one perceptual category of the structure concept and functions as an empirical question that can only be answered by studying the trait. The answer to the question is represented by the ‘*Object*’ of the RDF triple and corresponds with one of the values that are allowed for this category according to the structure concept. One such describing RDF triple represents a morphological datum – the smallest piece of morphological information possible.

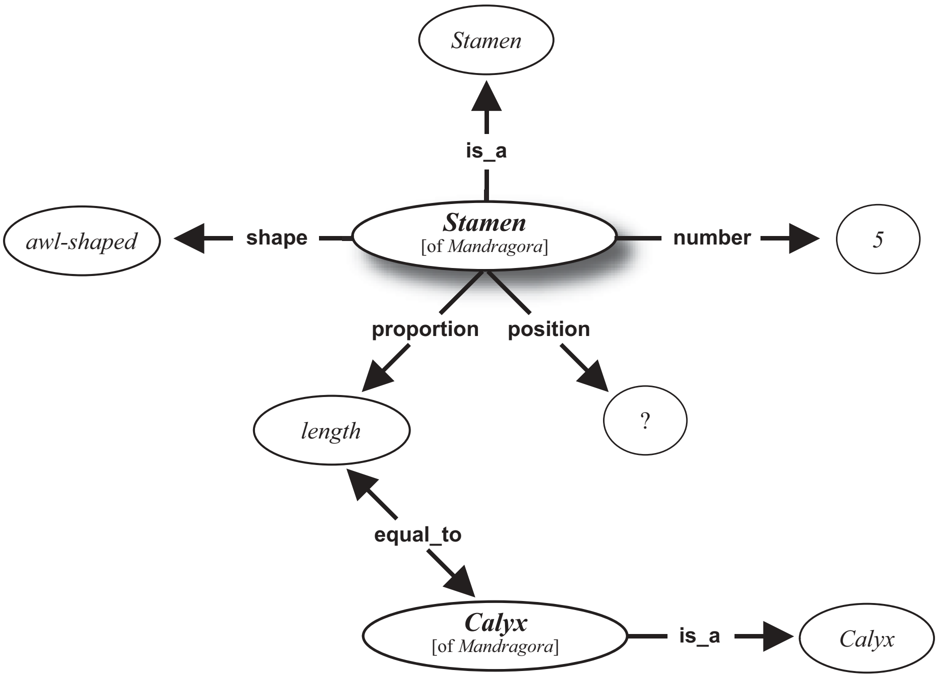

FIGURE 6. The part of the definition of the genus *Mandragora* that refers to its stamen, taken from Linnaeus's *Genera plantarum* (1737) and transformed into a RDF graph.

Which '**property'** is relevant for the description of a given entity is thereby controlled by the ontology via the specification of the *domain* and the *range* of each '**property'**. These questions can only be answered on the basis of observational judgments substantiated by experimental investigations and observations. When implementing an ontology in a data base, the advantage of both technologies is combined and the descriptions, which are based on terms and relations provided by the ontology, can be empirically substantiated by respective images from the data base.

A morphological ontology can provide a general morphological structure concept

Characteristics of biological objects

In biology, structures exist at all levels of organization, ranging hierarchically from the atomic and molecular to the cellular, tissue, organ, multicellular organism, population, and ecosystem level (see 'scalar hierarchy', Salthe 1985, 1993; 'levels of organization', Wimsatt 1976, 1994; 'cumulative constitutive hierarchy', Valentine & May 1996; 'Theorie des Schichtenbaus der Welt', Riedl 2000). Usually, a higher-level structure is composed of multiple copies of a lower-level structure. Thus, a morphological ontology has to cover all these different levels of organization, providing for each level the adequate terminology, without allowing for redundancies and inconsistencies.

Moreover, since morphological traits actively participate in specific types of processes, which is commonly understood as a property of the trait and referred to as its function, and since morphological traits also represent the result of morphogenetic processes, a morphological ontology has to cover relevant biological processes as well. The challenge here is to develop the ontology in such a way that it enables coherent representation of all the relevant inter-relationships between morphological traits and biological processes, thereby integrating structural, functional, and developmental aspects of morphological traits. This enterprise is far from being trivial.

Principles for developing a general morphological ontology

Beyond these conceptual challenges that have to be dealt with, a general morphological ontology should meet the following criteria:

- All morphological concepts should be, in principle, taxon-independent regarding their applicability. This is essential for establishing a high degree of comparability of morphological descriptions.
- A definition of a morphological kind should focus on its structural properties. This allows for unambiguous recognition of instances of the kind exclusively on the basis of morphological studies and does not require experimentation, as it would be the case with functional or developmental definitions.
- In some cases functional definitions will be inevitable, but must be clearly indicated as referring to active participation in a specific type of biological process (see example of '*Protonephridium*', Fig. 4b).
- All morphological concepts should be defined without reference to homology relations. This is essential in order to circumvent circular reasoning and is also required with respect to transparency and reproducibility of data generation.

A morphological ontology as a general morphological structure concept

The combination of a morphological ontology that meets the aforementioned criteria, imposed on a data base for morphological descriptions, would provide an integrative platform—although restricted to those particular data bases that use the ontology—within which comparative morphological studies through a broad taxonomic range would be possible in principle, since the ontology would guarantee a high degree of comparability of morphological data. In all fields in which morphological data are used, such morphological knowledge bases could take in a central methodological function comparable to GenBank for molecular data.

A premise for the success of such an approach for solving terminological problems in morphology is the development of a general structure concept for morphology. RDF ontologies, with their properties and with all their possible applications, represent the most promising tool for attempting to develop such a general morphological structure concept. Ontologies provide promising tools for the development of an easily and intuitively accessible terminology for morphology and provide a high degree of transparency of their basic underlying rules and axioms. Moreover, ontologies have the potential to provide a basis for establishing a general data standard not only for morphological data but for the entire field of biology (see Vogt *in press*), which would substantially facilitate all kinds of co-operations among the different fields in biology. Morphology as a whole would significantly benefit from this development, if it manages to participate in this already ongoing process. All it takes for its success are experienced morphologists who are willing to share their knowledge and who are willing to invest some of their time in helping to develop and improve a general morphological ontology.

Acknowledgments

I am very thankful to Prof. Alessandro Minelli for inviting me to give a talk at the meeting "*Updating the Linnaean Heritage: Names as tools for thinking about animals and plants*" of the Linnaean Society of London, held in Padova, Italy, in May 2008.

This study was supported by the German Research Foundation DFG (VO 1244/3–2). I am also grateful to the taxpayers of Germany.

References

Adanson, M. (1763) *Familles des plantes.* Vincent, Paris, 664 pp.
Aldrovandi, U. (1648) *Monstrorum historia.* Tebaldini, Bononiae, 401 pp.

Atran, S. (1990) *Cognitive foundations of natural history: Towards an anthropology of science*. Cambridge University Press, Cambridge, Mass, 360 pp.

