



HAL
open science

Understanding and managing Zoological Nomenclature in the era of Big Data and Open Science

Elie Mario Saliba

► **To cite this version:**

Elie Mario Saliba. Understanding and managing Zoological Nomenclature in the era of Big Data and Open Science. Systematics, Phylogenetics and taxonomy. Sorbonne Université, 2023. English. NNT : 2023SORUS460 . tel-04402685

HAL Id: tel-04402685

<https://theses.hal.science/tel-04402685v1>

Submitted on 18 Jan 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

MUSÉUM NATIONAL D'HISTOIRE NATURELLE
École Doctorale Sciences de la Nature et de l'Homme – ED 227

Année 2023

N°attribué par la bibliothèque

□□□□□□□□□□□□□□

THÈSE

Pour obtenir le grade de

DOCTEUR DU MUSÉUM NATIONAL D'HISTOIRE NATURELLE

Spécialité : Sciences du Vivant

Présentée et soutenue publiquement par

Elie Mario Saliba

Le 6 décembre 2023

**Understanding and managing
Zoological Nomenclature
in the era of Big Data and Open Science**

Sous la direction de : **Madame Annemarie OHLER, Professeure, MNHN**
Madame Régine VIGNES-LEBBE, Professeure, Sorbonne Université

JURY :

M. Robillard, Tony	Professeur, Muséum National d'Histoire Naturelle, Paris, France	Président
M. Pape, Thomas	Associate Professor, Natural History Museum of Denmark, Copenhagen, Danemark	Rapporteur
M. Polaszek, Andrew	Principal Researcher, Natural History Museum, Londres, Royaume-Uni	Rapporteur
M. Berendsohn, Walter	Professor, Botanischen Garten, Berlin, Allemagne	Examineur
M. Welter-Schultes, Francisco	Researcher, Indépendant	Examineur
Mme Ohler, Annemarie	Professeure, Muséum National d'Histoire Naturelle, Paris, France	Directrice de thèse
Mme Vignes-Lebbe, Régine	Professeure, Sorbonne Université, Paris, France	Directrice de thèse

Table of Contents

Figures	3
Tables	5
Acknowledgments	6
Notes.....	7
I. Introduction.....	8
I.1. Studying and naming the living world.....	9
I.1.1. Systematics	9
I.1.2. Taxonomy	10
I.1.3. Nomenclature	12
I.2. Key concepts in nomenclature	22
I.2.1. Scientific name.....	22
I.2.2. The nomenclatural process	29
II. Nomenclature and taxonomic databases.....	45
II.1. Nomenclature in TDWG standards and collection databases.....	50
II.1.1. Nomenclature in the Darwin Core.....	50
II.1.2. Nomenclature in collection catalogues	51
II.2. Current limits in digital nomenclature.....	56
II.2.1. Finding nomenclatural data	56
II.2.2. The structural interdependence of databases	57
II.2.3. The delay in updating data.....	59
II.2.4. The consensus fallacy in digital nomenclature.....	61
III. <i>Lognom</i> : Computerizing the rules of the Code	65
III.1. Objects, properties and processes in nomenclature.....	67
III.1.1. The concept of nomenclatural objects.....	67
III.1.2. Properties of names and works.....	68
III.1.3. The algorithmic structure of each stage of the nomenclatural process	77
III.2. The creation process of <i>Lognom</i>	82
III.2.1. Original specifications for <i>Lognom</i>	82
III.2.2 Entity-relationship model and database schema	85
III.2.3. The main function	92
III.2.4. <i>Lognom</i> 's structure and user interface	97
III.3. Efficiency and limits	108

III.3.1. Efficiency tests	108
III.3.2. Limits and possible solutions	109
IV. Reflections on zoological nomenclature	112
IV.1. Computerizing the Code	113
IV.1.1. Terms in use in both nomenclature and programming	113
IV.1.2. A classification of the rules.....	114
IV.1.3. Suggestions for a more computer-friendly Code	125
IV.2. Protaxa and apotaxa	128
IV.3. Nominal and taxonomic taxa	129
IV.4. Management of <i>nomina dubia</i> and unavailable names	132
IV.5. The management of the rank system.....	135
IV.6. The legal aspects of nomenclature	140
IV.6.1. Is nomenclature a legal system?.....	140
IV.6.2. Is nomenclature a Bona fide* system?.....	141
IV.6.3. What constitutes complexity in nomenclature	144
IV.6.4. Perception of nomenclature by the community	145
V. Perspectives and conclusion	148
V.1. Finishing the software	149
V.2. Managing the etymologies and establishing Lists of Available Names.....	151
V.3. Cooperative work, data sharing and zoological nomenclature standard	154
V.4. Towards a synergy with <i>ZooBank</i> ? Modeling of nomenclature as a whole...	157
V.4.1. Combining <i>ZooBank</i> and <i>Lognom</i>	157
V.4.2. Towards a modeling of nomenclature.	157
V.5. Conclusion.....	161
Glossary.....	163
Bibliography.....	165
Appendix 1: Data dictionary of <i>Lognom</i>	186
Appendix 2: List of the Articles of the Code and their treatment in <i>Lognom</i>	192
Appendix 3: Test of <i>Lognom</i>	220
Résumé.....	224
Abstract	225

Figures

Figure 1: The Löwenmensch.....	13
Figure 2: The pyramid of nomenclatural regulators as defined by Pavlinov (2021). ..	19
Figure 3: The main ranks and series in zoological nomenclature.	28
Figure 4: A simplified version of the nomenclatural process	31
Figure 5: The label and the legend of a specimen of grasshoppers on the MNHN online catalogue. Specimen MNHN-EH-EH16435.	55
Figure 6: <i>Pyrops oculatus</i> in FLOW (Bourgoin, n.d.), an extract of Walker (1850) and a picture of the type specimen of <i>oculatus</i>	56
Figure 7: <i>Bufo</i> <i>variabilis</i> in the GBIF (A), ITIS (B) and Amphibian species of the World (C) databases.	61
Figure 8: Properties of works and names throughout time; part 1 (1758–1930).	70
Figure 9: Properties of works and names throughout time; part 2 (1931–1949).	71
Figure 10: Properties of works and names throughout time; part 3 (1950–1960).	72
Figure 11: Properties of works and names throughout time; part 4 (1961–1985).	73
Figure 12: Properties of works and names throughout time; part 5 (1986–1999).	74
Figure 13: Properties of works and names throughout time; part 6 (2000–2012).	75
Figure 14: Properties of works and names throughout time; part 7 (2012–now).	76
Figure 15: Part of the steps in the test of availability.....	78
Figure 16: A simplified visualization of the process of testing validity.	81
Figure 17: Steps in the test of a <i>nomen oblitum</i> and <i>nomen protectum</i> under Art. 23.	82
Figure 18: Example of a simplified output for an imaginary genus of frogs.	84
Figure 19: Cardinalities in an entity-relationship model.....	86
Figure 20: Entity-relationship model of Lognom.	88
Figure 21: Structure of <i>Lognom</i> in Django, with the corresponding programming languages.	89
Figure 22: Database schema of <i>Lognom</i>	91
Figure 23: Screenshot of the home page of <i>Lognom</i> 's alpha version.	97
Figure 24: All the possible tests offered by <i>Lognom</i>	98
Figure 25: <i>Lognom</i> 's form for testing the availability of a published species series name.	102
Figure 26: Positive result of an availability test for a species name.	103
Figure 27: Negative result of an availability test for a species name.....	103
Figure 28: <i>Lognom</i> 's form for testing the validity of competing names.....	105

Figure 29: Result of a validity test for genus names, based on the example shown in Fig. 16.....	105
Figure 30: <i>Lognom</i> 's form for determining a genus name grammatical gender.....	106
Figure 31: Result for a test determining the grammatical gender of the name <i>Oryctolagus</i>	107
Figure 32: Nominal and taxonomic taxa at the species rank.....	130
Figure 33: A monotypic taxon and its corresponding nominal taxa.	131
Figure 34: A simplified vision of the prospective tool for the identification of valuable nomenclatural specimens in <i>Lognom</i>	150
Figure 35: A visualization of the consequences of two nomenclatural acts in an extract of a database aiming to model nomenclature, at a time t (above) and time t+1 (below).	160
Figure 36: A simplified view of the potential organization around a centralized nomenclatural database.	161

Tables

Table 1: Examples of the different kinds of names, in English for <i>Emberiza hortulana</i> and in French for <i>Thecadactylus rapicauda</i>	15
Table 2: Correspondence between terms used in botanical and zoological nomenclature.	32
Table 3: A classification of taxonomic databases focused on zoology, modified after Grenié et al. (2023).	48
Table 4: Important nomenclatural elements and their equivalent in Darwin Core.	51
Table 5: Structure of the sources for three different names of different animal groups in major databases.	63
Table 6: Example of a highly simplified entry in Lognom for an imaginary genus of frogs.	84
Table 7: Important nomenclatural elements and their equivalent in Darwin Core and <i>Lognom</i>	92
Table 8: Objectives of the major functions of Lognom.	93
Table 9: Types and relationships of the major functions of Lognom.	95
Table 10: Classification of the rules of the Code with the consequences on its potential computerization.	125
Table 11: Database cardinalities between names and taxa, modified after Remsen (2016).	128
Table 12: Vocabulary differences between taxonomic and nominal taxa.	131
Table 13: List of ranks in use in zoological nomenclature.	136
Table 14: Categories of names depending on their etymology.	152
Table 15: Examples of concepts that could form a new nomenclatural standard and their equivalencies.	156

Acknowledgments

I would like to express my deepest gratitude to Dr. Walter BERENDSOHN, Pr. Thomas PAPE, Dr. Andrew POLASZEK, Pr. Tony ROBILLARD and Dr. Francisco WELTER-SCHULTES for agreeing to serve on my doctoral jury.

I would also like to thank Pr. Isabelle MOUGENOT, Pr. Aaron BAUER and Pr. Éric GUILBERT who participated in my Ph.D. committee for their kind encouragement words and useful advice.

Words cannot express my gratitude to Pr. Annemarie OHLER and Pr. Régine VIGNES-LEBBE, who supervised my work. Few people could claim to be luckier than I was, to be able to work with such incredible women.

Many thanks to Pr. Alain DUBOIS for sharing his insights and incredible knowledge of nomenclature with me. I am also thankful to Sylvain BOUQUIN for his help with the deployment of *Lognom*, and his bits of advice and tips in programming. I am also indebted to Pr. Agnès DETTAÏ, Dr. Adeline KERNER, Pr. Violaine NICOLAS-COLIN, Pr. Thierry BOURGOIN, Pr. Valéry MALÉCOT, Dr. Marc PIGNAL and Pr. René ZARAGÜETA Y BAGILS for the help, useful discussions, support, and tips I got throughout the project. Many thanks to my dear colleague and nomenclature enthusiast Dr. Mario CUPELLO, for the insightful discussions we had.

I would like to express my deepest appreciation to the Doctoral School 227 team for funding this project and making it possible. Special thanks to the LIS and SPEC team members and the staff of the Reptiles and Amphibians lab (Laure, Mathilde, Jérôme, Stéphane and Yves). Very special thanks to Victoire KOYAMBA, our dearest librarian.

Lots of love to my dear Ph.D. student friends: Ana, Maya and Guillaume. You're the best, guys!

Finally, a huge thanks to my parents and my sister for their support, and to my aunt Nada who gave me my passion for science and languages. Lots of love to my friends all over the world whose moral support and long telephone conversations kept me going. Léa, Yara, Lynn and Joe, you are treasures in my life.

Notes

Any word or concept followed by an asterisk (*) is defined in the glossary that precedes the bibliography.

In the **blue boxes** are some thoughts on current debates in zoological nomenclature. In the **green boxes** are examples to illustrate the statements made in-text.

The International Code of Zoological Nomenclature (International Commission on Zoological Nomenclature, 1999) is designated by ‘the Code’, The International Code of Nomenclature for algae, fungi, and plants is designated by ‘the ICNafp’, and the International Commission on Zoological Nomenclature by ‘the Commission’.

Dates are written in the format Month-DD-YYYY. The bibliography is given under the APA7 guidelines.

Despite this document being written in American English, the spelling ‘catalogue’ rather than ‘catalog’ and the adjective ‘nomenclatural’ rather than ‘nomenclatorial’ will be used, to better reflect the dominant usage in the community.

I. Introduction

« S'il est vrai qu'on ne peut écrire correctement une langue dont on ignore la grammaire, il est tout aussi évident qu'on ne peut s'exprimer avec la précision désirable au sujet d'une science dont on ne connaît qu'imparfaitement le langage technique. »

Raphaël Blanchard, *Règles internationales de la Nomenclature zoologique*

I.1. Studying and naming the living world

I.1.1. Systematics

Systematics is a branch of biology that focuses on the study, description and inventory of biodiversity. It provides a framework for understanding the evolutionary history, classification, and diversity of life on Earth. The term ‘systematics’ is based on the Ancient Greek σύστημα, for organized body, and the current use of the word is inherited from Linnæus’ *Systema Naturae*. The discipline was defined by Simpson (1961) as “the study of the diversity of organisms and all their comparative and evolutionary relationships including their comparative anatomy, ecology, physiology and biochemistry”. This field is now usually subdivided between phylogenetics, a term derived from ‘phylogeny’, itself coined by Haeckel (1866), for the science that studies evolutionary relationships of living organisms, and taxonomy (see I.1.2.). In some cases, systematics and taxonomy are used interchangeably.

Rineau (2022) writes that, as a scientific field, systematics has to be based on a theory that provides causal interpretations for the classificatory systems it constructs. The theory of biological evolution explains that variations emerge in the living world, which can then be passed on to descendants, ultimately leading to the evolution of organisms. According to that theory, if life appeared once, all organisms on Earth are related to each other. Phylogenetics can use data from various sources, including morphology and molecular data, to construct phylogenetic trees that visually represent the evolutionary history of organisms (Wiley, 1981).

Cupello et al. (2023) clarifies the relationship between phylogenetics and taxonomy by saying that taxonomy is “the subfield of systematics dealing with the classification of organisms, i.e., the delimitation and categorization of taxa, as well as with the naming and identification of these taxa. In modern taxonomy, taxa are delimited based on the phylogenetic relationships of organisms, which are investigated by the other subfield of systematics, phylogenetics”. With these definitions in mind, it is easy to understand that a scientific classification should reflect relationships between taxa, and evolutionary theory constrains the form these relationships should take. This implies that there can only be one natural system, that systematists are trying to bring to light.

To be able to construct that classification, a deep understanding of the living world is necessary. The study of the characters of living beings was included for a long time in a

discipline called natural history¹. This discipline used to cover a multitude of current sciences such as, but not limited to, astronomy, geology, ethnology and medicine, as well as the study of plants, fungi and animals. Zoological natural history is the branch of that discipline that focuses on the latter. Aristotle (4th century B.C., around -350) and Pliny the Elder (77?) are responsible for some of the older works on the matter that reached the current times, namely *Περὶ τὰ ζῷα ἱστορίαι* (*Peri ta zoia historion – About the history of animals*) and *Naturalis historia* (Books VIII to XI). Natural history was then studied by quite a few scholars in the Western world, such as Albertus Magnus (1200(?)–1280), the ichthyologists Guillaume Rondelet (1507–1566), and Hippolito Salviani (1514–1572), Conrad Gessner (1516–1565), a Swiss scientist, Pierre Belon (1517–1564), considered sometimes a forefather of comparative anatomy, Ulysse Aldrovandi (1522–1605), an Italian scientist who inspired both Linnæus and Buffon (Gudger, 1934) and Albertus Seba (1665–1736), an Italian apothecary known for his *Thesaurus*, and curator of a collection that was later used by Linnæus. Eventually, natural history became partly synonymized with taxonomy.

1.1.2. Taxonomy

Exploring and understanding the vast diversity of living organisms is a fundamental pursuit of the biological sciences. At the heart of this endeavor lies the study of taxonomy, the science of classifying living organisms.

Taxonomy is derived from the Greek words τάξις (*taxis*), meaning order or arrangement, and νόμος (*nomos*), meaning law or rule. It was introduced as a concept by de Candolle (1813), but some consider it the oldest profession on Earth as the earliest representatives of mankind probably were taxonomists out of survival necessity (Chmielewski & Krayesky, 2013). Taxonomy encompasses a multifaceted approach to describing, identifying and classifying organisms. It thus contains many different subdisciplines, that have been divided in different ways, depending on the dimension of interest. In this manuscript, a distinction will be made between microtaxonomy (Mayr, 1982) which studies specimens and species, and macrotaxonomy (Mayr, 1969) which focuses on larger groups and higher-ranked taxa.

¹ In its original etymological sense, history (ἱστορίᾱ in Ancient Greek) is to be understood as “research” or “inquiry” (Liddell & Scott, 1940). Natural History includes other natural sciences, such as geology.

Approaches and Methods in Taxonomy

Many approaches have been used to classify living beings. Although these methods sometimes give rise to disputes over their effectiveness, or give way to accusations of obsolescence, they each provide part of the complex picture of Life.

One prominent approach in taxonomy is the study of morphology, which relies on the external (morphology) and internal (anatomy) characters of organisms. This method involves scrutinizing an organism's anatomical structures, such as its shape, size and color, to discern its place within a classification. It was the principal historical method that provided the foundation for most classifications.

Molecular taxonomy, a more recent approach, employs the analysis of proteins and genetic material, such as DNA and RNA, to establish similarities and evolutionary relationships among species. By comparing genetic sequences, scientists can uncover shared inherited molecular traits, often leading to reclassification based on molecular evidence.

Ecological taxonomy considers the ecological niches and habitats of organisms as a basis for classification. It takes into account an organism's interactions with its environment, helping to discern functional roles within ecosystems. This approach is particularly relevant for understanding species distributions (Woodbury, 1952).

Lastly, integrative taxonomy is an approach that chooses to combine multiple data sources, including morphology, genetics, ecology and behavior, to provide a comprehensive understanding of species relationships. By synthesizing diverse information, integrative taxonomy aims to provide a more holistic and accurate depiction of biodiversity (Dayrat, 2005).

This work does not intend to discuss the technicalities of the taxonomic process, but it is indispensable to keep in mind that taxonomy has an important instability burden linked to the variety of methods and approaches used in this field². Like any science, taxonomy benefits from progress in technology, methodology and epistemology. These positive changes reinforce the potential instability in the results produced by that science, namely

² An image I like to give when asked to describe how taxonomy works is that of an impressionist painting. It is easy to distinguish shapes and objects, but it is way more complex to draw boundaries, each person trying to do so will vary a little bit compared to the next one. To complicate matters, as there are different methodological approaches in taxonomy, the people who are asked to define these boundaries do not even agree on which tool to use to draw the lines.

classifications. Furthermore, taxonomy suffers from an intrinsic severe philosophical handicap. Trying to fit fuzzy notions like populations or species (Hey, 2001) into the strictly organized, hierarchal construction constituted by a classification, can be a real challenge.

Taxonomists are also responsible for the identification of specimens and their attribution to a taxon. That part of taxonomy benefits from diagnoses, descriptions, dichotomous keys, genetic profiles, and other such tools (Cupello et al., 2023). Like taxonomy, identification has benefited from major technological advances, particularly in computer science and artificial intelligence. The introduction of multi-access keys accessible online, e.g., Xper³ (Kerner et al. 2021) eased the process of identification (Murguía-Romero et al., 2021) and made it available and accessible to a larger audience. In the same vein, the development of AI tools for automatic identification such as Pl@ntNet (Joly et al., 2014) makes it even possible for non-specialists to recognize the fauna and flora that lives around them.

1.1.3. Nomenclature

If taxonomy constitutes the science of classification, nomenclature is the discipline responsible for naming the groups of living beings that have been defined. Naming animals, however, is one of the oldest endeavors undertaken by human societies, as interactions between them and animals predate history. Communicating ideas about these organisms quickly became a major issue for human societies, as animals constitute an important part of mankind's nutritional resources, potential threats and sources of materials for handicrafts. They also sometimes constitute important spiritual or cultural entities.

One of the first representations of animals as cultural ideas, in that case, small figurines, are the sculptures found in the Swabian Jura, in the locality where the 'Löwenmensch' (Lion-man) (Fig.1) was unearthed. These works of art go back to 40,000 BP (Museum Ulm, n.d.). Without access to the time ways of communication, they constitute a primitive physical trace of humanity's recognition of the animal kingdom on a cultural level.



Figure 1: The Löwenmensch. Picture by Dagmar Hollmann, distributed under CC BY-SA 3.0.

With the advent of writing, many naming systems have been recorded, usually shared by a given human population, in a given geographical area, timeframe and language. These names are called vernacular or common names, and can differ a lot even inside a given culture. It is possible for a language to use multiple vernacular names to designate the same entity, or to have none, and thus have to rely on loanwords: Giraffe, Jackal, Kangaroo or Orangutan are English (and French) borrowings from Arabic (زرافة), Indo-Persic (شغال; शृगाल), an aboriginal language called Guugu Yimidhirr and Malay respectively (Deroy, 1956). Moreover, some names may vary in their designation over time. In French, for example, the original name for a fox, “Goupil” or “Gupil” (e.g., in de Thaon³ [1121-1135⁴]), ultimately from the Latin *Vulpes*, became “Renart” then “Renard” by eponymy after the publication of a series of popular stories on a deceitful fox (joined today in one opus and called ‘le Roman de Renart’) by mostly anonymous authors starting in the 12th century. Vernacular names also change depending on the development stage of an animal, its sex, or caste for eusocial animals.

³ Sometimes spelled ‘de Thaün’ or ‘de Thau’n’

⁴ The exact date of publication is unclear

Vernacular names were sufficient for communication, as long as animals were viewed as utilitarian objects (a modern scientist might say “for their ecological anthropo-centered value”), inside a given culture. At this stage, which Pavlinov (2021) calls ‘folk nomenclature’, taxonomy was itself quite rudimentary. There was no need for any scientificity at that point.

Indigenous names in nomenclature

Three years ago, a call was made by two New Zealanders to modify the rules of the ICNafp (Turland et al. 2018) to “restore” indigenous names in taxonomy (Gillman & Wright, 2020). This would constitute, in fact, a philosophical return to folk nomenclature. A handful of articles have been published in response, e.g., Palma & Heath (2021), listing the main weaknesses of that proposal. While that particular issue will not be debated here, the main takeaway is that without respect for nomenclatural norms (see I.1.3.1.), in this instance, Universality and Monosemy, a nomenclatural system is doomed to fail, and thus, it is impossible to accept a system lacking the necessary elements to ensure its scientificity.

Currently, there are at least four different kinds of names given to animals (examples given in Table 1). The first one is the vernacular name in a given language, which people with little to no connection with taxonomy use. There might be multiple vernacular names for a single taxon. The second one is normalized names, which are names that are expressed in a vernacular language but associated with a given taxon. These names can be strictly associated with one taxon, the way a scientific name is, but can also cover a large group of taxa that are deemed to look similar enough. Normalized names are sometimes created and managed by international bodies for a given language, like the *International Ornithological Committee* which assigns English names to birds (Aves), or the *Commission Internationale des Noms Français des Oiseaux* which fills the same role for the French language. Normalized names will be used by people with an interest in natural science. They appear in identification keys that are destined for an amateur public, some legal texts on a national level, or on information panels in parks and zoos. The third category of names is vernacular scientific names, or translated names e.g., a French scientific name. These are most likely unknown to the general public and are usually a direct translation of the scientific name into a language. Most of the time, these names are proposed by scientists when describing a new species or genus, without knowing if it will reach general use in the population or with the amateur public, and without concern about older names given to the taxa. Moreover, vernacular scientific names are often

not even introduced in the language of the local human populations that are likely to interact with the animal. A translated scientific name is attached to strictly one taxon, but without any rule or committee to maintain and use them, ends up being an artificial construction most of the time. It can be used in some cases in place of a normalized name if the language lacks a name to designate a taxon, while needing to refer to it in a non-scientific context. The fourth and last category of names is scientific names, which will be the main focus of this work. These names are used when an unambiguous system of naming is crucial for a given work, and usually, aside from science and scientific work, will appear on international agreements and conventions for trade, hunting and conservation.

Table 1: Examples of the different kinds of names, in English for *Emberiza hortulana* and in French for *Thecadactylus rapicauda*.

Language	English	French
Vernacular name	Ortolan	Grand Mabouya collant
Normalized name	Ortolan bunting (see Gill et al. [2023])	None
Translated scientific name	Gardener bunting (not in use)	Thécadactyle à queue turbinée (in Dewynter et al. [2019])
Scientific name	<i>Emberiza hortulana</i> Linnæus 1758	<i>Thecadactylus rapicauda</i> (Houttuyn, 1782)

Scientific names, born from the joined heritage of Carolus Linnæus, Hugh Edwin Strickland, Raphaël Blanchard and Francis Hemming in zoology, are part of a discipline called ‘nomenclature’. Nomenclature itself is a polysemic term that can be defined as a subdiscipline of taxonomy whose aim is to determine how to create, manage and select the correct scientific name for a given taxon. Nomenclature also designates the list of scientific names that result from the application of these rules. In zoology, these names are governed by the *International Code of Zoological Nomenclature* (International Commission on Zoological Nomenclature, 1999), shortened to ‘the Code’ thereafter. The Code provides the rules, guidelines and

recommendations to correctly manage and use scientific names. It is written, managed, and promoted by the *International Commission on Zoological Nomenclature*, which will be referred to as ‘the Commission’ thereafter.

A few historical elements

Nomenclature as it is known today was introduced by Carl Linnæus, also known as Carl von Linné, originally for botany, in a book called *Fundamenta Botanica* (Linnæus, 1735; Malécot, 2008), as a series of aphorisms defining rules and guidelines to establishing names (Dayrat, 2010). Linnæus is credited with the creation of the binominal system, giving each species a name composed of two names, but his original idea was quite different from what is understood today in that system, as the current specific name is technically the descendant of Linnæus’ *nomen triviale*, a name first used for indexing, rather than his *nomen specificum legitimum* (Raynal-Roques & Roguenant, 2008), which he considered the specific name. For Linnæus, the *nomen specificum legitimum* was meaningful, as it provided the definition of the taxon (Pavlinov, 2021). With time, his practices evolved and the binominal system is now referred to as Linnean nomenclature.

The first few decades of debate around Linnean nomenclature were mostly discussed among botanists, except for Buffon, who was a critic of Linnæus’ system, especially of the binominality at the species level, as it was, in his opinion, an unwelcomed mix of taxonomy and nomenclature (Malécot, 2008). The discussion about the fact that names should be meaningful (like Linnæus intended) or devoid of meaning (what Pavlinov [2021] calls nominalism, the idea of a name being a label and nothing more), appeared fairly quickly. This debate was only still going almost a century later when Alphonse de Candolle introduced the *Lois de la nomenclature botanique*. Alphonse de Candolle himself will settle that debate in his *Nouvelles Remarques* by writing: “La désignation d’un groupe, par un ou plusieurs noms, n’a pas pour but d’énoncer les caractères ou l’histoire de ce groupe, mais de donner un moyen de s’entendre lorsqu’on veut en parler [The purpose of naming a group, by one or more names, is not to state the group’s characteristics or history, but to provide a way of understanding each other when we want to talk about it]” (de Candolle, 1883).

In zoology, the next big step after Linnæus was the Stricklandian Code (Strickland et al., 1843), written by a committee appointed by the British Association of the Advancement of Science (Melville, 1995). That Code introduced elements of what would become the principle

of Priority and the process of typification (Dayrat, 2010). However, as the principle of Priority was still open to too many interpretations, and with the absence of a general consensus in a fragmented community, the Strickland Code failed to bring taxonomists together (Dayrat, 2010). In 1858, Kiesenwetter proposed a Code (Kiesenwetter, 1858) to amend the Stricklandian Code (Melville, 1995) with a watered-down principle of Priority, and introduced notions such as a mandatory diagnosis in Latin, and the concept of combination (Malécot, 2008), but still failed to create a consensus.

Another wave of debates in the following years will focus on the Type vs. Description approach to associate names with taxa, and the limits of the principle of Priority. The first four *International Congresses of Zoology* ended up choosing 1758 as a starting date and saw the birth of the *International Commission on Zoological Nomenclature* in 1895 (Dayrat, 2010; Melville 1995). Blanchard's *Règles internationales de la nomenclature zoologique*, in French, English, and German, was published a decade later in 1905 (International Commission on Zoological Nomenclature [Blanchard et al.,] 1905), and constituted the first International Code. The first half of the 20th century will witness very few updates of that 1905 Code, until the election in 1936 of Francis Hemming as secretary of the Commission. In 1942, Hemming launched the *Bulletin of Zoological Nomenclature* and republished the first *Opinions* issued between 1910 and 1920, then in 1947, with Karl Jordan, founded the International Trust for Zoological Nomenclature (Hemming, 1995; Malécot, 2008). In the early 1950s, the drafting of a new Code, which would incorporate all the *Opinions* issued by the Commission in the meantime would start and will result, in 1961, in the first edition of *the International Code of Zoological Nomenclature*. From this point onward, most of the critical notions of nomenclature that have been debated during the previous century have been accepted by the overwhelming majority of the taxonomist community as well as the Code's sole authority on the question. The second edition of the Code was published in 1964, the third in 1983, and the fourth, currently in power, in 1999.

Throughout their historical modifications, the rules of nomenclature evolved to ensure the maximum possible freedom of thought and action for taxonomists. That property of nomenclature can be called theory-freedom and stems from the nominalist approach that was selected by history (see Pavlinov [2021] for a discussion on the limits of theory-freedom; see also Saliba [2022] for an analysis of that concept).

Despite this apparent independence, nomenclature and taxonomy are sometimes mixed up, as they are deeply related disciplines. They are however distinct in both their epistemological nature and their object of study. Taxonomy is a science, based on hypotheses constructed with data leading to the definition and delimitation of taxa. These hypotheses can be debated, supported or proven wrong. Nomenclature, however, is a scientific discipline⁵, consisting of the application of rules to manage names. Nomenclature necessitates a pre-existing taxonomy to be successfully applied, and in its current form (the Code), is applicable under any ranked taxonomic paradigm (Dubois, 2008b; Pyle & Michel, 2008; Pavlinov, 2021).

Nomenclatural norms

The Code defines nomenclature as “a system of names, and provisions for their formation and use”, a definition that can be applied to systems other than taxonomic nomenclature. The crucial point in that definition is the idea of “provisions”. Pavlinov (2021)⁶ proposes a hierarchy of these provisions, that he calls nomenclatural regulators (Fig. 2), with nomenclatural norms as the most inclusive element of a nomenclatural system.

Norms can be defined as necessary regulations of any nomenclatural system that make it possible for it to be recognized as rational and scientific (Saliba, 2022). They differ from Principles by the fact that there are no alternatives to a norm that can exist in a system without compromising the scientificity of the system, while multiple opposing principles can be proposed or even cohabit within a nomenclatural system. Norms do not constitute *conditiones sine qua non* for the operation of a nomenclatural system, but more of an ideal place towards which it should be directed. Principles and rules are practical tools used to achieve the aim of the norms.

At least four norms can be identified as playing a crucial role in underlying the Code’s nomenclatural system.

⁵ To be understood here as “a particular area of study, especially a subject studied at a college or university” (Cambridge Dictionary, n.d.)

⁶ Pavlinov describes the hierarchy in the legal order, putting norms at the top of the pyramid. Fig. 2 represents the hierarchy by considering the norms as the base of the pyramid, the latter getting narrower each time the element in it gets more precise. Even if the graphical representation varies, the core idea is the same.

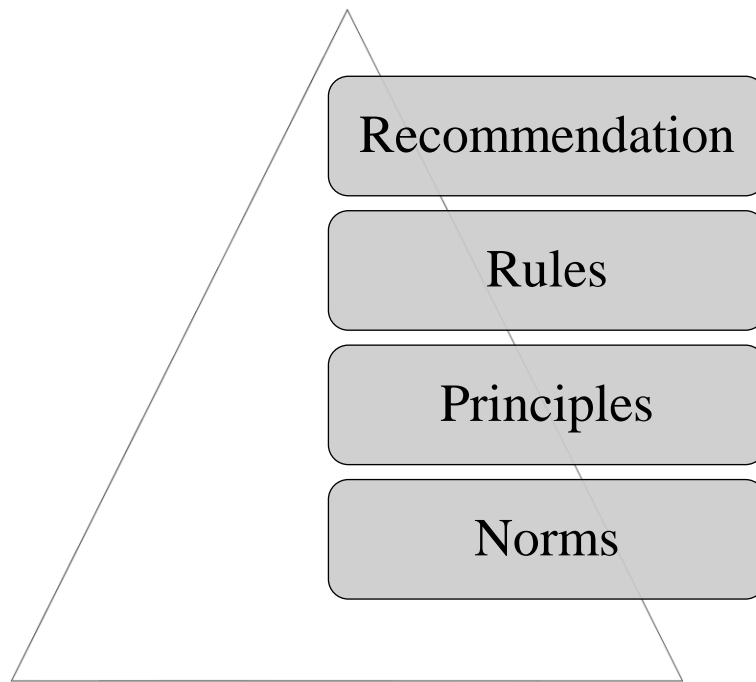


Figure 2: The pyramid of nomenclatural regulators as defined by Pavlinov (2021).

Monosemy

Monosemy, as Pavlinov (2021) defines it, is the unambiguous link between “one object [a taxon] and one name”. Technically speaking, monosemy designates the fact that the scientific name is unambiguously linked to one taxon and not the fact that a taxon has only one scientific name, but for the sake of this demonstration, the formula put forward by Pavlinov will be kept as a definition. Monosemy is one of the main regulators that distinguish vernacular names from scientific names, as the former are polysemic. An in-depth analysis of monosemy, the concept of “admissible polysemy*” and solutions to the issue raised by the latter are discussed in Saliba (2022). Monosemy takes form in the Code in the Principles of Homonymy, Synonymy, Typification, Priority and Usage, which all play a role in ensuring the unambiguous naming of taxa.

Universality

The norm of universality, as defined in this work, corresponds partly to Dubois’ (2005) “Principle of Universality” and “Principle of Univocality” and partly to the principle of “overall universality” *sensu* Pavlinov (2021), and has been discussed in detail in Saliba (2022). Universality corresponds to the fact that the overwhelming majority of taxonomists

follow the same nomenclatural system, independently of their country, culture or language of origin.

Universality is linked to every part of the Code as it is an underlying concept of the ‘international’ nature of the Code itself. It underpins the use of Latin as a universal language. It is also the core of the concept of Availability, meaning that a name has to be recognized as usable by every follower of the Code, whatever the author’s or the users’ of the name cultural background is. Part of the rules on Publication also are a consequence of this idea of Universality, as the accessibility of nomenclatural acts is essential to have a widely accepted system.

The moralistic temptation in nomenclature

That universal dimension of nomenclature is not an easy dimension to manage. Humans are deeply rooted in their culture and from that culture emerges a set of values, laws and taboos. That sometimes leads to the will to conform every cultural tool to an axiomatic moral reference, which by definition cannot be universal. In nomenclature, it takes the form of calls to delete eponymous names linked to people deemed undeserving of honors –for the time being, in the context of the ICNafp (see Hammer & Thiele [2021] or Smith & Figueiredo [2021] for proponents, Mosyakin [2022] for a rebuttal). The Commission (Ceríaco et al., 2023) took a stand against calling into question the ethics of eponymous naming preventively.

Universality became an essential factor for recognizing a nomenclatural system. The universality of nomenclature, in this case, is not a norm per se that ensures the scientificity of a system. A nomenclatural system can in fact exist and be scientific without being popular or widely used. However, the system must explicitly be intended to be universal to become a scientific nomenclature. Otherwise, it would be a particular case of normalized nomenclature, linked to one or a few people, or a given community. So, the intent of universality is what really constitutes the norm.

Regulation

Nomenclature as a system can only function when clear rules are established to make it possible for scientists all over the world to arrive independently to the same conclusion when

dealing with nomenclatural rules, especially when the scale of the system arrives to a point where internal, peer-to-peer communication becomes impossible⁷. This norm coincides with the general scientific property of reproducibility, as a way to ensure that no decision is taken arbitrarily. A regulating body, responsible for creating and maintaining the regulation system seems to always emerge when the scale of operation of that system becomes large enough (see IV.6.2.). In the case of zoological nomenclature, that leadership took form in the formation of the *International Commission on Zoological Nomenclature*. Regulation underlies the existence of the Code, as well as the role and influence of the Commission.

Permanence

The norm of Permanence can be paralleled by that of Universality. Universality represents the cultural and geographical continuity of a nomenclatural system. Permanence, on the other hand, is what enables the historical and temporal continuity of such a system. Information relating to the taxa being studied must be retrievable not only by contemporary colleagues but also by future scientists, otherwise, the nomenclatural system will generate polysemy. In the Code, this notion of Permanence plays an essential role in questions of publication, especially in the choice of publication medium, since it must be deemed perennial enough to survive for use by future generations. Permanence also underlies the relatively new notion of Registration, born with the digital media age, for which the introduction of *ZooBank* (Polaszek et al., 2005a) is a prime example in zoological nomenclature.

What about Stability?

The concept of Stability in nomenclature is a very tricky one to handle. The nomenclatural literature uses the term very often generally in the sense of an end goal to has to be reached (e.g., in the title of the book *Towards Stability in the Names of Animals* [Melville, 1995]). In a way, this would equate stability with the definition of a norm, since it is a goal towards which a nomenclatural system must strive. The peculiarity of the notion of stability is that it is in conflict with the nature of the objects, taxa, on which nomenclature applies. As a science, taxonomy is in perpetual motion. Taxonomic conclusions are constantly changing, and

⁷ In the beginning of the 20th century, it was still possible to fit most taxonomists, or representants of most taxonomic communities in a room, as illustrated by the discussions that took place around the Botanical Code (see Briquet et al. 1906).

nomenclature must adapt accordingly. So, in reality, it is much more a question of a weakened concept of Stability, reminiscent of the adage ‘as stable as possible, as unstable as necessary’. In practice, many types of instability occur in nomenclature, some of them having a direct link to taxonomy, like a change in the rank of a name or a modification of a taxon’s valid name, usually accompanied by an update in the list of synonyms. Other changes that might occur might be a reinterpretation of the allocation* of a name to a taxon, e.g., by the designation of a lectotype, the resurfacing of an older name or spelling correction mandated by the Code. Nomenclature can never claim to be completely stable, as this would mean that all taxonomic questions would be resolved forever, which could never happen in a world where Life is constantly evolving and scientific methods, concepts and approaches are changing. It can however aim to reach a state of balanced instability where unnecessary modifications are reduced to a minimum.

I.2. Key concepts in nomenclature

I.2.1. Scientific name

Scientific names, sometimes called nomina (Dubois, 2000; sing. nomen), are the main object of interest in nomenclature. From this point onwards, “names” will designate scientific names, unless specified otherwise. In the Code, scientific names are defined in the Glossary as

Scientific name: Of a taxon: a name that conforms to Article 1, as opposed to a vernacular name. The scientific name of a taxon at any rank above the species group consists of one name; that of a species, two names (a binomen); and that of a subspecies, three names (a trinomen) [Arts. 4 and 5]. A scientific name is not necessarily available.

An analysis of this definition leads to the following elements. Art. 1.2. refers to the scope of scientific names, and Art. 1.3. refers to exclusions, in other words, entities that cannot be named using scientific names. However, Art. 1.1. creates a circular definition (a scientific name is a name used in zoological nomenclature, zoological nomenclature being the “system of scientific names for animal taxa” [Glossary]). The definition by opposition to vernacular name, i.e. “A name of an animal or animals in a language used for general purposes as opposed to a name proposed only for zoological nomenclature” is also unsatisfactory as the concept of general purposes is difficult to interpret when confronted to the fact that some

scientific names are also used for general purposes. However, a very interesting point in the definition of vernacular name is the idea that a scientific name is indeed proposed for zoological nomenclature.

Interestingly, the Code's definition of a scientific name is slightly different in the French version and reads:

Nom scientifique: Nom d'un taxon qui est conforme à l'Article 1, par opposition à un nom vernaculaire. [...]

This first part of the definition in the French version makes it possible to read that it is the taxon that conforms to Art. 1. rather than the name. The definition thus makes a bit more sense.

Very few attempts have been made at the task of finding an absolute definition of a scientific name. This lack of a proper definition is mentioned in Rousse (2003), as a limit to proper formalization of nomenclature. In fact, an ambiguity resides in the distinction between a scientific name as an absolute and a scientific name within a given system. For the absolute definition, at least four elements separate a scientific name from the three other categories of names that have been described above:

- 1) Permanence: A scientific name should not vary significantly through time.
- 2) Universality: A scientific name should not vary through space or depending on the culture.
- 3) Monosemy: A given taxon is intended to have one and only one scientific name in a given system of codified nomenclature, and a scientific name is intended to apply to one and only one taxon.
- 4) Intent of scientificity: There is an explicit or implicit intent to enter the name in a scientific system of codified nomenclature.

Without surprise, three of the four elements fit perfectly with the four nomenclatural norms mentioned above. Unity in time, space and culture makes it possible to distinguish between a scientific name and vernacular, normalized and translated names, as it *de facto* excludes any

name in a vernacular language⁸. The explicit or implicit intent to enter the name in a system of codified nomenclature makes it possible to distinguish a name in a Latin text that does not aim to enter nomenclature, as it is just mentioned, and a name that is intended to have a scientific value.

In the context of the Code, a mention of the system itself is necessary, at least to the extent that the definition does not have to recognize alternative scientific names. I would propose the following definition for the Code's glossary:

Scientific name: A scientific name is a name that falls, implicitly or explicitly, within the frame of the present Code. It is applied to a taxon as defined in Art. 1 and is proposed with a permanent and universal aim. A scientific name may be deemed unavailable.

And the following definition for the concept in general:

Scientific name: A name applied to a taxon in the context of Systematics. It usually falls under the authority of an ensemble of rules and may be excluded or deemed inadmissible by said rules. These rules should be built on norms that ensure that the names proposed in their frame are as permanent, universal and monosemic as possible.

The use of Latin

Scientific names are sometimes considered to be Latin in language or Latinized⁹. Historically, a majority of names were derived from Classical Latin (e.g., *Felis sylvestris* Schreber, 1777, which means 'forest cat'), Ancient Greek (e.g., *Panthera* Oken, 1816, which is 'all the beasts of prey'), or a mix of the two, (e.g., *Miniopterus* Bonaparte, 1837, which is derived from the Latin 'minus' and the Greek 'pteron', for 'small wing'). Names can also be derived from

⁸ On the condition that vernacular languages do not take the role of a central scientific language like Latin did in Europe first, and then the world. If a *lingua franca* ended up being shared by all scientists for a period of time comparable to Latin (over the span of centuries), a sensible point could be made that the use of that language does not call into question the norms cited above.

⁹ Oddly enough, the Code does not say explicitly that names should be treated as Latin words, but as Latinized. This is consistent with the possibility of deriving a name from a modern language, or forming one with an arbitrary combination of letters (Art. 11.3.). Some people mistakenly still refer to scientific names as Latin names, but the strict respect of that tradition has in fact long been abolished if it ever existed (Brisson based *Giraffa* on an Arabic word, as early as 1762). The 1905 Code (International Commission on Zoological Nomenclature, 1905) already accepts explicitly names that are not of Latin or Greek origin.

modern languages or be an arbitrary combination of letters, as long as it is pronounceable¹⁰ when transliterated. Latin offered the advantage of being the *lingua franca* for scholars in European science when Linnæus founded modern nomenclature. Writing in Latin, and referring to taxa in that language made it possible for taxonomists to not be hindered by any translation effort at the time.