Appel, T. A. (1987) *The Cuvier–Geoffroy Debate: French biology in the decades before Darwin*. Oxford University Press US, New York, 305 pp.

Bacon, F. (1620) *Novum organum scientiarum*. J. Billium, Londinii.

Bard, J. (2003) Ontologies: formalising biological knowledge for bioinformatics. *BioEssays*, 25, 501–506.

Bard, J. & Rhee, S.Y. (2004) Ontologies in biology: design, applications and future challenges. *Nature Reviews Genetics*, 5, 213–222.

Beck, L.W. (1969) *Early German philosophy: Kant and his predecessors*. Harvard University Press, Cambridge, Mass, 567 pp.

Beckett, D. (2004) *RDF/XML syntax specification (revised)*. W3C recommendation published online 10 February 2004. Available from: http://www.w3.org/TR/2004/REC-rdf-syntax-grammar-20040210/ (August 2008).

Berners-Lee, T. (2005) *Primer: Getting into RDF & Semantic Web using N3*. Published online 29 June 2005. Available from: http://www.w3.org/2000/10/swap/Primer.html (August 2008).

Bittner, T., Donnelly, M. & Smith, B. (2004) Individuals, universals, collections: on the foundational relations of ontology. *In*: Varzi, A. & Vieu, L. (Eds), *Proceedings of the International Conference on Formal Ontology in Information Systems*. IOS Press, Amsterdam, pp. 37–48.

Blake, J. (2004) Bio-ontologies — fast and furious. *Nature Biotechnology*, 22, 773–774.

Bock, G.R. & Cardew, G. (eds.) (1999) *Novartis Foundation Symposium 222: Homology.* John Wiley & Sons, Chichester.

Boyd, R. (1999) Homeostasis, species, and higher taxa. *In*: Wilson, R. A. (Ed), *Species — New interdisciplinary Essays*. MIT Press, Cambridge, MA, pp. 141–185.

Boyle, R. (1660) *New experiments physico-mechanical: Touching the spring of the air and their effects*. H. Hall, Oxford.

Brazma, A. (2001) On the importance of standardisation in life sciences. *Bioinformatics*, 17, 113–114.

Brickley, D. (2004) *RDF Vocabulary Description Language 1.0: RDF Schema*. W3C recommendation published online 10 February 2004. Available from: http://www.w3.org/TR/rdf-schema/ (August 2008).

Brower, A.V.Z. & Schawaroch, V. (1996) Three steps of homology assessment. *Cladistics*, 12, 265–272.

Cambrensis, G. (1185) *History and topography of Ireland*.

Cardano, G. (1557) *De vita propria.*[First published 1643, Villery, Paris.]

Cesalpino, A. (1583) *De plantis libri XVI*. Georgius Marescottus, Florentiae.

Claridge, M.F., Dawah, H.A. & Wilson, M.R. (1997) *Species — The units of biodiversity.* The Systematics Association Special Volume Series 54. Chapman & Hall, London, 460 pp.

Coleman, W. (1964) *Georges Cuvier, zoologist: A study in the history of evolution theory.* Harvard University Press, Cambridge, Mass, 212 pp.

Cuvier, G. (1800-5) *Leçons d'anatomie comparée*. Baudoin, Paris.

Cuvier, G. (1817) *Le règne animal; distribué d'après son organisation; pour servir de base à l'histoire naturelle des animaux et d'introduction à l'anatomie comparée. Volume I*. Deterville, Paris.

Cuvier, G. (1825) *Discours sur les révolutions de la surface du globe et sur les changements qu'elles ont produit dans le règne animal*. G. Dufours et Ed. D'Ocagne, Paris et Amsterdam.

Daston, L. (1992) Objectivity and the escape from perspective. *Social Studies of Science*, 22, 597–618.

Daston, L. (1998) Fear and loathing of the imagination in science. *Daedalus*, 127, 16–30.

Daston, L. & Galison, P. (1992) The image of objectivity. *Representations*, 40, 81–128.

Daston, L. & Park, K. (1998) *Wonders and the order of nature, 1150–1750*. Zone Books, New York, 512 pp.

Dondi, G. (1372–1374) *De fontibus calidis agri Patavini consideratio.* [Published 1553, Iunta, Venetiis.]

Drake, S. (1957) *Discoveries and opinions of Galileo*. Doubleday, New York, 320 pp.

Eco, U. (1988) *Über Spiegel und andere Phänomene*. Carl Hanser Verlag, München/Wien, 261 pp.

Eigen, E. A. (1997) Overcoming first impressions: Georges Cuvier's types. *Journal of the History of Biology*, 30, 179–209.

Ereshefsky, M. (1997) The evolution of the Linnaean hierarchy. *Biology and Philosophy*, 12, 493–519.

Farber, P.L. (1976) The type-concept in zoology during the first half of the nineteenth century. *Journal of the History of Biology*, 9, 93–119.

Freudenstein, J. V. (2005) Characters, states, and homology. *Systematic Biology*, 54(6), 965–973.

Fristrup, K.M. (2001) A history of character concepts in evolutionary biology. *In:* Wagner, G. P. (Ed), *The character concept in evolutionary biology.* Academic Press, San Diego, pp. 13–35.

Galison, P. (1998) Judgement against objectivity. *In*: Jones, C. A. & Galison, P. (Eds), *Picturing science, producing art*. Routledge, New York, London, pp. 327–359.

Gene Ontology Consortium (2006) The gene ontology (GO) project in 2006. *Nucleic Acid Research*, 34, D322–D326.

Gingrich, O. (2004) *The book nobody read: Chasing the revolutions of Nicolaus Copernicus*. Walker Publishing Company, Inc., New York, 360 pp.

Hall, B.K. (1994) *Homology — The hierarchical basis of comparative biology.* Academic Press, New York, London, 504 pp.

Heintz, B. (2000) *Die Innenwelt der Mathematik. Zur Kultur und Praxis einer beweisenden Disziplin.* Springer, Wien, 318 pp.

Heywood, V. (1985) Linnaeus — the conflict between science and scholasticism. *In*: Weinstock, J. (Ed.), *Contemporary perspectives on Linnaeus*. University Press of America, Lanham, Maryland, pp. 1–16.

Jonston, J. (1657) *Historiae naturalis de quadripedibus libri.* J.J. Schipper, Amstelodami.

Jungius, J. (1678) *Isagoge phytoscopica*. Vagetius, Hamburg.

Kanz, K.T. (2002) Von der BIOLOGIA zur Biologie – Zur Begriffsentwicklung und Disziplingenese vom 17. bis zum 20. Jahrhundert. *In:* Hoßfeld, U. & Junker, T. (Eds), *Die Entstehung biologischer Disziplinen II – Beiträge zur 10. Jahrestagung der DGGTB in Berlin 2001*, Verhandlungen zur Geschichte und Theorie der Biologie, Volume 9. VWB Verlag, Berlin, pp. 9–30.

Kukla, R. (2006) Objectivity and perspective in empirical knowledge. *Episteme: A Journal of Social Epistemology,* 3, 80–95.

Lankester, E.R. (1870) On the use of the term homology in modern zoology, and the distinction between homogenetic and homoplastic agreements. *Annals and Magazine of Natural History,* (series 4) 6, 34–43.

Larson, J.T. (1967) Linnaeus and the natural method. *Isis*, 58, 304–320.

Linnaeus, C. (1735) *Systema naturae sive Regna tria naturae. Systematice proposita per Classes, Ordines, Genera, & Species*. Theodor Haak, Lugduni Batavorum.

Linnaeus, C. (1737) *Genera plantarum.* Salvius, Holmiae.

Linnaeus, C. (1751) *Philosophia botanica.* Kiesewetter, Stockholm.

Locke, J. (1689/1979) *An essay concerning human understanding*. Oxford University Press, Oxford.

Manola, F. & Miller, E. (2004) *RDF Primer,* W3C Recommendations published online 10 February 2004. Available from: http://www.w3.org/TR/rdf-primer/ (August 2008).

Mayden, R.L. (1997) A hierarchy of species concepts: the denouement in the saga of the species problem. *In*: Claridge, M. F., Dawah, H. A. & Wilson, M. R. (Eds), *Species: The units of biodiversity.* Chapman & Hall, London, pp. 381–424.

McGuiness, D.L. & van Harmelen, F. (2004) *OWL Web Ontology Language Overview.* W3C recommendation published online 10 February 2004. Available from: http://www.w3.org/TR/owl-features/ (August 2008).

Mery, J. (1718) *Sur un exomphale monstrueuse. HARS, Année 1716.* Paris.

Montecatini, U. da (1471) *De balneis mineralibus et artificialibus.* [Publ. 1950, Olschki, Firenze.]

Müller-Wille, S. (1999) *Botanik und weltweiter Handel. Zur Begründung eines Natürlichen Systems der Pflanzen durch Carl von Linné (1707–78).* Studien zur Theorie der Biologie, Band 3. Verlag für Wissenschaft und Bildung, Berlin.

Nagel, T. (1986) *The view from nowhere.* Oxford University Press, Oxford, 255 pp.

Newton, I. (1687) *Philosophiae naturalis principia mathematica*. Societas Regia, Londini.

Olivi, G.B. (1584) *De reconditis et praecipuis collectaneis ab onestissimo et solertissimo Francisco Calceolari Veronesi in Musaeo adservatis.* Zanfrettus, Venetiis.

Owen, R. (1843) *Lectures on comparative anatomy and physiology of the invertebrate animals delivered at the Royal College of Surgeons*. Longman, Brown, Green and Longmans, London, 689 pp.

Panchen, A.L. (1999) Homology – history of a concept. *In:* Bock, G. R. & Cardew, G. (Eds), *Homology, Novartis Foundation Symposium 222*. John Wiley & Sons, Chichester, pp. 5–18.

Pennisi, E. (2003) Modernizing the tree of life. *Science*, 300, 1692–1697.

Pigliucci, M. (2003) Species as family resemblance concepts: The (dis-)solution of the species problem? *BioEssays*, 25, 596–602.

Pleijel, F. & Rouse, G.W. (2000) Least-inclusive taxonomic unit: a new taxonomic concept for biology. *Proceedings of the Royal Society of London, Series B: Biological Sciences*, 267, 627–630.

Porter, T. (1992) Objectivity as standardization: The rhetoric of impersonality in measurement, statistics, and cost–benefit analysis. *Annals of Scholarship*, 9, 19–59.

Pullan, W. (2000) Structuring structure. *In*: Pullan, W. & Bhadeshia, H. (Eds), *Structure in science and art*. Cambridge University Press, Cambridge, pp. 1–8.

Ramírez, M.J., Coddington, J.A., Maddison, W.P., Midford, P.E., Prendini, L., Miller, J., Griswold, C.E., Hormiga, G., Sierwald, P., Scharff, N., Benjamin, S.P. & Wheeler, W.C. (2007) Linking of digital images to phylogenetic data matrices using a morphological ontology. *Systematic Biology,* 56, 283–294.

Ray, J. (1703) *Methodus plantarum emendata et aucta*. Smith and Walford, London.

Riedl, R. (2000) *Strukturen der Komplexität – Eine Morphologie des Erkennens und Erklärens*. Springer, Berlin, 367 pp.

Rieppel, O. (1993) The conceptual relationship of ontogeny, phylogeny, and classification – the taxic approach. *Evolutionary Biology*, 27, 1–32.

Russell, E.S. (1916) *Form and function: a contribution to the history of animal morphology.* John Murray, London, 394 pp.

Salthe, S.N. (1985) *Evolving hierarchical systems: their structure and representation*. Columbia University, New York, 343 pp.

Salthe, S.N. (1993) *Development and evolution: complexity and change in biology*. MIT Press, Cambridge, Mass, 368 pp.

Savonarola, M. (1448–1449) *De balneis seu thermis naturalibus omnibus Italiae*. [Publ. 1845, Andreas Belfortis, Ferrariae.]

Scotland, R.W., Olmstead, R.G. & Bennet, J.R. (2003) Phylogeny reconstruction: The role of morphology. *Systematic Biology*, 52, 539–548.

Shapin, S. (1996) *The Scientific Revolution*. University of Chicago Press, Chicago, 218 pp.

Shapin, S. & Schaffer, S. (1985) *Leviathan and the air pump: Hobbes, Boyle and the experimental life, including a translation of Hobbes, Dialogus Physicus De Natura Aeris*. Princeton University Press, Princeton NJ, 456 pp.

Spedding, J., Ellis, R.L. & Heath, D.D. (1857–1874) *The works of Francis Bacon*. 14 volumes. Longman, Green, Longman & Roberts, London.

Stamos, D.N. (2005) Pre-Darwinian taxonomy and essentialism – a reply to Mary Winsor. *Biology and Philosophy*, 20, 79–96.

Stein, L.D. (2003) Integrating biological databases. *Nature Reviews Genetics*, 4, 337–345.

Stevens, P.F. (1997) Mind, memory and history: How classifications are shaped by and through time, and some consequences. *Zoologica Scripta*, 26, 293–301.

Stevens, R., Goble, C.A. & Bechhofer, S. (2000) Ontology-based knowledge representation for bioinformatics. *Briefings in Bioinformatics*, 1, 398–414.

Strickland, H.E. (1840a) On the true method of discovering the natural system in zoology and botany. *Annals and Magazine of Natural History*, 6, 184–186.

Strickland, H.E. (1840b) Observations upon the affinities and analogies of organized beings. *Magazine of Natural History*, 4, 219–226.