This advantage provided by Latin faded when it slowly disappeared from scientific literature and school curricula (although describing taxa in Latin was made mandatory by the ICNafp until 2012 [McNeill & Turland, 2011]). The fact that Latin became a dead language offers however a new advantage to nomenclature. A dead language is unable to evolve, and thus will not go through a change of meaning, spelling, or grammatical rules anytime soon. This fact leads, in theory, to the possibility of eventually fixing (or deducing with invariable law) the ending of every name derived from Latin or treated as a Latinized word. Before 1999, it was also possible to deduce the root of a family-series name if the type-genus was derived from Latin (Art. 29.3.; 29.4.). As an example, *Strix* Linnæus, 1758, a genus of owls, based on a Latin word for a blood-sucking monster, has *Strigis* as a genitive. The root formed from that name is thus *strig-*, which served to create the family name *Strigidae* Leach, 1820.

¹⁰The pronounceability of a name is a subjective notion, highly depending on the language of origin of the reader. *Aaqrab*, that could be transliterated as *'akrab* or *Aaacrab* (from *عقرب*: scorpion) is perfectly pronounceable for an Arabic speaker, but is probably considered unpronounceable by a French one. There is however at least one scientific name derived from a similar sounding word, *Aafrita* Szwedo & Azar 2013, for a genus of fossil Hemiptera, derived from *عفريت* meaning 'Ifrit' [a demon of fire].

Gender agreement in nomenclature

During the writing of this manuscript, a fiery discussion took place on the *Taxacom* platform about the usefulness of gender agreement in zoological nomenclature (<https://lists.ku.edu/pipermail/taxacom/2023-June/subject.html#start>). Opponents to the retention of this Latin linguistic element cited the complexity of the corresponding grammatical rules and the fact that modern taxonomists are unfamiliar with Latin. They also mentioned the fact that some complex cases can be quite time-consuming, as well as the fact that gender agreement had no impact whatsoever on taxonomy. Proponents of gender agreement opposed the fact that only a small portion of names are problematic, Latin rules are not complicated to apply, as only a tiny subset of Latin grammar is necessary, and that complex cases are an issue that would be resolved eventually.

There are many ways to approach this problem, and many solutions have been proposed:

For the proponents of the abandonment of gender agreement:

- Reversing to the original spelling of the specific name
- Freezing the spelling (using the last spelling used before a given date)

For the proponents of the conservation of gender agreement:

- Having a centralized database or list of names with the nature of the specific name and the gender of the generic one.
- Trusting the taxonomists to determine the correct gender agreements as they did for the last two centuries.

I am personally opposed to the abandonment of grammatical agreement. The ability to eventually be able to manage perfectly a system is, in the long term, the best one to support, on the condition that it does not rely on overthrowing what exists for a meager advantage.

There are two main reasons I would like to submit for consideration:

Two approaches have been proposed to replace gender agreement: freezing the spelling at a given date, or referring back to the original spelling. I exclude usage as it is not an easily accessible and hardly objective approach. With the current gender agreement system, dynamic stability should be reached eventually.

For the frozen spelling, aside from some currently studied animals, the last modification of a binomen might be difficult to find, creating potential instability in a system that would have

lost its flexibility. The last modification might not be widely accepted, (e.g., *Aus bus* becoming *Ca ba* in 1979, is the last recombination and use of that name, but the taxonomic change was not followed by anyone, so the change to *Aus ba* would surprise most specialists of the group) while being compliant with the rules.

For the original spelling approach, the issue might arise in the case of the discovery of an older authorship for the name, with potentially another original combination, or a change in the gender of a name that has been stable for decades. For example, *Hylarana signata* (Günther, 1872), was introduced as *Polypedates signatus*, but only stayed in a masculine form for 10 years, before changing to the feminine form when transferred to *Rana* in 1882, and then *Hylarana* in 1938 (Frost, 2023c). It stayed feminine for more than 85 years. A sudden change to *Hylarana signatus* in databases, catalogues, zoos and field guides in the name of ‘stability’ might be puzzling to some, if understood at all. I would still, however, be more favorable to this approach than the former if gender agreement had to be left behind.

Another important argument that did not appear in the discussion about gender agreement is the fact that nomenclature is, in many ways, the storefront of taxonomy with both the general public and the governing bodies that fund research. That calls for the consideration of the rupture of stability for no scientific reason, as well as the importance of elegance given to scientific names. Seven of the ten most spoken languages (English, Mandarin, **Hindi**, **Spanish**, **French**, **Arabic**, Bengali, **Russian**, **Portuguese**, **Urdu** [Central Intelligence Agency, 2023]) are gendered, meaning that it would seem weird for their speakers to see a rupture of agreement in another gendered language. It might also end up backfiring on the discipline’s reputation if it seems that taxonomists are unable to apply basic language rules.

As for the solution, I am favorable to the approach supported by Commissioner Douglas Yanega, which consists of an acceleration of the natural flow towards a stable dynamic system, by having either a centralized reference for the gender of genus names and the grammatical nature of specific names, or a smaller community-driven effort that might rely on already present tools such as the submission of parts of the List of Available Names (LAN*), with the support of *Lognom* (see V.3.). A more radical solution is that for very complex cases, e.g., a name that might be presented as masculine in some dictionary and neuter in others, would be to introduce a rule in the Code that allows a taxonomist to set the

gender of a genus name in stone, in the same way a first Reviser would with multiple spellings. It would need to provide a clear pathway so that the average nomenclature user would not need anything more than a handful of dictionaries to solve a gender agreement case, but should lead to a satisfying and elegant dynamic stability.

Ranks and groups (series)

In zoological nomenclature, names are assigned to ranks and ranks belong to a group, that will be called series from now on (see III.4.1. for the rationale). This ranked approach of zoology is inherited from Linnæus (1758), who originally recognized the *species*, *genus*, *ordo*, *classis* and *regnum* ranks. The Code mentions quite a lot of ranks, from kingdom to infrasubspecific ranks. Seven ranks are deemed mandatory in classification, meaning that any taxon should be included in a higher taxon assigned to these ranks, even if the taxon is monotypic*. They are, from the most inclusive to the least inclusive, kingdom, phylum, class, order, family, genus, and species (in bold in Fig. 3).

In other words, on the one hand, a family has to be in an order, a class, a phylum, and a kingdom. On the other hand, the use of a taxon at the superfamily or subclass rank is left to the judgment of taxonomists. These ranks are thus considered optional.

Kingdom
Phylum
Class
Order
Superfamily
Family
Subfamily
Tribe
Subtribe
Genus
Subgenus
Species
Subspecies

Figure 3: The main ranks and series in zoological nomenclature.

The seven mandatory ranks are in bold, and each series is in a different color.

Zoological nomenclature covers the kingdom Animalia, and part of what was the kingdom Protista (Art. 1.1.1.). The other part is left to the ICNafp, but some ambiguous taxa can end up having a name under both the ICNafp and the Code. Corliss (1995) estimates that around 30,000 protistan names are in this situation.

One of the main differences between the ICNafp and the Code is the fact that zoological nomenclature recognizes nominal series (or groups). The first series is the species series that contains the subspecies and species rank (in yellow in Fig. 3). The second is the genus series, which counts the genus and subgenus rank (in green in Fig. 3). The third is the family series that covers every rank over the genus and under the superfamily (in blue in Fig. 3). Five ranks belonging to this series are explicitly mentioned by the Code and are, from least inclusive to most inclusive, the subtribe, the tribe, the subfamily, the family, and the superfamily. The family series can, in theory, be composed of as many ranks as deemed necessary for a given classification.

Interestingly, for Pavlinov (2021), the use of ranks is one of the cases where nomenclature, with the existence of mandatory ranks, impinges on the taxonomic freedom of scientists. The respect of a ranked system obliges taxonomists to introduce names and taxa that might force them to define ranked characters. A monotypic genus would need to be described with ‘generic characters’ that nothing logically differentiates from ‘specific characters’. Ranks are often attacked for their epistemological vacuity (Laurin, 2023), and are vastly accepted to be closer to a convention than a meaningful scientific construction.

It is difficult to disentangle the taxonomic and nomenclatural dimensions of ranks, as the latter has become such an integral part of post-Linnean taxonomy. However, it is at least as difficult to imagine a hierarchical classification system without naming the levels of the hierarchy, as it happens with a phylogenetic tree. Ranks do provide the advantage of creating a perfect partition of Life and are not sensitive to the addition or merging of taxa in the classification. Ranks will continue to exist as long as there is a need to provide an overall hierarchical classification of living things.

1.2.2. The nomenclatural process

In addition to the names that were discussed above, nomenclature manages works and nomenclatural acts. These three elements can be defined as nomenclatural objects (see III.1.1.). Once introduced, it is necessary to understand how they interact and influence each other (Fig. 4), and how to deal with them by attributing the correct status to each one. Code

compliance is the term that will be used to designate the states that allow a nomenclatural object to meet the Code-defined conditions of availability, validity, or other statuses associated with it.

The nomenclatural process interweaves elements and links their status. For example, a published and available work is a prerequisite for the validity of a nomenclatural act and the availability of a name contained within it. Each of the steps will be detailed thereafter. Some steps in the nomenclatural process are or were optional, such as the explicit designation of type specimen until 1999, or the registration of a work if it is published on paper. Each step of the process can in turn be divided into the verification of more precise elements. These elements have one or many Code-compliant state(s) and one or many Code non-compliant state(s) and are either independent or depend on prerequisites, such as the rank of the name, the date of publication or the medium of publication for a work. For example, the explicit mention of the date of publication in a work is a condition for the availability of the work only if the latter is published electronically (Art. 8.5.2.).

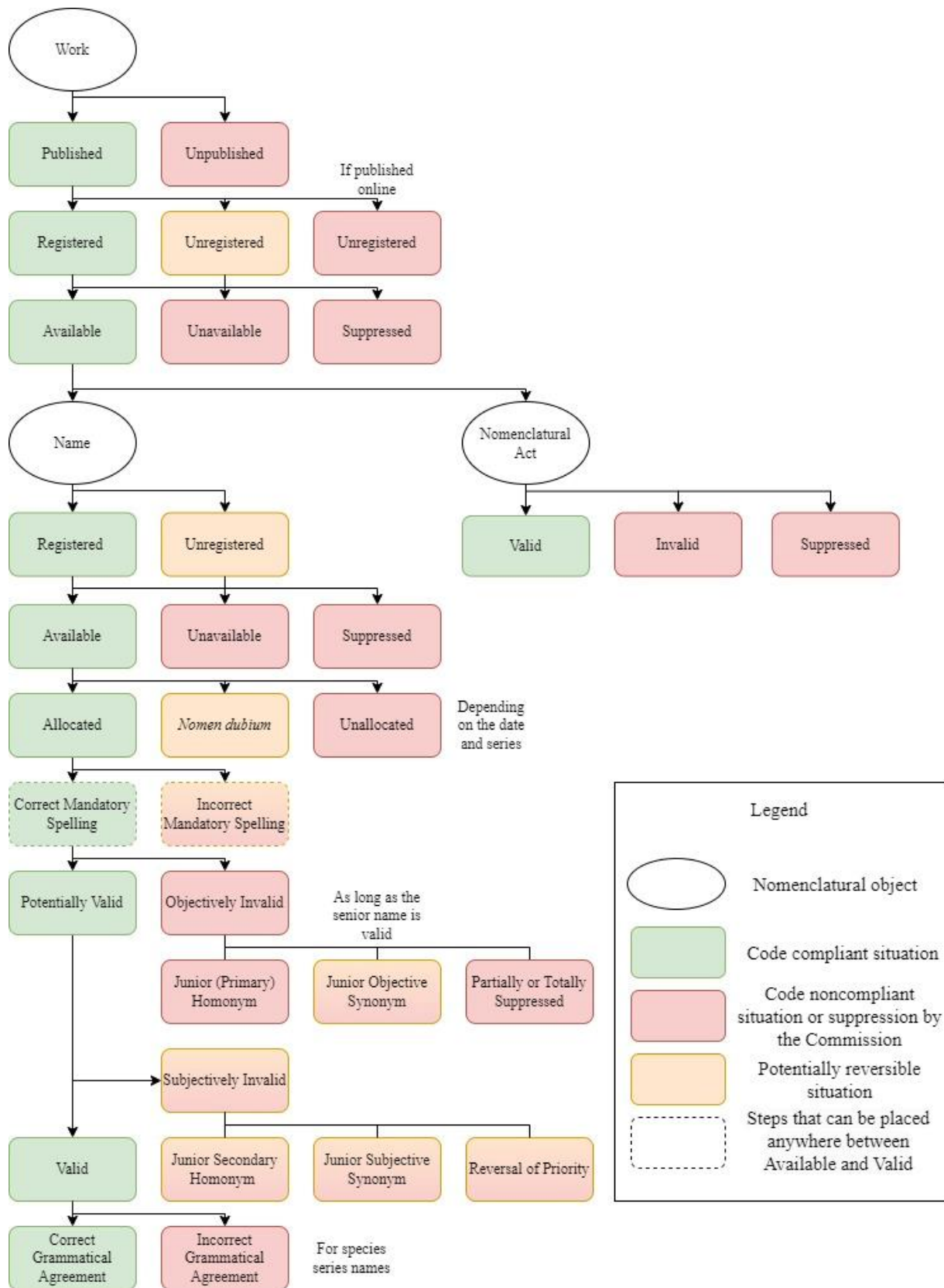


Figure 4: A simplified version of the nomenclatural process, after the Code's p. 123 figure and elements from Dubois (2011b).

Nomenclatural dialects and technical vocabulary

Over time, zoological and botanical nomenclature got further away from each other. Today, each of them has its own vocabulary (Table 2) and regulatory differences, which sometimes make it difficult to find the exact conceptual equivalencies between the two disciplines. These differences are called nomenclatural dialects by Pavlinov (2021). Some resources, such as Hawksworth (2010), compile the vocabulary of the different dialects of nomenclature. There are many, but the major ones can be summarized by the following points: conceptually, botanical nomenclature does not rely on the nominal taxon concept and the notion of series and therefore does not recognize the principle of Coordination. The latter is an essential point of the zoological Code as it makes it that with each name introduced, a new name is simultaneously introduced at each rank of a given nomenclatural series with the same date of publication and authorship. The ICNafp also regulates higher ranks more precisely and allows for infrasubspecific ranks, but forbids the use of tautonymous species names (same generic and specific name/epithet) which are permitted in the zoological Code.

Table 2: Correspondence between terms used in botanical and zoological nomenclature.

Terms used in botanical nomenclature	Terms used in zoological nomenclature
Binomial (see Saliba [2022] for a possible explanation of the difference between the two terms)	Binominal
Effective publication	Publication
Valid publication	Availability
Legitimate	Potentially valid
Correct	Valid
Autonym	Nominative taxon
Epitype	Neotype (partly)
Specific epithet	Specific name
Basionym (The first combination and rank under which a name was published)	Protonym (<i>sensu</i> Dubois [2000] or Pyle [2004]). Original combination, partly.

Publication and Availability

One of the main concepts in nomenclature is the notion of availability. The Code lists two elements that can bear this potential property: names and works. Availability means essentially that the object is accepted as a part of the system by the rules of nomenclature. Each of the two nomenclatural objects can be either available or unavailable, as there are no in-between states. In some cases, the availability of a name is questioned, and the Commission has to make a ruling to decide whether the nomenclatural object is in fact available or not.

Criteria of availability and publication change throughout history. These modifications of the rules constitute one of the major difficulties for the newcomers in nomenclature. Six principal periods can be distinguished: 1758–1930; 1931–1960; 1961–1985; 1986–1999; 1999–2012, and 2012–now. To be available, a work has to be published (*sensu* Code) and must contain a handful of elements, some depending on the publishing medium (Art. 8.; 9.).

In the case of nomenclatural works, one of the most remarkable evolutions of the rules has been the change of media deemed acceptable to be a support of publication. If the traditional printed paper was always an option, throughout history, other media were permitted at some points in history, such as optical discs, between 1986 and 2012 (Art. 8.4.). Since 2012, online publication in a format with “fixed content and layout” such as PDF/A is possible, if accompanied by a registration in *ZooBank*. This evolution in the possible medium of publication is always accompanied with the risk associated with the sustainability of the given medium. Nowadays, optical disc readers are getting rare and hard to find, and when a name is made available, it stays so as long as the nomenclatural system says so. This might create a possible difficulty in getting a hold of original publications. Fortunately, in the case of optical discs, it was proposed to transfer the data to a more sustainable medium (Dubois et al., 2021). With these kinds of difficulties, it comes as no surprise that the introduction of digital publication led to fiery debates on the durability of that medium (Dubois et al., 2013; International Commission on Zoological Nomenclature, 2014). Nevertheless, the taxonomist community widely took advantage of the opportunity to publish online.

As for names, not unlike works, they have to be both published in an available work and respect the articles of the fourth Chapter of the Code (Art. 10.–16.). Depending on the period, the exact elements necessary to make sure that a name is available vary. They all, nonetheless,

participate in the fact that the name has the necessary elements to make sure that its quality (compliance with the nomenclatural norms) is guaranteed. It comes as no surprise that the requirements got more and more precise and demanding with time, as issues and systemic flaws became more apparent. An example is the ban of the publication of a name as an explicit synonym of an already existing available name (Art. 11.6.), which was introduced with the 1961 first edition of the *International Code of Zoological Nomenclature*, which explains the leniency of the article for names introduced as synonyms before 1961. This rule prevents the gratuitous complexification of the synonymic loads, to best apply the norm of monosemy.

Availability can be understood as the recognition by a nomenclatural system of the belonging of a name or work to that given nomenclatural system. A name or work deemed unavailable is not to be taken into consideration. In practice, however, some unavailable names are still mentioned when associated with taxa in works deemed valuable (see IV.4.), and some unavailable or suppressed works can be used as a source for descriptions of taxa published subsequently.

Aside from an action by the Commission under Plenary Powers, a name that is available stays available. Availability is an immutable property of a name but can be open to interpretation. It is also an absolute property as the availability of a name does not depend on the availability of another name. For example, a specific name can be introduced with an unavailable generic name without challenging its own availability (Art. 11.9.3.1.).

There has been a clear effort made by the Commission to ensure that a name that was available at the time of publication stays available. That ended up manifesting in the preservation of availability criteria for a given period, even if the availability conditions change with time. That explains the different periods with each having their sets of rules, regulations and good practice guides. That effort participates in the complexity of the Code but respects the ethics of recognizing the contribution of past scientists.

Allocation and Typification¹¹

Singling out the notion of allocation and distinguishing it from that of availability is supported by the fact that allocation plays a crucial role in the current system of nomenclature by ensuring the correct application of names to taxa. If the exact modalities of allocation evolved throughout the different eras of nomenclature, its rationale, based on the type system, always stayed the same.

The expectations for the Code compliance when it comes to allocation varied quite a lot depending on the period. For example, the obligation for a species or subspecies name to have an explicit name-bearing type was only introduced in the 4th edition of the Code (Art. 16.4.). For the genus series, an explicit type-species designation has been mandatory only since 1930 (Art. 13.3.), even though the concept of a type-species was already present in the 1905 Code. However, the Code gives guidelines to identify one or multiple name-bearing types whenever the name was introduced, so, as a consequence, no species or genus series names actually exist within the frame of the Code without one or multiple associated type-specimens and a type-species respectively. In practice, even if all names have a traceable type, accessing or identifying that type specimen can be so complex, that it makes the name a *nomen dubium*.

Aside from the family series, where the absence of a type makes the name *de facto* unavailable (as the name is not based on an available genus series name and thus does not fulfill the conditions of Art. 11.7.1.1.), the same difficulty can emerge in the genus series and might necessitate a proper nomenclatural revision.

Nomina dubia are names whose link to the taxon they name is difficult to establish. Therefore, they are unable to compete with other available names for validity, which is why I, unlike the Code, do not consider them subjectively invalid (as that would suppose that they are potentially valid), but rather unable to be valid until they have a clear type allocated to them. They are partly stuck in a nomenclatural limbo¹² until the proper steps are taken (see IV.4.).

¹¹The two terms are mostly synonyms, at the exception that ‘allocation’ can be used for taxa over the family series. It will thus be preferred in this manuscript.

¹²If a *nomen dubium* is hardly usable at its original rank, it can still be used as a nominal type for a name of a higher rank.

The concept of Type in nomenclature

The type concept is defined, rather vaguely, by the Code, as:

Type: A term used alone, or forming part of a compound term, to denote a particular kind of specimen or taxon.

A more interesting definition is the one given by the Code for the name-bearing type entry:

Name-bearing type: The type genus, type species, holotype, lectotype, series of syntypes (which together constitute the name-bearing type), or neotype that provides the objective standard of reference whereby the application of the name of a nominal taxon can be determined.

What is crucial in this definition is the idea of an objective standard of reference. The name-bearing type exists to link the name to the taxon, and its role ends there. However, this property of the type specimen(s) is extremely important in nomenclature, as it anchors the whole abstract system of naming taxa in the physical reality.

At the microtaxonomic level, types are in fact specimens or parts of specimens that we can perceive with our senses. At higher levels, types become nominal taxa rather than specimens, but each nominal taxon is eventually connected, at least potentially, to a physical specimen. Unless a misidentification is at play, the only certainty when it comes to type specimens is that they belong to both the nominal taxon and its corresponding taxonomic taxon. Type specimens are not representative of the taxonomic taxon, nor do they have to display effectively several distinctive characteristics. They might have been used in the original description, or have been designated subsequently.

A lot of misinformation still exists about types in nomenclature, usually associated with persisting concepts linked to the term 'type' itself.

The use of the term 'Type' in nomenclature

The use of the term 'type' has long been the subject of controversy in nomenclature. The term emerges from the idea of typological definition, itself derived from platonic philosophy. In the same vein, Whewell (1840) writes "A type is an example of any class, for instance, species of a genus, which is considered as eminently possessing the characters of the class". That definition encapsulates well the basis of what is called 'typological thinking'.

The concept's nomenclatural meaning has nothing to do with this idea of a paragon or essentialist approach. It is closer to the image of an anchor, a way to stabilize a system of naming. Strickland (1845) explained it in these terms:

We may obtain a great amount of fixity, [...], by invariably selecting a type, to be permanently referred to as a standard of comparison. Every family, for instance, should have its type-subfamily, [...] and every genus its type-species. But it must not be supposed, [...], that these types really exist as such in nature; they are merely examples or illustrations selected for convenience to serve as permanent fixed points in our groups, whatever be the extent which we may give to their boundaries. By adhering to this notion of types we may often indicate these groups with greater precision than it is possible to do by means of definition alone.

This accusation of typological thinking, however, has never really disappeared, even though epistemological evolutions on taxon concepts have abandoned the idea of representativeness. It still lingers in the imagination of many, and even the Muséum National d'Histoire Naturelle writes on its collection page that types are "physical specimens that define the existence of animal and plant species" (Muséum National d'Histoire Naturelle, n.d.a), in an utmost essentialist approach of the role of type specimens.

Adding to this the fact that the term 'type' is used indistinctly to designate specimens that have an important role in the nomenclatural process (holotype, syntypes, lectotypes and neotypes) and specimens that play a role in the description or have taxonomic or historical value (type series, paratype, cotype, allotype) blurs the lines even more. In its definition of type, the Code excludes the mention of a particular nomenclatural value most likely

because of that ambiguity. Some concepts such as type series or type material are even harder to manipulate because there is an implied notion that these specimens are potentially and preferably selected to become holotypes and neotypes, and turn out to be *de facto* syntypes if no explicit designation of a type is made.

This leads to much confusion and elements such as paratypes are still treated by many museums as types in their own right, creating a bias in the number of nomenclatural types the institution actually hosts. ‘Type material’ is another interesting case, as it has been used for a long time in both zoology and botany to refer to the specimens described by the author (e.g., Jacot [1929] in zoology, Aedo et al. [2005] in botany), but are not defined by either of the two Codes. It is possible to suggest that this concept emerged from the historical ‘*échantillon-type*’ (e.g., in Bentham [1826] or Cosson & Germain [1840]) or the mix-up of the original material concept in botany and the type concept. In any case, no strict definition is provided for that concept, leading Federhen (2014) to (amusingly) define type material as a “type material [sic] of unknown type”.

To complicate matters more, in some cases, the nomenclatural and taxonomic material are indeed one and the same. Many species are described on a very small number of specimens, particularly in paleontology, for obvious reasons, but also in entomology, where 21 to 23 % of descriptions are based on a single specimen (Deng et al., 2019). In this case, a single entity is both the nomenclatural type and the totality of the original taxonomic material. Analysis and criticism of the term type is not a new phenomenon (Schuchert, 1897; Simpson, 1940; Farber, 1976), but the constation that the term polysemy causes confusion is still a current concern (Witteveen, 2016). Simpson (1940) called for the use of the term ‘onomatophore’ –from ὄνομα (onoma), name - φέρω (phero), I bear, I carry– as a replacement for type, and Dubois (2005) proposed ‘onymophoront’ –from ὄνομα (onoma), name, φέρω (phero), I bear and ὄν, ὄντος (on, ontos), being, individual– to replace type-specimen, as ways to overthrow the historical connotations.

Potential and full Validity

The notion of validity is the final step in the nomenclatural process. It is actually subdivided into two parts: potential validity and actual validity. In botany, these two stages have different names, respectively legitimacy and correctness. Two statuses play an important role in validity, homonymy and synonymy. Unlike availability, the notion of validity is relative, since a name without competition is de facto valid. Only a situation where several names compete calls for the application of validity rules. A name is thus potentially valid if it has not been declared permanently invalid. To be considered permanently invalid, a name must be a junior homonym or needs to be suppressed partially or completely by the Commission. A family series name can also become permanently invalid if its type genus is itself considered permanently invalid (Art. 39.).

Homonymy

Homonymy designates a situation where two available names introduced independently are deemed identical by the Code. At the species series level, homonymy is defined as two specific names that have exactly the same spelling, or have the same derivation and a similar spelling as defined by Art. 58. Two kinds of homonymy exist at the species level. On the one hand, primary homonyms are defined as the existence of two binominal names that are (or deemed under Art. 58.) the same. Secondary homonyms, on the other hand, are two species names that only share the same specific name, but, after a transfer, end up sharing their generic name too.

In the genus series, a single letter difference is sufficient to make sure that two names are not homonyms (Art. 56.2.).

In the family series, homonymy emerges when two names are founded on the same generic root (which would also make them objective synonyms), or on two generic names that differ only by their ending (Art. 55.3.). *Rana* and *Ranus* would both give the family Ranidae, creating a homonym. One way out is to modify the root to create a different name, in that case by introducing Ranaidae or Ranusidae.

Being a junior primary homonym makes a name objectively invalid. However, junior secondary homonyms are only invalid as long as they are treated as a member of the same

genus as the senior name (unless it was replaced by another name before 1961 [Art. 59.3.]). A name can be suppressed by the Commission, except for homonymy, meaning that despite a name being technically deemed unavailable, another name with the same spelling cannot be introduced, as it would still be considered a junior homonym of that suppressed name.

Synonymy

Synonymy refers to a situation where two or more names refer to the same taxon. There are two types of synonymies. The first emerges if the two synonyms are allocated to the same name-bearing type. In this case, synonymy is said to be objective, as it does not depend on any taxonomic decision or analysis. In formal terms, objective synonyms are the names of two nominal taxa that share the same name-bearing type. The second is subjective synonymy. This occurs when a taxonomic analysis establishes the unicity of several previously separate taxa. In terms of formalization, subjective synonymy is the fact that several nominal taxa are applied to the same taxonomic taxon.

On the one hand, being a junior objective synonym makes a name objectively invalid. In some cases, however, it can become valid, if, for example, the senior name is declared invalid for being a junior homonym. On the other hand, a junior subjective synonym is only invalid as long as the taxonomic taxon it names is considered the same as the taxonomic taxon the senior name is applied to.

Priority

To select a valid name between two or more competing ones, the Code relies on the Principle of Priority. Priority gives the upper hand to the older of the competing names. If, by chance, they were both proposed at the same time, but at different ranks, the name proposed at the higher rank would prevail over the one proposed at the lower rank (Art. 24.1.). If the two names are proposed at the same time, at the same rank, the action of an author, be it the author of the names or an independent one, is needed. A choice is made *a posteriori* to decide which of the two names takes precedence. This author is called the first reviser, and that act is regulated by the Principle of the First Reviser (Art. 24.2.).

Priority offers the advantage of objectivity when it comes to applying the rule. However, it is not always easy to determine the actual publication date of a work. The date indicated on the

cover may differ significantly from the actual date on which contemporaries could have had access to the work, and determining the actual date has become an exercise in itself, and even the subject of quite a few publications (e.g., Dupuis [1986]).

Priority also applies to nomenclatural acts (Art. 23.6.), and it is possible to consider that the principle that is applied to names is a particular instance of that rule, as the introduction of a new name is a nomenclatural act among others.

In practice, the principle of Priority is in fact a limited priority system, as it set to start in 1758, with the 10th edition of Linnæus' *Systema Naturae*, with the exception of the names published in Clerck's *Aranei Svecici* (Art. 3.). This arbitrary choice was made to have a partially stable system, as it is not threatened by the reappearance of names contemporary or older than Linnæus' and escapes endless debates on the authorship of historical names.

Usage

In some cases, the principle of Priority needs to be mitigated to ensure the stability of the nomenclatural system. Usage can be called upon to protect widely used junior names. If a senior name is found, considering that it should take precedence over a widely known junior one, the Commission may be asked to reverse the precedence of that senior name (Art. 81.2.3.). Since the introduction in the 4th edition of the Code of Art. 23.9., in certain cases, a taxonomist can declare a pre-1899 name a *nomen oblitum*, and the more recent name a *nomen protectum*, under the condition that the former has not been used after 1899 and that the latter is sufficiently used. It is interesting to note that the concept of *nomen oblitum* carries with it a certain paradox, since the name has been found, it is by definition not forgotten. In more classical cases, the Commission, when an appropriate inquiry has been made, can then decide to reverse the priority between two competing names, creating a subjectively invalid name, as the senior name could still be extracted from its invalidity status if it is not considered a synonym to the junior name anymore.

The validity of suprafamilial names is not directly regulated in the Code (Art.1.2.2.). In practice, a hybrid system of Priority and Usage is applied, leading to possible confusion and making it sometimes very difficult to establish what name to use for these ranks.

Registration

Registration can be seen as a supplementary step in the nomenclatural process. It is the most recent addition to the nomenclatural workflow.

Registration can be subdivided into two dimensions, indexing names on one side and indexing works and nomenclatural acts on the other. A handful of scientists have undertaken the colossal task of creating comprehensive lists of names and works by their predecessors and contemporaries before the information age. These works were the precursors of the modern conception of Registration.

As the task was daunting, some settled for a more reasonable part of the mission. Early on, Agassiz published the *Nomenclator Zoologicus* (1842–1847; see Bowley and Smith [1968]). In the 20th century, Schulze, a German scientist specialist of sponges, set the goal to compile the generic and subgeneric names of animals in the *Nomenclator Animalium Generum et Subgenerum* (1926-1954). Around the same time, over the Channel, Neave was creating the *Nomenclator Zoologicus* (1939-2004) with the same objective in mind (Polaszek & Michel, 2010). Both these projects outlived their founders, and the *Nomenclator Zoologicus* went on to become a database until its discontinuation in 2004. The Interim Register of Marine and Nonmarine Genera project inherited that endeavor, and extended it to the rest of Biota (Rees et al. 2020), hosting currently almost half a million generic names (Rees, n.d.), as well as names at other ranks.

Another scientist, namely Charles Davies Sherborn, decided on a more ambitious approach. Thirty years of compiling work led to the publication of the 11 volumes of the *Index Animalum* (Polaszek & Michel, 2010). By systematically going back to the source material, working extensively, and by himself (Evenhuis, 2016), Sherborn indexed around 420 000 names, with a remarkably low individual (non-systemic) mistake rate, of around 1-2% (Welter-Schultes et al., 2016) granting him the title of “Father of Biodiversity Informatics” (Polaszek & Michel, 2010).

In parallel, another historical repository for zoological names was the *Zoological Record*, created in London in 1864 under the name *The Record of Zoological Literature*. Rather than focusing on past names and works, the *Zoological Record* tries to provide real-time publication tracking, making sure that the average zoologist would stay updated on the newer

introduced names and taxonomic works. Founded under the leadership of Albert Günther, it was a commercial failure at first, and was about to be abandoned by the publisher when a community of zoologists decided to save it from disappearing by forming the Zoological Association. In 1900, the *International Catalogue of Scientific Literature* ended the preeminence of the *Zoological Record* as the reference but disappeared in 1921, indexing literature up until the First World War (Bridson, 1968). Between 1980 and 1988, the *Zoological Record* was resurrected and benefited from funding and support from BIOSIS and the Zoological Society of London. This enabled the update of the Record to that date. In 2004, Thomson bought BIOSIS, before merging with Reuters in 2008 (Reuters staff, 2018). This event marks the passage of the *Zoological Record* into private companies' hands. Currently, Clarivate is the company that manages the *Zoological Record*, which makes part of its data only available behind a paywall (Clarivate Analytics, n.d.). Fortunately, part of the data entered in the repository is publicly available in the ION (Index of Organisms Name; <http://www.organismnames.com/>) database.

The idea of having an online database where names could be registered was introduced during the discussions preceding the fourth edition of the Code (Bouchet, 1999). At the time, the *Zoological Record*, still under the control of BIOSIS, was considered the natural primary actor to host the data (Thorne, 2003). However, the creation of a repository by BIOSIS never came to fruition, as it became private property along with the *Zoological Record* it was in charge of¹³.

In the meantime, pressure to be able to introduce electronic publication (Knapp & Wright, 2010), as well as multiplying calls for the development of taxonomic databases and the systematic registration of new names (Patterson, 2003; Patterson et al., 2003) pushed the Commission to work on a repository of its own. The oldest workshop technical paper on the subject was published as early as 2005 (Polaszek et al., 2005b). The resulting database was introduced in 2008, under the name *ZooBank*, as an initiative of the Commission, as the latter recognized the need for a centralized platform that would facilitate the registration of new taxon names when published electronically.

¹³ The *Zoological Record* data (or part of it) was nonetheless added to *Zoobank* in its early years, until a majority of the Commission decided that it would be inappropriate to have a private commercial company host an open and freely accessible database.

ZooBank is an online repository and registry for names, works and nomenclatural acts, and the center of the 2012 amendment to the Code (International Commission on Zoological Nomenclature, 2012). It provides a platform for taxonomists to document their work following the rules set forth by the Code (Art. 8.5.3.). Unlike its historical predecessors, *ZooBank* relies on a community effort rather than the commitment of a small group or a single taxonomist. Since the very beginning of the *ZooBank* project, different visions for registration requirements were discussed (summarized in Pyle & Michel [2008]) and the path that prevailed was the one already discussed in the 2005 technical paper. It pushes for the mandatory registration of works and names published online, and an optional registration for traditionally published names and works.

While this decision was understandable from the point of view of social acceptability, it has led to nomenclatural data being scattered all over the Web. Without any deliberate move towards a centralized repository, the information ended up being spread across a multitude of taxonomic databases, which took a crucial role in the modern accessibility and management of taxonomic and nomenclatural data.

II. Nomenclature and taxonomic databases

« Seulement, il ne faut pas consacrer une erreur grave pour l'avantage médiocre de suivre une habitude. »

de Candolle, *Loi de la nomenclature botanique*

Databases are defined in many different ways, depending on the person one is talking to, the context, and the goal of the database. A permissive enough definition, given by Elmasri & Navath (2017) is that a database is “a logically coherent collection of data with some inherent meaning” that “represents some aspect of the real world”. The authors also add that “a database is designed, built, and populated with data for a specific purpose. It has an intended group of users and some preconceived applications in which these users are interested”.

In taxonomy, databases are essentially used to represent the classifications developed by taxonomists, the taxa derived from them, and the properties of these taxa. These databases are generally used by the community itself and enable taxonomists to keep abreast of research on the taxonomy of a group with which they are more or less familiar. Other notable users of taxonomic databases are biologists, ecologists, and collection managers. Collection databases are another major example of taxonomic databases.

Depending on the database, their primary intentions and the human and financial resources available, the information stored will differ. Some databases feature morphological and ecological information on taxa or global distribution data or reported occurrences of specimens belonging to a taxon. Other might list the references mentioning the taxon, in logonymic lists*, or provide the conservation status of a given taxon. Due to the sheer number of existing databases, and their differences in content and format, most of the elements that will be discussed will find their roots in examples rather than comprehensive data analyses. Most taxonomic databases were introduced in the last thirty years and contain a wealth of data that can easily be considered massive, approaching Big Data* territory. For example, on September the 5th, 2023, the Catalogue of Life (COL; [Bánki et al., 2023]), which aims to create a reference list of taxa, counts 5,021,561 names. The Global Biodiversity Information Facility (GBIF; gbif.org), which specializes in species occurrences, had 2 575 225 094 entries listed (GBIF, 2023).

The key to accessing data in most of these databases is the scientific name. Some databases (e.g., Amphibian Species of the World [Frost, 2023a]) allow the user to query the data using an English vernacular name, but many require a scientific name (e.g., COL) to find the information associated with a given taxon.

Purely nomenclatural databases are a rare occurrence, as nomenclatural data are seen as subservient to taxonomic information in most databases. Nomenclature indeed plays an essential role in the management of taxonomic databases. It enables the link between various data stored in databases and the taxa to which they refer. However, nomenclature is not treated in the same way in all databases, some aiming to put forward purely taxonomic and nomenclatural data, while, for others, the mention of this data is merely a means to an end. In all cases, nomenclature should make alignment between different databases possible, to provide a complete range of information available on a taxon, or the individuals that constitute it. If, in theory, nomenclature should be able to have that role, the reality is far more complex. Even putting aside ‘grey nomenclature’* (Minelli, 2017), and potential mistakes, different databases may not use the same taxonomy. That leads to increased difficulty when trying to connect them efficiently. Different taxa coverage (especially the inclusion of inter-Code homonyms), variable spellings of names and authors, and the use of different taxon concepts can also complicate the correct alignment of data (Remsen, 2016). The Global Name Architecture initiative (Pyle, 2016; <https://globalnames.org/>) was created to mitigate this difficulty. Linking names in different databases is still a very current subject (see Berendsohn [2023] or Bánki et al. [2023])

Grenié et al. (2023) formalizes a distinction¹⁴ between databases, based on the spatial scale and the taxonomic breadth, in other words, how large the most comprehensive included taxon is. This division separates worldwide generalist databases and smaller localized and specialized ones (Table 3). Each of these categories of databases works on a different scale, and face different kind of challenges. Other distinctions are possible, such as databases mostly focused on taxonomic data on one side (such as the Reptile Database [Uetz, 2023] or Fishbase [Froese & Pauly, 2023]), and databases providing mainly other types of data, that rely on a taxonomic backbone, such as the IUCN Red Lists of threatened species (International Union for the Conservation of Nature, 2022), that provides conservation statuses, or the GBIF, that centralizes occurrence data, on the other.

¹⁴ I developed a very similar distinction for a master (=postgraduate) course called “Problématiques actuelles en nomenclature zoologique”, but did not recognize the Medium Taxonomic Breadth. I only distinguished between the databases that take into account all of Biota or a subpart of it.

Table 3: A classification of taxonomic databases focused on zoology, modified after Grenié et al. (2023).

	Narrow or Medium Taxonomic Breadth	Wide Taxonomic Breadth (no taxonomic restriction)
Regional	<p>Animals; China CASD (http://zoology.especies.cn/)</p> <p>Coleoptera; Colombia Coleoptera de Colombia (https://sites.google.com/view/coleopol/inicio)</p> <p>Arthropoda; UAE Arthropod fauna of the UAE (http://uae.hymis.net/)</p>	<p>Sweden Dyntaxa (https://www.dyntaxa.se/)</p> <p>UK NBN (https://nbn.org.uk/)</p> <p>Europe PESI (https://www.eunomen.eu/portal/index.php)</p> <p>EUBON (https://cybertaxonomy.org/eubon/utis/1.3/)</p> <p>China SP2000CN (http://sp2000.org.cn/)</p> <p>Taiwan TaiCOL (https://taibnet.sinica.edu.tw/eng)</p> <p>France Taxref (https://inpn.mnhn.fr/programme/referentiel-taxonomique-taxref)</p> <p>Netherlands Taxainfo (https://taxainfo.nl/)</p> <p>Czech Republic Biolib (https://www.biolib.cz/en/main/)</p>
Global	<p>Amphibians ASW (https://amphibiansoftheworld.amnh.org/)</p> <p>Birds eBird/Clements (https://ebird.org/science/use-ebird-data/the-ebird-taxonomy)</p> <p>Fish FishBase (https://www.fishbase.in)</p> <p>Mammals MMD (https://www.mammaldiversity.org/)</p> <p>Reptiles ReptileDB (https://www.reptile-database.org/)</p> <p>Spiders World Spider Catalog (https://wsc.nmbe.ch/)</p> <p>Diptera Systema Dipterorum (http://www.diptera.org/)</p> <p>Echinoidea The World Echinoidea Database (https://www.marinespecies.org/echinoidea/index.php)</p> <p>Hymenoptera Atas Hymenoptera (http://www.atlashymenoptera.net)</p>	<p>BOLD (http://www.barcodinglife.org/)</p> <p>COL (https://www.catalogueoflife.org/)</p> <p>EOL (https://eol.org/)</p> <p>GBIF (https://www.gbif.org/)</p> <p>GNI (https://index.globalnames.org/)</p> <p>ION (http://www.organismnames.com/)</p> <p>ITIS (https://www.itis.gov/)</p> <p>IUCN Red List (https://www.iucnredlist.org/)</p> <p>NatServe (https://explorer.natur.eserve.org/)</p> <p>NCBI (https://www.ncbi.nlm.nih.gov/taxonomy)</p> <p>Neotoma (http://neotomadb.org/)</p> <p>OTL (https://opentreeoflife.github.io/)</p> <p>PBDB (https://paleobiodb.org/#/)</p> <p>Wikidata (https://www.wikidata.org/)</p> <p>Wikispecies (https://species.wikimedia.org/)</p>

In many of the listed databases, nomenclature plays a secondary role, and very few of them try to indicate purely nomenclatural information such as a specific name's grammatical nature. Most, however, include information that is necessary to be able to efficiently lead a taxonomic revision. In practice, the following data can be considered crucial for that kind of work:

- 1) Name, author, date and rank (the author + date ensemble is sometimes called authority, the three elements together form the nominal complex [Dubois, 2000]).
- 2) The reference to the original publication, and in it, the original combination of the taxon if the taxon is a species.
- 3) The type-specimen(s) label number(s) and location(s) for a species series name.
- 4) The type-species and type-genus (although the latter can be easily deductible most of the time) for a genus and family series name respectively.
- 5) The nomenclatural status of the name (available, *nomen dubium*, etc.).

In some cases, the following data may also be useful:

- 6) The current taxonomic status of a name (valid, secondary homonym, etc.).
- 7) The nomenclatural Code under which the name is considered –especially for ambireginal organisms.

A difficulty that emerges when dealing with taxonomic data online is that most databases are independently run, possess their particular data structure, and thus differ from the others in the nature and management of said data. There are as many databases structures as there are databases, each with their own strengths and weaknesses. As a result, nomenclatural data may be stored differently depending on the database, i.e., in plain text or in database fields, complicating the setting up of semantic equivalences of this information. The Reptile Database, for example, has a very complete set of information on species, including associated diagnoses and type specimens, but there is no information whatsoever on genera. Amphibian Species of the World has nomenclatural information on species, genera, families and subfamilies, and present these ranks in a classification, but tribes and subtribes are listed in plain text rather than integrated in the database. Further research is needed to get a complete picture of the nomenclatural data inventory present in each database. This endeavor is made harder by the fact that the schematics of these databases are not available most of the time.