Tournefort, J.P. de (1694) *Elemens de botanique*. Imprimerie Royale, Paris.

Valentine, J.W. & May, C.L. (1996) Hierarchies in biology and paleontology. *Paleobiology*, 22, 23–33.

Vogt, L. (*in press*) The future role of bio-ontologies for developing a general data standard in biology – Chance and challenge for zoo-morphology. *Zoomorphology*.

Vogt, L., Bartolomaeus, T. & Giribet, G. (*submitted*) The linguistic problem of morphology – Challenges to the future role of morphology in phylogenetics.

von Sachs, J. (1875) *Geschichte der Botanik*. Oldenburg, München.

Wang, X., Gorlitsky, R. & Almeida, J.S. (2005) From XML to RDF: how semantic web technologies will change the design of 'omic' standards. *Nature Biotechnology*, 23, 1099–1103.

Wheeler, Q.D. & Meier, R. (2000) *Species concepts and phylogenetic theory: a debate*. Columbia University Press, New York, 256 pp.

Whewell, W. (1840) *Philosophy of the inductive sciences*. J.W. Parker, London.

Wilson, R.A. (ed.) (1999) *Species — New interdisciplinary essays*. MIT Press, Cambridge, Mass, 349 pp.

Wimsatt, W.C. (1976) Reductionism, levels of organization, and the mind–body problem. *In:* Globus, G., Maxwell, G. & Savodnik, I. (Eds), *Consciousness and the brain: A scientific and philosophical inquiry*. Plenum Press, New York, pp. 202–267.

Wimsatt, W.C. (1994) The ontology of complex systems: Levels, perspectives, and causal thickets. *Canadian Journal of Philosophy*, Supplemental Volume 20, 207–274.

Winsor, M.P. (2003) Non-essentialist methods in pre-Darwinian taxonomy. *Biology and Philosophy*, 18, 387–400.

Winsor, M.P. (2006a) Linnaeus's biology was not essentialist. *Annals of the Missouri Botanical Garden*, 93, 2–7.

Winsor, M.P. (2006b) The creation of the essentialism story: an exercise in metahistory. *History and Philosophy of the Life Sciences*, 28, 149–174.

Zootaxa 1950: 153–163 (2008)
www.mapress.com/zootaxa/

ISSN 1175-5326 (print edition)
ZOOTAXA
ISSN 1175-5334 (online edition)

Principles for a names-based cyberinfrastructure to serve all of biology*

DAVID J. PATTERSON[1,3], SARAH FAULWETTER[2] & ALEXEY SHIPUNOV[1]

[1] *Biodiversity Informatics Group, Encyclopedia of Life, Marine Biological Laboratory, 7 MBL Street, Woods Hole, MA 02543, USA*
[2] *Institute of Marine Biology and Genetics, Hellenic Centre for Marine Research, 71003 Heraklion, Crete, Greece*
[3] *E-mail: dpatterson@eol.org*

** In: Minelli, A., Bonato, L. & Fusco, G. (eds) Updating the Linnaean Heritage:Names as Tools for Thinking about Animals and Plants. Zootaxa,* 1950, 1–163.

Table of contents

Abstract

The rapidly growing amount of biological data on the internet and the increasing need for large-scale analyses mandate improvements to the management of taxon-centric information. This information, traditionally managed by taxonomists, is now transforming into a web-based infrastructure. The complexity and narrative quality of the biological sciences require an information management framework that is sensitive to the scale, richness, character, and heterogeneity of the discipline. Given that the names of organisms offer us a nearly universal system for indexing biological data objects, a names-based cyberinfrastructure has the capacity to index the totality of available biological information and to aggregate taxon-centric data over a broad scale. In order to serve its role, this infrastructure should incorporate thirteen principles that are proposed here.

Key words: Biodiversity informatics, Taxonomic intelligence, Encyclopedia of Life, Data management, Taxonomy

Accepted by A. Minelli: 24 Oct. 2008; published: 5 Dec. 2008

Introduction

Shifts in the agenda of the biological sciences in the last two decades have been driven by diverse factors such as unifying molecular technologies, the challenges of climate change or, associated with the latter, the "*biodiversity crisis*" — the loss of biodiversity at all levels. In the latter area, both predictions and preparations for change will need analyses that integrate biospheric, economic, historical, social and geospheric information and on a scale that has previously not been considered. Biologists and conservationists need to synthesise different layers of understanding in order to "understand the whole" (Bisby 2000) and they require the availability of this information at an ever-increasing rate (De Carvalho *et al.* 2008). The extraction of understanding from the data will be empowered by an organisational framework that can interconnect biological information distributed in heterogeneous environments across the Internet. Ideally, as articulated by E.O. Wilson (2003), we should look to a future in which any piece of information on a species on the Internet will provide a gateway to all other information on the same species. With such interconnections in place, users will be able to index biological data objects, promote their atomisation into the smallest semantic parts and through those processes, create a vast communal pool of readily available taxon-centric data, making novel large-scale analyses possible.

Mayr (2004) correctly pointed out that biology is an unusual scientific discipline. The management of biological information will require solutions that are sensitive to the oddities of biology. Traditionally, biological knowledge has been catalogued and organised by taxonomists. From the time of Linnaeus — "*Filum ariadneum botanices est systema, sine quo chaos est res herbaria*" (Linnaeus 1751), taxonomy has effectively unifed biological knowledge and prevented its disarray. The declining numbers in the taxonomic community, as well as the growing deluge of biodata (termed "*the second bioinformatics crisis*" by Godfray, 2002), require us to come up with new ways of managing biodiversity information (Godfray 2002, Bortulus 2008). The vision presented here is one of a biological information management environment that embeds taxonomic and nomenclatorial thinking into the design of databases, data schemas, transfer protocols, applications, etc. with the intent of assembling an infrastructure capable of managing any piece of information of any type or size for any taxon.

Taxonomy as information management system relies on two elements that have long served the unique character of biology: names and hierarchies. Names are associated with most usable data objects and so are the common denominator of heterogeneous information coming from distributed sources. In the world of contemporary informatics, names can serve as metadata for all data objects that relate to taxa. That is, they can be used to form the foundation of an indexing system for all biology. Hierarchical arrangements, such as taxonomic, phylogenetic or any other kind of classification, specify relationships between elements by placing them into nested structures. They serve as ontologies which can be used to add higher levels of organization to the metadata and the data they include can be used to test hypotheses represented by the ontologies.

The use of names as metadata, of hierarchies as ontologies, and of both for data management are not usually considered to be part of taxonomy, and may run counter to principles of compliance with the nomenclatural codes. As an example, algorithmic indexing of digital data, such as the content of the Biodiversity Heritage Library (BHL, http://www.biodiversitylibrary.org) has to rely on name-recognition tools (e.g. Koning *et al.* 2005). Yet, many "names" in documents are obsolete or misspelled. If we are to index and recover all information, we need to catalog and cross-reference not only the code-compliant names but also the archaic and the misspellings. The implementation of this vision is not straightforward as a names-based information management has to overcome a number of problems, of which the most significant are the "many names for one taxon" (synonymy) and the "same name for many taxa" (homonymy) problems. Both confound the collation of all relevant information about the same species. In the case of synonyms, a simple search with one name will fail to find information linked to other names used for the same taxon, and in the case of homonyms, a search will draw together information on different, often unrelated, taxa. These problems have always been at the mainstream of the descriptive taxonomy *sensu* Godfray (Godfray, 2002). A names-based cyberinfrastructure therefore needs to emulate the practices of taxonomists who have traditionally managed information in this area — that is, the infrastructure must be "taxonomically intelligent". Taxonomically intelligent names-based information management has an enormous potential for the biological sciences, especially if those developments are designed to allow machine-to-machine dialog through the use of globally unique identifiers, standardised data schemas, and interoperable data transfer protocols (Page, 2006).

The *Encyclopedia of Life* (EOL, http://www.eol.org) is the first major integrative project within biology that is explicitly based on these principles. To be able to fulfil its goal of delivering Web pages for every species it must be able to automatically aggregate taxon-centric information across the full spectrum of biodiversity. EOL relies on names-based information management. We have identified and discuss here thirteen features that we believe a taxonomically intelligent names-based cyberinfrastructure must have if it is to be effective for all types of organisms, and for all pieces of information, past, present and future.

The principles

Inclusive

A names-based infrastructure that is intended for managing information about any and all forms of life, must be designed to include all entities that satisfy any definition of "life" — whether viruses (and even prions), both types of prokaryotes, protists, plants, animals, or fungi. This requires that the infrastructure move in a direction opposite to the current trend of fragmentation into subdisciplines. This fragmentation led to similar but mostly independent codes of nomenclature, whereas the goal of biodiversity binformatics should be a single unified system. All codes seek to establish stability and to remove ambiguity in the use of names. They foster these goals within the jurisdiction of each code, but the independence of the codes can promote ambiguity when it comes to organisms that do not fit comfortably within one particular code (the ambiregnal names; Corliss 1995). The emerging infrastructure must not only apply to all organisms, but also to respect all code-based nomenclatural practices inclusive of the more innovative PhyloCode (Cantino & de Queiroz 2000). This code departs from traditional nomenclatural practices by seeking to regulate names that depict monophyletic and holophyetic clades by explicitly using phylogenetic principles. Irrespective of the logic by which they are derived, the names act as metadata and will be organized within an ontology. A names based infrastructure must accommodate such schemas if it is to serve the advocates of this phylogenetically motivated nomenclature.

Comprehensive

If a system is to be capable of indexing any biological data object, it must be capable not only of accommodating some information about organism, but all information on all organisms. That will require the architecture to include any identifier that has been used to assign a taxonomic context to a data object. In biology, most identifiers are scientific names, but the approach must also embrace vernacular names, and surrogates (such as culture isolate numbers, sample or specimen numbers) that are used in place of names. In the future, automated indexing tools will analyse electronic repositories, identify the labels, and use the label to link the data object with the taxon. The success of these tools will depend on how well they handle mis-spelled names, obsolete names, differently abbreviated names (and authority information), names that have been distorted through OCR (Optical Character Recognition) errors or Web algorithms (e.g. the Flickr machine tag format (http://www.flickr.com) removes spaces between names, creating "Iguanaiguana" from *Iguana iguana*). Automated name-recognition tools work through recognition of known names and/or the discovery of unknown names (Leary *et al.* 2007). Their functioning is facilitated through the assembly of a pool of all known names in all of their forms which can then serve as the basis for recognition algorithms, and for the improvements of name-discovery rules. The need for such a structure had led to the creation of the uBio NameBank (http://www.ubio.org), currently with over 10,000,000 names for 1,800,000 species.

Taxonomically intelligent reconciliation

The most widespread problem in the use of names for indexing purposes is that there are many different names and variants of names for the same species. The conventional taxonomic solution uses nomenclatural principles to select the correct name for the taxon, to which some or all of the other code-compliant names (synonyms) that have been applied to the same taxonomic concept can be linked. This logic lies behind the Catalog of Life Partnership compilation (CoLP, http://www.catalogueoflife.org). However, a nomenclaturally based solution such as this cannot form the basis of an indexing system because many names and variants of names which are associated with data objects are not code-compliant and so will be excluded. Moreover, the nomenclatural solution provides for the correct names at the present time, but cannot be retrospectively applied to many older documents. A purely nomenclatural approach can not serve well the needs of other major initiatives, such as the Biodiversity Heritage Library. An alternative solution, which is adopted here, is to catalogue all of the name strings that have been used for an entity and group them together within a reconciliation groups. The members of each reconciliation group contain all of the names that have been used for a given taxon. A query starting with any of the names in a group can exploit the reconciliation group to explode the query so that it uses all names. Reconciliation thus improves recovery of records, especially with older data (Table 1).

TABLE 1. Recovery of records from PubMed (http://www.ncbi.nlm.nih.gov/pubmed), JSTOR (http://www.jstor.org) and Google (http://www.google.com) with 5 of the 15 known related names for the red spotted newt from eastern North America.

Name	Year of first use	Items in PubMed	Items in JSTOR	Items in Google
Notophthalmus viridescens	1965	377	281	31,900
Diemictylus viridescens	1959	36	38	2,180
Triturus viridescens	1949	99	280	14,100
Diemyctilus viridescens	1965	1	3	105
Diemyctylus viridescens	1964	4	70	1,830

Any names that refer to different taxa (i.e. are homonyms) can be disambiguated through the use of reconciliation groups. Disambiguation requires copies of the same name string that refers to different taxa (such as *Aotus*, or *Peranema*) to be placed in different reconciliation groups and to be annotated with a flag that alerts users (whether people or machines) to a need for special attention.

Reconciliation groups contain names in several categories. Lexical variants are alternative spellings of the same term. An example might be different yet code-compliant spellings of the same name: such as *Gerardia paupercula var. borealis* (Pennell) Deam and *Gerardia paupercula* (A.Gray) Britton subsp. *borealis* (Pennell) Pennell, mis-spellings (*Eugelna* vs. *Euglena*), binomials with different endings (like *Pomatomus saltator* and *Pomatomus saltatrix*) and even abbreviations like *Camp rotu* (for *Campanula rotundifolia*) widely used in plant ecology. All are unarguably variations of the same name and so are objectively linked. Reconciliation groups include so-called “objective synonyms” that include homotypic synonyms (*Pinus abies* Linnaeus and *Picea abies* (Linnaeus) H. Karsten), nomenclatural variants and combinations of names based on the same type material (e.g., *Pomatomus saltator*, *Temnodon saltator*, and *Gasterosteus saltatrix*), as well as subjective synonyms — names based on different types but accepted as synonyms in a particular treatment (e.g., *Dactylorhiza baltica* and *Dactylorhiza purpurella*). Because of their subjective component, reconciliation groups can be considered as a form of a taxonomic concept — a circumscription of the underlying biological meaning (see *Concept-capable* discussed below). Reconciliation groups must also extend to vernacular names because they too label data objects and in some environments are preferred over code-compliant names. Vernaculars will need to be placed within a linguistic, geo-referenced and script-based context given that the same names are often used for different organisms and the pattern of use depends on location. Reconciliation groups also include surrogates for names (such as culture identifiers, herbarium labels, nucleotide database IDs, etc.).