Fortunately, efforts have been made to promote a way to safely exchange data between databases.

II.1. Nomenclature in TDWG standards and collection databases.

In Biodiversity Informatics, the Taxonomic Databases Working Group (TDWG), known currently as BIS-TDWG (Biodiversity Information Standards-TDWG), is an international structure composed of biodiversity informaticians and biologists. One of TDWG's historical goals was to propose standards for use within the community. These standards are of multiple natures (Vignes-Lebbe, 2023). The ones that will be of major interest in this work can be considered guidelines for the way to organize and exchange information in digital formats, usually in databases. They are the result of a collegial effort to ease the findability and interoperability of biodiversity data.

The BIS-TDWG, since its creation in 1985, has worked on and adopted 23 standards which scope of application ranges from audiovisual data (e.g., Audiovisual Core [AC]) to descriptive data (e.g., Structured Descriptive Data [SDD]) (TDWG, n.d.). Historically, TDWG has been overwhelmingly composed of botanists, which might explain the preponderance of botanical standards. In nomenclature, the 'Authors of Plant Names', as well as the 'Plant Names in Botanical Databases' standards are notable (TDWG, n.d.). The 'Plant Names in Botanical Databases' standard (TDWG Minimum Standard for Names in Botanical Databases Subgroup, 1995) regulates the way a botanical scientific name is to be represented, i.e., which information is to be included and how. No equivalent exists currently in zoology, as names in botany have more elements that need to be added to the nominal complex¹⁵. However, nomenclatural data are also treated in other standards, such as the Darwin Core.

II.1.1. Nomenclature in the Darwin Core

The Darwin Core (Wieczorek et al., 2012) is one of the most well-known and used standards produced by the TDWG community. It is defined as a standard "intended to facilitate the sharing of information about biological diversity by providing identifiers, labels and definitions" (Darwin Core Maintenance Group, n.d.). The Darwin Core is not a database structure, but a framework to exchange information from a given database to another database. It partly palliates the difficulties of aligning data based on raw nomenclatural data

¹⁵Elements such as Author *ex. Author* for a name made available by a subsequent author, or the mention of the subspecific rank within the name (e.g., *Genus species* subsp. *subspecies*).

alone and provides a collection of fields to fill depending on the nature of data that have to be exchanged.

In the Darwin Core, many fields pertain directly or indirectly to nomenclature (Darwin Core Maintenance Group, 2023). Fortunately, all the essential elements cited above for basic taxonomic work are already taken into account in that standard (Table 4).

Table 4: Important nomenclatural elements and their equivalent in Darwin Core.

Element	Property in Darwin Core
Name	dwc:scientificname
Author	dwc:scientificNameAuthorship
Date	dwc:namePublishedInYear
Rank	dwc:taxonRank
Original publication	dwc:namePublishedIn
Original combination	dwc:originalNameUsage
Type status	dwc:typeStatus
Collection	dwc:collectionCode
Collection label number	dwc:catalogNumber
Nomenclatural status	dwc:nomenclaturalStatus
Taxonomic status	dwc:taxonomicStatus
Nomenclatural Code	dwc:nomenclaturalCode

However, the Darwin Core was by no means conceived and created as a format for purely nomenclatural data. The development of another standard of exchange might be necessary and could lead to an efficient way to extract and communicate data. Fortunately, some ontologies already exist, such as NOMEN (Yoder et al., 2017) or Zoonom (Saliba et al., 2021)¹⁶, and can be used as bases to construct a specialized format for purely nomenclatural data (see V.3.).

II.1.2. Nomenclature in collection catalogues

Naturalist collections hold immense scientific value as repositories of biological, geological and ecological information. These meticulously curated assemblages of specimens

¹⁶ It may come as a surprise, but the overlap between NOMEN and Zoonom is narrower than one can think, coming mainly from the fact that NOMEN covers more categories of unavailable names, and Zoonom has a larger collection of concepts for invalid names.

provide researchers with the means to study the diversity of life. By allowing for the direct observation and comparison of specimens from different periods and locations, these collections facilitate the tracking of evolutionary changes, the exploration of genetic and morphologic variation and the understanding of ecological dynamics (Holmes et al., 2016). In a rapidly changing world, naturalist collections offer a critical baseline for measuring shifts in biodiversity and provide essential insights into the impacts of environmental changes on ecosystems (Pellens, 2021).

When it comes to taxonomy, collections play a vital role in preserving specimens that have been seen and described by authors (the latter being referred to as the type series). The accessibility of such specimens ensures the application of the scientific principle of reproducibility (Miller et al. 2020). Collections are also the place where type specimens are usually stored, and where they can be preserved over the centuries. A global registry of naturalist collections (GRSciColl) is maintained by the GBIF (Grosjean et al., 2022). Collection catalogues databases are digital structures used to locate specimens in naturalist collections. These catalogues are the successors of large books and registers, where this information was manually stored. Collection catalogues have a major role to play in nomenclature since they make it possible for a taxonomist to locate type specimens, the original material studied by a predecessor, or the taxonomist's own previously studied specimens.

There are two levels of possible digitization when it comes to natural collection. The first one is the creation of a database containing purely technical information on the collection, essentially transcribing what was written in the logbooks. These databases usually include data associated with specimens in the collection, such as the locality of origin, collector, sex, and developmental stage. The second level is the digitation of the specimens themselves, as a way to provide taxonomic information to the user, through the use of scanned, 2D and 3D specimens. This approach can even lead to enabling new types of research (Soltis, 2017) and boost the ease with which current methods are applied (Lendemer et al., 2019).

Collection catalogs still suffer to a large extent from the difficulty of digitizing and entering into databases all the information required to manage a collection, particularly in large collections such as that of the Muséum National d'Histoire Naturelle, or the Natural History Museum in London. Both bigger and smaller collections however struggle to find the funds

and the policy-makers' support necessary to digitize their records and specimens, or even ensure their sustainability (Naggs, 2022).

The essential nomenclatural information that a taxonomist might want to find in a collection catalogue is the holotype or syntypes associated with a name, as well as complementary data about these specimens. In some cases, the type series, or paralectotypes may also prove to be useful for a neotype designation or to check potential misapplied names. However real difficulty might emerge when trying to find such information as the data are incomplete on many institutional databases, and collection catalogues are not conceived to be taxonomically curated. That might lead to confusing situations where the name in the catalogue differs from the name on the label and from the currently valid name.

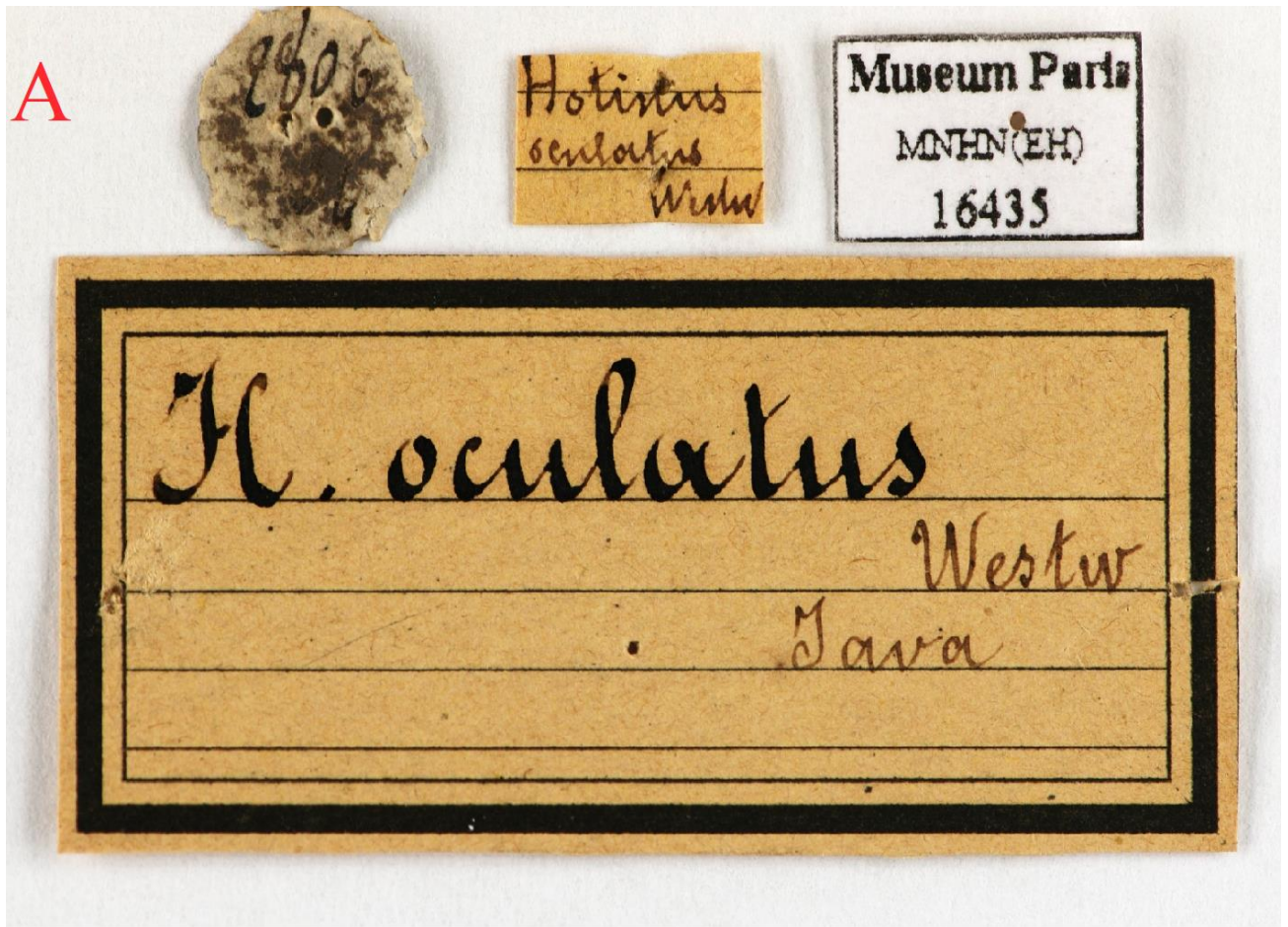
Example

In Fig. 5: A, is the label of a type specimen of planthopper (Fulguroidea) (Fig. 6: C). The name used is *Hotinus oculatus* Westwood, 1838. This is a recombination published in Walker (1850) (Fig. 6: B). It is in all likelihood the (or one of the) specimen(s) used to describe that species, as the original description mentions a specimen in the ‘Mus. Reg. Paris’ (Westwood, 1838). However, in the catalogue database (Fig. 5: B), the specimen is entered under the name *Fulgora oculata*. This is the original combination. The current valid name in FLOW (Fulgoromorpha Lists On the Web [Bourgoin, 2023]) is *Pyrops oculatus* (Fig. 6: A).

In other words, someone with the currently valid name would have to check the synonymy on a database such as FLOW to be able to efficiently search for the type specimen. In practice, on the FLOW database, the synonymy with *Hotinus oculatus* is missing, but the original combination is present.

An article by Nagai et al. (2017) gives 1839 as a description date for *Pyrops oculatus*, and places its type in the University Museum of Natural History of Oxford. That mistake might have originated from the fact that the Oxford catalogue (Oxford University Museum of Natural History, n.d.) has 1839 as the publication date for *Fulgora oculata*, and that Westwood worked for Oxford. In fact, that institution curates most of the types associated with Westwood’s names, but this one is an exception.

Furthermore, complementary research helped to determine that the first publication of *Fulgora oculata* is not in fact the one listed on FLOW (Westwood, 1838) but an older one. It should be “L’Institut Journal des académies et sociétés scientifiques de la France et de l’Étranger”, number 228, page 66, which is itself a report on a text by Westwood that will only be published a year later in the Transactions of the Linnean Society of London (Westwood, 1839). Fortunately, the text in L’Institut has the same date as the currently accepted date of publication of *Fulgora oculata*, so no change to the date is required.



SPECIMEN

MNHN-EH-EH16435

Collection *coll. générale*

Sex M

2806 34 (= 1834 n° 2806 : Insectes donnés à Mr. Audoin par Mr. Barthelemy de Manille; Indes Orientales)/ *Hotinus oculatus* Westw./ *H. oculatus* Westw., Java/ Muséum Paris, MNHN(EH) 16435

TAXONOMY

Phylum Arthropoda
 Class Insecta
 Order Hemiptera
 Family Fulgoridae
 Genus *Fulgora*
 Species *Fulgora oculata*
 Name *Fulgora oculata* Westwood, 1838

Figure 5: The label (A) and the legend (B) of a specimen of grasshoppers on the MNHN online catalogue. Specimen MNHN-EH-EH16435 (Muséum National d'Histoire Naturelle, n.d.b).

Photography by Laurent Fauvre.

Species

Pyrops oculatus (Westwood, 1838)

A

Basionym original publication

Westwood J. O. 1838 - Nouvelles diverses. Fulgores. *Annales de la Société Entomologique de France, Bulletin Entomologique* 2e trim. (6 juin 1837): xxxviii-xxxix [38].

▼ Chronological account

Fulgora oculata Westwood, 1838 previous combination of *Pyrops oculata* (Westwood, 1838) according to Burmeister (1845): 4

Pyrops oculata (Westwood, 1838) previous combination of *Pyrops oculatus* (Westwood, 1838) according to Walker (1851): 267

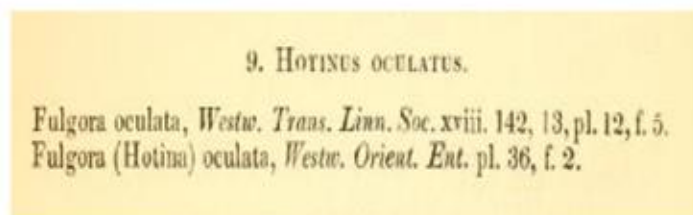
Fulgora oculata Westwood, 1838 *combinatio revivisco* according to Butler (1874): 99

Pyrops oculata (Westwood, 1838) *combinatio revivisco* according to Schmidt (1928): 124

Pyrops oculata (Westwood, 1838) previous combination of *Laternaria oculata* (Westwood, 1838) according to Metcalf (1947): 202

Fulgora oculata Westwood, 1838 *combinatio revivisco* according to Lallemand (1963): 77

Fulgora oculata Westwood, 1838 previous combination of *Pyrops oculatus* (Westwood, 1838) according to Nagai & Porion (1996): 25



B



Figure 6: *Pyrops oculatus* in FLOW (Bourgoin, n.d.)(A), an extract of Walker (1850) (B) and a picture of the type specimen of *oculatus* (C).

Photography by Laurent Fauvre.

II.2. Current limits in digital nomenclature

II.2.1. Finding nomenclatural data

One of the most delicate issues when it comes to nomenclature and databases is the fact that there is no centralized database with access to all nomenclatural data. As seen previously, some data are findable in databases that allow exporting or presenting data using

the Darwin Core standard, and other data are findable, despite a potential accrued effort, in collection catalogues. *ZooBank* also curates relevant data, as well as some databases like AnimalBase (AnimalBase Project Group, 2023) that carry valuable and rare data such as the grammatical information of “early” (1758 to 1790) names and their status regarding decisions by the Commission.

II.2.2. The structural interdependence of databases

Not all databases are created equal. Some of them are filled with data directly found in taxonomic literature, and others feed on existing databases.

In taxonomy, a first-degree database¹⁷ can be defined as a database whose sources are constituted directly by the scientific literature. Examples of first-degree databases include FLOW, Amphibian Species of the World and FishBase (Froese & Daly, 2023). First-degree databases are often maintained by an individual or a small group of people on their own and are often specialized in a reasonably sized taxon. This ensures that the team or person in charge can efficiently update the information in the database with data from the literature. The curators of these bases sometimes count on the fact that their colleagues will send relevant information their way or point out mistakes. This kind of database is usually the most complete and the most rapidly updated, and yet, suffers sometimes from a lack of funding. They also depend on one person or a small group of people, which can lead to technical fragility, and question the perennity of such structures.

Conversely, multiple-degree databases (second, third, fourth...) retrieve information from other databases, aggregate these data, and publish lists that are sometimes intended to be authoritative. Each time a database moves away from the literature, it gains a degree. A given database, for a given entry, can have a different degree of distance from the literature. In some cases, determining the exact degree of a database for a given entry can reveal itself to be quite difficult. Many aggregator databases have a redundancy in the data that they harvest to

¹⁷ Not to be confused with a primary and secondary database. Primary and secondary databases are notions used in bioinformatics and defined by Brooksbank and Cowley (2023) as the following: “**Primary databases** are [databases] populated with experimentally derived data such as nucleotide sequence, protein sequence or macromolecular structure. Experimental results are submitted directly into the database by researchers, and the data are essentially archival in nature. Once given a database accession number, the data in primary databases are never changed: they form part of the scientific record” and “**Secondary databases** comprise data derived from the results of analysing primary data. Secondary databases often draw upon information from numerous sources, including other databases (primary and secondary), controlled vocabularies [...] and the scientific literature. They are highly curated, often using a complex combination of computational algorithms and manual analysis and interpretation to derive new knowledge from the public record of science.”
All taxonomic databases are technically secondary databases.

construct their own taxonomies, and a system of priority is built into the system to choose between potentially conflicting data (Döring et al. 2023).

Example

In the GBIF taxonomic backbone, on July the 3rd, 2023, *Aglaophenia struthionides* (Murray, 1860) has its source indicated in the World Hydrozoa Database (Schuchert, 2023), which itself gives a literature source, viz. Murray (1860) for the original description. In that case, the GBIF taxonomic backbone is 2 degrees away from the literature (GBIF Backbone Taxonomy, 2023a). In the case of *Triturus marmoratus* (Latreille, 1800), GBIF cites ITIS, which itself cites Amphibian Species of the World (Frost, 2023a), in a 2019 version. Frost cites both Latreille (1800) and the first use of the current combination (Dunn, 1918). This second example sits three degrees away from the literature (GBIF Backbone Taxonomy, 2023e).

The most reputable multiple-degree taxonomic database is undoubtedly the Catalogue of Life, which aims to cover all living beings. The Integrated Taxonomic Information System (ITIS; www.itis.gov) also serves as a central focal point for several taxa.

As one can easily suppose, the higher the degree of a database, the higher the chances are that said database differs from the literature or contains mistakes. This can be explained by two factors: distance from the data (see II.2.3.) and reliance on a single source (see II.2.4.). The GBIF, for example, added names that are in the Official lists and indexes database¹⁸, but without filtering out completely the suppressed names and incorrect subsequent spelling. Some of them are thus marked as accepted in the database backbone, with usually no occurrences attached. Others are marked as doubtful and await further examination.

¹⁸ This database is not currently available to the public (Pyle, R. L., pers. comm.). I am thus unable to identify exactly where the mistake happened with these particular cases.

Examples

On September the 5th, 2023, *Lybia tresselata* Milne Edwards, 1834 which has been rejected as an incorrect subsequent spelling of *Grapsus tesselatus* Latreille, 1812 in Direction 37 (International Commission on Zoological Nomenclature, 1956), but is present in the GBIF backbone as an accepted species (GBIF Backbone Taxonomy, 2023c). However, *Pagurus* Berthold, 1827, which was deemed a *nomen nudum* in Opinion 472 (International Commission on Zoological Nomenclature, 1957), is marked as a doubtful name (GBIF Backbone Taxonomy, 2023d).

II.2.3. The delay in updating data

Each biodiversity database has a different schedule when it comes to updating data. It seems like, the lower the degree of the database, the quicker the updates are made, as long as the database curator is active. That, however, is not a general rule, as these databases are usually managed and curated by one person or a small team, who can get overwhelmed by the task or lack the necessary time to curate the database efficiently. On the other side of the scale, some high-degree databases are managed by so many people and cover so many taxa that the time between full updates can be very long. Updates can also be very inconsistent, with some taxa getting regular updates, and other taxa not getting any for a long time, depending, in between other factors, on the availability of the dedicated curator.

It is hard to truly quantify this delay in the data phenomenon or its impact on the faithfulness of a database to the literature, but it might explain part of the gap or differences in data that some databases admit to struggle with (Waller, 2022). In some cases, however, the delay can get quite important. It comes as no surprise that the higher the degree of the database, the longer the time needed for an update. The discrepancy that is thus created, between both the databases and between the databases and the literature can then generate both complexity and mistakes.

Example

In the GBIF taxonomic backbone, the varying toad, *Bufo variabilis* (Pallas, 1769) (GBIF Backbone Taxonomy, 2023b), originally described as *Rana variabilis* has its source in ITIS (Fig. 7: A) (Integrated Taxonomic Information System, n.d.). In turn, ITIS cites Amphibian Species of the World, exactly like in the example above, in its 2019 version (Fig. 7: B). However, *B. variabilis* is not considered a valid name in Frost's database (Fig. 7: C) and is included in the synonymy of *Bufo viridis* (Laurenti, 1768), that decision being taken following Dufresnes et al. (2019) (Frost, 2023b). There is thus at least a four years delay between the first-degree database (ASW) and the GBIF taxonomic backbone. As COL also uses ITIS as an intermediary for Amphibians, the synonymy is also missing there (Catalogue of Life, 2023). If a scientist following Dufresnes et al. (2019) interpretation of the taxon submits occurrence data on what would have been *Bufo variabilis* to the GBIF, it will be put under *Bufo viridis*, as it would have been the name used by the scientist. If the two species are separated again, meticulous work will have to be done to reconstitute which occurrence belongs to which species—that is considering that the specialists are aware of that possible mix-up.

Fortunately, the GBIF backbone also relies on data coming from Plazi (Agosti et al., 2020). Plazi maintains a digital taxonomic literature repository and tracks taxonomic treatments. That leads to the fact that, in some cases, new information in articles published in journals covered by Plazi, is directly transmitted to the backbone, without having to rely on intermediary databases. It makes it more complicated to navigate the structure of the backbone, but overall, enables the submission of data on relatively new species. However, not all publications are treated by Plazi, and many taxonomic changes still fly under the radar.

A *Bufotes variabilis* (Pallas, 1769)

source: [The Integrated Taxonomic Information System](#)

Varying Toad In English Basionym: *Rana variabilis* Pallas, 1769

B *Bufotes variabilis* (Pallas, 1769)

Taxonomic Serial No.: 1103415

Other Source(s): [Amphibian Species of the World: an Online Reference v6, database \(version 6.0\)](#)

Source: 2019

Acquired: Frost, Darrel R. 2019. Amphibian Species of the World: an Online Reference. Version 6 (12 September, 2019). Electronic

Notes: Database accessible at <http://research.amnh.org/herpetology/amphibia/index.html> American Museum of Natural History, New York, USA

Bufotes variabilis

Reference for:

C *Bufotes viridis* (Laurenti, 1768)

Class: [Amphibia](#) > Order: [Anura](#) > Family: [Bufonidae](#) > Genus: [Bufotes](#) >

Species: [Bufotes viridis](#)

Bufo viridis [Laurenti, 1768, Spec. Med. Exhib. Synops. Rept.](#): 27, 111. Type(s): Including frog figured on pl. 1, fig. 1 of original publication. Type locality: "inter fissuras, seu cavernas murorum obscuras Viennae", Austria.

Rana variabilis [Pallas, 1769, Spicilegia Zool., Fasc. 7](#): 1, pl. 6, figs. 1, 2. Types: Including frogs "in Edleri Museo" as well as frog figured in tab. 6, figs. 3 and 4, according to the original; if type material is lost, noted as the frog illustrated from Lübeck according to [Stöck, Günther, and Böhme, 2001, Zool. Abh. Staatl. Mus. Tierkd. Dresden, 51](#): 259. Type locality: "Lubecae" = Lübeck, Germany. Stated by [Pallas, 1777, Naturgesch. Merkwuerd. Thiere, 7](#) [Pallas' own German translation of [Pallas, 1769, Spicilegia Zool., Fasc. 7](#)], to have been originally coined as a synonym of *Bufo viridis* Laurenti, 1789, although there is no mention of Laurenti's work in the original description. Synonymy by [Dufresnes, Mazepa, Jablonski, Oliveira, Wenseleers, Shabanov, Auer, Ernst, Koch, Ramirez-Chaves, Mulder, Simonovo, Tiutenko, Kryvokhyzhar, Wennekes, Zinenko, Korshunov, Al-Johany, Peregontsev, Masroor, Betto-Colliard, Denoël, Borkin, Skorinov, Pasynkova, Mazanaeva, Rosanov, Dubey, and Litvinchuk, 2019, Mol. Phylogenet. Evol., 141](#): 19

Figure 7: *Bufotes variabilis* in the GBIF (A), ITIS (B) and Amphibian Species of the World (C) databases (Aug-25-2023).

II.2.4. The consensus fallacy in digital nomenclature

Zoological nomenclature relies on two main principles to determine a valid name: Priority and Usage. Priority gives prominence to the oldest of two or more names, called the senior name, while Usage gives prominence to the one name that is the most used. The opposition between priority and usage has been a regular topic of debate among taxonomists, as Dubois (2011b) attests.

Priority is a quantitative notion that offers a very small space for interpretation. Data collection for priority is not always easy, but is minimal. The only necessary information to solve the vast majority of the cases is the date of publication of the competing names. Finding this information can be difficult, but, at the end of the day, priority solves the vast majority of cases. Usage, however, can be trickier to establish. Some cases are clear, with an overwhelming indication that one of the competing names is undoubtedly the one to use. The more difficult cases can quickly become problematic, as it is impossible to have access to all the necessary information to establish usage. Within these cases, it can be tempting to rely on databases as a way to provide information on usage, like for example, which spelling is the most in use, with the implication that the spelling found on the higher number of databases is the one mostly in use, or that the more ‘authoritative’ database reflects more closely the practices in the community.

There is, however, an important limit to this approach that results from the elements that were developed until now. The fact that the databases, especially the ones that are of higher degrees, use data from other databases, leads to the possible propagation of mistakes (Code-wise), that will seem consensual to the non-informed eye. In the case of usage, databases can create a false sense of consensus, derived from the fact that a single source (usually a first-degree database) will be the overwhelmingly used digital source for a taxon group (Table 5). An involuntary mistake by the original database curator would then be propagated until it is corrected by either the curator or an attentive colleague. Emig et al. (2015) provide another example of the propagation of a mistake in databases, that ends up being included in several publications.

In the three examples that are cited below, the databases surveyed are among the most comprehensive. The three species were chosen at random in Arthropoda, Chordata and Echinodermata, to reflect different groups in Animalia. All three of them have a different first-degree database that is in charge of the taxon, but they all rely heavily on that given database. Aside from the Index of Organism Names, which belongs to Clarivate and thus draws its information from the Zoological Record, almost all the other databases rely directly or indirectly on that first-degree database.

Table 5: Structure of the sources for three different names of different animal groups in major databases.

A) *Acanalonia caymanensis* Fennah, 1971 (Fulgoromorpha)

Database	Presence of the name	Source cited by the database	Link
FLOW	Present	Fennah (1971)	https://flow.hemiptera-databases.org/flow/?db=flow&rank=species&id=22303&card=taxon&page=explorer&lang=en
GBIF	Present	FLOW	https://www.gbif.org/species/8362658
COL	Present	FLOW	https://www.catalogueoflife.org/data/taxon/8SSG
ITIS	Missing		
EOL	Present	FLOW	https://eol.org/pages/46924934/names
ION	Present	Zoological Records	http://www.organismnames.com/details.htm?lsid=2310938
Wikispecies	Missing		
IRMNG	Missing		

B) *Cynisca nigeriensis* Dunger, 1968 (Amphisbaenidae)

Database	Presence of the name	Source cited by the database	Link
The Reptile Database	Present	Dunger (1968)	https://reptile-database.reptarium.cz/species?genus=Cynisca&species=nigeriensis
GBIF	Present	The Reptile Database	https://www.gbif.org/species/5227637
COL	Present	The Reptile Database	https://www.catalogueoflife.org/data/taxon/339HF
ITIS	Present	The Reptile Database	https://www.itis.gov/servlet/SingleRpt/SingleRpt?search_topic=TSN&search_value=1125512#null
EOL	Present	The Reptile Database	https://eol.org/pages/795160/names
ION	Present	Zoological Records	http://www.organismnames.com/details.htm?lsid=674022
Wikispecies	Present	The Reptile Database	https://species.wikimedia.org/wiki/Cynisca_nigeriensis
IRMNG	Missing		

C) *Calveriosoma gracile* (A. Agassiz, 1881) (Echinoidea)

Database	Presence of the name	Source cited by the database	Link
World Echinoidea Database	Present	Agassiz (1881)	https://www.marinespecies.org/echinoidea/aphia.php?p=taxdetails&id=414208
GBIF	Present	World Echinoidea Database	https://www.gbif.org/species/2279020
COL	Present	World Echinoidea Database	https://www.catalogueoflife.org/data/taxon/Q4ZF
ITIS	Missing		
EOL	Present	World Echinoidea Database	https://eol.org/pages/4710471/names
ION	Present	None indicated	http://www.organismnames.com/details.htm?lsid=2289285
Wikispecies	Missing		
IRMNG	Present	World Echinoidea Database Codes for Australian Aquatic Biota https://www.cmar.csiro.au/caab/	https://www.irmng.org/aphia.php?p=taxdetails&id=10887239

Data collected on Aug-26-2023

In conclusion, taxonomic databases harbor a wealth of information that may have seemed unimaginable a few decades ago. They represent a capital component of the current taxonomic workflow and an amazing showcase for the discipline. They are, however, not exempt from weaknesses. Relying heavily on the work of a small number of people to feed almost the entire system makes databases susceptible to potential slowness, the occasional mix-up and mistakes, and creates a false idea of usage.

It is also important that, in the long term, the purely nomenclatural information gets extracted efficiently from these databases, and gets unified in the same way as taxonomic information is today. This information would then be able to benefit from multiple controls that would make it possible to locate and inspect any inconsistencies. For that to be done, the rules of the Code must be analyzed, and computerized, which might result in better curated nomenclatural data in the long run.

III. *Lognom*:
Computerizing the rules of the Code

The profusion of nomenclatural data available on the web can feel overwhelming at times, especially when trying to assess the quality of these data. At the same time, the traditional nomenclatural process and the application of the Code's rules are not always easy or within the grasp of the average newcomer.

A way of computerizing the rules of nomenclature had to be devised to alleviate some of that difficulty. After all, an important part of the taxonomists' workflow has benefited from advances in computer technology and there is no reason that the application of nomenclatural rules stays outside this trend.

The angle chosen for this project was the creation of a web application, called *Lognom*, for both 'nomenclatural **log**book' in English, and 'logiciel de **nomenclature** [nomenclature software]' in French. It is free to use and is part of the open science movement.

Lognom aims to answer its users' nomenclatural questions. It does not rely on the authority of an expert or a database consensus, but calculates the nomenclatural status of names and works depending on the data supplied by its users. It also includes several tools to answer some minor nomenclatural questions.

Creating *Lognom* necessitated a series of steps into uncharted territory, since no equivalent software or application currently exists. As a result, the web application had to be started from the ground up, by first identifying which entities should be stored in a database for this computerization project. Then, an in-depth analysis of the Code was carried out, to flesh out the properties of the identified objects. Specifications were then drawn up, leading to the definition of the database and functions of the future software. *Lognom* could then be implemented. At the time this document is written, *Lognom* is in a prototypical version, which could also be described as an elaborate Proof of Concept, waiting for the finishing of its tests and the fine-tuning of some details. An alpha version is already available at <http://lognom.infosyslab.fr/home/>. The beta version should be running in a couple of months. Any update on its situation will be available at the following address: <https://infosyslab.fr/index.php/en/resources/software/lognom/>.

III.1. Objects, properties and processes in nomenclature

III.1.1. The concept of nomenclatural objects

The first step in the preliminary work was the identification of nomenclatural objects. The latter are the meeting point between nomenclature itself and its translation to an algorithmic logic. Nomenclatural objects are the concepts that are manipulated to apply the rules of nomenclature.

In the context of this work, to be recognized as a nomenclatural object, an essential element that must be respected is the fact that the object itself or its properties influence nomenclatural decision-making. The properties linked to an object must all depend on the object, and not on another property of the object¹⁹.

This research singled out the following objects:

- Names (which can be nominal types)
- Taxa
- Works
- Type-specimens
- Ranks
- Series
- Nomenclatural acts

The notion of nominal type can be merged with the name, since a name, in addition to its properties, can be the nominal type of another name, while retaining its own properties. Each of these seven objects has its own properties (for example, a name, a work and a nomenclatural act have an availability status, a type specimen has a collection number, certain ranks have obligatory endings, and a taxon has a nature—extant, extinct, ichnotaxon, etc.), except for series. These properties will be listed and discussed at length. Conversely, something like the author of a name does not, strictly speaking, possess any property, and is therefore not isolated as a nomenclatural object²⁰. A date is not a nomenclatural object for the

¹⁹ This, in fact, is a simple application of part of the 2nd normal form of database management, which states that “a nonkey field [=property] must provide a fact about the key [=object], the whole key, and nothing but the key” (Kent, 1983)

²⁰ In theory, the fact that an author is anonymous or unknown can be considered properties of the author, and thus allow for its consideration as a nomenclatural object. In practice, it is not necessary to isolate this information in fields, as indicating that the author’s name is ‘Anonymous’ or ‘Unknown’ is sufficient to deduce these facts.

same reasons, nor are type-localities or first reviewers as both of them do not have properties that would influence the decision-making process. The objects that were singled out during this analysis ended up becoming the entities visible in the entity-relationship model presented in III.2.2.

III.1.2. Properties of names and works

After defining the nomenclatural objects that will be of interest, it was necessary to determine their properties. This was not a simple task, since it involved listing every element which, directly or indirectly, can have a consequence on the status of the objects. In order to determine these properties, the Code was analyzed, article by article, to highlight the existence of useful properties. Examples of such properties for names are the fact that a name is conditionally proposed, is associated with a type, gets assigned to a given rank, or has a potentially anonymous author. These properties were then listed and associated with the relevant possible states. Not unlike in taxonomy, where characters of taxa have character states, nomenclatural objects have property states for each of their properties. For a conditional proposal sensu Art. 11.5.1. or Art. 11.9.3.6., a name can be either conditionally proposed, unconditionally proposed, or, for a species name, proposed with a conditional combination. For its allocation, a species series name can have an explicit holotype at the time of publication, syntypes, or simply miss that information altogether.

As the rules of nomenclature change over time, in order to represent these properties and their associated potential states, it has been necessary to construct a specific diagram for each period that has specific rules. Figures 8 to 14 show the properties of only names and works (and nomenclatural acts), the two main objects, in seven parts. To read these figures, the starting points are the central objects, in a gray oval. Then, one must continue by working a way through each line to its terminal, passing by the status that will be affected, to end on a property and its potential states, in the square boxes. The status of each property can be of several types: in compliance with the Code (in green), not in compliance with the Code (in red), in a situation not recommended by the Code (in yellow), in an indecisive situation (orange) or in a situation requiring the intervention of a first reviser, the Commission, or additional elements (in purple). For each figure, the period is indicated in the top right-hand corner and in the legend.

This representation of each property makes it easy to follow the contents and colors of the boxes over time, which correspond to changes in the rules. For example, an anonymous author was accepted until 1950. On any figure, by starting from the name and going straight up, one arrives at the box containing the author. That property has two states, named and anonymous. Until 1950, as publication by named and anonymous authors resulted in an available name, so both states are colored green. After 1950, the state ‘anonymous author’ is changed to red, reflecting the guidelines given by Art. 14.

This early and lengthy step in the analysis process enabled the drawing up of a list of properties to be attached to names and works. But the benefits of this mapping is not limited to a simple list, since it also enables the visualization on a single illustration of the totality of possible states for each of these properties, for a given timeframe, and their consequences on the statuses of the relevant objects. The properties that were singled out during this analysis ended up forming the properties visible in the entity-relationship model in III.2.2., for names and works, and the states of these properties led to the conception of the algorithmic logic presented in III.1.3.

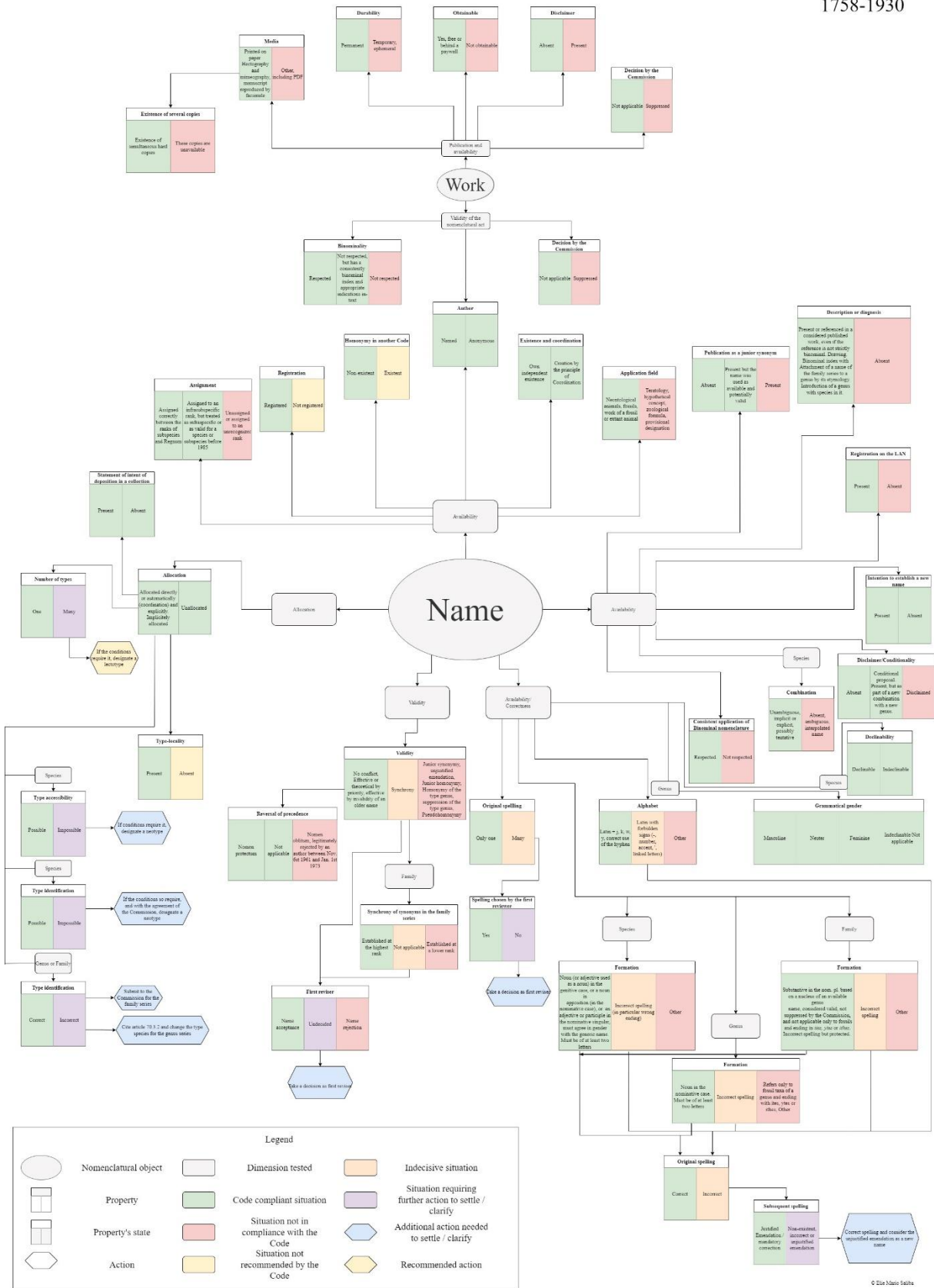


Figure 8: Properties of works and names throughout time; part 1 (1758–1930).



Figure 9: Properties of works and names throughout time; part 2 (1931–1949).

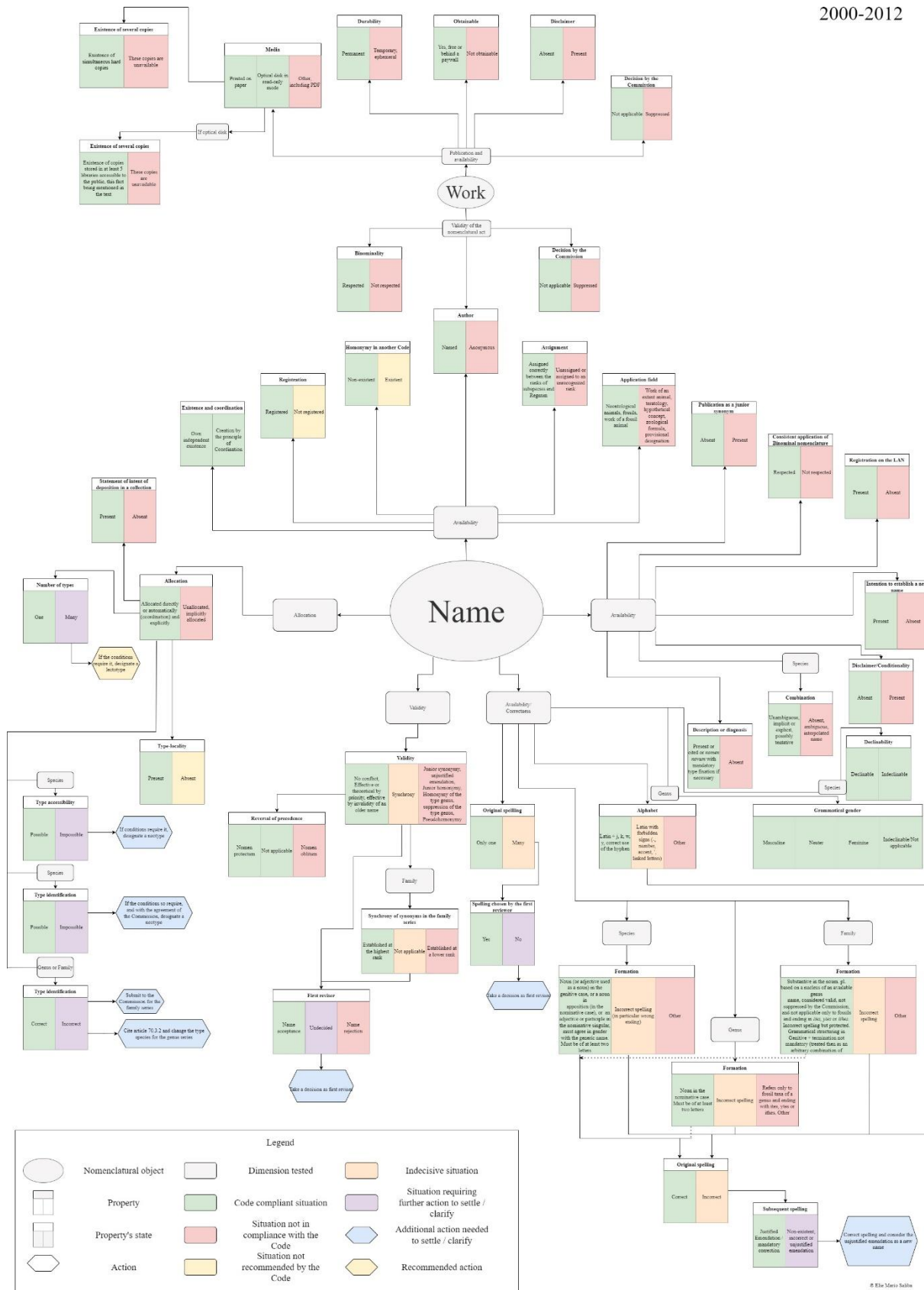


Figure 13: Properties of works and names throughout time; part 6 (2000–2012).