Names and their relationships with each other within the reconciliation groups can be annotated by flags to distinguish what kind of name the string refers to, the nature of the relationship among names, the provenance of assertions (Smith believes this is a junior subjective synonym of that), or to indicate the nomenclaturally correct name.

Taxonomically intelligent disambiguation

A names-based infrastructure must be capable of discriminating among the different uses of identical name strings for different taxa. Without this, automated systems have a high risk of confounding information on hemihomonyms — homonyms assigned to different taxa subject to different codes (Kluge 2000) like *Oenanthe* (the plant) with *Oenanthe (*the bird). One step for this class of homonyms is to mark all such names with code identifiers (as it is common practice in linguistics and philosophy). For homonyms falling under the jurisdiction of the same code (such as *Argus*, used for spiders, molluscs, birds and various insects) other means of disambiguation, such as providing the taxonomic context as discussed above, may be employed. These names should have a flag that alerts users and systems to the homonym problem. Any action involving one of these names must initiate a process of disambiguation that will lead to the association of a data object with the correct reconciliation group. Disambiguation is essential for automatic names-finding and indexing tools that will frequently encounter spelled-alike abbreviations (such as *C. marina*) that may refer to many different taxa. Name-recognition tools will require rule sets that can clarify the intention from the context in which the string appears. Scientific names can also be disambiguated by reference to broader taxonomic categories (*Peranema* Pteridophyta is not the same as *Peranema* Protista), by the naming authority (*Peranema* Dons vs. *Peranema* Dujardin), or by key words that associate with the target taxa (i.e. the co-occurrence of terms like “frond” or “spores”, or the names of other fern genera or species would indicate that the *Peranema* refers to the plant and not the protist). However, in many cases, the rules may be insufficient to disambiguate taxa, and in such cases, the indexing must be vetted and finalized through the community of experts to ensure that data objects are correctly classified. Vernacular names can be disambiguated by reference to scientific names that are used in conjunction with them, or through their linguistic or geographic context.

Concept-capable

Kennedy and co-workers (2005) argue that names are imprecise flags of “taxonomic concepts”. We may use the same name but have different opinions about what it refers to (e.g., does the name *Gorilla beringei* include the individuals that others refer to as *Gorilla beringei graueri* or not?). Taxonomic concepts refer to the scope of application of the name of a taxon. This can be done through bibliographic references ("*sensu* Smith 1900"), references to specimens, or comparisons of different taxonomies that include or exclude other taxa. The challenges of concept management and the supremacy of concepts over names have led some to dismiss the value of a names-based infrastructure (Berendsohn 1995, Kennedy *et al.* 2005). Yet, a purely concept-based indexing system will be ineffectual because names are predominantly applied without clear indication of the concept to which they refer (Agnarsson & Kuntner 2007, Bortulus 2008). The best solution would be a marriage of a names-based with a concept-based management system. Taxonomic concepts can be incorporated within a names-based infrastructure in several ways. The inclusion of subjectivism within reconciliation groups offers one solution, multiple classifications (see below under "*Hierarchical structure*") offer a second, and placement of taxonomic concepts within a particular hierarchy offers a third.

Hierarchical structure

Classifications are important components of taxonomy because they represent hypotheses of the evolution of taxa or indicate relatedness. Hierarchical arrangements of taxa also provide a useful structure for biological data management (Kennedy 2003) as they may serve as ontologies defining relationships among (metadata) elements. Classifications can be exploited to disambiguate homonyms, browse content, drill towards more taxonomically precise groups or expand searches. They permit hierarchical aggregation of data, such that when a search is made on "Diptera", the settings can specify 'find me all data objects with the word "Diptera" associated with them', but also 'find me all data objects that carry the name of any fly or group of flies'. Hierarchical searches furthermore have significant value in being able to compare phylogenetic hypotheses by providing metrics of the consistency of data objects and their metadata with competing hypotheses.

Neither the entire species inventory nor the tree of life have been assembled, nor will they be. As a result, many different classifications will coexist and none of them is correct (Yoon & Rose 2001). A single, static classification will not be able to serve the needs of all users. Therefore, our management system has to be able to represent multiple, evolving hierarchies to reflect these different opinions about how organisms are related (see 'Phylogenetic'), eventually forming a graph of overlapping hierarchies (Kennedy 2003).

Phylogenetic structure

The hypothesis that all known life is inter-related through ancestor-descendant relationships remains unfalsified. That grounds the principle that seeks to incorporate our understanding of those relationships within our cyberinfrastructure. Closely related taxa share higher proportions of their genome, and can be expected to share a high proportion of their attributes. This provides a logical basis for the hierarchical organization within the system. It also allows us to infer and predict properties before they have been reported, an important feature as biology shifts towards datacentricity. The elimination and transformation of taxa that do not fully reflect our understanding of phylogenetic relationships characterizes the evolution of taxonomy. A cyberinfrastructure that can mimic the trend towards monophyly and holophyly will be more powerful and will gain acceptance from its user community. As noted under 'hierarchical', this trend should not be accompanied by a unitary point of view. Rather, the infrastructure can be designed to promote a process towards an architecture unified by phylogenetic relationships, and that process can be expressed in allowing stakeholders to change the composition and relationships of taxa.

Distributed organization

Information on names (whether lists of names, nomenclatural status, relationships) is located in many different on-line sources, all with their own specific purpose, taxonomic territory and user community (Table 2). There are many additional repositories of names and name associated information targeted on particular taxonomic groups (e.g., Antbase: http://www.antbase.org; CrustaceaNet: http://www.crustacea.net; Millipeet http://www.fieldmuseum.org/research_collections/zoology/zoo_sites/millipeet/), on particular habitats (such as the OBIS, ERMS, WoRMS and APHIA initiatives that address marine taxa), names relating to particular geographic regions, or lists holding vernacular names. This decentralisation of names providers has several advantages — it captures the enthusiasm of individuals by allowing them to identify closely with projects, it creates stability through redundancy, richness through diversity, relevance through purpose, and spreads the tasks among many players. A names-based infrastructure that can bring together the strengths of this distributed richness will outperform one that does not. On the other hand, information coming from different sources is usually heterogeneous in structure and value, so federation (integration of independent operations) is not straightforward. Creating a single point of access to this distributed information will help in organising information about, or attached to these names. This integration process can be addressed through devices like the Global Names Index (see *Up-to-date*, below) and by ensuring consistency through normalisation (see *Interoperability*, below).