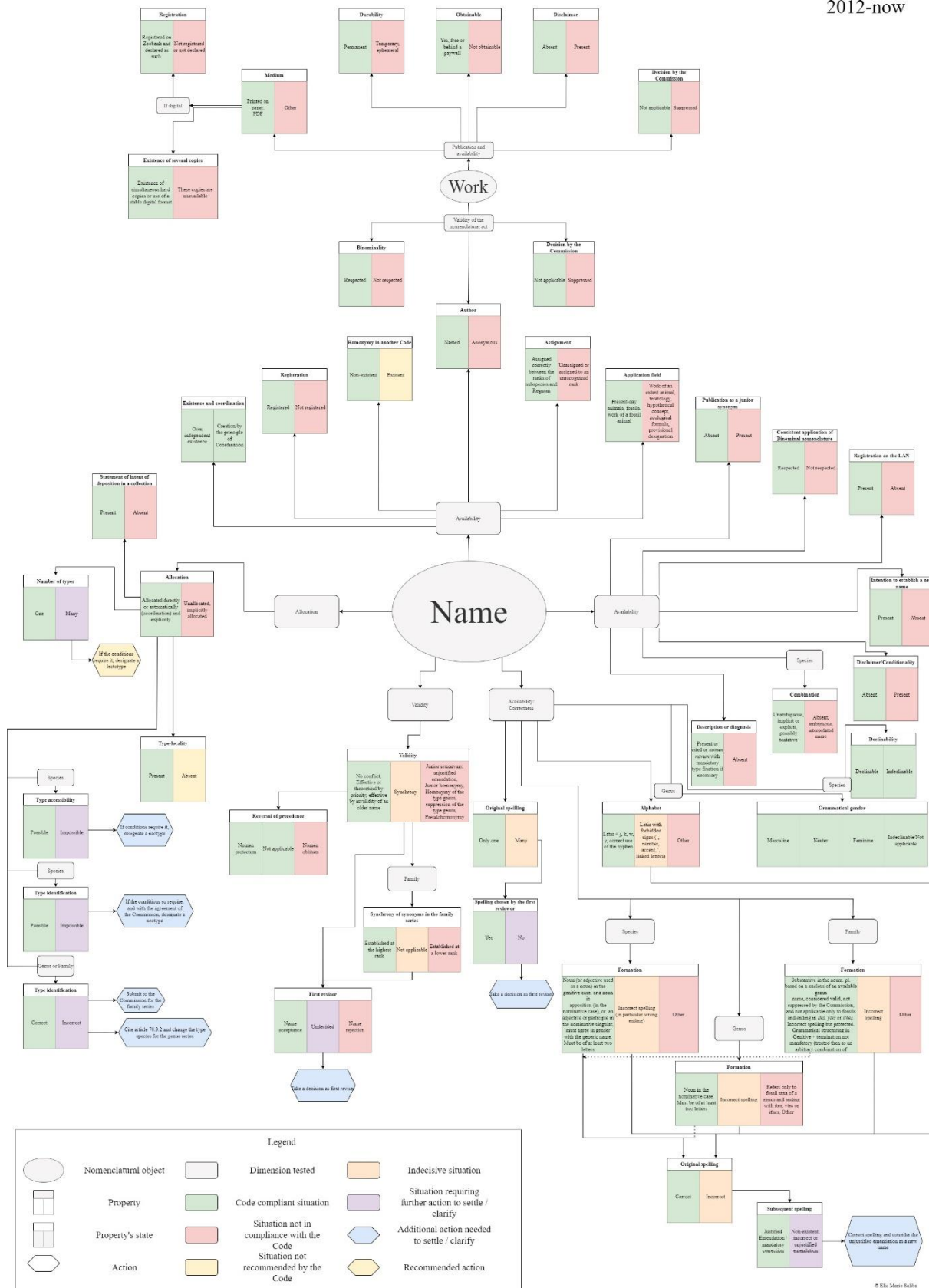


Figure 14: Properties of works and names throughout time; part 7 (2012–now).

Unsurprisingly, the next step required the transformation of this static visualization of the properties of names and works into an algorithm-like process

III.1.3. The algorithmic structure of each stage of the nomenclatural process

Once all the properties of the names and jobs had been identified, different approaches to establishing the logic needed to computerize the Code were considered, depending on the stage of the nomenclatural process they play a role in. The logic presented hereunder is behind the functions presented in III.2.3.

Availability

To determine availability, there are two logical possibilities. The first is to check that all the elements enabling availability are present and presuppose that the name is unavailable until proven otherwise. The competing approach is based on the idea that a name is considered available at the outset, and then check that it does not meet any condition that would make it unavailable. The second option is the one that has been chosen.

To apply this logic, the algorithm has to follow a step-by-step path, passing through each of the boxes that have been defined above. If only one unavailability condition is met, then the name is considered unavailable and an explanation for its unavailability is added. A name may be unavailable for a multitude of reasons, so it is necessary to check each step to ensure that the most complete account of its exact status is presented at the end. As an example, in Fig. 15, a fictional name follows the blue path. The name starts as being available. Since it was introduced in 2006, the set of rules that will be applied to it are those corresponding to Fig. 13 (Part 6). The name has a known author and has been published as valid. As these two conditions for its availability are fulfilled, the name, at this stage, is still available, and no action has to be taken. For its allocation, the name is associated with syntypes rather than a holotype, a situation which is considered not ideal by the Code, so that should prompt a warning message explaining that fact. Still, that does not render the name unavailable. However, as the name is introduced without the explicit mention of its novelty (Art. 16.1.), it becomes unavailable, and receives an explicit mention of the reason for its unavailability.

Validity

The treatment of validity follows a different logic compared to the one considered for availability. Firstly, validity is not an evaluation based on the accumulation of conditions, but a relative property of a name when put in a competition against others. This means that multiple names are given at the starting point, and that the order chosen for the steps is important. The strategy adopted for this part was to first determine which name has priority, then to check whether an element made this name (and its competitors) objectively or subjectively invalid. If the name designated as valid has a property that renders it invalid, then the second highest-priority name will take its place. In the example shown in Figure 16, *Genus* Author, 1790 is the oldest name, but when homonymy is checked, it turns out to be a junior homonym of another name. This makes it objectively invalid and *Genustres* Author, 1879 takes its place as the frontrunner for validity. It is then possible to move on to the next names for that property, and then the remaining properties. Like with availability, a message is added to inform the user of the reasons for a name subjective or objective invalidity. If none of the names is potentially valid, a message explains that no available name can be used for this taxon, and that the introduction of a new name is necessary.

The alternative approach would have been to test each name for every property that renders it invalid, and apply the principle of Priority only to the potentially valid names. The issue with that second approach lies with the reversal of precedence. As this particular property is relative to another name, it is necessary to know if the name that has had its precedence reversed is the valid one by virtue of priority alone, before checking for consequences linked to the reversal of priority. In the configuration selected, the workflow is more streamlined, and the reversal of precedence is the penultimate test undergone by the names.

The last control aims to recognize the kind of synonym a name is. For this to be made possible, the valid name needs to be determined first, which explains why synonymy is the last step in the validity process.

Other questions

In contrast to the two previous situations, to answer minor nomenclatural questions, the logic does not follow a particular generalized pattern, but varies according to the test and the associated question. One example only is provided in this document, the determination of a *nomen oblitum* under Article 23.9.

The logic underlying this test is quite simple, and is reminiscent of the one used for an available name. It is presented in Figure 17. The names entered as a *nomen oblitum* and *nomen protectum* respectively start by being a considered as such. As long as the conditions listed in article 23.9. are met, the senior name continues to be considered a *nomen oblitum* and the junior name a *nomen protectum*. If a single condition is not met, the names lose their respective tested statuses, and an explanation is provided to the user.

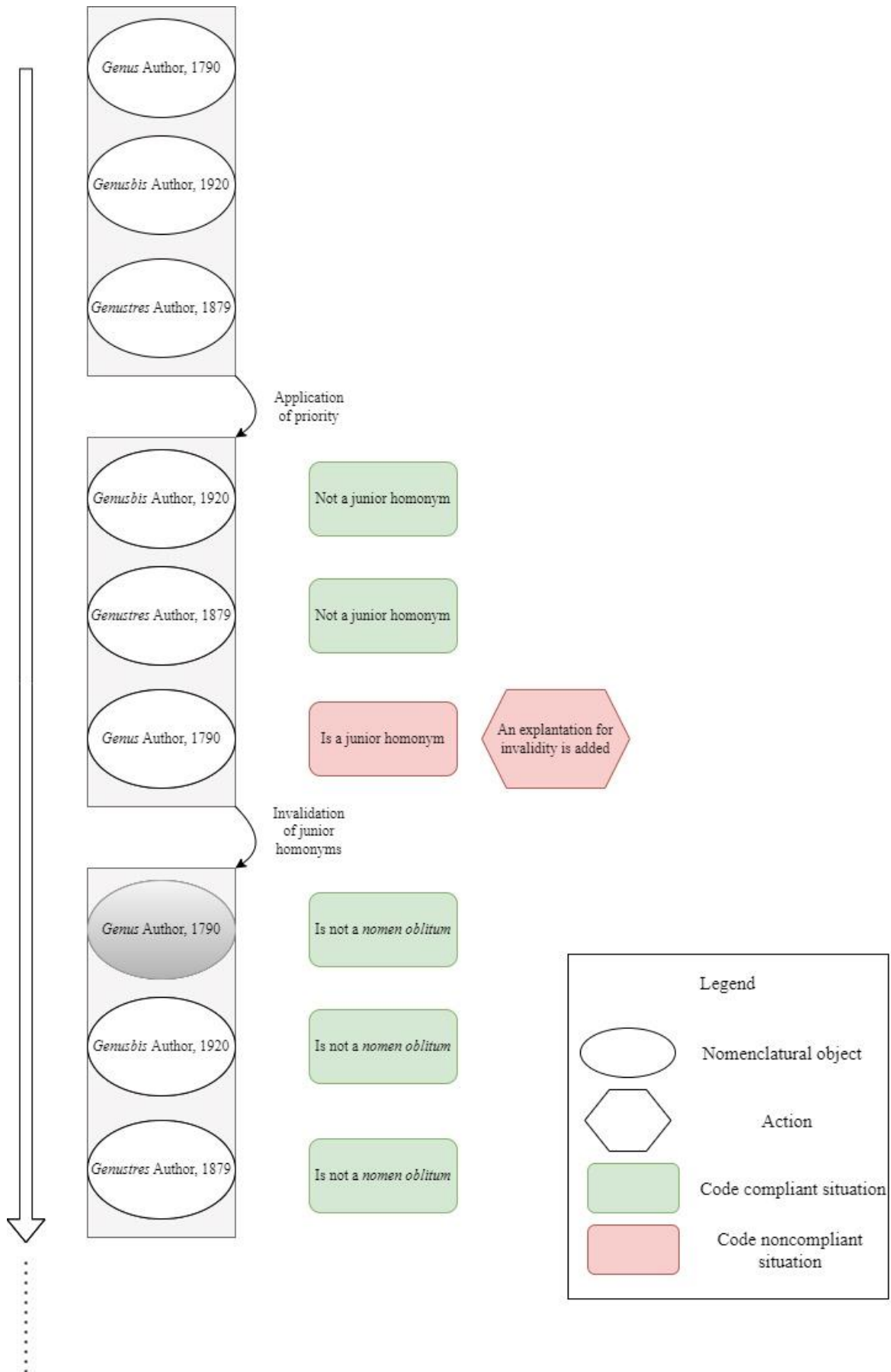


Figure 16: A simplified visualization of the process of testing validity.

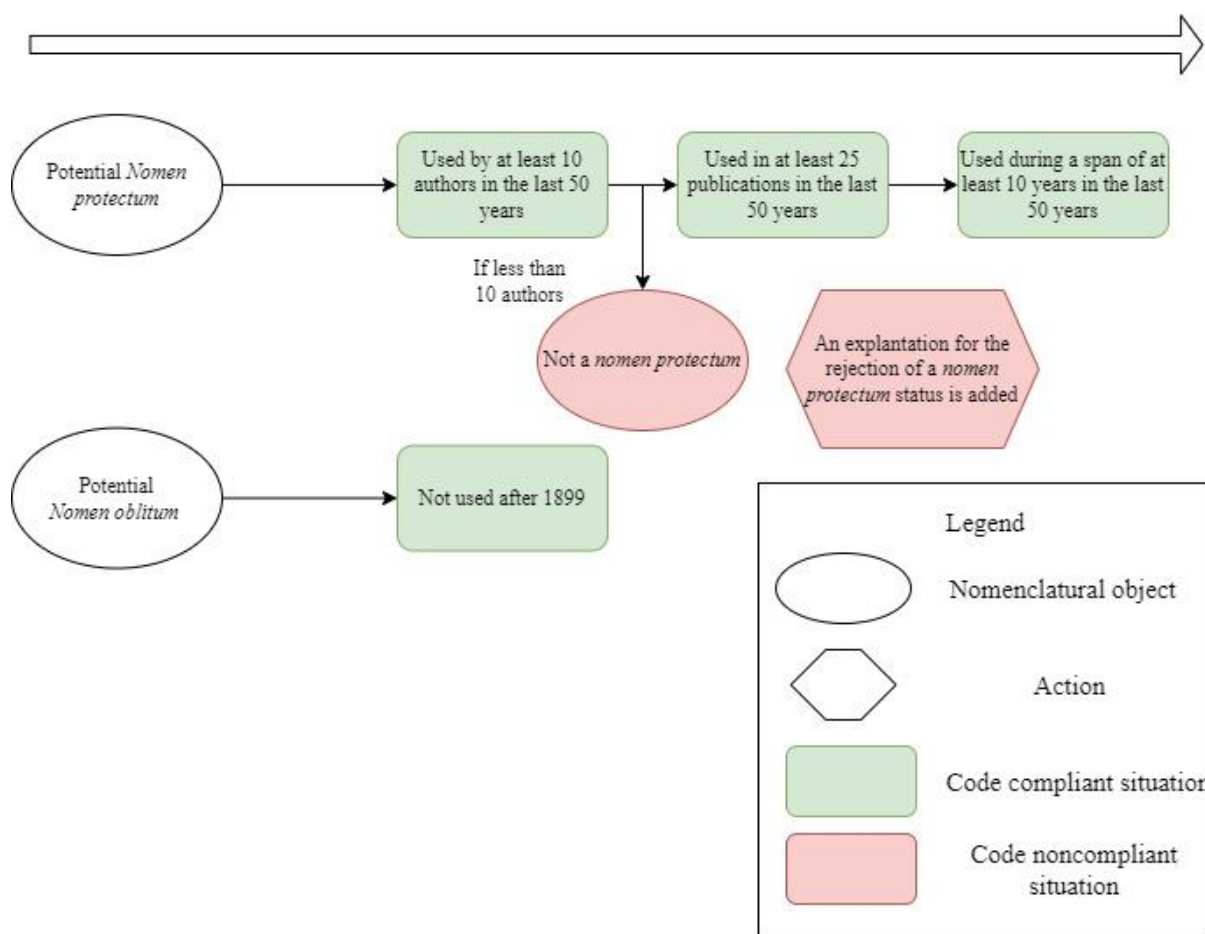


Figure 17: Steps in the test of a *nomen oblitum* and *nomen protectum* under Art. 23.9.

III.2. The creation process of *Lognom*

Before launching the programming step, and after the preliminary analysis stage, a bill of specification for *Lognom* was created. This bill served as guidelines in the early stages of development, and the ideas drifted gradually when confronted with difficulties or unexpected options.

III.2.1. Original specifications for *Lognom*

We are looking to build a web app that supports nomenclatural data and enables the rules of the Code to be applied. The aim of this software is to help in nomenclatural decision-making, within the framework of a given taxonomy, and to save the time and energy of the user. It must store a given nomenclatural data structure in memory, and make it possible to share these data. The software will also include a short interactive guide to ensure that as many Code recommendations as possible have been followed when creating a name.

Target audience

This software is intended for use by a public essentially made up of professional scientists and researchers, or students in training. A basic knowledge of nomenclature is a prerequisite for the use of the software, but expanding its audience to neophytes might be considered.

Exclusions and limitations

This software is not intended to:

- Support nomenclatural systems other than the International Code of Zoological Nomenclature.
- Propose a or frame taxonomy. As stipulated by the Code, the decisions relevant to taxonomy are left to the author. The creation of a computerized nomenclature linked to a taxonomy does not constitute, by any means, a legitimization of that taxonomy.
- Manage ranks above the Superfamily rank, as these are not currently regulated completely by the Code.
- Search for or generate nomenclatural data from existing databases. This could potentially be done afterwards, but the web app itself should rely only on data entered by the user.
- Resolve nomenclature conflicts not explicitly covered by the Code, or take any decision regarding the interpretation of a given Article.
- Automatically generate names or ranks from a phylogenetic tree or classification.
- Dispense with the need for careful prior data collection. As with all data management software, performance is closely linked to the quality and completeness of the data provided. The intended approach is one that favors indecision over the possibility of making mistakes.

Main components

The software will consist of two "essential" parts on the user side, and three optional parts (depending on the time available for their implementation):

1. An interface comprising all relevant data to be able to assess the availability status of a given name or work. It should take the shape of one or multiple forms, keeping in mind that the user should have to provide as little data as possible to make the correct decision.

Table 6: Example of a highly simplified entry in *Lognom* for an imaginary genus of frogs.

Name	Author(s)	Date	Type (collection number for the species, type-species or type-genus if relevant)	Homonymy (none by default)	Date associated with the homonymy
<i>Frogus</i>	Dubois	1982	<i>Rana sylvestra</i>	None	N/A
<i>Dafdaa</i>	Ohler	2015	<i>Rana dafdaa</i>	None	N/A

2. An interface for submitting various data on a given taxon, to determine validity. This interface can take the form of a potentially interactive table form reminiscent of a CSV file (Table 6). Each row will be given a name, and each column will be given the properties of interest and necessary for making nomenclatural decisions. On output, the software should return the valid name for the taxon, as well as synonyms, their nature (objective and subjective), and the reasons for synonymization (Fig. 18). Naturally, all the nomenclatural information associated with the name will also be displayed in its proper form, i.e., author, date and associated name-bearing specimen.

<p>Genus <i>Frogus</i></p> <p><i>Frogus</i> Dubois, 1982</p> <p>Type species: <i>Rana sylvestra</i></p> <p><i>synonym: Dafdaa</i>, Ohler 2015: junior subjective synonym</p> <p>Type species: <i>Rana dafdaa</i></p>
--

Figure 18: Example of a simplified output for an imaginary genus of frogs.

If it is not possible to decide on the validity of the names, then the missing information necessary to do so will be mentioned, to enable the user to complete the necessary data entry.

If the decision can be made, but the information remains incomplete, then a warning message should appear.

3. An interface for checking the elements necessary to confer availability before proposing a new name or work. This interface manages one name at a time, and proposes to review one by one the various nomenclature elements required or recommended before publishing a new name. The user can check boxes corresponding to each of these elements, and receive guidance on whether these elements are mandatory or optional.

4. (Optional) An interface for managing dates and date conflicts

5. (Optional) An interface for managing authors and authorship conflicts.

Aesthetics and ergonomics

The aesthetics of the software should be sober but welcoming, inspired by old manuscripts.

Ergonomics are essential to motivate the public to use the software, and to this end, the following guidelines elements will be focused on:

- Prefer, in this order and whenever possible, Booleans (checkboxes), then drop-down lists and finally manual entry.
- Make it possible and easy to modify or call up previously entered data.
- Provide help boxes and advice each time that they might be needed.
- Develop a F.A.Q. page.

Testing and feedback

The software will be tested in two stages. First, with specialists from a group with in-depth knowledge of nomenclature, to identify any shortcomings, errors or bugs in terms of algorithms, then with a more general public, to consider the aspects relating to comprehension and ergonomics, maybe by comparing *Lognom* to the use of the paper or online Code.

III.2.2 Entity-relationship model and database schema

After determining the objectives, the objects concerned, the properties of these objects and the form that the process should take, *Lognom*'s entity-relationship diagram (Fig. 20) was

drawn. An entity-relationship diagram is used to visualize how ‘entities’ such as people, objects or concepts relate to each other. In Fig. 20, rectangles represent entities (which are our objects) with their list of attributes (which are the objects’ properties). These entities are linked by a relationship that explains the semantic link between two entities. Some entities are specialization of more inclusive entities, and inherit their properties in addition to their own specialized properties. This relation is signified by the presence of an arrow indicating the inheriting entity. On top of these elements appears the cardinality, which corresponds to the numerical attributes of the relationship between two entities (see Fig. 19).

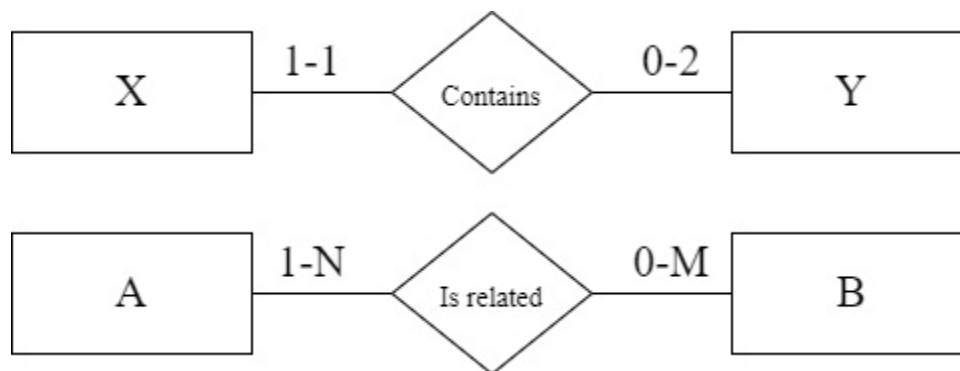


Figure 19: Cardinalities in an entity-relationship model.

To read the *Lognom* entity-relationship model cardinality:

- an instance of X is contained in one and only one instance of Y (1-1 is read one-to-one);
- an instance of Y contains zero to two instances of X (0-2 is read zero-to-two);
- an instance of A is related to at least one to multiple instances of B (1-N is read one-to-many);
- an instance of B may be related to one or more instances of A (0-M is read zero-to-many).

Unsurprisingly, this entity-association diagram includes the main objects mentioned above (names, works, type specimens, taxa, ranks and series). Nomenclatural acts are not included (see V.4.2.). In addition to these objects are:

- the software user, which represents the different users of *Lognom*;
- the nomenclatural file, which will serve to organize the different names and works and taxa into projects, so that it will still be possible to manage the data even if the number of objects to get around becomes important;
- the bibliography, that will provide supplementary information for each work, by going further than the critical information needed for purely nomenclatural reasons.

Moreover, an analysis of the properties of the links between names and taxa revealed a number of peculiarities. These peculiarities will be discussed again in Chapter 4. Essentially, when a taxon is introduced, it is linked to one and only one name with the properties listed (that would exclude homonyms, as they have different authors, date of publication, types, etc.). That taxon is called protaxon by Dubois (2005), as it is the taxon when first introduced. However, when trying to determine the validity of a name, it is more than likely that the taxon has a link with several names, as it is now the subsequent accepted use of that taxon, or an ergotaxon (Dubois, 2005). However, this ergotaxon does not bear the same properties that a protaxon in the context of answering nomenclatural questions. Elements such as the nature of the taxon (extant animal, fossil, ichnotaxon, work of an animal, etc.) are properties that are fixated when the taxon is introduced and the name made available. On the other hand, for the validity step, no consideration of the nature of the final taxon is necessary²¹ to make the decision. For all these reasons, protaxa and ergotaxa are differentiated in the entity-relationship model.

²¹Even if some elements, like the fact that the name of an ichnotaxon cannot compete for priority with the name of a classical taxon (Art. 23.7.3.), might make it seem like the nature of the taxon is important at the apotaxon stage, these properties are introduced simultaneously with the taxon description rather than subsequently.

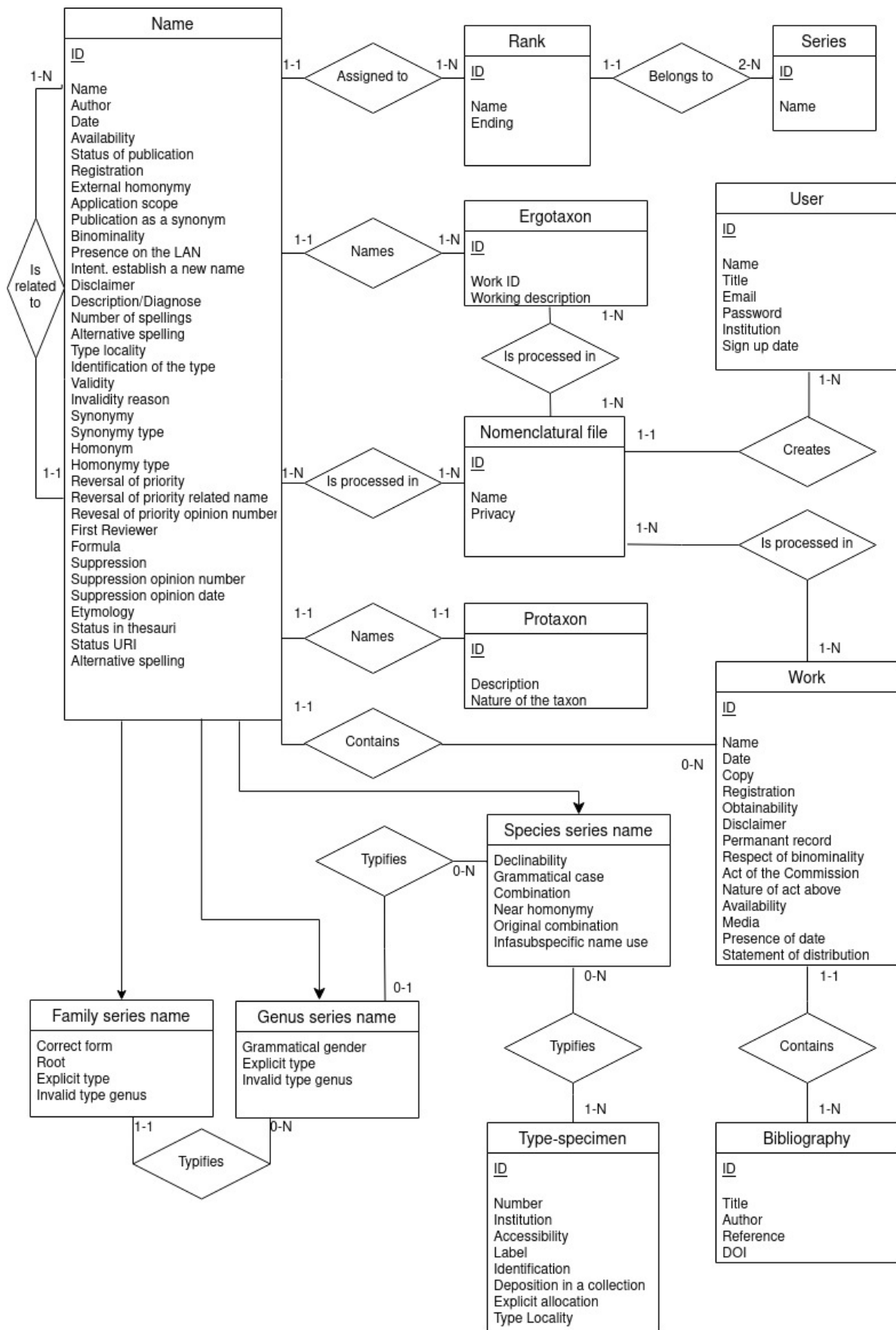


Figure 20: Entity-relationship model of *Lognom*.

Implementation

After all this important preliminary work on paper, the next step dived into the programming stage. As mentioned earlier, *Lognom* was coded using the Django framework. Django is a “high-level Python web framework that encourages rapid development and clean, pragmatic design [of websites and web applications]. [...]. It is free and open-source” (Django Software Foundation, n.d.). Python is the main programming language used in Django, and the framework includes the management of the database, the forms, the functions (back end) and the user interface (front end). All dimensions of the software were developed by myself. The database is managed with a database management system called SQLite (<https://www.sqlite.org/index.html>), which is free and open-source. The front end was created ‘manually’ using a combination of HTML, CSS and Javascript (Fig. 21).

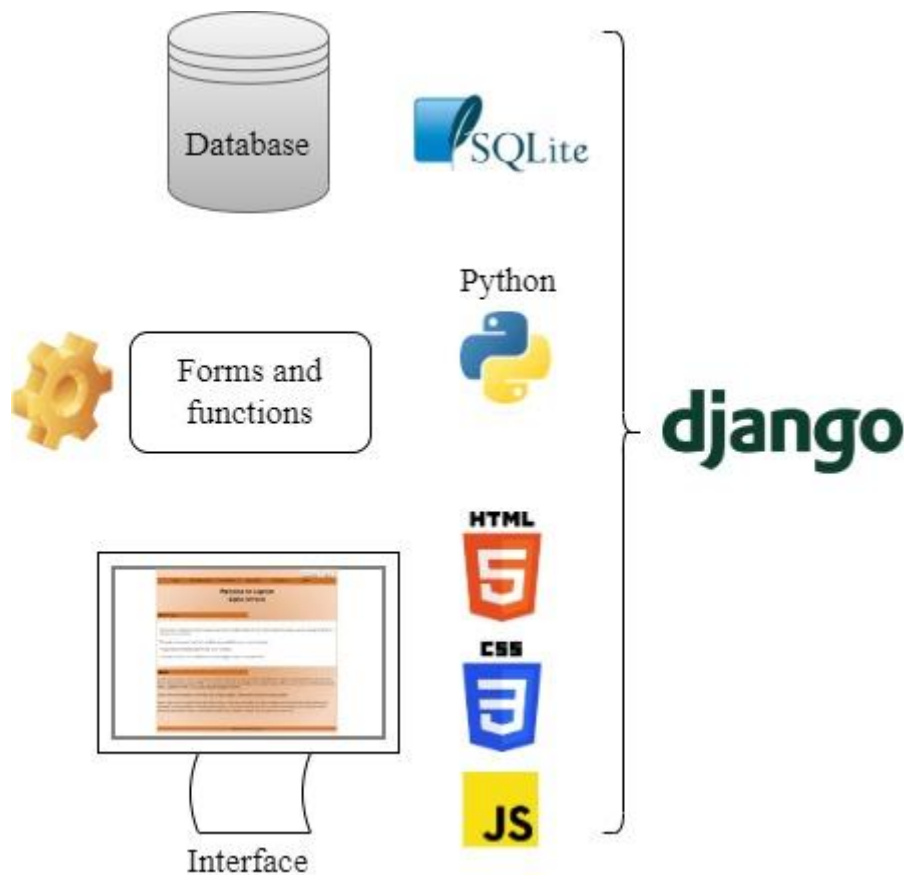


Figure 21: Structure of *Lognom* in Django, with the corresponding programming languages.

The logos for HTML and CSS were made by the World Wide Web Consortium. The other logos are in the public domain or under an open license.

One of the very first steps in programming under the Django framework is to create the database in the ‘models’ file. The database schema is visible in Fig. 22, and is a translation of the entity-relationship model presented above. Each rectangle represents an object, and is

called a table. Properties are now called fields, and are represented by the columns of the table. Relationships are translated into links between these tables using foreign keys. They make it possible to call in a field of a table an entry of another table. Some other elements vary between the entity-relationship model and the database schema:

- Django has an integrated user model, which makes it easier to create systems to sign up, log in and log out. To complete this integrated model, a supplementary table, called Profile, was added.
- When a relation is many-to-many, a supplementary table is added, called a join table, that will collect foreign keys from both sides. A join table between table 1 and table 2 will be called table1xtable2.
- The table for type-specimen is called Onymophoront, to avoid using type (see IV.1.1). The join table between Onymophoront and Name is called Taxomen, a concept defined as “The permanent association between a [name] and an onomatophore [=name-bearing type] allowing objective, non-ambiguous and stable allocation of [names] to taxa.”²² (Dubois, 2000).
- There are two tables that deal with nominal types. In the first situation, the name that constitutes the nominal type is already entered in the Name table. In that case, the user can designate it directly as a type name. In the second situation, the user does not want to enter the name and all its properties entirely and just wants a quick answer. In that case, an entry in the Type Name table is created, which contains only the critical elements. It is a way of avoiding the process of creating a full entry for a name just to use it as a type name.

An in-depth guide to each field in the database is provided in Appendix 1.

All the elements considered critical in Chapter 2 are deliberately present in *Lognom*'s database. The equivalencies are given in Table 7. That ensures that the software should be able, in the long run, to use part of the data already present online, and that alignments using the Darwin Core standard are possible to make.

²² I first understood that definition in the sense that a taxomen is the association itself, rather than a combination of the three elements (name, type and the association). Hence the name of this table. I was mistaken, as a taxomen is a very close synonym to a nominal taxon, but I still think that naming that association is useful in the context of formalization and theoretical nomenclature.

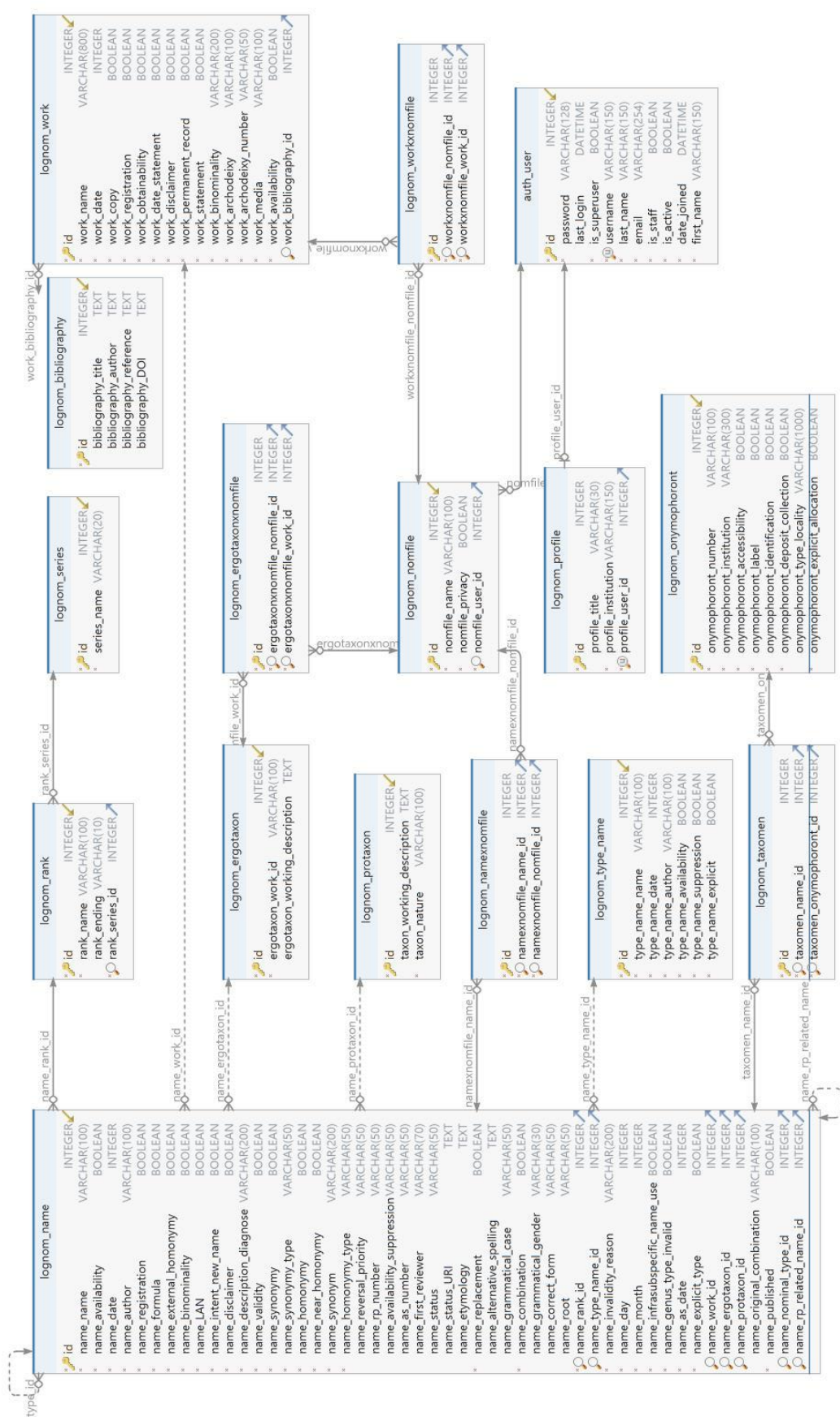


Figure 22: Database schema of Lognom.

Table 7: Important nomenclatural elements and their equivalent in Darwin Core and *Lognom*.

Element	Property in Darwin Core	<i>Lognom</i> field
Name	dwc:scientificname	name_name
Author	dwc:scientificNameAuthorship	name_author
Date	dwc:namePublishedInYear	name_date
Rank	dwc:taxonRank	name_rank
Original publication	dwc:namePublishedIn	name_work
Original combination	dwc:originalNameUsage	name_original_combination
Type status	dwc:typeStatus	name_type_name, name_name_type, name_onymophoront
Collection	dwc:collectionCode	onymophoront_collection
Collection label number	dwc:catalogNumber	onymophoront_number
Nomenclatural status	dwc:nomenclaturalStatus	name_availability and many others
Taxonomic status	dwc:taxonomicStatus	name_validity and many others
Nomenclatural Code	dwc:nomenclaturalCode	N/A (Always ICodeZN)

III.2.3. The main function

The only step left that is needed to get the digital nomenclatural process up and running is the addition of functions that apply the rules and treatments in the way described in III.1.3 and illustrated in Figs. 15, 16 and 17. Each function plays a role in a different test. The role and explicit objective of each function of *Lognom* are presented in Table 8.

Moreover, there are three types of functions in the software. The first type is main functions, which run the main tests regarding availability and validity. These functions call upon other functions to work efficiently. The second type of function is the auxiliary functions, whose role is to provide controls as the program runs in the main functions. They are called upon by the main functions. The last type of functions is the one created for tools that help resolve precise nomenclatural questions. They work independently, and are not connected to the other functions. A summary of the relationships of the major functions of *Lognom* is provided in

Table 9. The web application also includes a number of minor functions whose role is displaying static pages, as well as managing the users' accounts, nomenclatural files or the classical login/logout features. The full program for each function is commented and will be shared if requested.

Table 8: Objectives of the major functions of *Lognom*.

Function	Part of	Objective
controlHomonymy()	Homonymy under Art. 58.	Determines if two specific names are homonyms under Art. 58.
genusGender()	Determination of a genus series name gender	Determines the gender of a Genus name
nomenOblitum()	Reversal of precedence under Art. 23.9.	Determines if Art. 23.9. can be applied
nameAvailabilitySpecies(), nameAvailabilityGenus(), nameAvailabilityFamily()	Availability of published names	Tests the availability of published names adapted to each series
nameValidity()	Validity	Tests the validity of names
publishSpeciesName(), publishGenusName(), publishFamilyName()	Availability of unpublished names and works	Tests the availability of to-be-published names adapted to each series, and to-be-published works
publishWork()	Availability of unpublished works	Tests the availability of to-be-published works
workAvailability()	Availability of published works	Tests the availability of published works
availabilityMain()	Availability of published names	Tests the availability of published names for all the common criteria of the three series.
controlAlphabet()	Spelling	Checks the spelling and corrects the use of banned characters

controlEnding()	Spelling	Checks the ending of family-series names
controlFossilEnding()	Spelling	Checks the ending of fossil taxa ending with 'ithes', 'ites' or 'ytes'
nomenclaturalAvailabilityStepName()	Availability of published and unpublished names	Checks the effect on availability for one property of a name
nomenclaturalAvailabilityStepWork()	Availability of published and unpublished works	Checks the effect on availability for one property of a work
validityStep()	Validity	Checks the effect on validity for all relevant properties of each competing name

Table 9: Types and relationships of the major functions of *Lognom*.

Function	Type of function	Relationship to other functions
controlHomonymy()	Tool	None
genusGender()	Tool	None
nomenOblitum()	Tool	None
nameAvailabilitySpecies(), nameAvailabilityGenus(), nameAvailabilityFamily()	Main function	Calls for nomenclaturalAvailabilityStepName(), controlAlphabet(), controlEnding(), controlFossilEnding() availabilityMain()
nameValidity()	Main function	Calls for validityStep()
publishSpeciesName(), publishGenusName(), publishFamilyName()	Main function	Calls for nomenclaturalAvailabilityStepName(), controlAlphabet(), controlEnding(), controlFossilEnding()
publishWork()	Main function	Calls for nomenclaturalAvailabilityStepWork()
workAvailability()	Main function	Calls for nomenclaturalAvailabilityStepWork()
availabilityMain()	Auxiliary function	Called upon by nameAvailabilitySpecies(), nameAvailabilityGenus(), nameAvailabilityFamily()
controlAlphabet()	Auxiliary function	Called upon by nameAvailabilitySpecies(), nameAvailabilityGenus(), nameAvailabilityFamily(), publishSpeciesName(), publishGenusName(), publishFamilyName()
controlEnding()	Auxiliary function	Called upon by nameAvailabilityFamily(), publishFamilyName()
controlFossilEnding()	Auxiliary function	Called upon by nameAvailabilitySpecies(), nameAvailabilityGenus(), nameAvailabilityFamily(), publishSpeciesName(), publishGenusName(), publishFamilyName()

nomenclaturalAvailabilityStepName()	Auxiliary function	Called upon by availabilityMain(), nameAvailabilitySpecies(), nameAvailabilityGenus(), nameAvailabilityFamily(), publishSpeciesName(), publishGenusName(), publishFamilyName()
nomenclaturalAvailabilityStepWork()	Auxiliary function	Called upon by publishWork(), workAvailability()
validityStep()	Auxiliary function	Called upon by nameValidity()

III.2.4. Lognom's structure and user interface

Lognom is currently online, in an alpha version (Fig. 23). At this stage, it is closer to a proof of concept than to a totally finished product. *Lognom* welcomes visitors on a 'home' page, but the rest of the site is inaccessible without creating an account and logging in. That mandatory registration makes it possible to save the data entered by the users for subsequent uses. The home page includes a section for announcing the latest updates, and an overall explanation of the context of nomenclature. That summary links to the Commission's website and the Code online.

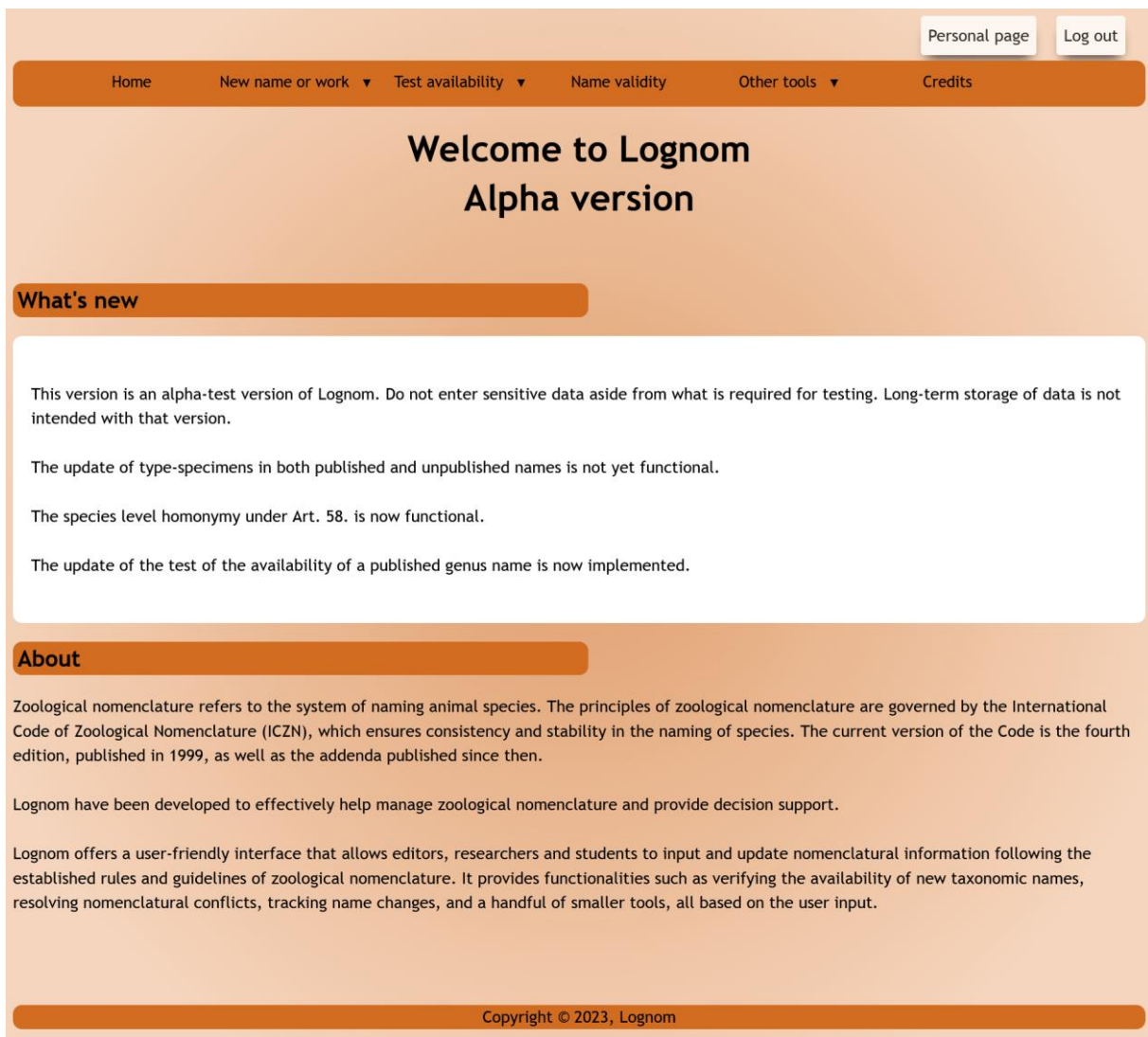


Figure 23: Screenshot of the home page of *Lognom*'s alpha version.

Once registered, logged in, and after creating a new project, if necessary, the user can start browsing the horizontal bar to select the tests he or she wishes to run. The possibilities are listed below (Fig. 24).

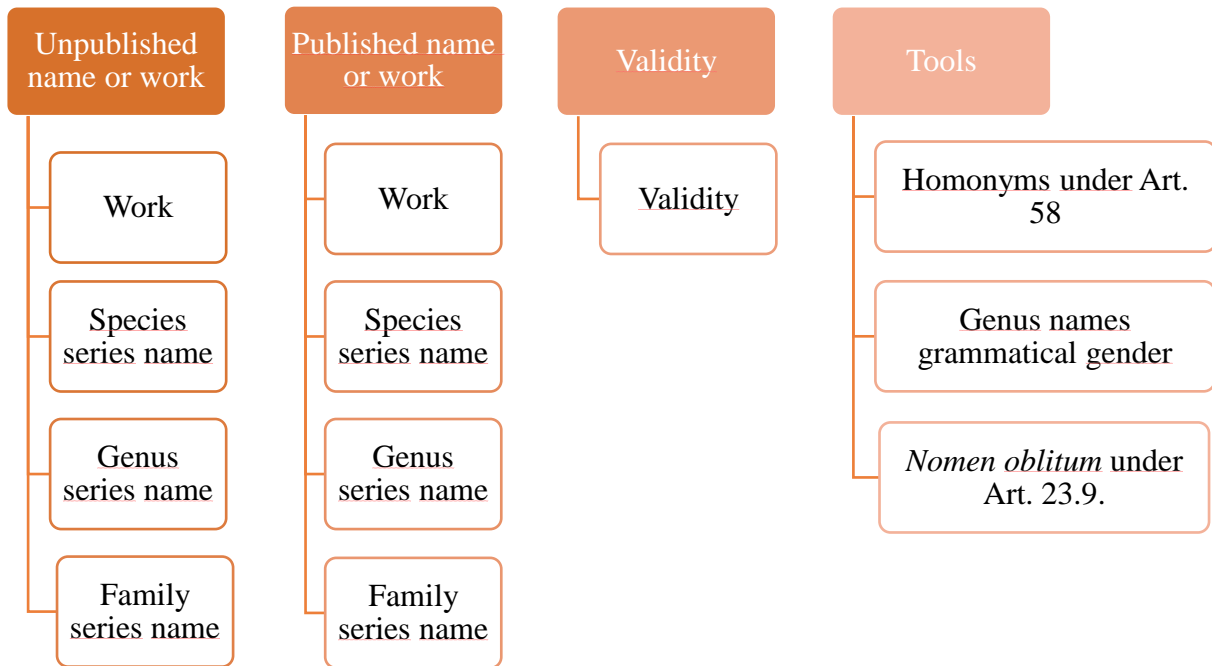


Figure 24: All the possible tests offered by *Lognom*.

Availability of new names and works

The first type of test is to check the potential availability status of a work or a name before it gets published. Due to the availability conditions differences between the three series, they are separated into three pages. At this stage, the names or work entered have not been published yet, so it is still possible to warn the user about some unrecommended situations and recommend optimal choices. Some examples include the fact that it is better to have one type specimen for a species name than several, or reminding the user to register a work or name on *ZooBank*, even if it is not mandatory in the Code. The tests rely on forms that have to be filled by the user, containing mainly checkboxes, dropdown lists and some text fields to fill. As stipulated in the specifications, the first two widgets are the favored means to ask the user for information as they are quicker to fill, easier to interact with and less likely to cause mistakes. Checkboxes are the logical widgets to use for a Boolean value, which reflects a simple dichotomous choice. Dropdown lists are also quite useful, and are in fact used in case of a choice to be made between mutually exclusive scenarios. Technically, all entries that are dropdown lists could have been checkboxes, but considering that these choices are

mutually exclusive, it is optimal to use the list widget. Fields to fill manually are left for the information that is not gatherable otherwise. Together, these form elements collect all the relevant data to make it possible to check the availability of a name or work following the logic presented in Fig. 15.

Each name and work is saved in a nomenclatural file. This allows the management of names and works as their number grow. A nomenclatural file could be created for each article a taxonomist prepares, or be used to limit the scope of the names and works stored in it to a taxon, or a geographical area (e.g., “Flies of Greece”, “Neotropical dung-beetles” or “Delphinidae of the Southern Atlantic Ocean”). To copy from a nomenclatural file to another, one simply has to update the unpublished work or name and select the new nomenclatural file he or she wishes to copy the name or work into. Any modifications done in the data about a name or work in any of the nomenclatural files will be copied to the others, as putting a name or work in different files does not create duplicates.

All the forms for availability tests, be they published or unpublished names and works, are subdivided into multiple sections, with the exception of unpublished works, which form is quite small. Typically, a name form will have a section for the name nominal complex, one about the main availability parameters, a section on the work, one on the taxon, another on actions by the Commission (if published), a section on the grammar and finally one on the type name or type specimen. The work form has to adapt to the information given in the work medium field. As a consequence, some parts will appear or be hidden if they are relevant or not to the work’s medium.

The unpublished names and works are intended primarily for authors who would like to describe new taxa but do not necessarily possess the nomenclatural skills needed to ensure that their names will be available. Colleagues that are specialists of other biological disciplines or student taxonomists are examples of potential users. Another audience who could benefit from this section of the application is reviewers and editors. They can use the software while going through the manuscripts entrusted to them, and check that they have not accidentally overlooked an error or forgotten about a recommendation they could make to the authors.

When updating an unpublished name, it is possible to indicate that the name has been published and send an unpublished name to the published names pile.

Availability of published names and works

The second collection of tests is about the availability of already published names or works. Once again, the test is subdivided into the three series. This test requires the longest and most complex forms of all of the tests included in *Lognom*. The species series availability test for an already published name is, in fact, the longest of all the forms in the software (Fig. 25).

The principle is similar to checking the availability of an unpublished name or work, at the difference that the date of publication is now required to make sure that the conditions of availability that are enforced for that particular time period are respected.

Some recommendations are still mentionable at this stage, but most of them are not. This test is to be used after the publication and cannot prevent undesirable nomenclatural decisions taken when introducing a name, as most recommendations are no longer relevant. This test main goal is to check efficiently the availability of a name that might be encountered by a taxonomist, without having him or her going back to the Code to verify which conditions are in force for that period. It is meant mainly for taxonomists and student taxonomists that are unfamiliar with the rules of the Code. These forms can also be used as a way of collecting all the possible relevant data that might be useful for a nomenclatural database, or for future possible developments. This is why some fields such as the etymology of a name, or institution hosting a type specimen are proposed, even if they do not play a direct role in the decision-making process. The idea is that they might still prove useful for subsequent uses, so they might as well be included.

In both pre-published and published names tests, a check of the permitted characters is made, as well as a suffix test for family series names. If characters with accents are used, the corresponding character without an accent is used to replace the forbidden character. If no equivalent character exists, i.e., a character from a different alphabet or a special character such as ‘!’, they are deleted. Numbers from 1 to 20, as well as 30, 40, 50, 60, 70, 80, 90, 100 and 1000 are translated into their cardinal equivalent in Latin, in the correct gender declension

for 1, 2 and 3. AnimalBase (AnimalBase, n.d.) suggests a list of prefix translation for numbers for published names. These should be added eventually.

The software returns one of two answers. Either the name is available (Fig. 26), or it is not (Fig. 27). In the case of a negative output, all the reasons that render the name unavailable are listed, with the corresponding article in the Code, so that the user might check directly if he or she is not sure about its interpretation. In both cases, elements, such as spelling mistakes or mandatory emendation, that do not affect the availability of a name but induce a change are also mentioned, with the relevant corresponding articles. Lastly, recommendations also appear if the user can still apply them at this stage.

[Personal page](#) [Log out](#)

Home
New name or work ▾
Test availability ▾
Name validity
Other tools ▾
Credits

Checking the availability of a species-series name

General information

Store in project

Name rank

Genus


Species

Generic name in original combination

Specific name form in original combination

Name of the author(s)

Date of publication of the name



About the work

Choose a work

Keep in mind that a name published in an unavailable work is itself unavailable.

About the name

The name follows the Principle of Binominality

The intention to create a new name is present

The specific name is introduced in an unambiguous generic combination

There is a disclaimer stating the name should not enter nomenclature.

The name has an equivalent spelling to another species name and/or is a junior zoological homonym

The name is registered in Zoobank

The name is present in the List of Available Names, if there is one for the given taxon.

The name is a new replacement name

The name is

The name has a related

About an act by the Commission

Action of the Commission

Year of the decision

Case number

About the name grammar and etymology

The name grammatical case

The name grammatical gender

The etymology of the name

The name is a zoological formula

About the taxon

The nature of the taxon

About the type specimen

The collection number of the type-specimen (write 'None' if there is no type-specimen, '!' if a type specimen exists but has no number)

The institution where the type-specimen is conserved

The type specimen type locality name

The type specimen is explicitly designated in the work

The type-specimen is accessible

The type-specimen is identifiable

A statement of intent that the type-specimen will be (or is) deposited in a collection and the name and location of said collection are explicit in the text.

Copyright © 2023, Lognom

Figure 25: Lognom's form for testing the availability of a published species series name.



Figure 26: Positive result of an availability test for a species name.

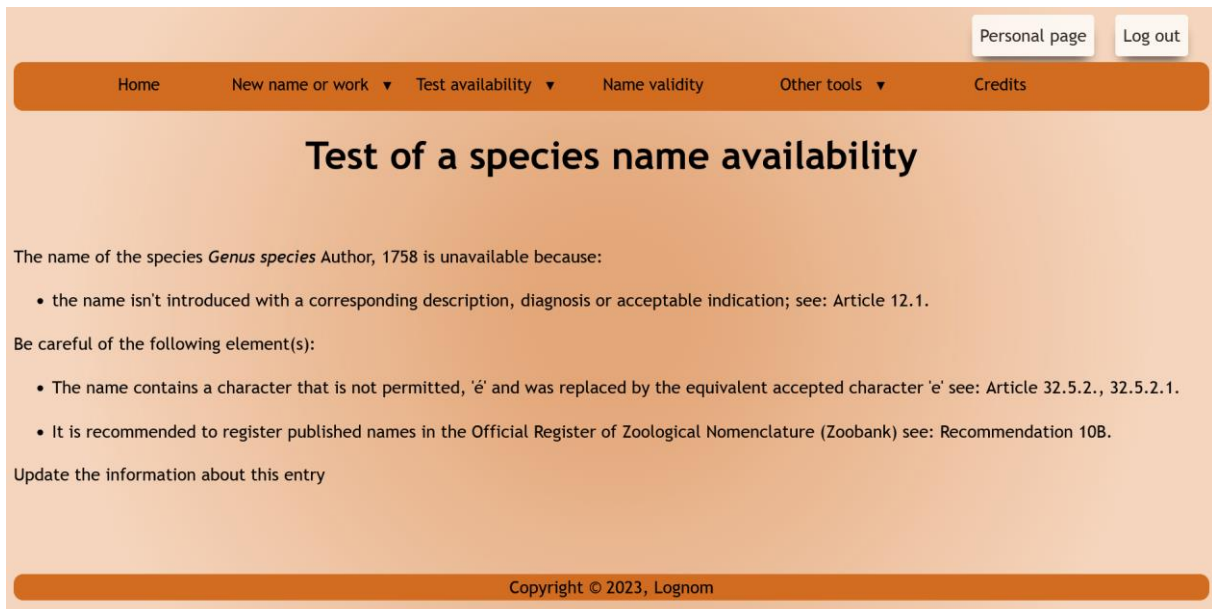


Figure 27: Negative result of an availability test for a species name.

Validity

The third test included in *Lognom* focuses on validity (Fig. 28). It is mandatory that the names that will be tested are registered in the database as either an unpublished name or as a published one. The unpublished name will have the current year as a publication date in regard to the calculation of priority.

To make use of this test, the user chooses at which rank the validity test should take place.

The subsequent choices of names will be thus limited to all the names in the relevant series,

so as to not create a dropdown list of choices of names that would be too long and irrelevant. The nomenclatural file is also required, to limit the names that should be considered, once again as a way to reduce the length of the first drop-down list. Otherwise, all the names entered by the user would appear, quickly becoming unmanageable. A name is given to the result of that test so that the user is able to retrieve it and update it in case the taxon changes or a competing name is added or taken off the list. At the current stage, the update of a validity test is not yet implemented.

The next step for the user is to fill the columns that constitute the form. The last entry, alternative spellings, is to store *lapsus calami*, original spellings for family series names before the introduction of mandatory endings, or other spelling variants that do not constitute a new name. These names are not processed, but simply reprinted to provide a complete picture of the synonymy.

The software returns the valid name (Fig. 29) and calculates the other name's synonymy type when compared to the valid name. If any of these names is not invalid because of the simple application of the principle of Priority, all the relevant reasons that make these names invalid, or that deny them from their potential validity are listed. Moreover, if a name is unavailable, it will appear as so at this stage, but will still be mentioned apart, underneath the list of competing available names.

Figure 28: *Lognom*'s form for testing the validity of competing names.

Name	Month	Day Homonymy	Reversal of Priority	Related name	Case number	First Reviser	Alternative spelling
<input type="checkbox"/> -----	December	31	<input type="checkbox"/>	Not applicable	-----	Not applicable	-----
<input type="checkbox"/> -----	December	31	<input type="checkbox"/>	Not applicable	-----	Not applicable	-----
<input type="checkbox"/> -----	December	31	<input type="checkbox"/>	Not applicable	-----	Not applicable	-----

Genustres Author, 1879
Genustres Author, 1879 **Type-Species:** *Genus speciestres* Author, 1758 **Status:** This is the valid name
Genusbis Author, 1920 **Type-Species:** *Genusbis speciesbis* Author, 1758 **Status:** This name is a junior subjective synonym of the valid name
Genus Author, 1790 **Type-Species:** *Genus species* Author, 1758 **Status:** This name is a junior homonym of an older name; this name is a subjective synonym of the valid name

Figure 29: Result of a validity test for genus names, based on the example shown in Fig. 16.

Other tools

The last component of *Lognom* is the tools that solve minor nomenclatural enquiries. With the first tool, the users can check if two specific names that have the same derivation and meaning should be treated as homonyms under Art. 58. This tool is quite straightforward, as it only asks for two specific names as entries. It is not, however, a complete tool in terms of checking each of the cases listed by the Code, as it is only possible to compute part of the

situations that are put forward by Art. 58. The missing elements are listed on the page of the test. These difficult rules have in common the fact that they rely more on an interpretation of the word than on a simple computable pattern such as a doubling of consonants. Recognizing a connecting vowel in a compound name (Art. 58.12.), for example, necessitates the interpretation of the two parts of that name to find which letter should be compared to which (e.g. *nigrincinctus* & *nigrocinctus* vary by a connecting vowel, but in the imaginary specific names *leuciana* and *leucoana*, ‘i’ and ‘o’ are different enough to not fall under Art. 58.). The second tool is about the grammatical gender of genus series names. This tool (Fig. 30) supports the user in determining the grammatical gender of a genus name by guiding him or her to different possible outcomes. The answer depends on linguistic factors such as the genus name and language of origin. Sadly, it cannot provide a definitive answer in most cases, but should be able to tell the user the next step he or she should take, or what resources he or she has to reach for (Fig. 31).

The last tool consists of a test that calculates the necessary parameters to declare two names *nomen oblitum* and *nomen protectum* under 23.9., following the logic presented in Figure 17. It does not present major challenges on the programming side, as it is based on simple numerical comparisons. Art. 23.9. provides an interesting quantitative approach to the notion of usage, and the difficulty is more in finding the data than applying the rule (see IV.2.).

Figure 30: *Lognom*'s form for determining a genus name grammatical gender.

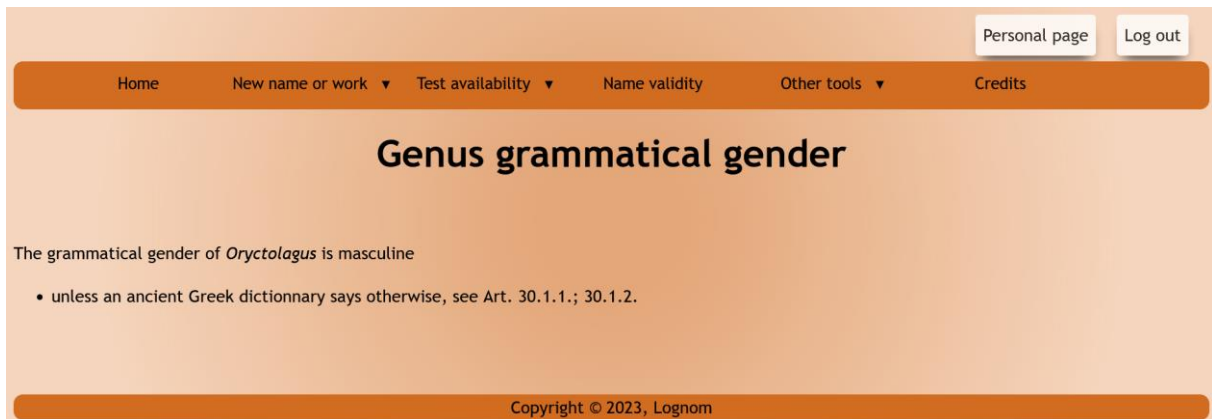


Figure 31: Result for a test determining the grammatical gender of the name *Oryctolagus*.

Statistics on the computerization of the rules

Out of the 668 rules contained in the 90 articles of the Code (recommendation excluded), 480 contain direct instructions. Currently, a bit under half (230/480) of these instructions have been computerized (see Appendix 2 for a detailed report, rule by rule). Some of them might be added eventually, such as the management of type specimens and homonymy. However, at least 25 of them present major obstacles to their full computerization. Some of them rely on concepts that are difficult to include, and thus end up being translated into the software as texts or warnings, and others rely on semantic interpretation that makes it necessary to have an intervention by a human mind to conclude the process. These elements are discussed in detail in the next chapter (see IV.1.2.).

Help and advice

Throughout the software, message blocks appear whenever the user hovers over certain parts of the screen. In the availability tests, these hints and tips are indicated by a dotted underlining throughout the forms. In the validity section, help is provided by hovering over the column headings. These indications are to be included as much as possible, as stipulated in the specifications, to minimize the user's possible errors when handling the software. The use of the Code Glossary may prove particularly useful in recalling and explicating nuances that are not always evident in the Code's text itself.

Moreover, a video tutorial should be made available to help with the first uses of the software, along with a frequently asked questions page (or two) to answer any questions users may

have. In the case of a double F.A.Q. page, one would be reserved for purely technical questions related to the software, such as how to copy a name from one project to another, and the second for the addition of specific elements to help users make decisions when filling in forms. This could take the form of examples or guidance on the interpretation of open-ended notions, such as the notion of conditionality, or the elements that make up an acceptable description. Through this page of advice and assistance in the interpretation of nomenclatural elements, the Commission could directly issue instructions to homogenize possible interpretations of fuzzy concepts.

III.3. Efficiency and limits

III.3.1. Efficiency tests

Software effectiveness tests are still ongoing. They are of two kinds. The first takes the form of testing by specialists. These testers are taxonomists with experience in nomenclature and taxonomy. They are given the task of simply entering a few cases they know about to see if the software reacts in the way they expect. Given the nature of these testers, it is more efficient to let them ‘mess around’ with the software without guiding them, as they are more likely to find mistakes and identify shortcomings. Feedback on ergonomics and requests for changes are also collected.

The second set of tests concerns users who are not nomenclature specialists. The pool of testers includes students, young researchers and scientists whose specialties make them users of taxonomy and/or nomenclature, without requiring an in-depth understanding of the nomenclatural practice and rules. This test takes the form of a comparison between the use of the software and the use of the online Code. To do this, the tester is given a sheet with seven nomenclature questions that can be solved by *Lognom* or by the Code online. These questions are in English or French, depending on which language the tester is most comfortable with. The two versions of the questionnaire can be found in Appendix 3. Each tester must answer three questions using the Code online and three questions using the software, the choice of questions being made by a randomizing algorithm.

For this category of tests, two dimensions are measured: the ability to answer questions correctly, and the time used to determine the answers. Testers are given a short explanation of the tools available in *Lognom*, as well as an explanation of how the Code online works, and its search function, as most of them have never used it before. Then it is up to them to answer the

questions. They may abandon a question if they wish, given that they spend at least 10 minutes fruitlessly searching for the answer.

The results of these tests will be communicated as soon as they are concluded. By anticipation, the following scenarios, or a combination of them, can be imagined:

1. the users obtain on average better results, in less time, by using the software, meaning that its initial objective of supporting the work of taxonomists is fulfilled;
2. the testers manage to answer the questions correctly more often with the software than with the Code online, but it takes longer to answer with the software, meaning that its ergonomics need to be improved;
3. the users make more mistakes with the software than with the Code, meaning that the wording of the elements of the software needs to be revised to make them clearer, for users less experienced with the technical jargon;
4. a particular or a handful of questions have worse results in general with the software, in which case it is crucial to understand why.

As testing progresses, comments and questions from users will continue to be collected in order to improve *Lognom*'s ergonomics and ease of use.

III.3.2. Limits and possible solutions

The process of creating *Lognom* brought its share of unknowns, mistakes and failures. With no equivalent software to be inspired by, part of the programming endeavor was based on a trial-and-error approach. As a result, there are many choices that were made during the creation of this software, which, in hindsight, were not optimal. Furthermore, The Code can become really complicated, and it is really hard to imagine that no errors went undetected in this preliminary version of the web application. The tests should find some of them, but some rare or complex situations might still go under the radar for a while. A way to signal these potential issues should be implemented, via direct contact with the software manager, and, when the software becomes more complete, move it to a GitHub issue-raising model.

In fact, the most obvious challenge for *Lognom* at this stage is to be completed. It still lacks an important part of the Code's rules, but the majority of this problem has more to do with a lack of time than with structural impossibility (see IV.1). A caveat has to be mentioned, however, for the management of homonymy. Secondary homonyms are not a problem to computerize, as they appear on the same nomenclatural files in *Lognom*. However, it is

impossible to automatize the process of finding primary homonyms without a huge number of names already in the system, or an exhaustive reference system outside of the software. It is not the rules on homonymy themselves that are difficult to transform into algorithms, but that this endeavor is pointless without a critical mass of data. So, homonyms are probably going to be put on hold, and the rules might be computerized on the side, in a library of functions, rather than directly into *Lognom*'s backend.

The objectives that were aimed for during the conception stage were to have a tool that would aim for a 100% success rate, and be easy and quick to use at the same time. It also had to answer the question of the possibility to transform the rules into algorithms. If it is not fundamentally impossible to make sure that each of these objectives is accomplished, they do not work in synergy, and favoring one complicates the achievement of the others. If an exhaustive approach to nomenclature is tried, rather than choosing to cover the most common situations, the forms get longer, to the point where they probably feel overwhelming. The same issue arises when recommendations are included, as they make the web app more relevant, but make the forms even lengthier. To palliate this issue, the choices by default will be left available, as they might make most entries quicker. Once the scientific research phase of this project is over, some fields, such as the presence in the LANs, whose relevancy is debatable²³, might be taken off to make the forms lighter.

The doubled system for nominal types stems from a clear will to make it easier for the user to use the software, even *en passant*, without having to spend an afternoon researching all the information on a type-species or name if his or her aim is to have an answer on a genus series name. The same is true for a family series name. A consequence of that choice, however, is a weakening of the general robustness* of the data. It also leads to a significant complication of the functions given that many instructions are doubled in the program, as both possible entries must be treated at each step of the availability and validity processes.

Furthermore, *Lognom* calls for people to start from scratch, which might be discouraging for a lot of potential users. There are some names and works that arguably should be integrated directly by default for all users, (Linnæus', Clarck's, and other 'old' or 'legacy' authors, as

²³ There is only on part of a LAN that was effectively adopted, the generic and species names of the Rotifera phylum (Segers et al., 2012; International Commission on Zoological Nomenclature, 2019). A strong case may be made that the names present in this LAN should be added by default in the database.

well as the names and works that have been ruled over by the Commission). This would lead to both a simplification of the forms, as the parts on actions by the Commission would be taken down and treated without any input from the user, but also provide some bricks to build on and free the users from the weight of starting from square one.

IV. Reflections on zoological nomenclature

« Ce qui fait donc que de certains esprits fins ne sont pas géomètres, c'est qu'ils ne peuvent du tout se tourner vers les principes de géométrie. Mais ce qui fait que des géomètres ne sont pas fins, c'est qu'ils ne voient pas ce qui est devant eux et qu'étant accoutumés aux principes nets et grossiers de géométrie, et à ne raisonner qu'après avoir bien vu et manié leurs principes, ils se perdent dans les choses de finesse où les principes ne se laissent pas ainsi manier. [...]

Mais les esprits faux ne sont jamais ni fins ni géomètres. »

Blaise Pascal, *Pensées*

IV.1. Computerizing the Code

IV.1.1. Terms in use in both nomenclature and programming

In programming, certain terms are considered keywords. These keywords can be functions, methods or designated entities already integrated into the language used. As it happens, at least three words used in nomenclature are similar to programming keywords. These three words are ‘class’, ‘type’ and ‘group’.

The first one, class, is both a suprafamilial rank and the keyword used to create an object in Django. Fortunately, as the software is currently limited to the ranks completely regulated by the Code, class as a rank has still to make an appearance in *Lognom*.

However, the concept of ‘type’ is used in the software, which is why it has been replaced by synonyms (onymophoront for type-specimen), when possible, or more complex constructions (type_name), so that the chances of introducing an error due to the homonymy between the keyword and the nomenclatural concept are reduced to a minimum. For the user, that change is invisible, as in the parts of the software that do not require an imperative language (Python), but declarative one (HTML), the term ‘type’ appears as expected.

For the word ‘group’, on the other hand, criticism of the word goes further than a simple homonymy between the Code and several programming languages. One argument made by Dubois (2000) is that the use of the word ‘group’, in both English and Romance languages, creates a similarity between fairly different notions such as ‘groupe espèce’ (species group) in the sense of nomenclatural series, and ‘groupe d’espèces’ (species group or group of species) in the sense of taxonomic or ecological entities, or even as a possible superspecific rank. There is also a genus series rank that has been used called group (Brunner von Wattenwyl, 1878), which further adds to the confusion.

The French version of the Code was probably written being aware of that possible mix-up and uses the word “niveau” (level). This word, however, defined by the Larousse dictionary as “une position dans une hiérarchie [a position in a hierarchy]” (Larousse, n.d.), evokes ranks more than groups, and erases implicitly the existence of several simultaneously managed ranks under the principle of Coordination. The latter being one of the most difficult concepts to grasp in the zoological Code, any terminology change that would help in its understanding should be encouraged. The term ‘series’, on the other hand, implies multiple similar entities

related to each other, making the principle of Coordination a bit more transparent in the discourse.

IV.1.2. A classification of the rules

In the process of trying to create *Lognom*, the analysis of the Code's rules (in the sense of 'mandatory instructions') and recommendations led to the distinction between different properties of these regulations. In this taxonomy of the rules, some essential characteristics made it possible to distinguish between categories that do not necessarily behave the same way when computerized. The properties are summarized by the following six elements:

- Role: the fact that a rule defines the properties of a nomenclatural object or gives instructions that modify the status of a nomenclatural object.
- Independence: the fact that a rule does or does not need another rule to be complete.
- Temporality: the fact that a rule applies differently or indifferently during different periods.
- Code clarity: the fact that the Code is able to provide a clear and computable definition of the concepts used in a rule.
- User freedom of interpretation: the fact that the rule is more or less open to interpretation by the user.
- Access to the answer: the ease with which a user can find the elements necessary to apply the rule.

Each of these properties has different states, usually two, and will be explored shortly thereafter. A summary is provided in Table 10. Different combinations of these states can equate to the same computerization process (or lack thereof), but a given computerization process can be shared by different combinations of states. It is also particularly important to distinguish between two levels of computerization. The first level is simple computerization, and is the one which takes place in *Lognom*. It seeks to create the conditions where an algorithm can help make a decision, but still depends on human input. This level can itself be subdivided into total and partial computerization. The former means that a given input generates an absolute and unquestionable answer, and that the human intervention is limited to the initial input. The latter means that the final decision is given back to the user, increasing the number of human interventions to two.

Total computerization opens the door to the prospect of the second level of computerization, automation, i.e., removing the human from the equation. With the right tools, this should lead to the creation of wide-range controls over existing data, or to fully automated management of nomenclature. In both cases, this dimension goes beyond the objectives of this work, but are important perspectives that require reflection.

If there is an optimal combination for computerization and automation, there is not one given combination that would be “the worst”, but multiple combinations that render ineffective the total transformation of a rule into a computer program.

For each of these properties, a small explanation and examples will be given. Like any taxonomy, where the exact interpretation of a character state is open to discussion and debate, some rules might be classified in another way than the one provided.

Role

The first element that allows us to classify rules is the role that the rule plays. Some rules define the properties of a nomenclatural object, for example, a name. These rules do not act on the status of these nomenclatural objects, nor give directions to the user. These rules will be called ‘defining rules’.

These rules do not have a direct role in the process of nomenclatural decision-making; however, they are necessary to understand other rules, construct the database or choose the right data to enter in the process.

Example of a defining rule

42.1. Definition

The genus group, which is next below the family group and next above the species group in the hierarchy of classification, encompasses all nominal taxa at the ranks of genus and subgenus (see also Articles 10.3 and 10.4).

The second type of rule when it comes to their roles is ‘directing rules’. Directing rules give straight instructions that pertain to the status of a nomenclatural object.

Example of a directing rule

11.2. Mandatory use of Latin alphabet

A scientific name must, when first published, have been spelled only in the 26 letters of the Latin alphabet (taken to include the letters j, k, w and y); the presence in a name when first published of diacritic and other marks, apostrophes or ligatures, or a hyphen, or a numeral in a compound species-group name, does not render the name unavailable (for corrections, see Articles 27 and 32.5.2).

Directing rules have a capital role in the process of nomenclatural decision-making.

Depending on the choices made, they can be integrated as part of the computerization as conditions, controls, or in the structure of the database. In the case of the example above, in *Lognom*, a specific control in the function ‘controlAlphabet’ checks each character against a list of the 26 permitted ones, to make sure that only Latin characters remain in the name.

The third type of rule is ‘application rules’. These rules exist only as a way to regulate or limit the scope of other rules. They provide indirect instructions that can apply to both defining rules and directing rules.

Example of an application rule

55.1. Application of the Principle of Homonymy

The Principle of Homonymy applies to all family-group names, including names of ichnotaxa at the family-group level.

In terms of computing, exactly like defining rules, application rules are necessary to understand other rules and limit or expand the field of action of a directing rule. None of these three categories make it particularly difficult *per se* to transform the rules into algorithms, but the role they play is quite different. An application rule will be translated into a condition in

the algorithms, while the defining rules might end up in a help box, or as a mention in text of the relevant information.

Independence

The second element that characterizes a rule is its necessity to call upon another rule to be understood and applied. An independent rule may however call for a definition, or for an article that specifies the application field, or articles that list mutually exclusive cases.

Example of an independent rule

11.9.1. A species-group name must be a word of two or more letters, or a compound word (see Article 11.9.5), and, if a Latin or latinized word must be, or be treated as,

11.9.1.1. an adjective or participle in the nominative singular (as in *Echinus esculentus*, *Felis marmorata*, *Seioptera vibrans*), or

11.9.1.2. a noun in the nominative singular standing in apposition to the generic name (as in *Struthio camelus*, *Cercopithecus diana*), or

11.9.1.3. a noun in the genitive case (e.g. *rosae*, *sturionis*, *thermopylarum*, *galliae*, *sanctipauli*, *sanctaehelena*, *cuvieri*, *merianae*, *smithorum*), or

11.9.1.4. an adjective used as a substantive in the genitive case and derived from the specific name of an organism with which the animal in question is associated (as in *Lernaeocera lusci*, a copepod parasitic on *Trisopterus luscus*).

In the case of Art. 11.9.1., the reference to Art. 11.9.5. links to a definition. Otherwise, the rule does not refer to or include another rule. Each of the elements described in Art. 11.9.1.1. to 11.9.1.4. are a list of mutually exclusive cases, as, for example, a name cannot simultaneously be a name in the genitive case (Art. 11.9.1.3.) and a name in apposition in the nominative singular (Art. 11.9.1.2).

Some rules are subordinated to other rules. These rules are subordinate rules and their overarching rule is a superordinate rule. Subordinate rules are usually easy to spot as they

have an additional number added to their article number compared to their superordinate rule. A rule that is subordinated can be independent if its application does not require referral to the superordinate rule. However, a superordinate rule cannot be independent, as it does not make sense on its own. All superordinate rules and part of subordinate rules are thus ‘dependent rules’.

Example of a dependent superordinate rule

10.1. General conditions to be met

A name or nomenclatural act is available, and takes authorship and date, only when it has satisfied the provisions of this Article and, when relevant, of Articles 11 to 20 (for date and author see Articles 21 and 50). A name may be ruled to be available by the Commission [Arts. 78-81] if these conditions are not fully met.

Example of a dependent subordinate rule

32.5.2.4.1. If any of the separate parts is an abbreviation of a name (or part of the name) of a place or a saint, it is to be written in full and united without any intervening mark.

This rule is subordinated to:

32.5.2.4. In a compound species-group name of which the first part consists of an abbreviation in Latin letters, or a Latin letter or a number of Latin letters qualifying the second part, whether or not separated by punctuation or a hyphen, the parts are to be united as follows.

In terms of programming, dependent rules are translated into code by the addition of a potential condition. They are less simply translatable into an instruction and might need additional controls. In the example above, it is first necessary to check if the name contains an abbreviation before applying the correction needed. However, a rule being dependent, independent, superordinate or subordinate is not a decisive element when it comes to its potential computerizability.

Temporality

This property of the rules is quite straightforward. Either a rule applies indiscriminately to all periods in the nomenclature, or it applies to only part of the seven time periods that were defined in the last chapter (see III.1.2)

If they are active across time, these rules can be called ‘permanent rules’. These permanent rules present no computerization challenge since they are integrated into the system without any conditions on the date of publication.

Example of a permanent rule

8.2. Publication may be disclaimed

A work that contains a statement to the effect that it is not issued for public and permanent scientific record, or for purposes of zoological nomenclature, is not published within the meaning of the Code.

Conversely, ‘time-specific rules’ must be integrated with a date condition. If this condition is not met, the rule does not apply. This does not constitute a complex case to program, as it takes the form of a simple condition based on the name’s date of publication.

Example of a time-specific rule

16.1. All names: intention of authors to establish new nominal taxa to be explicit

Every new name published after 1999, including new replacement names (nomina nova), must be explicitly indicated as intentionally new.

However, some time-specific rules, such as 11.6.1., are not linked to the date of publication, but give a certain time frame in which the name could have been used as available, valid, or rejected. These rules are not easy to program, not because they are time-constrained, but because they rely heavily on user research (see Access to the answer).

Example of a time-specific rule with a time frame

11.6. Publication as a synonym

A name which when first published in an available work was treated as a junior synonym of a name then used as valid is not thereby made available.

11.6.1. However, if such a name published as a junior synonym had been treated before 1961 as an available name and either adopted as the name of a taxon or treated as a senior homonym, it is made available thereby but dates from its first publication as a synonym

Code clarity

The notion of Code clarity covers the use of concepts that are more or less defined in the Code. Some rules are based on clear concepts, and these rules can be qualified as ‘simple’. Others rely on concepts that are difficult to apprehend, even for the human mind. These rules are called ‘complex’.

Simple rules can be computerized and have the potential to be fully automated. It is possible, for example, to imagine the Principle of Binominality verification being entrusted to an algorithm responsible for identifying whether a species name is indeed in two names, that of a subspecies in three, and that a name including a subgenus does indeed have one more element than expected for a species or a subspecies. These rules, in general, have a way of being reduced to a quantitative calculation of some sort, or a pretty clear decision tree as shown for Art. 23.9. (see Fig. 17).

Example of a simple rule

23.9. Reversal of precedence

In accordance with the purpose of the Principle of Priority, its application is moderated as follows:

23.9.1. prevailing usage must be maintained when the following conditions are both met:

23.9.1.1. the senior synonym or homonym has not been used as a valid name after 1899, and

23.9.1.2. the junior synonym or homonym has been used for a particular taxon, as its presumed valid name, in at least 25 works, published by at least 10 authors in the immediately preceding 50 years and encompassing a span of not less than 10 years.

Conversely, a complex rule requires interpretation, because the Code is missing a very detailed definition of a concept it relies on. Such a rule cannot be totally computerized or automated and must be interpreted by a human in order to be applied. These rules' outcomes are the most potentially open to debate, and disagreement may arise as to their exact application. In addition to the example given, the concept of usage, or notions such as the identifiability of a type-specimen or its accessibility, are also complex rules and recommendations as they are open to interpretation.

Example of a complex rule

15.1. Conditional proposal

A new name or nomenclatural act proposed conditionally and published after 1960 is not thereby made available. A new name or nomenclatural act proposed conditionally and published before 1961 may be available [...].

In this example, the concept of conditionality is difficult, if not impossible, to translate into elements that would be comprehensible by a computer. Even for humans, in certain cases, conditionality may be complicated to handle and define.

In some cases, the Code chooses to introduce rules that contradict some of its principles, such as with Art. 35.5., where both Priority and Coordination are subdued by the intervention of Usage.

35.5. Precedence for names in use at higher rank

If after 1999 a name in use for a family-group taxon (e.g. for a subfamily) is found to be older than a name in prevailing usage for a taxon at higher rank in the same family-group taxon (e.g. for the family within which the older name is the name of a subfamily) the older name is not to displace the younger name.

This ultimately weakens the very notion of the automaticity of rules, not in the sense of their computability, but in the sense of their logical interpretation. Weakening two of the Code's core concepts with a vague notion ends up undoing part of the usefulness of these principles. In that particular case, someone might see in a paper or a database a mistake regarding the principle of Coordination, and ask for a correction, only to be told that it is because that choice reflects the prevailing usage at a certain moment, which can be difficult to establish or verify, and can even change with time.

User freedom of interpretation

The fifth dimension that is useful to analyze the Code's rules is the freedom of interpretation given to the user of the rules. This dimension is mostly outside the scope of computerization, at least as defined for *Lognom*, since the step at play is the data extraction from the literature, common practices or type-specimens rather than interpreting the Code's formulations. This dimension is better represented by a spectrum compared to discrete states. The rules that leave a lot of freedom to the user mostly coincide with complex rules. However, the opposite is not always true. Some rules might be very straightforward and rely on well-defined concepts, but end up leaving a lot of space for interpretation.

Example of a simple rule that leaves a lot of freedom to the user

35.1. Definition

The family group encompasses all nominal taxa at the ranks of superfamily, family, subfamily, tribe, subtribe, and any other rank below superfamily and above genus that may be desired

If too much freedom is left to the user, it will be difficult to fully automatize the Code. In the case of the rule presented in the example, it makes it impossible for the software to anticipate all possible ranks that could constitute the family series (see IV.5.).

Access to the answer

The final dimension for classifying nomenclature rules is the difficulty of finding information to comply with the rules.

Some rules only require consultation of the original work in which the name or the nomenclatural act is introduced. This situation is the most widespread and simplest, and explains the importance given to the original publication. These rules can be called ‘directly accessible rules’.

Example of a directly accessible rule

8.4.2. Works on optical disc

To be considered published, a work on optical disc must be issued, in read-only memory form, after 1985 and before 2013, and

8.4.2.1. if issued before 2000, must contain a statement that any new name or nomenclatural act within it is intended for public and permanent scientific record and that the work is produced in an edition containing simultaneously obtainable copies [...]

However, there are two other levels of complexity for this dimension. Some rules, that can be named simply ‘indirectly accessible rules’, might call for the consultation of one or more additional documents. Spelling rules, for example, request the use of dictionaries. The application of the principle of First Reviser is also an example of a generalized recourse to a second document. Indications and decisions by the Commission are other examples that will require the use of other documents. The determination of the date of publication is another case of a less accessible rule. This approach makes it extremely difficult to fully automate rules but poses no particular difficulty for the human brain as long as these other sources are accessible.