Interoperability

Data flow between various names providers requires the adoption of standards, schemas and transfer protocols to facilitate the machine-to-machine dialog. Various standards exist in the field of biodiversity. The Biodiversity Information Standards group (TDWG, http://www.tdwg.org) is a key player in the development of standards for data exchange in different fields of biodiversity and promotes the deployment of Life Science Identifiers (LSIDs) to serve as globally unique identifiers (GUIDs) of taxonomic names. In order to capture the information which the data providers hold, a names-based infrastructure needs to be compliant with industry standards by adopting the current schemas, by serving LSIDs or other GUIDs through agreed data exchange standards. Furthermore, a names-based architecture should promote the usage of RDF (Resource Description Framework) formatted data and ontologies to facilitate semantic data exchange and retrieval. Many well-established databases would not be able to convert to new standards with ease, and in these cases, the solution will need to export and import data through “abstract layers” that transform data from one schema or format to another.

Up-to-date

Taxonomy is an evolving discipline. New taxa are continuously being discovered, new relationships are being described, and taxa are split, merged or renamed to reflect the most current knowledge about the evolution and relatedness of organisms. This continuously evolving knowledge has to be reflected by a dynamic names-based cyberinfrastructure. Excepting molecular biology, which involves the submission of published genetic sequences to central registries, the dissemination of biocentric information is not organised centrally. The Global Names Index, of which a prototype was established in 2008 is an emerging federated web services environment that dynamically interconnects an array of names partners. Names partners may include authoritative nomenclatural sources (such as ZooBank, the International Plant Index — IPNI, Index Fungorum, the Universal Virus Database — ICTVdB) or other repositories of authoritative information (such as the Catalogue of

Life Partnership). The names partners link to a common index through Web Services that automatically keeps the index appraised of changes in each participating database. Partners unable to provide appropriate web services can pass simple names lists into a hosting service that informs the index on their behalf. The common index provides a searchable and machine accessible environment that keeps all partners up to date on new names and associated metadata.

TABLE 2. A selection of on-line resources providing names information.

Project name	URL	Description
AlgaeBase	http://www.algaebase.org	Names of terrestrial, freshwater and marine algae
Deutsche Sammlung von Mikroorganismen und Zellkulturen (DSMZ)	http://www.dsmz.de/microorganisms/bacterial_nomenclature.php	Lists of names of eubacteria and archaebacteria that are compliant with the *International Code of Nomenclature of Prokaryotes*
Index Algarum	http://ucjeps.berkeley.edu/INA.html	Names of terrestrial, freshwater and marine algae
Index Fungorum	http://www.indexfungorum.org	Index of all code-compliant fungus names
Index Nominum Genericorum	http://www.botany.si.edu/ing	Compilation of generic names for organisms covered by the International Code of Botanical Nomenclature
Integrative Taxonomic Information System (ITIS)	http://www.itis.gov	Taxonomic information on plants, animals, fungi, and microbes, mostly of North America
International Plant Names Index (IPNI)	http://www.ipni.org	Names of genera and species of seed-bearing plants with their place of publication
List of Prokaryotic Names with Standing in Nomenclature	http://www.bacterio.cict.fr/allnames.html	Lists of names of eubacteria and archaebacteria that are compliant with the *International Code of Nomenclature of Prokaryotes*
micro*scope	http://microscope.mbl.edu	Information on the biodiversity of microbes
Nomenclator Zoologicus	http://uio.mbl.edu/NomenclatorZoologicus	Compilation of genera and subgenera in zoology from 1758 to 2004
The Catalogue of Life Partnership (CoLP)	http://www.catalogueoflife.org	An incomplete catalogue of all known species of organisms on Earth
uBio	http://www.ubio.org	Assembles all name strings ever used for organisms in literature and Internet, mainly for indexing purposes
Universal Virus Database (ICTVdb)	http://www.ictvdb.rothamsted.ac.uk	Approved virus names, linked to virus descriptions
World Register of Marine Species (WoRMS)	http://www.marinespecies.org	Comprehensive list of names of marine organisms, including information on synonyms
ZooBank	http://www.zoobank.org	Intended as the official registry of Zoological Nomenclature

Participation

The assembly and maintenance of a names-based infrastructure requires schemas, rules and algorithms to automate processes. Yet biology is not a “units and rules” science within which the totality can be derived from a sum of all of the parts, nor is it as ‘rectangular’ as informaticians might like it to be (where ‘rectangular’ refers to data in columns and rows). Biology has an inherent narrative component, and the elements of our understanding, whether the taxonomic perspectives, or the -ologies that transect the discipline, lack the atomic character of many other sciences, and so require an interpretative approach. The historical assembly of the science has often been by form of a social narrative in which personas have played a significant role in determining what is ‘true’. An infrastructure for biology will not be one based on a few simple principles. Rather, it must handle data objects deriving from a complex, layered, inconsistent and sometimes unpredictable system. With tens of millions of names to manage, the nuancing of the system can only be achieved through active community involvement. Taxonomic experts who are willing to act as custodians for a clade will shoulder the responsibility of keeping the information up to date and of refining the crude algorithm-based approach to suit the nature of the discipline.

At another level, we can promote the evolution of the infrastructure so that it grows to become more appropriate to the task. By placing the concepts, algorithms, and tools into a communal open-source environment and by opening up content through APIs (Application Programming Interfaces) and other web services, we create the foundation for an evolutionary process to come into play and facilitate the emergence of cybertaxonomy as a cornerstone of the discipline. Furthermore, to encourage participation, an attribution system should be provided for all types of contribution to the system (e.g., provision of data, taxonomic editing of clades).

Authoritative

Despite the shapelessness of much of biology, there are better practices and poorer practices. Nomenclatorial aspects of taxonomic practices are usually regulated by the International Codes of Nomenclature following a framework of rules and recommendations that provide a certain level of structure and reliability to the subject. Nevertheless, the accumulation of data from heterogeneous sources will reveal errors and inconsistencies or even introduce new problems (e.g., algorithmically created errors, OCR errors, unicode conversion errors or erroneous “names” introduced by automated name-recognition tools).