Example of an indirectly accessible rule

10.7. Availability of names not listed in a relevant adopted Part of the *List of Available Names in Zoology*

No unlisted name within the scope of an adopted Part of the List of Available Names in Zoology is available, despite any previous availability [Art. 79.4.3].

The final category of rules in this category, the ‘inaccessible rules’, are those that call for the verification of potentially all of the relevant literature, in order to find proof of a consensus. These rules are extremely complex, both for the human and the machine, since no one can claim to have access to every archive in taxonomy. The question of usage is, of course, the first to come to mind, in all its forms and application scope.

Example of an inaccessible rule

29.3. Determination of stem in names of type genera

29.3.1.1. If the stem so formed ends in -id, those letters may be elided before adding the family-group suffixes. If, however, the unelided form is in prevailing usage, that spelling is to be maintained, whether or not it is the original spelling.

Table 10: Classification of the rules of the Code with the consequences on its potential computerization.

Property	State	Consequence of the computerization
Role	Defining rule	Defines the properties of nomenclatural objects
	Directing rule	Gives the instruction that the functions will enforce
	Application rule	Structures directing rules by providing a precise application scope
Independence	Independent	None
	Superordinate	Provides condition for the subordinate instructions
	Subordinate	Requires the validation of certain conditions to be applied
Temporality	Permanent	None
	Time-specific	Creates a condition that depends on the publication date
Code clarity	Simple	Has the potential to be fully automated
	Complex	Necessitates human intervention
User freedom of interpretation	Limited	Has the potential to be fully automated
	More or less important	Necessitates human intervention
Access to the answer	Directly accessible	Has the potential to be fully automated
	Indirectly accessible	Is difficult to automatize
	Inaccessible	Is impossible to automatize

IV.1.3. Suggestions for a more computer-friendly Code

There are many improvements that the Code could incorporate to make it more amenable to computerization. It is capital, however, to bear in mind that the Code's primary public are human beings, with human intellectual capacities, and not machines. The aim should not become a voluntary 'dumbing down' of the Code in order to satisfy computers. The driving force behind nomenclature should always be based on respecting the scientificity of the system and accommodating the practical and theoretical requirements of the taxonomic community.

With this warning in mind, it is still possible to introduce some improvements to the Code to enable a better furthering in the computerization of the nomenclatural process.

The first recommendation, as mentioned in Chapter 1 about grammatical agreement, is to provide a solution for situations that become too complex for even human decision-making. As a matter of fact, if a human is unable to make a decision given the data at his or her disposal, a computer will not be able to make a choice either, since it is impossible to provide it with the elements needed to decide on this complex situation. In the case of *Lognom*, that would happen if the user is incapable of knowing whether to tick a box or which value to choose in a dropping list. A possible solution would be to provide some conditions to ensure that a situation is indeed inextricable, and then offer the possibility of modifying the *status quo* via a First Reviser act.

This answer, however, would not work for rules that rely on finding historical elements, such as the use of a name before or after a given date. There are some solutions to this particular case, some of them being easier to apply than others. Taxonomists can choose to list such names, spellings and acts from one of these periods that they wish to see conserved, and after a certain point in time, exclude any nomenclatural object that is not on these lists. That might entail an authoritarian fixation of comprehensive lists of the status of various nomenclatural objects, for example, all infrasubspecific names that should be conserved under Art. 45.6.4.1., with the willful exclusion of nomenclatural objects that are not included at the time of adoption of these lists. This approach is reminiscent of the LANs idea but on a way smaller scale that might make it more viable. From a programming point of view, that would lead to a simplification of the process but entail a greater complexity in the management of data. These lists will form shareable datasets that need to be added to open databases and software such as *Lognom* to make sure that the rules are correctly applied. However, in the long run, it would ensure greater stability as an upheaval in the statuses of these nomenclatural objects will become less likely.

Another solution would be to never introduce these complex or inaccessible rules again, and simply wait out the problem and hope that eventually, all literature will become easily available. The issue with this solution is obviously that there is no guarantee that it will ever happen. The last, more radical one, would be to simply abandon these rules, but that would constitute an unjust *a posteriori* massive sidelining of names, which is a dangerous precedent.

Moreover, it is contrary to an implicit but capital part of the ethos of the Code that wishes to respect as much as possible the work done by predecessors, and keep the disruption of this system to a minimum. The use of exceptions is understandable, as respect for past colleagues and the refusal (or minimization) of *a posteriori* modification of the rules with retroactive consequence play a major role in the social acceptability of the Code. However, every time the Code introduces exceptional situations, it complicates the decision-making process, adding branches to the decision-making tree presented before. If priority is given to a simplification of the Code, it is essential to reduce exceptions to the minimum possible.

A second recommendation would be to define or abandon concepts that are not easily quantifiable. The concept that comes to mind here is obviously predominant usage. As demonstrated in Chapter 2, databases do not necessarily provide a clear image of what the real usage is, and taxonomists might be unable to easily determine this usage, when the gap between the concurrent names, spellings or allocations starts to be too narrow and thus complex to manage. In the same vein as the situation described above, if the human brain cannot process this information, it will be extremely difficult, if not impossible, to inculcate in a computer. Even when providing a clearer, numbered, and quantifiable framework, such as with Art. 23.9., the problem mentioned above reappears, a usage relies on having access to all the literature at a given moment. That issue reappears with names above the superfamily, and rules will be necessary eventually, for the nomenclatural system of the Code to stay cohesive and relevant to the community.

In all cases, there is no doubt that in order to fully computerize, simplify, and implement the rules of Code in an algorithm, a deep reflection on how nomenclature can be translated into a program is necessary. In this work, the approach that is provided is in the form of a web app that produces status-determining calculations. Other approaches are possible, like producing controls to identify and filter errors and incoherences in already existing databases or creating a library of functions that might answer more punctual questions. Fortunately, as it was made in with open-science context in mind, part of the algorithms developed for *Lognom* can be reused and would make it easier to implement these other solutions.

This being said, the task of computerizing and managing nomenclature cannot be undertaken without in-depth theoretical reflections on the Code. What is the best way to represent nomenclatural concepts? Is it possible to limit the uncertainty permitted by the Code without

impeding taxonomic freedom? Should the access and generalization of new technologies and efficient structures in biodiversity informatics lead to the modification of our approach to some traditional concepts? These reflections on both the causes and consequences of decisions taken by the Commission, the Code, and the taxonomic community have implications that potentially go beyond the strict framework of science and the rational discourse on nomenclature. They need to include social considerations and imply a real reconsideration of the relationship taxonomists have with nomenclature.

IV.2. Protaxa and apotaxa

In the course of the creation of the software, an interesting element (that was briefly discussed in Chapter 3) appeared. The kind of relationship that links a name to a taxon can be both 1-1 (one to one; one name is applied to only one taxon; one taxon has one and only one name), and N-1 (many to one; one name is applied to only one taxon; one taxon can have multiple names applied to it). This fact is not trivial from a programming perspective, but it is not trivial in the context of onymology* either.

Remsen (2016) started an analysis of that cardinality dimension and proposed the following classification (Table 11, 3 first columns):

Table 11: Database cardinalities between names and taxa, modified after Remsen (2016).

Cardinality	Abbreviation	Example	In <i>Lognom</i>
One to One	1-1	Stable taxon	Protaxon
Many-to-One	N-1	Synonyms	Apotaxon
One-to-Many	1-N	Homonyms / Polysemes	N/A
Many-to-Many	N-N	Taxon Concept	N/A

It is important to note that for Remsen (2016) synonyms contain recombinations, orthographic variants and synonyms as classically understood.

In the case of a 1-1 link, the name is in fact linked to a protaxon (Dubois, 2005), which corresponds to the taxon when first introduced by an author, and opposed to an apotaxon, which is a subsequent use of a taxon (Dubois, 2005). An ergotaxon is a particular case of apotaxon, as it is the accepted taxon in a given complete taxonomy. Remsen's stable taxa (which are not clearly defined in his article) are most likely protaxa in their large majority, or

apotaxa, when their taxonomic definition has not moved, nor their placement in the classification (see Bourgoïn et al. [2021], for an explanation about the importance of the place of a taxon in a classification). This taxon cannot have more than the name it was introduced with applied to it. Even in the case of multiple original spellings, there is only one name present, just different forms of that one name, at least with the way nomenclature is modeled in *Lognom*. That explains why there is a strict 1-1 relation in *Lognom*'s database when it comes to the taxon at the availability stage.

However, when the protaxon becomes an apotaxon, and if its definition or classification varies, the potential for multiple names to be applied to it appears. These names are synonyms (sensu Remsen [2016]; with the exception that in *Lognom*, orthographic variations [except unjustified emendations] do not constitute independent instances of names), and this situation corresponds to the second line of Table 11. That introduces an N-1 relation, and is visible in the N-1 relation in *Lognom*'s database when it comes to the ergotaxon used at the validity stage. If the taxon at the validity stage has only one name, it is an exceptional case of a 1-1 relation that corresponds to Remsen's stable taxon,

Lognom does not manage homonyms in the way suggested by Remsen (2016). Two homonyms in *Lognom* will correspond to two different entries in the Name table, which will each have its own protaxon (1-1), and act like other names for apotaxa (N-1). As a matter of fact, this relationship (1-N) should not appear in a purely nomenclatural database. The same goes for the taxon concept or polysemes, as the 1-N or N-N relationship is explained by the management of multiple taxonomies or meanings of a given taxon. These two cases are specific to taxonomic databases. It is crucial to be aware of these structural difference between nomenclatural databases (such as *ZooBank* or *Lognom*'s database) and taxonomic databases (such as FLOW), to manage their data in an optimal way. These differences also reflect the two different realities constituted by nominal and taxonomic taxa.

IV.3. Nominal and taxonomic taxa

One of the most complex elements of zoological nomenclature to handle is the concept of the nominal taxon. The Code defines this concept as:

Nominal taxon: A concept of a taxon which is denoted by an available name (e.g. Mollusca, Diptera, Bovidae, *Papilio*, *Homo sapiens*). Each nominal taxon in the

family, genus or species groups is based on a name-bearing type (although in the latter two groups such a type may not have been actually fixed).

The nominal taxon is, in sum, a name and if relevant, its type. It would be interesting to add to this definition the link between the name and the type, as a way to complete the picture, but also to better model it. This leaves us with the fact that two kinds of nominal taxa exist, i.e., the name, the type specimen, and the link between the two for a species series nominal taxon, and a type nominal taxon, a name and the link between the two for a genus and family series nominal taxon. One way of formalizing a *nomen dubium* is by considering that it is impossible to identify the nominal taxon to which the name belongs, whether because the link between the name and the type is unclear, or because the type itself is lost or defective. That leads to cutting off the relationship between the name and the physical reality of the world.

Taxonomic taxa, on the other hand, are what most taxonomists understand when they use the word taxon. The common link between taxonomic taxa and nomenclatural taxa is the name-bearing type. Both these taxa can evolve completely independently otherwise (Fig. 32). Two nominal taxa that are deemed to apply to the same taxonomic taxon are considered synonyms, either objective if they share a type (Fig. 32: A & C), or subjective if they do not (Fig. 32: B & D). Two taxonomic taxa can be syntaxic (Dubois, 2005) if they share their circumscription (Fig. 32: A & B), either because they have the same definition or because one's definition is contained in another's, or they can partially overlap (Fig. 32: C), or not at all (Fig. 32: D).

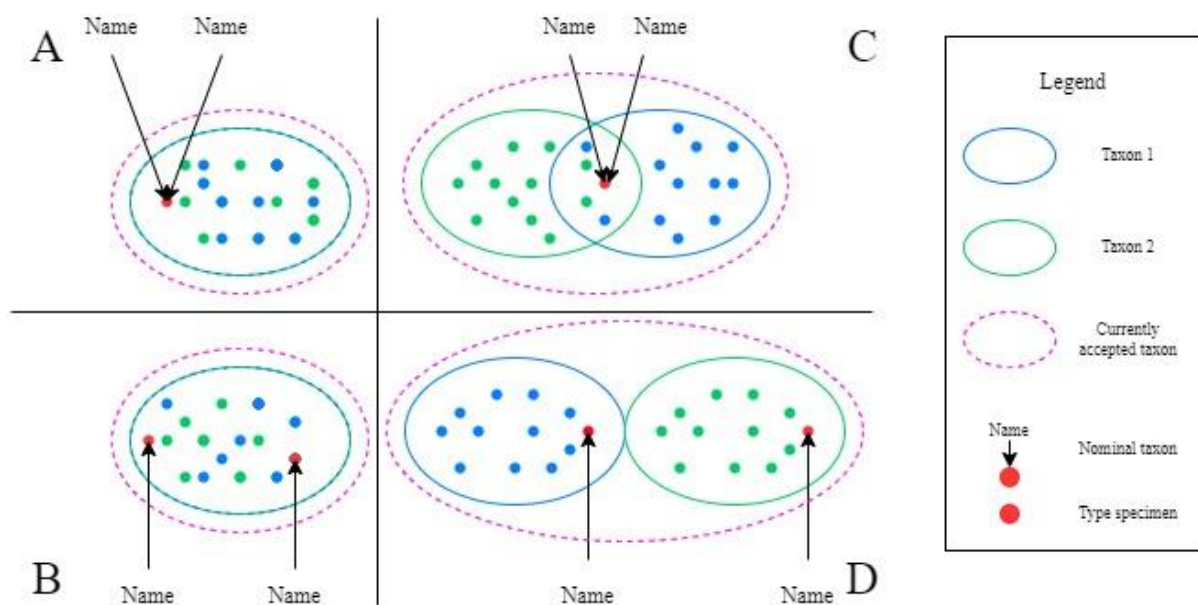


Figure 32: Nominal and taxonomic taxa at the species rank.

This differentiation between taxonomic and nominal taxa also shows why compulsory ranks partly force the hand of scientists. A monotypic taxon actually corresponds to two nominal taxa related to a single taxonomic taxon (Fig. 33). It is, therefore, necessary to artificially create an additional definition (indicated by the blue dotted lines in Fig. 33), without adding to the extension of the taxon, to respect the Code's requirements.

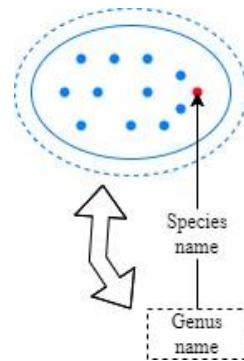


Figure 33: A monotypic taxon and its corresponding nominal taxa.

It is possible to represent both types of taxa in a database, and that distinction has been discussed (see Berendsohn [1995]). As previously demonstrated, they have different properties and can interact in many different ways. It is therefore necessary to separate their existence in databases, but also to be aware of the different elements that relate to each dimension (Table 12).

Table 12: Vocabulary differences between taxonomic and nominal taxa.

	Taxonomic taxon	Nominal taxon
The first form of the taxon	Protaxon	Protonym (sensu Pyle [2004] or Dubois [2000])
Subsequent use of the taxon	Apotaxon	Chresonym* ²⁴
- With a different definition than the original	Taxon <i>sensu</i> Author (polysemes)	Emendations, multiple spellings
- Potential taxon (Berendsohn, 1995)	Taxon <i>secundum</i> Author (Berendsohn, 1995)	No name available or use of insertions such as cf.
Accepted use of the taxon	Ergotaxon (Dubois, 2005)	Valid name and corresponding type

²⁴For some reason that I have not been able to determine, the Inventaire National du Patrimoine Naturel (INPN) considers chresonyms to be variations in the authority of a name (INPN, n.d.), rather than its original definition. That kind of definition drift happens quite often in taxonomy. 'Hemihomonyms', for example, have been defined as names of suprafamilial ranks resembling names of genera that are not included in these taxa (Starobogatov, 1991), but they currently also refer to inter-series or inter-Code homonyms (Shipunov, 2011).

Lognom's database contains only nomenclatural information, and the taxonomic data required to run the tests. This distinguishes this database and its structure from those of most taxonomic databases, which represent taxon-centric data. In the case of nominal taxa, the definition of the taxon will be given by the type specimen, not the taxonomic definition. This approach is supported by Pyle et al. (2022), and can be called type-centric. This type-centric method gives a powerful tool to track the evolution of a name, through tools such as the Taxon Name Usage* (TNU), but does not include the tracking of taxonomic elements, such as a taxon's place in the classification. The alternative approach is the one favored by most taxonomic databases, and theorized by Bourgoïn et al. (2019), where nomenclatural information is subordinated to taxonomy, at the risk of sacrificing some of the quality of the nomenclatural information. There is no doubt that, to represent efficiently both types of information, it is necessary to separate the taxonomic and nomenclatural dimensions in different databases, or at least to treat them separately in two independent levels in the same database, as managed, for example, in *Index Dipterorum* (Pape & Evenhuis, 2023).

IV.4. Management of *nomina dubia* and unavailable names

In the era of databases, the management of unavailable names and *nomina dubia* needs to be at least partially reformed.

Unavailable names are not normally part of nomenclature. They are not recognized as belonging to it. The difficulty lies in the fact that there are several types of unavailable names, with different scientific values, but only two will be of interest thereafter.

The first category of unavailable names is those that are not intended to enter a scientific nomenclature system, because, for example, they are simply listed, with no information attached. Examples of this type of names are names in indexes as meant in Art. 11.4.3., which carry no associated information or a clear taxon to which the name is allocated.

The second category of unavailable names is the one that deserves more interest than what it is getting today. These are names that are unavailable because they do not comply with one or more articles of the Code despite the fact that they are unambiguously linked to a taxon, and are associated to valuable taxonomic or biological information. This excludes *de facto* both

nomina dubia (if unavailable²⁵) and *nomina nuda*, but comprises, for example, a name considered unavailable because, while being introduced after 1999, lacks the clear intent of the introduction of a new name (Art. 16.1.), or a name published online, without the required registration in *ZooBank* (Art. 8.5.3.).

The primary aim of nomenclature being to be able to share information about taxa and not lose it in a complex and increasingly large and diverse literature system, it is quite sensible to include these names in our catalogues. Otherwise, the longer time passes, the greater the chance will be to lose all that data in the mists of time. In the age of digitization, imagining that repositories take up the charge to include at least these valuable unavailable names is not a stretch, but must be set up with clear directions. Unavailable names must be clearly indicated as unavailable, and must not be included graphically in synonymies, as they cannot compete to become the validating name. They might, however, be presented in chresonymic repertories, while still indicating that the name is unavailable.

An objection to this approach is the fact that it is not always easy to determine what constitutes valuable information. On the one hand, an exhaustive approach could lead to citing too many names and complicating the work of both database managers and taxonomists, as digging up these forgotten names is not an easy task. On the other hand, a minimal approach could lead to losing valuable information, but also missing the opportunity to find a possible compromise solution for the electronically published names that are unavailable. A possible solution would be to recommend adding valuable unavailable names to synonymies during a taxonomic revision, and trusting taxonomists in recognizing the names that deserve to join their available counterparts. For recent names, especially if the author is still active, making these names available in a subsequent publication, with a reference to the ‘original publication’ should be strongly considered²⁶.

Nomina dubia, when available, can become a real headache in some groups, and in many instances, are treated like unavailable names, cast aside and ignored. *Nomina dubia* differ from unavailable names by the fact that they represent more of a taxonomic issue than a pure

²⁵ *Nomina dubia* are technically available names. What I mean by using that expression here is that some unavailable can nonetheless be correctly allocated.

²⁶ It would not be absurd to adopt the Name author *ex. author* notation used in botany to solve the question of unavailable names whose authors are no longer active. This would probably increase the social acceptability of making them available at a later date, with a different author.

nomenclatural one. Strictly speaking, a *nomen dubium* is the name part of a nominal taxon that lacks a clear link to a type (or types) belonging to a single given taxonomic taxon, meaning that the nominal taxon-taxonomic taxon equivalency is absent.

There are, in fact, three different categories of *nomina dubia*:

- A *nomen dubium* that has no current designated type, whether the type is lost or is known to exist somewhere but is inaccessible. Dubois (2011b) calls these *nomina dubia* ‘anaptonyms’.
- A *nomen dubium* allocated to syntypes which are thought to be specimens of different taxa. These are called ‘synaptonyms’ by Dubois (2011b).
- A *nomen dubium* that has a type-specimen that has been damaged or is unidentifiable. This last category of name is called ‘nyctonym’ by Dubois (2011b).

In almost all cases, being a *nomen dubium* is a transitory state for a name, that can be solved with a neotype designation or a lectotypification. However, one of the more complex cases is when the description associated with a *nomen dubium* is not clear enough to determine to which taxon the name is supposed to be applied to.

The issue with these *nomina dubia*, is that the more time passes since the author’s era, the more difficult it becomes to determine what they might possibly have referred to, unless some additional information appears, through the rediscovery of a book by said author or a contemporary colleague. They exist in a sort of nomenclatural limbo, and complicate the work of taxonomists and database curators. Furthermore, the more time passes, the more threatening to stability the management of *nomina dubia* becomes. In many cases, the use of a senior ex-*nomen dubium* has no impact overall on the stability of the name, if only a handful of people would be affected by that change, but the longer the name stays a *nomen dubium*, the more it becomes complicated to reinstate it as a valid name.

If taxonomists decide to get rid of complex *nomina dubia* in the group they are a specialist in, the traditional elegant method is to put them in synonymies where priority would make them a junior objective synonym of the currently valid name. In the case where that is not possible, because the *nomen dubium* is the senior name, it can be declared a *nomen oblitum*, under Art. 23.9., or submitted to the Commission for suppression or reversal of precedence. However, that would create a lot of confusion in the long run, and transfer the difficulty of management of *nomina dubia* to the difficulty of management of *nomina oblita* and reversal of precedence, in addition to creating a heavy workload for the Commission for a minimal benefit.

A *nomen dubium* can also be recycled in the description of a new taxon. In many groups, high number of new taxa are described, and the (re)use of these names stuck in a limbo would lighten the synonymy load (an example is the long discussed *nomen dubium Ixalus lateralis* Anderson, 1871 that has been resolved by a neotype designation in Humtsoe et al. [2008]). A more radical, and potentially controversial, approach would be to declare names unavailable if considered a *nomen dubium* for more than a certain span of time (e.g., 100 years). A softer similar approach would be to introduce a notion of permanent loss of potential validity, like a junior homonym, but that should be accompanied by a clear framing of when a *nomen dubium* should be declared permanently invalid, and unable to compete for seniority.

In all cases, a solution that would put the management of *nomina dubia* under the direct responsibility of specialists of the relevant group is to be preferred, as they have the best understanding of the potential allocation of the name. It is nonetheless necessary but to clearly encourage their disappearance by using them and reallocating them.

IV.5. The management of the rank system

A recurrent critique addressed to the Code is the fact that it limits its scope for many of its rules to the names in between the subspecies and the superfamily ranks (see. Rowe & Gauthier [1992]; Dubois [2004]), when it is not the complete concept of ranked nomenclature that is put under scrutiny (Minelli, 2000; Bertrand et al., 2006), or flat out deemed obsolete (Laurin, 2023). However, this classical issue surrounding the ranked system is not the one that will be of interest in this document, neither will the issue of finding a common definition for these concepts.

The subject of interest will be the absence of a reference for the introduction of different ranks. In other words, aside from the main ranks defined by the Code, there is total freedom to define as many ranks as deemed necessary by a taxonomist, especially starting in the family series and above. An issue that might arise is the fact that a given rank name is used by different taxonomists at different places in the hierarchy, leading to possible confusion when trying to configure or populate a database²⁷.

²⁷ It is noteworthy that many databases choose not to take into account even regulated ranks such as the tribe or subtribe. The reasons mentioned are a lack of time, resources, or an explicit disinterest in providing such information. As an example, the World Spider Catalog explains that the team running it “has limited financial

A starting point to illustrate that reflection will be the ranks listed by the Wikipedia page on taxon ranks ('Taxonomic rank', 2023). All the ranks listed were then searched for, and, when found in the literature, compiled below (Table 13), with an example of use. Moreover, ranks that were found in the bibliography while trying to fill this first table were also added. There is no doubt that this survey of the ranks is not exhaustive, but it is sufficient to visualize the main issues.

Table 13: List of ranks in use in zoological nomenclature.

SMALL CAPITAL: key rank; **Bold**: rank that is repeated in the hierarchy; Underlined: rank which position in the hierarchy is unclear; *Italics*: mandatory rank

†A point can be made, following Erna Aeschl's position in Opinion 2420 (International Commission on Zoological Nomenclature, 2018) "that the taxonomic terms 'section', 'division' and 'superspecies' are given in quotation marks [in the Code], because they are not nomenclatural terms" and that subgenus is the only possible rank in the genus series, other than the genus rank.

resources and these are more or less sufficient to cover the present running costs. Including and adapting names of the family- and genus-group other than genus and family names is beyond [its] current capability. In addition, the World Spider Catalog is a tool for taxonomy, not for phylogenetics of spiders [...]. Finally, [the team] feel[s] that creating too many new superspecific names in the age of highly dynamic phylogenies brings more confusion than stability into spider taxonomy [...] (Spider Web Catalog, n.d.).

Rank	Example	Source or example of use
<i>KINGDOM</i>	Animalia Linnæus, 1758	Classical rank
Subkingdom	Adictyozoa Cavalier-Smith, 1993	Cavalier-Smith (1993)
Infrakingdom	Euglenozoa Cavalier-Smith, 1981	Cavalier-Smith (1993)
Parvkingdom	Ciliomyxa Cavalier-Smith, 1993	Cavalier-Smith (1993)
Superphylum	Opalomyxa Cavalier-Smith, 1993	Cavalier-Smith (1993)
<i>PHYLUM</i>	Mollusca Linnæus, 1758	Classical rank
Subphylum	Pseudociliata Cavalier-Smith, 1993	Cavalier-Smith (1993)
Infraphylum	Sporozoa Leuckart, 1879	Cavalier-Smith (1993)
Parvphylum	Chondrichthyes Huxley, 1880	Accepted by Catalogue of Life (Bánki et al., 2023)
Microphylum	Not found	
Gigaclass	Actinopterygii Klein, 1885	Accepted by Catalogue of Life (Bánki et al., 2023)
Megaclass	Tetrapoda [Authorship debated, see Sues (2019)]	Accepted by Catalogue of Life (Bánki et al., 2023)
Superclass	Tetrapoda [Authorship debated, see Sues (2019)]	Guyer & Slowinski (1991)
<i>CLASS</i>	Cricoconarida Fisher, 1962	Classical rank
Subclass	Diapsida Osborn, 1903	Osborn (1903)
Infraclass	Symmetrodonta Simpson, 1925	Averianov & Lopatin (2011)
Subterclass	Acteonimorpha Schrödl, 2017	Bouchet (2018)
Parvclass	Ornithurae Haeckel, 1866	Livezey & Zusi (2007)
Superdivision	Chondrostei Müller, 1844	Benton (2005)
DIVISION	Ginglymody Cope, 1871	Benton (2005)
Subdivision	Eurypterygii Rosen, 1973	Benton (2005)
Infradivision	Osteoglossomorpha Greenwood et al., 1966	Benton (2005)
Superlegion	Prototheria Gill, 1872	Averianov & Lopatin (2011)
LEGION	Australosphenida Luo et al., 2001	Averianov & Lopatin (2011)

Sublegion	Boreosphenida Luo, Cifelli & Kielan-Jaworowska 2001	Benton (2005)
Infralegion	Tribosphenida McKenna, 1975	McKenna & Bell (1997)
<u>Supercohort</u>	Theria Parker & Haswell, 1897	McKenna & Bell (1997)
<u>COHORT</u>	Marsupialia Jameson, 1818	McKenna & Bell (1997)
<u>Subcohort</u>	Carinatae Merrem, 1813	Benton (2005)
<u>Infracohort</u>	Neognati Benton, 2005	Benton (2005)
DIVISION	Terrestrognathes Livezey & Zusi, 2007	Livezey & Zusi (2007)
Subdivision	Dendrognathes (Verheyen, 1961)	Livezey & Zusi (2007)
SECTION	Anomalognathes Garrod, 1874	Livezey & Zusi (2007)
Subsection	Trogones Livezey & Zusi, 2007	Livezey & Zusi (2007)
Gigaorder	Casichelydia Gaffney, 1975	Gaffney et al. (1987)
<u>Megaorder</u>	Cryptodira Gray, 1825	Kear (2003)
<u>Magnorder</u>	Ameridelphia Szalay, 1982	McKenna & Bell (1997)
<u>Grandorder</u>	Dasyuromorphia Gill, 1872	McKenna & Bell (1997)
<u>Capaxorder</u>	Kayentachelydia Gaffney et al., 1987	Gaffney et al. (1987)
<u>Mirorder</u>	Altungulata Prothero & Schoch, 1989	McKenna & Bell (1997)
<u>Hyperorder</u>	Not Found	Listed on Wikipedia
Superorder	Ichthyopterygia Owen, 1840	Kear (2003)
SERIES	Amniota Haeckel, 1866	Benton (2005)
<i>ORDER</i>	Characiformes Goodrich, 1909	Classical rank
Parvorder	Genasauria Sereno, 1986	Sereno (1986)
Nanorder	Thyreophora Nopcsa, 1915	Sereno (1986)
Hypoorder	Euornithopoda Sereno, 1986	Sereno (1986)
Minorder	Euryopoda Sereno, 1986	Sereno (1986)
Suborder	Casichelydia Gaffney, 1975	Bour & Dubois (1986)
Infraorder	Anguimorpha Fürbringer, 1900	Kear (2003)
Parvorder	Anthropoidea Mivart, 1864	McKenna & Bell (1997)
Minorder	Paracryptodira Gaffney, 1975	Bour & Dubois (1986)
Micrororder	Polycryptodira Gaffney, 1804	Bour & Dubois (1986)

Nanorder	Urocryptodira Bour & Dubois, 1986	Bour & Dubois (1986)
<i>Gigafamily</i>	Dryomorpha Sereno, 1986	Sereno (1986)
<i>Megafamily</i>	Dryosauroidea Milner & Norman, 1984	Sereno (1986)
<i>Grandfamily</i>	Styracosterna Sereno, 1986	Sereno (1986)
<i>Hyperfamily</i>	Chelonioides Oppel, 1811	Bour & Dubois (1986)
Superfamily	Vespoidea Latreille, 1802	Classical rank
Epifamily	Testudinoidae Batsch, 1788	Bour & Dubois (1986)
<i>FAMILY</i>	Apidae Latreille, 1802	Classical rank
Subfamily	Batagurinae Gray, 1869	Classical rank
Infrafamily	Geoemydinei Theobald, 1868	Bour & Dubois (1986)
Supertribe	Euceriti Latreille, 1802	Engel (2005)
TRIBE	Meliponini Lepeletier, 1836	Classical rank
Subtribe	Lebintina Robillard, 2004	Classical rank
Infratribe	Drosophiliti Rondani, 1856	Grimaldi (1990)
<i>Group (genus)</i>	Turpiliae Brunner von Wattenwyl, 1878	Brunner von Wattenwyl (1878)
<i>Supergen</i>	Bolitoglossa Duméril, Bibron, and Duméril, 1854	Lombard & Wake (1986)
<i>GENUS</i>	<i>Acheta</i> Linnæus, 1758	Classical rank
Subgenus	<i>Equus Hippotigris</i> C. H. Smith, 1841	Classical rank
SECTION	<i>Crinocheta</i> Legrand, 1946	Art. 10.4. of the Code†
DIVISION	Not found as an explicit rank	Art. 10.4. of the Code†
Superspecies	<i>Ornithoptera</i> (superspecies <i>priamus</i>) (Linnæus, 1758)	Art. 6.2. of the Code†. See also ultraspecies, Artenkries and semispecies as discussed in Bernardi (1957).
<i>SPECIES</i>	<i>Telicota augias</i> Linnæus, 1763	Classical rank
Subspecies	<i>Panthera tigris sondaica</i> (Temminck, 1844)	Classical rank
Variety, form, and other infrasubspecific ranks		Classical ranks (now excluded by the Code)

Some ranks appear at different levels of the hierarchy (in bold), such as Parvorder, used as both including and included in a suborder. Other ranks exist at a similar level, but it is not possible to know how to hierarchize them (underlined), as they are introduced in different taxonomies.

Dubois (2006) provides an in-depth analysis of the subject, but since that paper, many new ranks emerged, deepening the complexity of the question. It would be sensible, following in part Dubois (2006), that the Code provides at least a skeleton of its choice of ‘primary key ranks’ (the current mandatory ranks of the Code, in italics and small capitals in the table) and the secondary key ranks (the ranks that are in between the primary key ranks, but that are optional, in small capitals in the table), as well as standardized prefixes. Dubois (2006) rejects the use of ranks that are in use at different levels (viz. Cohort, Division, Section, Branch and Series), and proposes a list of prefixes to create subsidiary ranks.

There is no doubt that a list should be also provided by the Code, to make it possible to create a homogeneous ranking system. Otherwise, it will never be possible to produce databases covering the entire Animal world that can efficiently exchange information with each other. The freedom to use ranks as an author sees fit is indispensable to the practice of classification, but the possibility of chaotically introducing homonymous ranks at multiple points of the hierarchy can only lead to the impossibility, in the long term, of effectively managing the classification process. Moreover, there can be no efficient computerization of the Code and its rules as long as this question is not settled, which explains why *Lognom* is restricted to the 5 main ranks in the family series, and the species/subspecies and genus/subgenus ranks in the two lower series.

IV.6. The legal aspects of nomenclature

IV.6.1. Is nomenclature a legal system?

Nomenclature, being a regulatory system, has a lot in common with Law. The structure of the Code is reminiscent of the structure of legal texts. First the name of the ensemble of rules, the ‘Code’ is evocative of books of Law such as the French ‘Code Civil’²⁸. The format and structure of the Code, divided into Chapters and Articles, also reminds the reader of a text of Law. The Commission serves as both the legislative and judiciary power in

²⁸The Anglo-Saxon Law texts tend to not be called Codes. The Civil Code of California is a notable exception.

the domain, having itself a status defined in Chapter 7 of the Code, as well as an overarching constitution (International Commission on Zoological Nomenclature, 2020), that is ratified by the International Union of Biological Sciences.

Another component of the legislative world that is sometimes overlooked when comparing nomenclature to Law is the role played by notaries. The latter are an essential piece of the legal environment, as they are in charge of certifying the authenticity of documents and ensuring the lawfulness and truthfulness of some legal acts. The French law defines notaries as “public officers, established to receive all acts and contracts to which the parties must or wish to make these authentic, attached to the acts of public authority, and ensure the date, retain the deposit, deliver a Grosse [enforceable copy] and Expedition [authentic copy]” (Notaires de France, n.d.). Within the framework of nomenclature, it is *ZooBank* that plays a role close to that of a notary, since its purpose is to list nomenclatural acts and be a repository for nomenclatural information. This dimension of *ZooBank* might become even more substantiated in the upcoming years, if the Commission decides to give the platform more authority, by, for example, giving precedence to a spelling that has been registered on it, in case of multiple original spellings.

However, the resemblance with a national legal system seems limited to these points. In fact, Nomenclature does not have any form of the executive branch, capable of enforcing the rules, punishing outlaws, or excluding wrongdoers. Even if the Code is considered a standard, without any enforcement, it is bound to the goodwill of its users. These elements, added to its worldwide reach, bring nomenclature closer to international Law. The latter also has almost no power to enforce the rules, and relies a lot on the good faith of the participating countries or communities to function.

IV.6.2. Is nomenclature a Bona fide* system?

Bona fide, or good faith, is a legal term designating an action (usually a purchase or a decision) made in an honest, genuine way, without fraud or deception. This notion can be extrapolated to designate a system where a form of tacit pact of trust is present between actors of said system, to the mutual benefit of all of them (Kolb, 2000). *Bona fide* plays a capital role in international Law and has been theorized in that field for a long time (Kolb, 2000). Houben (1967) writes that “the principle of good faith (...) is generally recognized as expressing a fundamental concept underlying the entire structure of the international public order”. Van

Bijnkershoek (1737) even states that without good faith, international law would fall apart. Adherence to a set of international laws or international agreements is based on the goodwill of each state or community, making them incredibly fragile. Paradoxically, this system provides a lot of guarantees, as willful compliance may deliver better results than any external enforcement.

The Code works within a very similar system of good faith, as the general adherence to the Code relies on a mix of tradition, consensus, and the general will of taxonomists to participate in a system that leaves them with enough freedom to do their work as they see fit. The Code also operates thanks to the general acceptance of the idea that a common set of rules is beneficial to taxonomy as a science, and taxonomists as a community.

Bona fide systems have some interesting properties. They are usually deeply linked with the notion of reputation and are applied within a smaller community or whenever the legitimate use of force is impossible. In international Law, the use of force is prohibited, as it is in science (outcasting wrongdoers is nonetheless a possibility). The notion of reputation and accountability, however, is a very important one for scientists, and allows this system to maintain itself. No one, including the Commission, wants to be responsible for the downfall of 260 years of Linnean nomenclature. To maintain this *bona fide* sentiment, nomenclature relies on a consensus subtext that is commonly seen in science. However, Gavrilets et al. (2016) suggest that many of the elements that support these notions are based on preconceived ideas rather than realities.

Firstly, the authors explain that ideas such as consensus, which in nomenclature is essential to the discipline itself, or called upon in some cases, such as suprafamilial names, cannot operate in large groups. To be precise, the notion of a large group for Gavrilets et al. (2016) starts at 8 individuals or more, but that effect is already present in groups of 3 people. Informal leaders emerge and start to make the decisions, which may take the form of a pundit for a taxonomic group, or, more currently, authoritative databases. That reality might generate frustration in good faith users of nomenclature, especially if the leaders' decisions are deemed arbitrary.

Another interesting point in this article is the idea that stubborn individuals are more likely to emerge as leaders, sometimes creating bimodal consensus. This element explains the creation of alternative Codes and is particularly reflected in the debates that happened at the beginning of the 20th century, in both botany and zoology, or currently with the PhyloCode (Cantino &

de Queiroz, 2020). The higher the number of ‘stubborn people’ (or -if the extrapolation is correct- persistent schools of thought) that are present in a discussion, the harder it is to reach a consensus. That is relevant for a lot of debates in nomenclature as the discipline is obligated to find a global (i.e., geographically universal and taxonomically exhaustive) solution to every question. That fact makes the system less likely to be fully consensual, and at least creates a latency in the process of reaching that consensus.

Lastly, Gavrilets et al. (2016) give an important role to reputation, by explaining that with agreeableness and persuasiveness, reputation is a key element for the emergence of leaders. That dimension is visible in the fiery tone that some nomenclatural debate can take, as a way to damage the authors’ (or their school of thought) reputation. This question of reputation is directly linked to the *bona fide* dimension, as no one would be willing to follow a system in good faith if the reputation of its proponents or leaders is bad.

In de Candolle’s *Lois de la nomenclature botanique* (de Candolle, 1867), Art. 2 reads “Les règles de la nomenclature ne peuvent être ni arbitraire ni imposées. Elles doivent être basées sur des motifs assez clairs et assez forts pour que chacun soit disposé à les accepter. [Nomenclature rules cannot be arbitrary or imposed. They must be based on reasons that are clear and strong enough for everyone to be inclined to accept them]”. That sentence represents one of the first formalizations of the idea of willingness that characterizes adherence to the nomenclatural rules. That reliance on good faith is also visible in the fact that the threat to stop following the Code is waved by the disgruntled, since at least the Règles (Anonymous, 1905). This constantly brings back into the discussion the idea that the simple fact that some people walk away from the rules would cause the system to collapse for others, creating a form of ‘blackmail’. As long as the dissidents’ reputation is good enough, and their number important enough, the threat is to be taken seriously. It is not always a bad thing, as this is the normal process of large-scale consensus reaching. However, the idea that a consensus is easy to reach, or that it is the natural state of nomenclature denies part of how the discipline works.

This question of good faith also leads to a dimension of the law that is never discussed in nomenclature: the letter and spirit of the law. Nomenclature tends to have a very legalistic approach, i.e., close to the letter rather than the spirit of the law. This first way of interpreting a text gives greater weight to the wording of a rule, rather than to the will of the legislator or the aim of the system. An example in the context of the Code is, when describing numerous

new species, the declaration by the author that several nouns introduced in a publication would undoubtedly be synonyms. This does not contravene the letter of Art. 11.6. since there is no explicit designation of a precise name, but is contrary to the spirit of this article, since its very purpose is to avoid the deliberate overloading of logonymic lists²⁹. It can be considered a bad-faith move.

Respecting the spirit of the law is less important in nomenclature than in constitutional law, but it is part of the good faith of the participants. Without constraints to maintain order, it is crucial not to let an overly legalistic interpretation of the Code get in the way of the good functioning of the system. When a decision must be taken or an issue addressed by the community, trying to understand why the rules were written a certain way should help separate real unresolved questions from loopholes, as to not spend time and energy on useless debates.

In conclusion, nomenclature is most likely a *bona fide* system, and failure to understand the inherent advantages and weaknesses of that reality leads to a lot of misunderstanding when it comes to the functioning of the discipline. Good faith creates the condition for nomenclature to function, but also generates a lot of limitations that one must be conscious of when trying to discuss any idea that challenges the *status quo*.

IV.6.3. What constitutes complexity in nomenclature

The similarities between nomenclature and the Law leads to the expectation that the notion of complexity in former is, by nature, very similar to the notion of complexity in the latter. White (1992) defines legal complexity as “the amount of information that must be collected and processed for lawyers to evaluate a case and litigation to proceed”.

In nomenclature, Dubois (2005) defines the “Principle of Simplicity” as a principle which highlights the fact that unnecessary complications should not be included in an efficient nomenclatural system. For Dubois (2005), the complication comes from exceptions, which in turn can be referred to the previous definition, if interpreted as an increase in the amount of information that is needed to apply the nomenclatural rules. But complexity can emerge in additional ways. It can appear in the sheer quantity of regulatory information applying to a particular field, viz. the size of the Code. The first International Code by Blanchard was less

²⁹Example inspired by the Sharkey et al. (2023) controversy. Even if this rule was first introduced in another context, the underlying norm, monosemy, is the same.

than 20 pages long in French, while the current Code is around 100 pages in its French version, Glossary excluded. Another dimension of complexity may lie in the user's difficulty in understanding the rules, regulations and concepts that make up a domain. This complexity may be generated by the fact that the rules are too imprecise, that the vocabulary is open to multiple interpretations, or that the regulator's original intention is not clear. Without surprise, the same dimensions as those explored when analyzing the Code's rules reappear (IV.1.2.). In many cases, complexity as a notion might vary widely from one individual to another. In other words, what seems complex and harsh for one person may seem simple, and even enjoyable to another. A good example of that phenomenon is the discussion around grammatical agreement. For someone whose native language comprises grammatical agreement, and for whom searching for Latin elements in dictionaries does not seem like an overwhelming task, spending some time finding the correct agreement of a species name will not feel nearly as tedious as it is for another person. The same goes for determining if a name is, in fact, a *nomen oblitum* in the sense of Art. 23.9. The exercise of putting together all the necessary elements is indeed reminiscent of the tedious work of a jurist, compiling elements to be put forward to defend a position.

The goal of simplifying the Code or nomenclature in general is a respectable motivation. The discipline reaches more and more people, while full-time taxonomists are becoming a rare species and the teaching of taxonomy and nomenclature is disappearing from curricula. In practice, however, simplification is likely to be a Sisyphean task, since the rules exist for a reason. Some might be modified with little impact, and even a consensus, but would probably not be a major factor in the simplification process anyway.

If simplification is to be achieved, over and above the recommendations made at the beginning of this chapter, it could also involve optimizing the organization of nomenclatural data searches by encouraging initiatives that supply important data in nomenclature, in a clear, precise, and easily verifiable way (see V.3. and V.4.). A task that can be subdivided into simple tasks becomes a feasible task, even if it was very complex to begin with.