Especially because of the peculiarities and widespread relevance of biology, and because of the dependence of a names-based architecture on taxonomy, it will be critical that devices are in place to facilitate continuing improvements in quality (accuracy, precision and completeness of the data environment). These will allow the system to evolve towards authority. Devices to support automatic quality control and quality assurance should be implemented, such as simple consistency checks, algorithms implementing nomenclatural rules, loops that return information from users to providers with the intent of improving fitness for purpose, clear indication of the quality status of a name (e.g., “vetted by expert”, “coming from an authoritative source”, “unverified status”), as well as devices that allow the hypotheses (such as the hypothesis ‘Chromista’) to be tested for consistency against bodies of all indexed data). Transparency and documentation of all elements will help the system to grow into a trustful source of information. No algorithm will ever be able to capture all deviations from general trends such that a system that combines algorithmic solutions to the challenges of scale, integration with clade custodianship by experts, together with devices to allow co-existence of multiple points of view, will more likely achieve the authority that we seek.

Scalable

One of the crucial requirements for the success of a names-based cyberinfrastructure is scalability. In the near future, the online availability of information and the increasing interconnection of data sources will result in rapidly growing numbers of names, relationships between them and data objects attached to them. As an example, new technologies of pyrosequencing have the capacity to generate millions of records of the diversity and abundance of species within ecosystems (Sogin *et al.* 2006)—within a matter of hours. The evolution of molecular technologies will result in growing accumulations of full genomes and community metagenomes. The Global Biodiversity Information Facility (GBIF, http://www.gbif.org) is currently setting itself the goal of indexing over a billion specimen records. Conservatively, we need to plan for 10^{12} data objects. The architecture of the underlying infrastructure must work with extremely large amounts of data, provide effective indexing and management of names, be stable, and permit effective discover and fast retrieval of date.

Conclusions

The EOL informatics group assisted by the PROPE-taxon initiative of the EU Network of Excellence MarBEF (Marine Biodiversity and Ecosystem Functioning, http://www.marbef.org) has been promoting the development of a names-based infrastructure with the properties described above. The components that are in place include repositories of names such as uBio's NameBank, sources of authoritative information — from the nomenclators to the aggregators, the adoption of standards (TDWG) and the emergence of a dynamic networking of names providers (GNI). The release of the Encyclopedia of Life in February 2008 exploited a prototype of this environment, proved that this approach is feasible and that it will contribute to a comprehensive and authoritative management of biological information at large. We are now in the process of building a system with the properties identified here.

Acknowledgements

The authors acknowledge support by the MARBEF EU funded Network of Excellence (contract no. GOCE-CT-2003-505446). Support was also received from ECOSUMMER (FP6-2004-Mobility-2-020501-2). The EOL team acknowledges support from the John D. and Catherine T. MacArthur and the Alfred P. Sloan foundations. Christos Arvanitidis and David Shorthouse are kindly acknowledged for valuable comments on this manuscript.

References

Agnarsson, I. & Kuntner, M. (2007) Taxonomy in a changing world: seeking solutions for a science in crisis. *Systematic Biology*, 56, 531–539.

Berendsohn, W. (1995) The concept of “potential taxa” in taxonomic databases. *Taxon*, 44, 207–212.

Bisby, F.A. (2000) The quiet revolution: biodiversity informatics and the internet. *Science*, 289, 2309–2312.

Bortulus, A. (2008) Error cascades in the biological sciences: the unwanted consequences of using bad taxonomy in ecology. *Ambio*, 37, 114–118.

Cantino, P.D., & de Queiroz, K. (2000) *PhyloCode: a phylogenetic code of biological nomenclature.* Available from: http://www.ohiou.edu/phylocode (accessed 20 June 2008).

Corliss, J.O. (1995) The ambiregnal protists and the codes of nomenclature: a brief review of the problem and of proposed solutions. *Bulletin of Zoological Nomenclature*, 52, 11–17.

De Carvalho, M.R., Bockmann, F.A., Amorim, D.S., Brandão, C.R.F., de Vivo, M., de Figueiredo, J.L., Britski, H.A., de Pinna, M.C.C., Menezes, N.A., Marques, F.P.L., Papavero, N., Cancello, E.M., Crisci, J.V., McEachran, J.D., Schelly, R.C., Lundberg, J.G., Gill, A.C., Britz, R., Wheeler, Q.D., Stiassny, M.L.J., Parenti, L.R., Page, L.M., Wheeler, W.C., Faivovich, J., Vari, R.P., Grande, L., Humphries, C.J., DeSalle, R., Ebach, M.C. & Nelson, G.J. (2008) Taxonomic impediment or impediment to taxonomy? A commentary on systematics and the cybertaxonomic-automation paradigm. *Evolutionary Biology*, 34, 140–143.

Godfray, H.C.J. (2002) Challenges for taxonomy. *Nature*, 417, 17–19.

Kennedy, J. (2003) Supporting taxonomic names in cell and molecular biology databases. *OMICS: A Journal of Integrative Biology*, 7, 13–16.

Kennedy, J.B., Kukla, R., Paterson, T. (2005) Scientific names are ambiguous as identifiers for biological taxa: their context and definition are required for accurate data integration. *In:* Ludäscher, B. & Raschid, L. (Eds.), *Data integration in the life sciences. Proceedings of the Second International Workshop 2005, San Diego, CA, USA.* Springer-Verlag, Berlin, pp. 80–95.

Kluge, N.J. (2000) *Modern systematics of insects. Part I. Principles of systematics of living organisms and general system of insects, with classification of primary wingless and paleopterous insects*. Lan', S. Petersburg, 332 pp.

Koning, D., Sarkar, I.N. & Moritz, T. (2005) Taxongrab: extracting taxonomic names from text. *Biodiversity Informatics*, 2, 79–82.

Leary, P.R., Remsen, D.P., Norton, C.N., Patterson, D.J. & Sarkar, I.N. (2007) uBio RSS: Tracking taxonomic literature using RSS. *Bioinformatics*, 23, 1434–1436.

Linnaeus, C. (1751) *Philosophia botanica in qua explicantur fundamenta botanica cum definitionibus partium, exemplis terminorum, observationibus rariorum*. G. Kiesewetter, Stockholm.

Mayr, E. (2004) *What makes biology unique?* Cambridge University Press, Cambridge, U.K., 246 pp.

Page, R.D.M. (2006) Taxonomic names, metadata and the semantic web. *Biodiversity Informatics*, 3, 1–15.

Sogin, M.L., Morrison, H.G., Huber, J.A., Welch, D.M., Huse, S.M., Neal, P.R., Arrieta, J.M. & Herndl, G.J. (2006) Microbial diversity in the deep sea and the underexplored "rare biosphere". *Proceedings of the National Academy of Sciences*, 103, 12115–12120.

Wilson, E.O. (2003) The encyclopedia of life. *Trends in Ecology & Evolution*, 18, 77–80.

Yoon, N. & Rose, J. (2001) An information model for the representation of multiple biological classifications. *In:* Alexandrov, V.N., Dongarra, J.J., Juliano, B.A., Renner, R.S. & Tan, C.J.K. (Eds.), *Computational Science - ICCS 2001: International Conference, San Francisco, CA, USA.* Springer-Verlag, Berlin, pp. 937–946.