IV.6.4. Perception of nomenclature by the community

An interesting element that emerged since I have been studying nomenclature was the fact that some subjects have been coming back and forth, and appear repeatedly in the literature, at completely different periods of history. One of the most common subjects, that

emerged during the time of my Ph.D., with the Kaiser et al. (2013)/Hoser (2015) (see also Wüster et al. [2021]) controversy and the Sharkey et al. (2021; 2023) initiative is the issue of bad taxonomy³⁰. The first case concerns a controversy arising from the introduction of multiple names available in a journal edited by the author, without peer review, and additionally worsened by accusations of nomenclatural harvesting (Denzer & Kaiser, 2023). The second is the mass description and naming of new Hymenoptera species, with occasional renaming of already known species, based on barcoding, a taxonomic practice considered dubious by many (see Zamani et al. [2022]).

All too often, the nomenclature is still perceived as the consecration of taxonomic work, with the underlying idea that only a ‘good’ taxon ‘deserves’ a name. Meiers et al.’s (2021) article is even titled “A re-analysis of the data in Sharkey et al.’s (2021) minimalist revision reveals that BINs [Barcode Index Numbers] **do not deserve names**”. The “nomenclatural stability” is set as a goal to be preserved in the title of Wüster et al. (2021). Concepts such as ‘taxonomic vandalism³¹’ or ‘aspidonyms*’ were created to discuss these issues.

A corollary to this view is that bad taxonomy should not have the privilege of naming a taxon, and that the Code should act as a barrier against names coined in works deemed insufficient, ethically ambiguous or outright unscientific. This line of thought has, however, to face the fact that it is not the Code’s role to regulate the quality of taxonomy, merely to verify that the minimum is provided to be able to associate a name with a taxon. I hypothesize that, since the Code resembles a legal text, it is tempting for many to conflate its form and function, and use its authority to try to ‘govern’ taxonomic practices. Still, that only partly explains why authors seem to care so much about names.

Nomenclature sometimes becomes the crystallizing point of a lot of tension. In a serious paper, Brochu (2012) writes about the possible rejection of a *nomen dubium*:

³⁰ One point I find particularly interesting is that, after reading about these two cases in particular, I asked several of my fellow Ph.D. students’ colleagues to give me their criteria for a good taxonomy. Many of them are paleontologists, so the conversation did not focus on the subject of the contribution of genetics or Barcoding. While everyone had a fairly clear picture of what a bad taxonomy was, it ended up being seriously difficult for us to come up with objective, timeless and universal criteria for a good taxonomy. One of the criteria that emerged, which I found particularly interesting, was that a good taxonomy was easy for contemporaries or successors to evaluate, adopt and criticize. It is still nonetheless a very subjective criterion.

³¹ This concept, born in zoological taxonomy (Anonymous, 2007), was also adopted by mycologists (Loizides et al., 2022). However, its definition of a taxonomic vandal varies from name stealer (or nomenclatural harvester [Denzer & Kaiser, 2023]) to ‘author’ of ‘bad taxa’ (Páll-Gergely et al., 2020).

I do not say this lightly; Pristichampsus is widely used. Its elimination may lead to confusion in the literature and, in all likelihood, my own violent death at the hands of my colleagues. Taxonomic stability [sic] should be promoted, and I would rather not be killed. [...] In the absence of such material, we are obliged to set the name aside.

Beyond the smile it may create, this extract exemplifies well the sheer amount of emotional involvement that nomenclature may cause.

Many explanations have been proposed to explain that attachment to names. The mihi-itch approach (Evenhuis, 2008) [also known as mihilism [Bruun, 1950; Dubois, 2008a] or Mihisucht [see Pape (1993)], places the author's ego and pride in creating names and describing species at the center stage. Aescht (2018), proposes that even though names are accepted as labels whose semantics do not matter (c.f. Pavlinov's (2021) nominalist approach ['a name is just a label']), they retain an irreducible semantic and emotional value. That explains why their replacement by a number or electronic ID makes us uncomfortable (Aescht, 2018). Pure attachment to habits and traditions should be also acknowledged when trying to have a full picture of the relationship between taxonomists and nomenclature. Another element is the fact, that, being a *bona fide* system with a lenient governance, nomenclature is tied to personal notions such as reputation, generating a lot more heated debates and violent attacks than taxonomy, which relies more, in theory, on cold, impersonal arguments.

For all these reasons, nomenclature is in a category of its own in science. It may not be a science itself, but it is imbued with certain of its codes, while blending them with elements of law and history. Nomenclature also has an important sociological and inter-human dimension that cannot be ignored. This explains why sub-optimal solutions are sometimes favored, rather than more logical but divisive ones. Nevertheless, these are also the elements that give its appeal to this discipline. It is simply to be hoped that the theoretical and philosophical aspects of nomenclature will continue to be subjects of discussion and be allowed to grow, and that debates on these matters and more down-to-earth, practical issues will remain peaceful and constructive. If the taxonomic community aims to build a more efficient system, it will be necessary to go beyond partisan battles and senseless polemics. There are still many challenges to overcome in nomenclature, one being that the expertise in that field is slowly dying out, so every reasonable contribution and attempt to make things better should be heard.

V. Perspectives and conclusion

“I cannot separate the aesthetic pleasure of seeing a butterfly and the scientific pleasure of knowing what it is.”

Vladimir Nabokov

The development of the prototype of a web application that manages the rules of the nomenclatural process proved to be an interesting endeavor, sparked interesting thoughts, and opens numerous doors. If it is not currently possible to have a fully automatic system, multiple improvements to both *Lognom* and the nomenclatural process may lead to a smoother experience for the average taxonomist.

V.1. Finishing the software

The next step in computing the Code, which may seem obvious to the reader at this stage, would be to complete the software by adding the rules that it still lacks and completing the parts that are still underway. The finalization of *Lognom* could require the development of an additional part, currently missing, to support the rules concerning the management of type specimens. This would make it possible to integrate a whole new collection of forms, linked to the decisions linked to the designation of lectotypes, neotypes or simply *a posteriori* determination of the original type of an available name that did benefit from any explicit designation. The identification of specimens that have nomenclatural value and the exact way of labeling them should also be added (Fig. 34), to differentiate clearly specimens with a nomenclatural value from specimens with only a taxonomic or historical value. The development of this part of the software would have much to gain from exchanges with curators, collection managers and technicians to ensure that it covers the realm of practical possibilities that happen in natural history collections.

Moreover, *Lognom* should end up integrating the ranks above the superfamily, but that might imply an additional reflection on the computability of the rules that apply to these names, as they still lack a sufficient regulatory framework to be computerized efficiently (see IV.5.).

Feedback from testers is also an important source of information about the direction the software should take. Experienced testers will be able to pinpoint weaknesses and potential errors in the software. These shortcomings can then be corrected. The second series of tests will demonstrate the software's potential for inclusion in taxonomists' workflows. It will also identify the software's limitations in terms of ergonomics, clarity and usability.

The direction the software has to take is to evolve from a functioning prototype to a finished product that would become a must in the habitual workflow of a taxonomist. For that to be achieved, the improvement of the software's ergonomics and accessibility is a priority. A

better streamlining of the forms by their reduction to a strict minimum wherever possible is a desirable upgrade. The aim is to avoid leaving the user with a feeling of insurmountable work to be done to enter a simple name. In the same vein, the addition and availability by default of certain names of indisputable historical importance, such as the names of Linnæus and Clerk, is also under consideration.

Form		Answer
The specimen was included in the taxon by the author in the original description of the taxon The specimen was designated as a type or holotype by the author of the taxon in the original description The specimen is the only one that have been designated as a type ...	✓ ✓ ✓	The specimen is a holotype. It should be labeled as such and made recognizable (see Rec. 72D.)...
The specimen was included in the taxon by the author in the original description of the taxon The specimen was designated as a type or holotype by the author of the taxon in the original description The specimen is the only one that have been designated as a type ...	✓ X X	The specimen is not a nomenclatural type, but is part of the type series, and should labeled as 'paratype' (see Rec. 73D.)...

Figure 34: A simplified vision of the prospective tool for the identification of valuable nomenclatural specimens in *Lognom*.

Another possible, and undoubtedly very useful upgrade would be the addition of internal management of names that have received a judgment from the Commission, based on the indexes and official lists of available names and the Bulletin of Zoological Nomenclature. That addition could provide a real boost to both the software's usability and simplicity, as well as its relevance to the busy taxonomist. It might be made possible by the fact that this data is already partly available, under the management of Commissioner Evenhuis (Pyle, p.c., August 27, 2023).

Lastly, one of the major improvements that *Lognom* should undergo is that it should become usable by a beginner with little knowledge of nomenclature and by someone already familiar with the field. This could lead to the development of F.A.Q., the addition of text guides and figures to help the novice user get started with the nomenclatural rules.

V.2. Managing the etymologies and establishing Lists of Available Names

Another way the web application could be upgraded is by incorporating a better treatment of etymology. Asking the user for elements on the etymology seemed pointless at the first stage of the development of *Lognom*, as it does not influence the availability status of a name. At this point, the etymology of a name is an aimless field in the database, that is only used as a way to retain information. However, some elements, such as the grammatical gender and grammatical nature of genus and species series names have been already integrated, and serve to ensure that some basic spelling rules are respected. There are two possible evolutions for that dimension of the software:

The first possible improvement is the inclusion of the categories of names as defined by Jobling (2010). Beyond these nine categories, two others can be added, one proposed by Cupello (2022) and a new category, Gelotonym, defined in this work (Table 14).

Table 14: Categories of names depending on their etymology.

Category of name	Definition	Source	Example
Morphonym (not to be confused with morphonym <i>sensu</i> Dubois [2000], meaning a form of a name)	Name based on the morphology and physical characteristics of the animal	Jobling (2010)	<i>Bufotes viridis</i> , <i>Microptilotis albonotatus</i>
Eponym	Name based on the name of a real person or a mythical character (I would exclude fictional characters, given the recent debates)	Jobling (2010)	<i>Pelophylax bergerii</i> , <i>Ogcocephalus darwini</i> , <i>Abudefduf whitleyi</i> , <i>Acteon</i>
Autochtonym	Name based on the vernacular name of the animal in a language (even if the language is a classical one)	Jobling (2010)	<i>Panthera leo</i> , <i>Aquila</i> , <i>Gekko</i> , <i>Giraffa</i> , <i>Okapia</i>
Toponym	Name based on a geographical location	Jobling (2010)	<i>Malpolon</i> <i>monspessulanus</i> , <i>Sitta</i> <i>europaea</i> , <i>Acanthocinus</i> <i>sinensis</i>
Taxonym (not to be confused with Taxonym <i>sensu</i> Pavlinov [2014], meaning scientific name)	Name based on the resemblance or proximity with another taxon	Jobling (2010)	<i>Ammobatooides</i> , <i>Sphecodes</i> , <i>Pseudomalus</i> , <i>Hylarana</i>
Bionym	Name based on the natural habitat of the animal	Jobling (2010)	<i>Monticola</i> , <i>Speleomantes</i> , <i>Sylvicola</i> <i>fenestralis</i>
Ergonym (Not to be confused with Ergonym)	Name derived from the action, habits, or	Jobling (2010)	<i>Solifugae</i> , <i>Monticola</i> <i>explorator</i> , <i>Stercorarius</i>

<i>sensu</i> Dubois [2000], meaning valid and correctly spelled name used in multiple ergotaxonomies*)	behavior of the animal.		<i>Parasiticus</i>
Phagonym	Names based on the eating habits of the animal.	Jobling (2010)	<i>Myrmecophaga</i> , <i>Piscivoravis</i> , <i>Ophiophagus</i> , Folivora
Phononym	Names derived from the sound produced by the animal	Jobling (2010)	<i>Bombus</i> , <i>Mirafa javanica</i> <i>cantillans</i>
Koultonym	Names derived from anything related to works of art such as musical compositions, paintings, films, and fictional books.	Cupello (2022)	<i>Ateuchus pastoralis</i> , <i>Sycorax</i> , <i>Salticus</i> <i>lilliputanus</i> , <i>Pimonia</i> <i>cthulhu</i> , <i>Zoosphaerium</i> <i>darthvaderi</i> , <i>Abraracourcix</i>
Gelotonym (from γέλως, laughter and ὄνυμα name)	Names derived from a wordplay, a joke, an anecdote, or that are completely random.	Original	<i>Pieza pi</i> , <i>Mini ature</i> , <i>Galae donut</i> , <i>Agra</i> <i>cadabra</i> , <i>Vini vidivici</i>

This division would make it possible to collect data for statistics on scientific names, as a support discussion on the use of any of these categories, as it is currently happening with eponyms (Guedes et al. 2023). It would lead to better visibility of the consequences of changes when discussing etymologies, and give an idea of the shift in fashion and habits of the taxonomic community.

The second possible development would be to collect the exact etymology and original language of each scientific name. This has several advantages, some of which may go beyond the scope of the software. One advantage would be to be able to develop advanced grammar-checking tools using ancient language dictionaries, or guides developed precisely for

nomenclature (AnimalBase, n.d.; Vendetti & Garland, 2019). This would be an ambitious project, but by its very nature, the grammar of classical languages is fixed and thus having a checking tool that helps decide the gender of genus names, or the grammatical nature of species names might not be as impossible as it appears. However, the complexity entailed by the creation of said tool is much higher than the one used for *Lognom*. It would most probably end up being a semi-automatic process rather than a fully automatic one, as it necessitates the manipulation of semantics and context rather than a rather simple if/else logic.

The second advantage brought by the collection of etymologies is that, together with the information already present in *Lognom*, this clarification of etymology would make it possible to participate in the formation of LANs.

Almost all the information needed to compile a list of available nouns is already collected in the web application (name, author, date, associated work and bibliography, associated type, etymology). An advantage of the digital medium is that it makes changes quite dynamic and makes it possible to establish evolving lists. That would boost the potential adoption of LANs in some communities, but it would necessitate collaborative work to be done.

V.3. Cooperative work, data sharing and zoological nomenclature standard

Another key component of working online that may greatly benefit the ulterior version of *Lognom* is the implementation of the possibility to share work and data with colleagues. This is essential to *Lognom*'s open science objective. Entering the data for the complete nomenclature of a group is not an easy task. It might even seem completely overwhelming. However, taxonomists who might be interested in that kind of work might be encouraged if trusted with a given task or motivated to take over partially done work. This can be achieved in two ways, the first being the possibility of exchanging partial or complete files containing the nomenclatural data of a given group, and the second being the possibility of working simultaneously on the same data. The latter is quite harder to implement, as, aside from the purely programming technicalities, the simultaneous work on interdependent data would need a strict system of warnings and controls. If a genus is wrongly deemed unavailable by a user, that decision could cascade on the availability of a family series name it typifies, and lead to a severe change in the validity for at least two taxa.

That explains why a double system (online and offline) of file exchange should be developed first, with limited reach of subsequent users on the original data.

The export of files locally should be made available in widely used formats, such as CSV or SQL. When online, the option to publish data as well as access to said data should be instated, to encourage importing and building upon, as well as reviewing the work already produced by colleagues. This aspect of data exchange is already partially implemented in *Lognom*, with the possibility of making a nomenclatural project public for every other user to find and reuse. Another possible course of action that could be implemented is the creation of a data exchange standard based on purely nomenclatural data (Table 15). This standard could be based on existing ontologies, such as NOMEN (<https://github.com/SpeciesFileGroup/nomen>) or Zoonom (<https://skosmos.loterre.fr/th63/en/>). The two ontologies do differ quite a bit, resulting in the coverage of almost all currently described nomenclatural situations. Moreover, NOMEN covers a much broader domain than zoological nomenclature, as it encompasses the five biological nomenclatural Codes. If needed, that would make it possible to push towards a more generalist standard, covering other nomenclature codes, even if this means having to manage some form of necessary redundancy. This structural redundancy is, as a matter of fact, already present and managed in the NOMEN ontology by the addition of the Codes' acronyms before the given property. Depending on the timeframe, the standard could then be submitted to the Commission for control, and to check that it is correctly adapted to the 5th edition of the Code, or can be built upon the 5th edition directly.

Table 15: Examples of concepts that could form a new nomenclatural standard and their equivalencies.

Concept	Zoonom	NOMEN	Associated field in <i>Lognom</i>	Associated field value in <i>Lognom</i>
Scientific name	Nomen	ICZN name	name_name	Any string
<i>Nomen nudum</i>	Gymnonym	ICZN nomen nudum	name_description_diagnosis	"The name lacks an accompanying description or a diagnosis"; "The name references a description or a diagnosis in a work that is considered not published"
Described as variety or form after 1960	None	ICZN variety or form after 1960	name_rank	name_rank: "variety", "form" name_date: >1960
Registered spelling of a name	Delograph	None	None	None

Lognom should be able to take advantage of this kind of standard to export, tag, and archive data entered by users. *ZooBank* could also benefit from that data structure, and even become the main user (see V.5.).

It will then be possible to submit this standard to the TDWG community and discuss several possible scenarios for the future. With access to the expertise of the TDWG community, a bigger standard for the exchange of nomenclatural data can be collegially constructed. Moreover, gateways with existing standards are a possibility that would greatly improve the whole biodiversity ecosystem, especially when combined with access to the data dispersed in literature databases, in particular Biodiversity Heritage Library (BHL; <https://www.biodiversitylibrary.org/>). BHL is without doubt the largest depository of old

taxonomic literature. The initiatives taken to automatize data extraction from this incredibly rich resource coupled with the Global Name Architecture could give a great boost to the collection of nomenclatural data (Richards, 2020).

V.4. Towards a synergy with *ZooBank*? Modeling of nomenclature as a whole

V.4.1. Combining ZooBank and Lognom

ZooBank, being the database of the Commission and the “Official Register of Zoological Nomenclature” (International Commission on Zoological Nomenclature, 2012), has a capital role as a central and permanent structure hosting nomenclatural data. However, two main obstacles still compromise this aim of being the focal point of zoological nomenclature online. The first is the difficult integration of historical names into the database. This difficulty was noted early after *ZooBank* was made available (Pyle & Michel, 2010). The second is the difficulty encountered by the community in correctly applying the rules introduced by the 2012 amendment on electronic publication (Dubois, in preparation). In both these situations, *Lognom* can play a role in the improvement of *ZooBank*. For the first issue, a simplified transfer of data from *Lognom* to *ZooBank* could lead to the creation of another data source for the Commission’s database. That would be true for most data that would be deemed interesting for *ZooBank*, including what could evolve under the fifth edition. If a taxonomist builds even a small file on *Lognom* and chooses to do so, it could be cascaded directly into *ZooBank*, and only the fraction of data that is still needed specifically for *ZooBank* would have to be completed.

Moreover, the integration of controls and forms inspired by *Lognom* directly into *ZooBank* could have multiple advantages. Mainly, it could prevent some honest mistakes, verify that the work and name are indeed available, or at least that some elements necessary are present. On a smaller scale, it might help with *ZooBank*’s notarial work by making sure that some critical debated points are mentioned when registering the name, such as making sure that the newly created name is not a homonym or an unjustified emendation.

V.4.2. Towards a modeling of nomenclature.

Once all the elements presented so far are taken into account, a new project makes a logical apparition. Depending on the directions planned for databases in the future, in particular for *ZooBank*, imagining total modeling of nomenclature does not seem unattainable.

For the sake of a thought experiment, it would be interesting to imagine for a moment since Linnæus, a nomenclature database existed. This database would have been filled consciously by every taxonomist with all the relevant nomenclatural data throughout History. The question is, what would that database structure be, and how well would it be able to reconstitute, 250 years after the fact, the exact history of decisions made by taxonomists? A strong argument could be made that this database would have a structure quite similar to that of *Lognom*, but would differ in that it would not seek to apply the rules efficiently, but to retain the raw data and evolution of the entire history of nomenclature. With a sufficiently well-thought-out basic structure, which would take into account all nomenclatural objects, as defined in III.1.1., and properties discussed in III.1.2., it would even be possible to reconstitute all nomenclatural acts, and gradually recreate digitally the whole history of nomenclature. That database could benefit from the development of the grammatical controls discussed above.

The modeling of nomenclatural acts

Lognom differs from *ZooBank* by the fact that it does not consider nomenclatural acts as objects, as the nature of the data in *Lognom* makes it possible to treat nomenclatural acts as transitional objects. It is not the aim of *Lognom* to do such things, but it is interesting to try to figure out how this notion of nomenclatural act can be modeled through a modification of nomenclatural data in a database that would resemble that of *Lognom*. Currently, when modeled, nomenclatural acts are treated in *ZooBank* as an object by proxy (R. L. Pyle, personal communication, 08/13/2023). Nomenclatural acts are not treated per se in the Taxonomer model (Pyle, 2004), but were eventually introduced in *ZooBank* as a particular case of TNU (R. L. Pyle, personal communication, 08/13/2023).

What is meant by a transitional object is that it is possible to “reconstitute” a nomenclatural act by analyzing the way the data would evolve in the database (and verifying that the work in which it is published is available). For the sake of that thought experiment, let us consider that invalid nomenclatural acts are ignored by users and do not enter the system. In practice, any new name added to the database (in the form of a new entry in the ‘Name’ table [Fig. 35]) would be a reflection of a real-world introduction of a new name. An *a posteriori* determination of a type specimen can also be deduced by the fact that a new foreign key (i.e., the creation of a link between two tables, or via a joint table) linking the name table to the

type specimen table is created (Fig. 35, in green). A designation of a lectotype would be the suppression of multiple foreign keys save one that would be the lectotype (Fig. 35, in red). A neotype would replace an old foreign key with a new one. The destruction or loss of a type specimen would result in the disappearance (or indication of disappearance, as to not lose the information) of that entry rather than a change in the foreign key link. A choice by a First Reviser between multiple spellings would be equivalent to the suppression of a name, and the placement of the spelling in the “alternative spelling” of the concurrent name with a change on the first reviser status of said name. An intervention by the Commission would also be visible in the database, as an addition of data in the relevant columns dedicated to the Commission

However, there are some limits to this approach. The first resides in the fact that there is no current easy access to the entirety of the changes that ever happened. In its current form, for example, *Lognom* can only manage the original and current combination of a species name. However, a point has to be made that if the modeling of nomenclature as a whole was ever tried, a correct data scheme should make it possible to visualize nomenclatural acts indirectly as described above, rather than as their own entities.

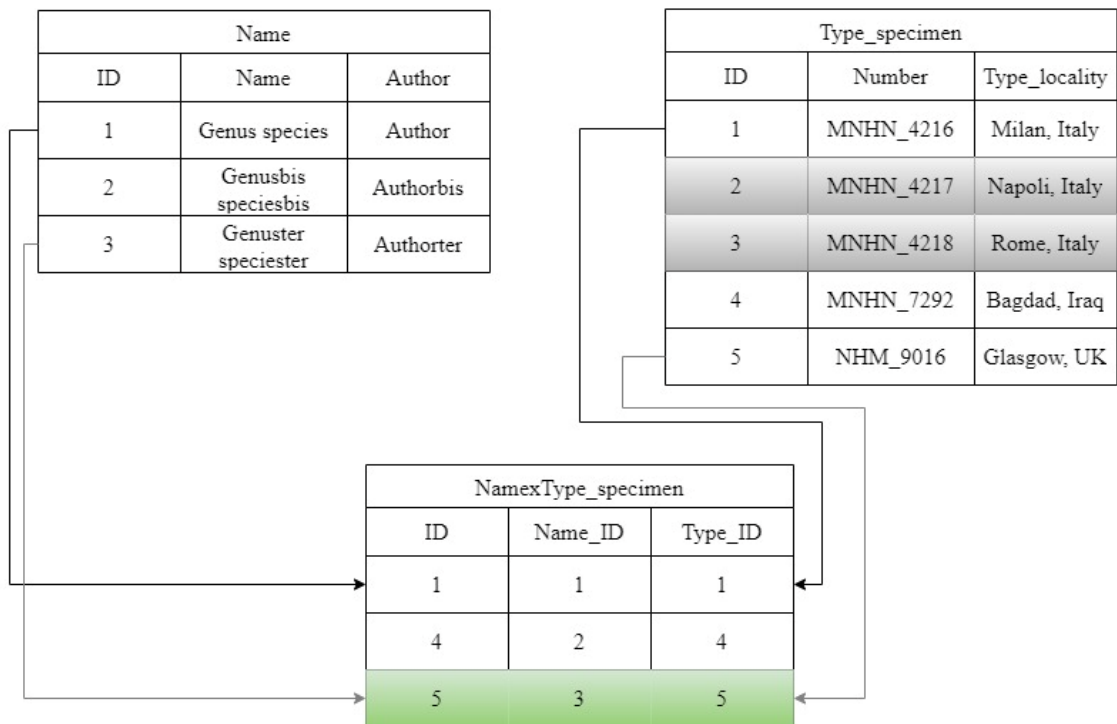
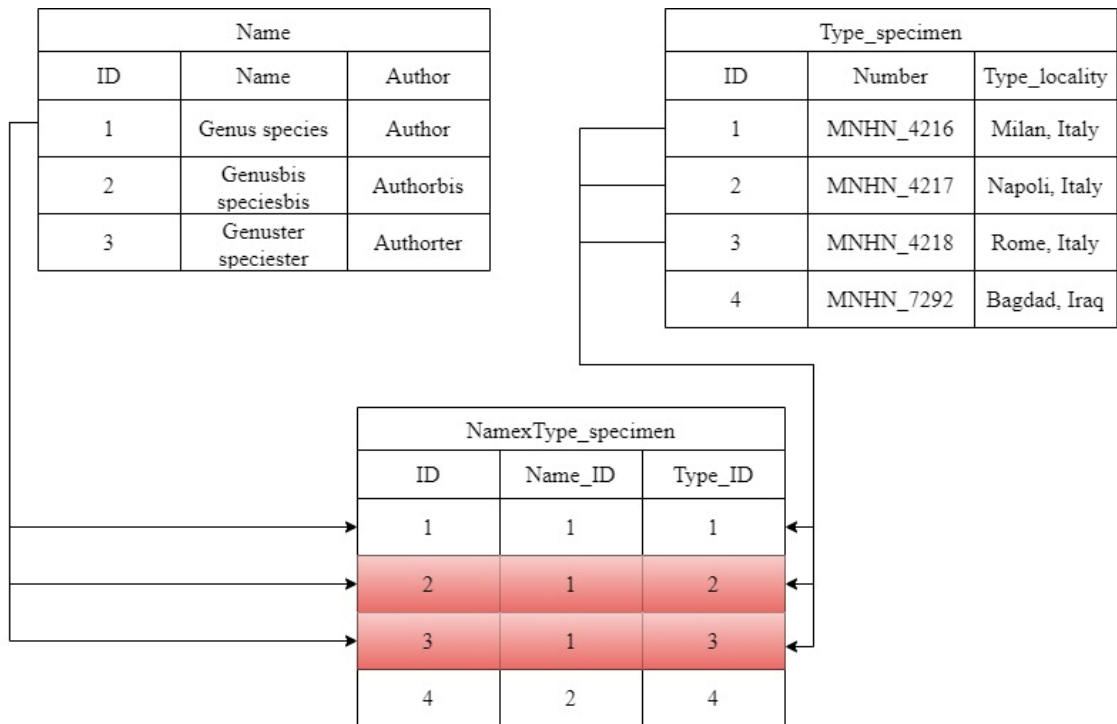


Figure 35: A visualization of the consequences of two nomenclatural acts in an extract of a database aiming to model nomenclature, at a time t (above) and time $t+1$ (below).

The arrows indicate links between the tables. In green: An *a posteriori* determination of a type specimen (addition of an entry in the Type_specimen table and a link in the join table). In red: A designation of a lectotype (suppression of two links in the join table). The two type specimens stripped of their functions are shown in grey.

The data necessary to model nomenclature are already partly available. *ZooBank* would naturally be the host of these data, not only because it makes sense as it is the Commission's database, but also because it was planned for it to take this direction since its conception (Pyle & Michel, 2010). In this case, *Lognom* (or rather algorithms similar to its core functions) would be appointed in a data-quality control role, the collection databases would provide data on the type specimens, taxonomic databases would provide raw nomenclatural data and BHL could provide missing data on old names with progress in text recognition and data retrieval (Fig. 36). In return, *ZooBank* could provide data to be integrated by default in *Lognom*, as a way to simplify the work of the users of the software, as they would have access to names with all their properties. The benefits of a standard to exchange nomenclatural data seem obvious due to the very centralized nature of this database. The wider possible participation by external actors and data providers is crucial.

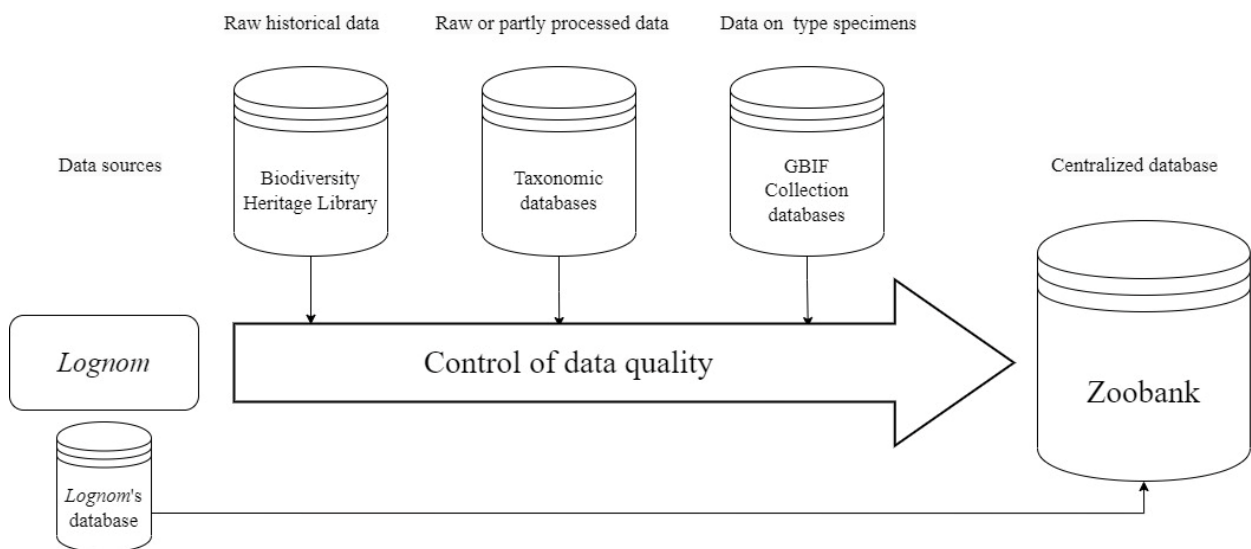


Figure 36: A simplified view of the potential organization around a centralized nomenclatural database.

V.5. Conclusion

The main question developed in my thesis was the computerization of the rules of the Code. The preliminary analysis enabled the extraction of the nomenclatural objects, the clarification of their properties and the conceptualization of the algorithmic process of nomenclature. Illustrations of the various elements of the analysis show both the complexity and the logic that imbues the majority of the rules. A functional software, *Lognom*, has been developed and is available for use. Almost half of the instructions given by the Code have

been successfully integrated into *Lognom*. The remaining half should be implemented in the future. Some dimensions of the Code proved impervious to the computerization, either because of their complexity or because of the difficulty of interpretation they entail. The rules of nomenclature should be as logical, easy to apply and unambiguous as possible. Recommendations to reach that goal have been provided, as well as a classification of the rules in terms of their potential for computerization and automation.

The next step is to work on completing the software with the rules of the current (4th edition) Code in order to be ready to adapt the developed software to the 5th edition of the Code. Nevertheless, it might be more efficient to work directly on adding the new rules and nomenclatural workflow to *Lognom*, with an emphasis put on the modified rules, as an additional way to familiarize the public with the changes. Moreover, in the long run, and with enough data already entered into the software (either by users or taken from the potential future central repository of zoological nomenclature) it should be possible to statistically model the impact of a rule change, even if the effect is retrospective.

This would imply a close collaboration with the Commission and, in particular, the Code Editorial Committee, which is in charge of drafting the new Code. That would ensure a strict application of the rules and leave no room for any misinterpretation or misunderstanding of the text.

Numerous challenges are to be faced before the rules of nomenclature can be considered fully accessible to the scientific community. We live in a time where taxonomic and nomenclatural expertise is not sufficiently valued, while being essential to face the century of extinctions (Dubois, 2011a). By making data on biodiversity accessible, nomenclature is one of the main links between our work and the rest of humankind. It provides continuity between the knowledge of the past, research in the present time, and the valuable information we share with the public and we pass on to our successors. Making sure that the set of skill necessary to guarantee that this discipline continues to thrive and grow remains goes beyond purely academic considerations. It is our duty, as taxonomists, to protect and nurture this legacy.

Glossary

Admissible polysemy: Cases in nomenclature in which total or partial polysemy is permitted by the different Codes. Concept defined in Pavlinov (2021); see Saliba (2022) for an analysis of the issue.

Allocation: Nomenclatural act by which a scientific name becomes attached to a taxon or several taxa through a system of type specimens or names, or a phylogenetic definition. Term coined by Dubois (2005).

Aspidonym: “New, scientifically and ethically acceptable names proposed to overwrite pre-existing names”. Term proposed by Wüster et al. (2021)

Big Data: A data size that is too large or complex to be dealt with by traditional data-processing application software.

Bona Fide: Latin expression meaning “in good faith”. In Law, *bona fide* is an action made (usually a purchase or a decision) in an honest, genuine way, without fraud or deception, regardless of its outcome. Traditional concept in Law.

Chresonym: In biodiversity informatics and nomenclature, a chresonym is the citation of a taxon name, usually a species name, in a publication subsequent to its original description. Term coined by Smith & Smith (1972).

Data robustness: A dataset’s robustness is the general level to which it can withstand changes in the methods used for collecting and integrating the data without losing any of the information content, statistical validity, or scientific significance. It describes the caliber of the data gathered and whether it is weak or strong.

Ergotaxonomy: A classification used in a certain work by a given author. Term coined in Dubois (2005).

Grey nomenclature: Scientific names that exist outside of ‘Linnean Nomenclature’. In practice, they are scientific names that are voluntarily excluded by their author(s) from the International Codes. Concept defined by Minelli (2017).

LAN: List of Available Names. Defined in Art. 79., the LAN in zoology is a project that aims to record every available name and nominal taxa with their spelling, date and authorship for names and name-bearing-type for nominal taxa in a unique list. The LAN is subdivided in parts and has to be adopted by the Commission.

Logonymic list: Synonymy list as used in taxonomic works. Any list of scientific names, including synonyms, subsequent spellings and/or chresonymic lists of a taxon. Term coined by Dubois (2000) to distinguish these lists from the synonymy list *sensu stricto* which includes only the available names given to a taxon.

Monotypic taxon: A taxon that contains only one immediately subordinate taxon. Term coined by Mayr and Ashlock (1991).

Onymology: The study of the concepts and theory of biological nomenclature, the philosophy of nomenclature. Term coined by Dubois (2000). Synonym: Taxonymy *sensu* Pavlinov (2021).

Taxon Name Usage: Term used in Biodiversity Informatics, describing the mention of a scientific name in a work. The definition does not seem to be fully formalized.

Bibliography

- Aedo, C., Fiz, O., Alarcón, M. L., Navarro, C., & Aldasoro, J. J. (2005). Taxonomic revision of *Geranium* sect. *Dissecta* (Geraniaceae). *Systematic Botany*, 30(3), 533–558.
<https://doi.org/10.1600/0363644054782260>
- Agassiz, A. (1881). Report on the Echinoidea dredged by H.M.S. Challenger during the years 1873–1876. In *Report on the Scientific Results of the Voyage of H.M.S. Challenger during the years 1873–76. Zoology 3* (part 9), 1–321 + pl. 1–66.
- Agassiz, L. (1846) *Nomenclatoris zoologici index universalis, continens nomina systematica classium, ordinum, familiarum et generum animalium omnium, tam viventium quam fossilium, secundum ordinem alphabeticum unicum disposita, adjectis homonymiis plantarum*. Jent & Gassmann. i–viii + 1–393.
- Aesch, E. (2018). Was lebt, „birgt“ einen Namen: Bionomenklatur als universelles Verständigungsmittel? In P. Ernst, S. Gaisbauer, A. Greule, & K. Hohensinner (eds.) *Namenforschung im Spannungsfeld zwischen Wissenschaft und Öffentlichkeit* (pp. 469–487). Vulpes.
- Agosti, D., Guidoti, M., Catapano, T., Ioannidis-Pantopikos, A., & Sautter, G. (2020). The standards behind the scenes: Explaining data from the Plazi workflow. *Biodiversity Information Science and Standards*, 4. <https://doi.org/10.3897/biss.4.59178>
- AnimalBase Project Group. (n.d.). *Gender agreement in zoological nomenclature*. AnimalBase. <https://wwwuser.gwdg.de/~fwelter/changeable.htm>
- AnimalBase Project Group. (2023). *AnimalBase. Early zoological literature online*. www.animalbase.uni-goettingen.de
- Anonymous [‘R. L.’]. (1905). Règles internationales de la Nomenclature zoologique. *Nature*, 71(1849), 534. <https://doi.org/10.1038/071534a0>
- Anonymous. (2007). Vandalism in taxonomy. *Koleopterologische Rundschau*, 77, 38.
- Aristotle (4th century B.C.). *Περὶ τὰ ζῷα ἱστορίαι* [History of animals]. Translated to French by J. B. Saint-Hilaire (1883). Hachette. Available online at <https://remacle.org/bloodwolf/philosophes/Aristote/tableanimaux.htm>
- Averianov, A. O., & Lopatin, A. V. (2011). Phylogeny of triconodonts and symmetrodonts and the origin of extant mammals. *Doklady Biological Sciences*, 436(1), 32–35.
<https://doi.org/10.1134/s0012496611010042>
- Bánki, O., Döring, M., & Jeppesen, T. S. (2023). Name IDs and name matching for Catalogue of Life: existing services and prospects. *Biodiversity Information Science and Standards*, 7, Art. 111662. <https://doi.org/10.3897/biss.7.111662>

- Bánki, O., Roskov, Y., Döring, M., Ower, G., Hernández Robles, D. R., Plata Corredor, C. A., Stjernegaard Jeppesen, T., Örn, A., Vandepitte, L., Hobern, D., Schalk, P., DeWalt, R. E., Keping, M., Miller, J., Orrell, T., Aalbu, R., Abbott, J., Adlard, R., Adriaenssens, E. M., et al. (2023). Catalogue of Life checklist (annual checklist 2023). *Catalogue of Life*. <https://doi.org/10.48580/dfs>
- Bentham, G. (1826). *Catalogue des plantes indigènes des Pyrénées et du Bas Languedoc*. Huzard. 1–127.
- Benton, M. J. (2005). *Vertebrate Palaeontology* (3rd ed.). Blackwell. I–XI + 1–455.
- Beolens, B., Watkins, M., & Grayson, M. (2011). *The eponym dictionary of reptiles*. Johns Hopkins University Press. xiii + 296 pp.
- Berendsohn, W. G. (1995). The concept of “potential taxa” in databases. *Taxon*, 44(2), 207–212. <https://doi.org/10.2307/1222443>
- Berendsohn, W. G. (2023). Use cases for scientific name identifiers and name matching: Progress report from the TETTRIs project. *Biodiversity Information Science and Standards*, 7, Art. 109666. <https://doi.org/10.3897/biss.7.109666>
- 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(9-10), 224–250. <https://doi.org/10.3406/bsef.1957.20372>
- Bertrand, Y., Pleijel, F., & Rouse, G. W. (2006). Taxonomic surrogacy in biodiversity assessments, and the meaning of Linnaean ranks. *Systematics and Biodiversity*, 4(2), 149–159. <https://doi.org/10.1017/s1477200005001908>
- Blanchard, R., Maehrenthal, F. von & Stiles, C. W. (1905). *Règles internationales de la nomenclature zoologique adoptées par les Congrès International de Zoologie. International Rules of Zoological Nomenclature. Internationale Regeln der Zoologischen Nomenklatur*. Rudeval. 1–57.
- Bouchet, P. (1999). Recording and registration of new scientific names: a simulation of the mechanism proposed (but not adopted) for the International Code of Zoological Nomenclature. *The Bulletin of Zoological Nomenclature*, 56, 6–15. <https://doi.org/10.5962/bhl.part.23018>
- Bouchet, P. (2018, February 1). *Acteonimorpha*. World Register of Marine Species. <https://www.marinespecies.org/aphia.php?p=taxdetails&id=1057250>

- Bour, R., & Dubois, A. (1986). Nomenclature ordinale et familiale des Tortues (Reptilia) Note complémentaire. *Bulletin Mensuel de la Société Linnéenne de Lyon*, 55(3), 87–90. <https://doi.org/10.3406/linly.1986.10749>
- Bourgoin, T. (n.d.). *Pyrops oculatus* (Westwood, 1838). Fulgoromorpha Lists On the Web. Retrieved September 2, 2023 from <https://flow.hemiptera-databases.org/flow/?page=explorer&db=flow&lang=en&card=taxon&rank=species&iid=17494>
- Bourgoin, T. (2023, June 26). FLOW (Fulgoromorpha Lists On the Web): a world knowledge base dedicated to Fulgoromorpha. Version 8. <https://hemiptera-databases.org/flow/>
- Bourgoin, T., Bailly, N., Zaragüeta, R., & Vignes-Lebbe, R. (2021). Complete formalization of taxa with their names, contents and descriptions improves taxonomic databases and access to the taxonomic knowledge they support. *Systematics and Biodiversity*, 19(7), 738–746. <https://doi.org/10.1080/14772000.2021.1915895>
- Bourgoin, T., Vignes-Lebbe, R., & Bailly, N. (2019). Visualisation of taxonomic knowledge: Exploring and reporting taxonomic data, training students in taxonomy. *Biodiversity Information Science and Standards*, 3, Art. e37730. <https://doi.org/10.3897/biss.3.37730>
- Bowley, D. R., & Smith, H. M. (1968). The dates of publication of Louis Agassiz's Nomenclator Zoologicus. *Journal of the Society for the Bibliography of Natural History*, 5(1), 35–36. <https://doi.org/10.3366/jsbnh.1968.5.1.35>
- Bridson, G. D. R. (1968). The Zoological Record—A centenary appraisal. *Journal of the Society for the Bibliography of Natural History*, 5(1), 23–34. <https://doi.org/10.3366/jsbnh.1968.5.1.23>
- Briquet, J., Ginzberger, A., Schiffner, A., Von Weizierl, T., Von Wettstein, R., & Zahlbruckner, A. (1906). *Verhandlungen des Internationalen botanischen Kongresses in Wien 1905. Actes du Congrès internationale de botanique tenu à Vienne (Autriche) 1905*. Gustav Fisher. 1–261.
- Brochu, C. A. (2012). Phylogenetic relationships of Palaeogene ziphodont eusuchians and the status of *Pristichampsus* Gervais, 1853. *Earth and Environmental Science Transactions of the Royal Society of Edinburgh*, 103(3–4), 521–550. <https://doi.org/10.1017/s1755691013000200>
- Brooksbank, C., & Cowley A. (2023, January). Bioinformatics for the terrified: An introduction to the science of bioinformatics. https://doi.org/10.6019/TOL.BioinfTer-c_2016.00001.1

- Bruun A.F. (1950). The Systema Naturae of the twentieth century. *Science*, 112, 342–343.
- Brunner Von Wattenwyl, C. (1878). *Monographie der Phaneropteriden*. F. A. Brockhaus.
<https://doi.org/10.5962/bhl.title.9035>
- Cambridge Dictionary. (n.d.). Discipline. Retrieved September 30, 2023, from
<https://dictionary.cambridge.org/dictionary/english/discipline>
- Candolle, A. de (1867). *Lois de la nomenclature botanique*. H. Georg. 1–64.
<https://doi.org/10.5962/bhl.title.139042>
- Candolle, A. de (1883). *Nouvelles remarques sur la nomenclature botanique : Supplément au commentaire du même auteur qui accompagnait le texte des Lois*. H. Georg. 1–76.
<https://doi.org/10.5962/bhl.title.127596>
- Candolle, A. P. de (1813). *Théorie élémentaire de la botanique, ou exposition des principes de la classification naturelle et de l'art de décrire et d'étudier les végétaux*. Déterville. I–VIII + 1–527. <https://archive.org/details/b29309426/mode/2up>
- Cantino, P. D. & de Queiroz, K. (2020). *International code of phylogenetic nomenclature (PhyloCode): A phylogenetic code of biological nomenclature*. CRC Press. 1–189.
- Catalogue of Life. (2023, August 17). *Bufo variabilis*.
<https://www.catalogueoflife.org/data/taxon/NPMR>
- Cavalier-Smith, T. (1993). Kingdom protozoa and its 18 phyla. *Microbiological Reviews*, 57(4), 953–994. <https://doi.org/10.1128/mr.57.4.953-994.1993>
- Central Intelligence Agency (2023, July 15). *The World Factbook*. <https://www.cia.gov/the-world-factbook/countries/world/#people-and-society>
- Ceríaco, L. M. P., Aesch, E., Ahyong, S. T., Ballerio, A., Bouchard, P., Bourgoïn, T., Dmitriev, D., Evenhuis, N., Grygier, M. J., Harvey, M. S., Kottelat, M., Kluge, N. J., Krell, F., Kojima, J., Kullander, S. O., Lucinda, P., Lyal, C. H. C., Pyle, R. L., Rheindt, F. E., . . . Pape, T. (2023). Renaming taxa on ethical grounds threatens nomenclatural stability and scientific communication. *Zoological Journal of the Linnean Society*, 197(2), 283–286. <https://doi.org/10.1093/zoolinlean/zlac107>
- Chmielewski J. G., & Kraysky, D. (2013). *General botany laboratory manual*. AuthorHouse, 1–286.
- Clarivate Analytics. (n.d.). *About*. ION: Index to Organism Names. Retrieved September 18, 2023, from <http://www.organismnames.com/about.htm>
- Corliss, J. O. (1995). The ambiregnal protists and the codes of nomenclature: a brief review of the problem and of proposed solutions. *The Bulletin of Zoological Nomenclature*, 52, 11–17. <https://doi.org/10.5962/bhl.part.6717>

- Cosson, E., & Germain, E. (1840). *Observations sur quelques plantes critiques des environs de Paris*. L. Bouchard-Huzard. I–VIII, 1–68.
- Cupello, M. (2022). *Ateuchus dung beetles and their evolution (Coleoptera: Scarabaeidae): A new model in speciology*. PhD thesis - Universidade Federal do Paraná, Curitiba, xxiii + 1–1826.
- Cupello, M., Silva, F. A. B., & Vaz-De-Mello, F. Z. (2023). The taxonomic revolution of New World dung beetles (Coleoptera: Scarabaeidae: Scarabaeinae). *Frontiers in Ecology and Evolution*, 11, Art. 1168754. <https://doi.org/10.3389/fevo.2023.1168754>
- Darwin Core Maintenance Group (n.d.). Darwin Core home page. *Biodiversity Information Standards (TDWG)*. Retrieved September 1, 2023 from <https://dwc.tdwg.org/>
- Darwin Core Maintenance Group. (2023, July 7). List of Darwin Core terms. *Biodiversity Information Standards (TDWG)*. <http://rs.tdwg.org/dwc/doc/list/2023-07-07>
- Dayrat, B. (2005). Towards integrative taxonomy. *Biological Journal of the Linnean Society*, 85(3), 407–415. <https://doi.org/10.1111/j.1095-8312.2005.00503.x>
- Dayrat, B. (2010). Celebrating 250 dynamic years of nomenclatural debates. In Polaszek A. (ed.) *Systema Naturae 250 – The Linnaean Ark* (pp. 185–240). CRC Press. <https://doi.org/10.1201/ebk1420095012-c17>
- De Thaon, P. (1121-1135). *Le Bestiaire*. [Accessed through a reedition and translation by Wright T. (1841), made available online by D. Badke at this address: <https://bestiary.ca/etexts/wright-bestiary-of-philippe-de-thaon.pdf>]
- Deng, J., Guo, Y., Cheng, Z., Lu, C., & Huang, X. (2019). The prevalence of single-specimen/locality species in Insect taxonomy: An empirical analysis. *Diversity*, 11(7), Art. 106. <https://doi.org/10.3390/d11070106>
- Denzer, W., & Kaiser, H. (2023). Naming and gaming: The illicit taxonomic practice of ‘nomenclatural harvesting’ and how to avoid it. *Journal of Zoology*, 320(3), 161–168. <https://doi.org/10.1111/jzo.13061>
- Deroy, L. (1956). *L’Emprunt linguistique*. Presses Universitaires de Liège. 427–470. <https://doi.org/10.4000/books.pulg.665>.
- Dewynter, M., Massary, J.-C. de, Bochaton, C., Bour, R., Ineich, I., Vidal, N., & J. Lescure. (2019). Liste taxinomique de l’herpétofaune dans l’Outre-mer français : III. Département de la Martinique. *Bulletin de la Société Herpétologique Française*, 169, 53–82.
- Django Software Foundation. (n.d.). *Django*. Django Project. Retrieved October 2, 2023, from <https://www.djangoproject.com/>

- Döring, M., Grojean, M., Russel, L. A., & Svenningsen, C. (2023, May 17). *Data use club practical sessions: Name matching and the GBIF Taxonomic Backbone*.
<https://www.gbif.org/event/42jhz2tRyrjzTF5B8pClKa/data-use-club-practical-sessions-name-matching-and-the-gbif-taxonomic-backbone>
- Dubois, A. (2000). Synonymies and related lists in zoology: general proposals, with examples in herpetology. *Dumerilia*, 4 (2): 33–98.
- Dubois, A. (2004). The higher nomenclature of recent amphibians. *Alytes*, 22(1–2), 1–14.
- Dubois, A. (2005). 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. (2006). 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. (2008a). 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. (2008b). Phylogenetic hypotheses, taxa and nomina in zoology. *Zootaxa*, 1950(1), 51–86. <https://doi.org/10.11646/zootaxa.1950.1.7>
- Dubois, A. (2011a). Taxonomy in the century of extinctions: taxonomic gap, taxonomic impediment, taxonomic urgency. *TAPROBANICA: The Journal of Asian Biodiversity*, 2(1), 1–5. <https://doi.org/10.4038/tapro.v2i1.2702>
- Dubois, A. (2011b). The International Code of Zoological Nomenclature must be drastically improved before it is too late. *Bionomina*, 2(1), 1–104.
<https://doi.org/10.11646/bionomina.2.1.1>
- Dubois, A., Aneesh, P. T., Bauer, A. M., Ceriaco, L. M. P., Daniel, G. M., Dellapé, P. M., Engel, M. S., Frétey, T., Löbl, I., Lorvelec, O., Marinov, M., Ohler, A., Schmitt, M., Whittington, A. E., Young, M. T., & Aescht, E. (2021). Optical discs in zoological nomenclature: problems and proposed solution. *Bionomina*, 24(1), 1–8.
<https://doi.org/10.11646/bionomina.24.1.1>
- Dubois, A., Crochet, P. A., Dickinson, E.C., Nemésio, A., Aescht, E., Bauer, A. M., Blagoderov, V., Bour, R., De Carvalho, M. R., Desutter-Grandcolas, L., Frétey, T., Jäger, P., Koyamba, V., Lavilla, E.O., Löbl, I., Louchart, A., Malécot, V., Schatz, H., & Ohler, A. (2013). Nomenclatural and taxonomic problems related to the electronic publication of new nomina and nomenclatural acts in zoology, with brief comments on

- optical discs and on the situation in botany. *Zootaxa*, 3735(1), 1–94.
<http://dx.doi.org/10.11646/zootaxa.3735.1.1>
- Dufresnes, C., Mazepa, G., Jablonski, D., Oliveira, R. C., Wenseleers, T., Shabanov, D. A., Auer, M., Ernst, R., Koch, C., Ramírez-Chaves, H. E., Mulder, K. P., Simonov, E., Tiutenko, A., Kryvokhyzha, D., Wennekes, P. L., Zinenko, O., Korshunov, O. V., Al-Johany, A. M., Peregotsev, E. A., . . . Litvinchuk, S. N. (2019). Fifteen shades of green: The evolution of *Bufo* toads revisited. *Molecular Phylogenetics and Evolution*, 141, Art. 106615. <https://doi.org/10.1016/j.ympev.2019.106615>
- Dunger, G. T. (1968). The lizards and snakes of Nigeria. Part 5: The amphisbaenids of Nigeria including a description of 3 new species. *Nigerian Field*, 33(4), 167–192.
- Dunn, E. R. (1918). The collection of Amphibia Caudata of the Museum of Comparative Zoology. *Bulletin of the Museum of Comparative Zoology*. 62, 445–471.
- Dupuis, C. (1986). Dates de publication de l'«Histoire naturelle générale et particulière des crustacés et des insectes» (1802-1805) par Latreille dans le « Buffon de Sonnini ». *Annales de la Société Entomologique de France*, 22(2), 205–210.
<https://doi.org/10.1080/21686351.1986.12278421>
- Elmasri, R., & Navathe, S. (2017). *Fundamentals of database systems* (Vol. 7). Pearson.
- Emig, C. C., Bitner, M. A., & Alvarez, F. (2015). Scientific death-knell against databases? Errors induced by database manipulations and its consequences. *Carnets de Géologie*, 15(16), 231–238. <https://doi.org/10.4267/2042/57949>
- Engel, M. S. (2005). Family-Group names for bees (Hymenoptera: Apoidea). *American Museum Novitates*, 2005(3476), 1–33. [https://doi.org/10.1206/0003-0082\(2005\)476](https://doi.org/10.1206/0003-0082(2005)476)
- Evenhuis, N. L. (2008). The “Mihi itch” — a brief history. *Zootaxa*, 1890, 59–68.
<https://doi.org/10.11646/zootaxa.1890.1.3>
- Evenhuis, N. L. (2016). Charles Davies Sherborn and the “Indexer’s Club.” *ZooKeys*, 550, 13–32. <https://doi.org/10.3897/zookeys.550.9697>
- Farber, P. L. (1976). The type-concept in zoology during the first half of the nineteenth century. *Journal of the History of Biology*, 9(1), 93–119.
<https://doi.org/10.1007/bf00129174>
- Federhen, S. (2014). Type material in the NCBI Taxonomy Database. *Nucleic Acids Research*, 43(D1), D1086–D1098. <https://doi.org/10.1093/nar/gku1127>
- Fennah R. G. (1971). Fulgoroidea from the Cayman Islands and adjacent areas. *Journal of Natural History* 5, 299–342.

- Froese, R., & Pauly, D (eds.). (2023, February). FishBase.
<https://fishbase.mnhn.fr/search.php?lang=English>
- Frost, D. R. (2023a, June 26). Home page. *Amphibian Species of the World: An Online Reference*. Version 6.2. American Museum of Natural History.
<https://amphibiansoftheworld.amnh.org/index.php>. doi.org/10.5531/db.vz.0001
- Frost, D. R. (2023b, August 1). *Bufotes viridis*. *Amphibian Species of the World: An Online Reference*. Version 6.2. American Museum of Natural History.
<https://amphibiansoftheworld.amnh.org./Amphibia/Anura/Bufonidae/Bufotes/Bufotes-viridis>
- Frost, D. R. (2023c, August 1). *Hylarana signata*. *Amphibian Species of the World: An Online Reference*. Version 6.2. American Museum of Natural History.
<https://amphibiansoftheworld.amnh.org/Amphibia/Anura/Ranidae/Hylarana/Hylarana-signata>
- Gaffney, E. S., Hutchison, J. H., Jenkins, F. A., & Meeker, L. J. (1987). Modern turtle origins: The oldest known cryptodire. *Science*, 237(4812), 289–291.
<https://doi.org/10.1126/science.237.4812.289>
- Gavrilets, S., Auerbach, J., & Van Vugt, M. (2016). Convergence to consensus in heterogeneous groups and the emergence of informal leadership. *Scientific Reports*, 6(1), Art. 29704. <https://doi.org/10.1038/srep29704>
- Gill, F., Donsker, D., & Rasmussen, P. (Eds.). (2023). *International Ornithological Committee World Bird List*. Version 13.1. <https://doi.org/10.14344/IOC.ML.13.1>.
- Gillman, L. N., & Wright, S. D. (2020). Restoring indigenous names in taxonomy. *Communications Biology*, 3, 609. <https://doi.org/10.1038/s42003-020-01344-y>
- Global Biodiversity Information Facility (GBIF). (2023). *GBIF home page*.
<https://www.gbif.org/>
- Global Biodiversity Information Facility Backbone Taxonomy (GBIF). (2023a). *Aglaophenia struthionides* (Murray, 1860). <https://www.gbif.org/species/2267365> Checklist dataset
<https://doi.org/10.15468/39omei>
- Global Biodiversity Information Facility Backbone Taxonomy (GBIF). (2023b). *Bufotes variabilis* (Pallas, 1769). <https://www.gbif.org/species/9787242>. Checklist dataset
<https://doi.org/10.15468/39omei>
- Global Biodiversity Information Facility Backbone Taxonomy (GBIF). (2023c). *Lybia tresselata* Milne Edwards, 1834. <https://www.gbif.org/species/8373211> Checklist dataset
<https://doi.org/10.15468/39omei>

- Global Biodiversity Information Facility Backbone Taxonomy (GBIF). (2023d). *Pagurus* Berthold, 1827. <https://www.gbif.org/species/7457172> Checklist dataset <https://doi.org/10.15468/39omei>
- Global Biodiversity Information Facility Backbone Taxonomy (GBIF). (2023e). *Triturus marmoratus* (Latreille, 1800). <https://www.gbif.org/species/2431886> Checklist dataset <https://doi.org/10.15468/39omei>
- Grenié, M., Berti, E., Carvajal-Quintero, J., Dädlow, G. M. L., Sagouis, A., & Winter, M. (2023). Harmonizing taxon names in biodiversity data: A review of tools, databases and best practices. *Methods in Ecology and Evolution*, *14*(1), 12–25. <https://doi.org/10.1111/2041-210X.13802>
- Grimaldi, D. A. (1990). A phylogenetic, revised classification of genera in the Drosophilidae (Diptera). *Bulletin of the American Museum of Natural History*, *197*, 1–139.
- Grosjean, M., Høfft, M., Gonzalez, M., Hahn, A., & Robertson, T. (2022). GRSciColl: What is the Global Registry of Scientific Collections and how to contribute? *Biodiversity Information Science and Standards*, *6*. <https://doi.org/10.3897/biss.6.93875>
- Gudger, E. W. (1934). The five great naturalists of the sixteenth century: Belon, Rondelet, Salviani, Gesner and Aldrovandi: A chapter in the history of Ichthyology. *Isis*, *22*(1), 21–40.
- Guedes, P., Alves-Martins, F., Arribas, J. M., Chatterjee, S., Santos, A. M. C., Lewin, A., Bako, L., Webala, P. W., Correia, R. A., Rocha, R., & Ladle, R. J. (2023). Eponyms have no place in 21st-century biological nomenclature. *Nature Ecology and Evolution*, *7*(8), 1157–1160. <https://doi.org/10.1038/s41559-023-02022-y>
- Guyer, C., & Slowinski, J. B. (1991). Comparisons of observed phylogenetic topologies with null expectations among three monophyletic lineages. *Evolution*, *45*(2), 340–350. <https://doi.org/10.1111/j.1558-5646.1991.tb04408.x>
- Hammer, T. A., & Thiele, K. R. (2021). Proposals to amend Articles 51 and 56 and Division III, to allow the rejection of culturally offensive and inappropriate names. *Taxon*, *70*, 1392–1394.
- Hawksworth, D. L. (2010). *Terms used in bionomenclature: The naming of organisms (and plant communities)*. Global Biodiversity Information Facility. 1–216.
- Hey, J. (2001). The mind of the species problem. *Trends in Ecology and Evolution*, *16*(7), 326–329. [https://doi.org/10.1016/s0169-5347\(01\)02145-0](https://doi.org/10.1016/s0169-5347(01)02145-0)
- Holmes, M. W., Hammond, T. T., Wogan, G. O. U., Walsh, R. E., LaBarbera, K., Wommack, E. A., Martins, F., Crawford, J. C., Mack, K. L., Bloch, L. M., & Nachman, M. W.

- (2016). Natural history collections as windows on evolutionary processes. *Molecular Ecology*, 25(4), 864–881. <https://doi.org/10.1111/mec.13529>
- Hoser, R. T. (2015). Best Practices in herpetology: Hinrich Kaiser's claims are unsubstantiated. *Australasian Journal of Herpetology*, 25, 39–52.
- Houben, P. (1967). Principles of international law concerning friendly relations and co-operation among States. *American Journal of International Law* 61(3), 703–736. <https://doi.org/10.2307/2197463>
- Humtsoe, L. N., Bordoloi, S., Ohler, A., & Dubois, A. (2008). Rediscovery of a long known species, *Ixalus lateralis* Anderson, 1871. *Zootaxa*, 1921(1), 24–34. <https://doi.org/10.11646/zootaxa.1921.1.2>
- Integrated Taxonomic Information System (ITIS). (n.d.). *Bufo variabilis* (Pallas, 1769). Retrieved August 25, 2023 from https://www.itis.gov/servlet/SingleRpt/SingleRpt?search_topic=TSN&search_value=1103415#null
- International Commission on Zoological Nomenclature. [Blanchard, R., Maehrenthal, F. von & Stiles, C. W.] (1905). *Règles internationales de la nomenclature zoologique. International rules of zoological nomenclature. Internationale Regeln der zoologischen Nomenklatur*. Rudeval.
- International Commission on Zoological Nomenclature. (1956). Completion of the entries relating to the names of certain genera of the order Decapoda (class Crustacea) placed on the *Official List of Generic Names in Zoology* in the period up to the end of 1936. *Opinions and Declarations rendered by the ICZN, 1D(D2)*, 47–82.
- International Commission on Zoological Nomenclature. (1957). Opinion 472: Addition to the Official List of Generic Names in Zoology of the generic name *Pagurus* Fabricius, 1775, with *Cancer bernhardus* Linnæus, 1758, as type species (class Crustacea, order Decapoda). *Opinions and Declarations rendered by the ICZN, 16(13)*, 213–276.
- International Commission on Zoological Nomenclature. (1999). *International code of zoological nomenclature* (4th edition). International Trust for zoological Nomenclature. I–XXIX + 1–306.
- International Commission on Zoological Nomenclature. (2012). Amendment of Articles 8, 9, 10, 21 and 78 of the International Code of Zoological Nomenclature to expand and refine methods of publication. *ZooKeys*, 219, 1–10. <https://doi.org/10.3897/zookeys.219.3944>

- International Commission on Zoological Nomenclature. (2014). Zoological nomenclature and electronic publication—a reply to Dubois *et al.* (2013). *Zootaxa*, 3779(1), 3–5. <https://doi.org/10.11646/zootaxa.3779.1.2>
- International Commission on Zoological Nomenclature. (2018). Opinion 2420 (Case 3678) – Proposal to affirm that ‘Formenkreis’ is not a genus-group taxonomic rank with potential availability: not approved. *The Bulletin of Zoological Nomenclature*, 75, 267–269. <https://doi.org/10.21805/bzn.v75.a057>
- International Commission on Zoological Nomenclature. (2019). Opinion 2430 – Parts of the List of Available Names in Zoology for phylum Rotifera: accepted. *The Bulletin of Zoological Nomenclature*, 76, 74–76.
- International Commission on Zoological Nomenclature. (2020). *Constitution of the International Commission on Zoological Nomenclature*. Available at <https://www.iczn.org/assets/About/5d5a317ab3/ICZN-Constitution-2022.pdf>
- International Union for the Conservation of Nature. (2022, February). *The IUCN Red List of threatened species. Version 2022-2*. <https://www.iucnredlist.org>.
- Inventaire National du Patrimoine Naturel (INPN). (n.d.). *Glossaire - C*. Retrieved October 19, 2023, from <https://inpn.mnhn.fr/informations/glossaire/liste/c>
- Jacot, A. P. (1929). American oribatid mites of the subfamily Galumninae. *Bulletin of the Museum of Comparative Zoölogy at Harvard College* 69(1), 3–36.
- Jobling, J. A. (2010). *Helm dictionary of scientific bird names*. Christopher Helm. 1–432.
- Joly, A., Goëau, H., Bonnet, P., Bakić, V., Barbe, J., Selmi, S., Yahiaoui, I., Carré, J., Mouysset, E., Molino, J., Boujemaa, N., & Barthélémy, D. (2014). Interactive plant identification based on social image data. *Ecological Informatics*, 23, 22–34. <https://doi.org/10.1016/j.ecoinf.2013.07.006>
- Kaiser, H., Crother, B. I., Kelly, C. M., Luiselli, L., O’Shea, M., Ota, H., Passos, P., Schleip, W. D., & Wüster, W. (2013). Best practices: in the 21st Century, taxonomic decisions in herpetology are acceptable only when supported by a body of evidence and published via peer-review. *Herpetological Review*, 44, 8–23.
- Kear, B. P. (2003). Cretaceous marine reptiles of Australia: a review of taxonomy and distribution. *Cretaceous Research*, 24(3), 277–303. [https://doi.org/10.1016/s0195-6671\(03\)00046-6](https://doi.org/10.1016/s0195-6671(03)00046-6)
- Kent, W. D. (1983). A simple guide to five normal forms in relational database theory. *Communications of the ACM*, 26(2), 120–125. <https://doi.org/10.1145/358024.358054>

- Kerner, A., Bouquin, S., Portier, R., & Vignes-Lebbe, R. (2021). The 8 years of existence of Xper3: State of the art and future developments of the platform. *Biodiversity Information Science and Standards*, 5, Art. e74250.
<https://doi.org/10.3897/biss.5.74250>
- Kiesenwetter, H. (1858). Gesetze der entomologischen Nomenclatur. *Berliner Entomologische Zeitschrift*, 2(3–4), xi–xxii.
<https://doi.org/10.1002/mmnd.18580020316>
- Knapp, S., & Wright, D. (2010). E-publish or perish? In Polaszek, A. (ed.) *Systema Naturae 250 – The Linnaean Ark* (pp. 83–94). CRC Press.
<https://doi.org/10.1201/ebk1420095012-c8>
- Kolb, R. (2000). *La bonne foi en droit international public*. Geneva Graduate Institute Publications, I–XLIII + 1–758 <https://doi.org/10.4000/books.iheid.2253>
- Larousse. (n.d.). Niveau. In *Larousse en ligne*. Retrieved September 21, 2023, from <https://www.larousse.fr/dictionnaires/francais/niveau/54687>
- Latreille, P. A. (1800) [An. VIII]. *Histoire naturelle des Salamandres de France, précédée d'un tableau méthodique des autres reptiles indigènes*. Villiers. I–XLVII+1–58.
- Laurin, M. (2023). The PhyloCode: The logical outcome of millennia of evolution of biological nomenclature? *Zoologica Scripta* [currently only available online].
<https://doi.org/10.1111/zsc.12625>
- Lendemer, J. C., Thiers, B. M., Monfils, A., Zaspel, J. M., Ellwood, E. R., Bentley, A., LeVan, K. E., Bates, J. M., Jennings, D., Contreras, D. L., Lagomarsino, L. P., Mabee, P. M., Ford, L., Guralnick, R., Gropp, R. E., Revelez, M., Cobb, N. S., Seltmann, K. C., & Aime, M. C. (2019). The extended specimen network: a strategy to enhance US biodiversity collections, promote research and education. *BioScience*, 70(1), 23–30.
<https://doi.org/10.1093/biosci/biz140>
- Liddell, H.G., & Scott, R. (1940). *A Greek-English Lexicon*. Clarendon Press.
- Linnæus, C. (1735). *Fundamenta botanica, quae majorum operum prodromi instar theoriam scientiae botanices by breves aphorismos tradunt*. Salomon Schouten. 1–35.
- Linnæus, C. (1758). *Systema naturae per regna tria naturae, secundum classes, ordines, genera, species, cum characteribus, differentiis, synonymis, locis. Editio decima, reformata* (tome 1). Laurentii Salvii. I–IV +1–824.
- Livezey, B. C., & Zusi, R. L. (2007). Higher-order phylogeny of modern birds (Theropoda, Aves: Neornithes) based on comparative anatomy. II. Analysis and discussion.

- Zoological Journal of the Linnean Society*, 149(1), 1–95.
<https://doi.org/10.1111/j.1096-3642.2006.00293.x>
- Loizides, M., Alvarado, P., Moreau, P., Assyov, B., Halasů, V., Stadler, M., Rinaldi, A., Marques, G., Zervakis, G. I., Borovička, J., Van Vooren, N., Grebenc, T., Richard, F., Taşkın, H., Gube, M., Sammut, C., Agnello, C., Baroni, T. J., Crous, P. W., . . . Bellanger, J. (2022). Has taxonomic vandalism gone too far? A case study, the rise of the pay-to-publish model and the pitfalls of *Morchella* systematics. *Mycological Progress*, 21(1), 7–38. <https://doi.org/10.1007/s11557-021-01755-z>
- Lombard, R. E., & Wake, D. B. (1986). Tongue evolution in the lungless salamanders, Family Plethodontidae IV. Phylogeny of plethodontid salamanders and the evolution of feeding dynamics. *Systematic Zoology*, 35(4), 532. <https://doi.org/10.2307/2413113>
- Malécot, V. (2008). Les règles de nomenclature - Histoire et fonctionnement. *Biosystema*, 25 41–76.
- Mayr, E. (1969). *Principles of systematic zoology*. McGraw Hill. 1–428.
- Mayr, E. (1982). *The growth of biological thought: Diversity, evolution, and inheritance*. Harvard University Press. 1–974.
- Mayr, E., & Ashlock P. D. (1991). *Principles of systematic zoology* (2nd edition). McGraw Hill. 1–428.
- McKenna, M.C., & Bell, S.K. (1997). *Classification of Mammals above the species level*. Columbia University Press. i–xii, 1–631.
- McNeill, J. H., & Turland, N. J. (2011). Major changes to the *Code of Nomenclature*-Melbourne, July 2011. *Taxon*, 60(5), 1495–1497. <https://doi.org/10.1002/tax.605030>
- Meier, R., Blaimer, B. B., Buenaventura, E., Hartop, E., Von Rintelen, T., Srivathsan, A., & Yeo, D. C. J. (2021). A re-analysis of the data in Sharkey et al.'s (2021) minimalist revision reveals that BINs do not deserve names, but BOLD Systems needs a stronger commitment to open science. *Cladistics*, 38(2), 264–275.
<https://doi.org/10.1111/cla.12489>
- 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. I–VIII + 92.
- Miller, S., Barrow, L. N., Ehlman, S. M., Goodheart, J. A., Greiman, S. E., Lutz, H. L., Misiewicz, T. M., Smith, S. M., Tan, M., Thawley, C. J., Cook, J. A., & Light, J. E. (2020). Building natural history collections for the twenty-first Century and beyond. *BioScience*, 70(8), 674–687. <https://doi.org/10.1093/biosci/biaa069>

- Minelli, A. (2000). The ranks and the names of species and higher taxa, or a dangerous inertia of the language of natural history. In M. T. Ghiselin & A. E. Leviton (eds.), *Cultures and institutions of natural history: Essays in the history and philosophy of science* (pp. 339–351). California Academy of Sciences.
- Minelli, A. (2017). Grey nomenclature needs rules. *Ecologica Montenegrina*, 7, 654–666.
<https://doi.org/10.37828/em.2016.7.31>
- Mosyakin, S. L. (2022). If “*Rhodes* -” must fall, who shall fall next? *Taxon*, 71(2), 249–255.
<https://doi.org/10.1002/tax.12659>
- Murguía-Romero, M., Serrano-Estrada, B., Ortiz, E., & Villaseñor, J. L. (2021). Taxonomic identification keys on the web: tools for better knowledge of biodiversity. *Revista Mexicana de Biodiversidad*, 92, Art. e923592.
<https://doi.org/10.22201/ib.20078706e.2021.92.3592>
- Murray, A. (1860). Description of new Sertulariadae from the Californian Coast. *Annals and Magazine of Natural History* 3(5), 250–252.
- Muséum national d’Histoire naturelle. (n.d.a). *Bases de données scientifiques naturalistes*. Retrieved September 21, 2023, from <https://www.mnhn.fr/fr/bases-de-donnees-scientifiques-naturalistes>
- Muséum national d’Histoire naturelle. (n.d.b). *Specimen MNHN-EH-EH16435*. Collection : Insectes - Hémiptères (EH). Retrieved August 31, 2023 from <https://science.mnhn.fr/institution/mnhn/collection/eh/item/eh16435?listIndex=1&listCount=2>
- Museum Ulm. (n.d.). *Der Löwenmensch*. Retrieved June 20, 2023 from <http://www.loewenmensch.de/index.html>
- Nagai, S., Porion, T., & Audibert, C. (2017). *Pyrops nishiguroi* n. sp., a new Fulgoridae from Philippines, with some notes on the *Pyrops oculatus* group (Hemiptera: Fulgoromorpha). *Faunitaxys*, 5(1), 1–5.
- Naggs, F. (2022). The tragedy of the Natural History Museum, London. *Megataxa*, 7(1), 85–112. <https://doi.org/10.11646/megataxa.7.1.2>
- Notaires de France. (n.d.). *Notary’s role*. Retrieved September 9, 2023 from <https://www.notaires.fr/en/notaire/role-notaire-and-his-principal-activities/role-notaire>
- Osborn, H. F. (1903). The reptilian subclasses Diapsida and Synapsida and the early history of the Diaptosauria. *Memoirs of the AMNH*, 1(8), 449–519.

- Oxford University Museum of Natural History. (n.d.). *Fulgora oculata*. Collections Online. Retrieved September 14, 2023, from <https://www.oumnh.ox.ac.uk/collections-online#/item/oum-catalogue-368333>
- Páll-Gergely, B., Hunyadi, A., & Auffenberg, K. (2020). Taxonomic vandalism in malacology: comments on molluscan taxa recently described by N. N. Thach and colleagues (2014–2019). *Folia Malacologica*, 28(1), 35–76. <https://doi.org/10.12657/folmal.028.002>
- Palma, R. L., & Heath, A. C. G. (2021). Science versus vernacular: Should some taxa of animals and plants be renamed according to ‘indigenous’ practices? *Bionomina*, 22, 1–7. <https://doi.org/10.11646/bionomina.22.1.1>.
- Pape, T. (1993). Die Mihiisucht - a senior synonym of the mihi itch? *Antenna*, 17, 117–118.
- Pape, T., & Evenhuis, N. L. (2023). Systema dipterorum. *Biodiversity Information Science and Standards*, 7, Art. 111959. <https://doi.org/10.3897/biss.7.111959>
- Patterson, D. J. (2003). Progressing towards a biological names register. *Nature*, 422, 661. <https://doi.org/10.1038/422661a>
- Patterson, D. J., Remsen, D., & Norton, C. (2003). Comment on zoological record and registration of new names in zoology. *The Bulletin of Zoological Nomenclature*, 60, 297–299.
- Pavlinov, I. Y. (2014). Таксономическая номенклатура: книга 2. От Линнея до первых кодексов [*Taxonomic nomenclature. Book 2. From Linnaeus to first Codes*]. KMK Scientific Press. 1–223.
- Pavlinov, I. Y. (2021). *Taxonomic nomenclature: What's in a name – Theory and history*. CRC Press. 1–265. <https://doi.org/10.1201/9781003182535-1>
- Pellens, R. (2021). Les collections naturalistes: Une ressource essentielle pour la science du XXI^e siècle. In R. Pellens (ed.) *Les collections naturalistes dans la science du XXI^e siècle : Une ressource durable pour la science ouverte* (pp. 7–18). ISTE éditions.
- Pliny the Elder (77). *Naturalis Historia, Books VIII to XI*. Translated to French by Émile Littré (1848), Dubochet. Available online at <http://remacle.org/bloodwolf/erudits/plineancien/livre8.htm>
- Polaszek, A. [and coauthors] (2005a). A universal register for animal names. *Nature*, 437(7058), 477. <https://doi.org/10.1038/437477a> [Note: the co-authors are not listed as it is classically done for article, but rather given in the Supplementary Information. The list of coauthors include Co-authors are D. Agosti, M. Alonso-Zarazaga, G. Beccaloni, P. de Place Bjørn, P. Bouchet, D. J. Brothers, the Earl of Cranbrook of the

- International Trust for Zoological Nomenclature, N. Evenhuis, H. C. J. Godfray; N. F. Johnson; F.-T. Krell, D. Lipscomb, C. H. C. Lyal, G. M. Mace, S. Mawatari, S.E. Miller, A. Minelli, S. Morris, P. K. L. Ng, D. J. Patterson, R. L. Pyle, N. Robinson, L. Rogo, J. Taverne, F. C. Thompson, J. van Tol, Q. D. Wheeler & E. O. Wilson]
- Polaszek, A., Bouchet, P., Evenhuis, N. L., Krell, F., 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. *The Bulletin of Zoological Nomenclature*, 62, 210–220. <https://repository.si.edu/handle/10088/17055>
- Polaszek, A., & Michel, E. (2010). Linnaeus–Sherborn–ZooBank. In Polaszek A. (ed.) *Systema Naturae 250 – The Linnaean Ark* (pp. 163–172). CRC Press. <https://doi.org/10.1201/ebk1420095012-c15>
- Pyle, R. L. (2004). Taxonomer: A relational data model for managing information relevant to taxonomic research. *Phyloinformatics*. 1, 1–54.
- Pyle, R. L. (2016). Towards a global names architecture: The future of indexing scientific names. *ZooKeys*, 550, 261–281. <https://doi.org/10.3897/zookeys.550.10009>
- Pyle, R. L., Bailly, N., & Remsen, D. (2022). Modeling taxon concepts: A new approach to an old problem. *Biodiversity Information Science and Standards*, 6, Art. e93927. <https://doi.org/10.3897/biss.6.93927>
- Pyle, R. L., & Michel, E. (2008). ZooBank: Developing a nomenclatural tool for unifying 250 years of biological information. *Zootaxa*, 1950(1), 39–50. <https://doi.org/10.11646/zootaxa.1950.1.6>
- Pyle, R. L., & Michel, E. (2010). ZooBank: Reviewing the first year and preparing for the next 250. In Polaszek A. (ed.) *Systema Naturae 250 – The Linnaean Ark* (pp. 173–184). CRC Press. <https://doi.org/10.1201/ebk1420095012-c16>
- Raynal-Roques, A., & Roguenant, A. (2008). Nommer et classer les plantes : un parcours de trois siècles. In D. Prat, A. Raynal-Roques, & A. Roguenant (Eds.), *Peut-on classer le vivant ? Linné et la systématique aujourd’hui*. Société Linnéenne de Lyon.
- Rees, T. (Compiler). (n.d.). *Index*. Interim Register of Marine and Nonmarine Genera. Retrieved September 21, 2023, from <https://www.irmng.org/index.php>
- Rees, T., Vandepitte, L., Vanhoorne, B., & Decock, W. (2020). All genera of the world: An overview and estimates based on the March 2020 release of the Interim Register of Marine and Nonmarine Genera (IRMNG). *Megataxa*, 1(2), 123–140. <https://doi.org/10.11646/megataxa.1.2.3>

- Remsen, D. (2016). The use and limits of scientific names in biological informatics. *ZooKeys*, 550, 207–223. <https://doi.org/10.3897/zookeys.550.9546>
- Reuters staff. (2018, January 31). Factbox: A chronology of Thomson Reuters Corp. *Reuters*. <https://www.reuters.com/article/us-thomsonreuters-f-r-blackstone-factbox-idUSKBN1FK085>
- Richard, J. M. (2020). Improving taxonomic name finding in the Biodiversity Heritage Library. *Biodiversity Information Science and Standards*, 4. <https://doi.org/10.3897/biss.4.58482>
- Rineau, V. (2022, November 30). *Systématique*. E-Systematica. <https://e-systematica.org/2022/11/30/systematique-valentin-rineau/>
- Rousse, G. (2003). *Vers une ontologie de la systématique: Les objets-métiers du domaine*. PhD thesis, Université Pierre et Marie Curie - Paris 6, Paris. i–viii + 1–204.
- Rowe, T., & Gauthier, J. A. (1992). Ancestry, paleontology, and definition of the name Mammalia. *Systematic Biology*, 41(3), 372–378. <https://doi.org/10.1093/sysbio/41.3.372>
- Saliba, E. M. (2022). Reflections on Pavlinov (2021): Understanding and expanding the study of the history and philosophy of nomenclature, a comprehensive state of the art. *Bionomina*, 31, 88–96. <https://doi.org/10.11646/bionomina.31.1.5>
- Saliba, E. M., Vignes-Lebbe, R., Dubois, A., & Ohler, A. (2021). ZooNom: An online thesaurus for alleviating ambiguity in the terminology of zoological nomenclature. *Zootaxa*, 5071(1), 42–50. <https://doi.org/10.11646/zootaxa.5071.1.2>
- Schuchert, C. (1897). What is a type in natural history? *Science*, 5(121), 636–640. <https://doi.org/10.1126/science.5.121.636>
- Schuchert, P. (ed.). (2023, July 3). World Hydrozoa Database. <https://www.marinespecies.org/hydrozoa>. <https://doi.org/10.14284/357>
- Segers, H., De Smet, W. H., Fischer, C. E. C., Fontaneto, D., Michaloudi, E., Wallace, R. L., & Jersabek, C. D. (2012). Towards a List of Available Names in zoology, partim phylum Rotifera. *Zootaxa*, 3179(1), 61. <https://doi.org/10.11646/zootaxa.3179.1.3>
- Sereno, P. (1986). Phylogeny of the bird-hipped dinosaurs (order Ornithischia). *National Geographic Research* 2(2), 234–256.
- Sharkey, M. J., Baker, A. J., McCluskey, K., Smith, A., Naik, S., Ratnasingham, S., Manjunath, R., Perez, K., Sones, J. E., D’Souza, M., St Jacques, B., Hajibabaei, M., Whitfield, J. B., Arias, D., Solís, M. A., Metz, M. A., Burns, J. M., Zúñiga, R., Phillips-Rodríguez, E., . . . Janzen, D. H. (2023). Minimalist revision of *Mesochorus*

- Gravenhorst, 1829 (Hymenoptera: Ichneumonidae: Mesochorinae) from Área de Conservación Guanacaste, Costa Rica, with 158 new species and host records for 129 species. *Revista de Biología Tropical*, 71(S2), 1–174.
<https://doi.org/10.15517/rev.biol.trop..v71is2.56316>
- Sharkey, M. J., Janzen, D. H., Hallwachs, W., Chapman, E. G., Smith, M. A., Dapkey, T., Brown, A. D., Ratnasingham, S., Naik, S., Manjunath, R., Perez, K., Milton, M. A., Hebert, P. D. N., Shaw, S. R., Kittel, R. N., Solís, M. A., Metz, M. A., Goldstein, P. Z., Brown, J. W., . . . Burns, J. M. (2021). Minimalist revision and description of 403 new species in 11 subfamilies of Costa Rican braconid parasitoid wasps, including host records for 219 species. *ZooKeys*, 1013, 1–665.
<https://doi.org/10.3897/zookeys.1013.55600>
- Shipunov, A. (2011). The problem of hemihomonyms and the on-line hemihomonyms database (HHDB). *Bionomina*, 4(1), 65–72. <https://doi.org/10.11646/bionomina.4.1.3>
- Simpson, G. G. (1940). Types in modern taxonomy. *American Journal of Science*, 238(6), 413–431. <https://doi.org/10.2475/ajs.238.6.413>
- Simpson, G. G. (1961). *Principles of Animal Taxonomy*. Columbia University Press. 1–247.
- Smith, G. F., & Figueiredo, E. (2021). “Rhodes-” must fall: Some of the consequences of colonialism for botany and plant nomenclature. *Taxon*, 71(1), 1–5.
<https://doi.org/10.1002/tax.12598>
- Smith, H. M., & Smith, R. B. (1972). Chresonymy ex Synonymy. *Systematic Zoology*, 21(4), 445. <https://doi.org/10.2307/2412440>
- Soltis, P. S. (2017). Digitization of herbaria enables novel research. *American Journal of Botany*, 104(9), 1281–1284. <https://doi.org/10.3732/ajb.1700281>
- Starobogatov, Y. I. (1991) Problems in the nomenclature of higher taxonomic categories. *The Bulletin of Zoological Nomenclature*, 48(1), 6–18.
- Strickland, H. E. (1845). On the progress and the present state of ornithology. *Report of the fourteenth meeting of the British Association for the Advancement of Science for 1844*, 170–221.
- Strickland, H.E., Philipps, J., Richardson, J., Owen, R., Jenyns, L., Broderip, W.J., Henslow, J.S., Shuckard, W.E., Waterhouse, G.R., Yarrell, W., Darwin, C., & Westwood J.O. (1843). Report of a committee appointed « to consider the rules by which the Nomenclature of Zoology may be established on a uniform and permanent basis ». *Report of the twelfth meeting of the British Association for the Advancement of Science for 1842*, 105–121.

- Sues, H.-D. (2019). Authorship and date of publication of the name Tetrapoda. *Journal of Vertebrate Paleontology*, 39(1), Art. e1564758.
<https://doi.org/10.1080/02724634.2019.1564758>
- Taxonomic Databases Working Group Minimum Standard for Names in Botanical Databases Subgroup. (1995). Plant Names in Botanical Databases (standard).
<http://www.tdwg.org/standards/106>
- Taxonomic Databases Working Group (TDWG). (n.d.). Standards. Retrieved August 25, 2023 from <https://www.tdwg.org/standards/>
- Taxonomic rank. (2023, September 12). In *Wikipedia*.
https://en.wikipedia.org/wiki/Taxonomic_rank
- Thorne, J. (2003). Zoological Record and registration of new names in zoology. *The Bulletin of Zoological Nomenclature*, 60(1), 7–11.
- Turland, N. J., Wiersema, J. H., Barrie, F. R., Greuter, W., Hawksworth, D. L., Herendeen, P. S., Knapp, S., Kusber, W.-H., Li, D.-Z., Marhold, K., May, T. W., McNeill, J., Monro, A. M., Prado, J., Price, M. J., & Smith, G. F. (eds.). (2018). *International Code of Nomenclature for algae, fungi, and plants (Shenzhen Code) adopted by the Nineteenth International Botanical Congress Shenzhen, China, July 2017*. Regnum Vegetabile 159. Koeltz Botanical Books. <https://doi.org/10.12705/Code.2018>
- Uetz, P. (ed.). (2023, May 7). The Reptile Database. <http://www.reptile-database.org>
- Van Bijkershoek, C., & (1737). *Quaestiones juris publici libri duo*. Translation by F. Tenney (1930) in *A photographic reproduction of the Edition of 1737, with a list of errata and a portrait*. Clarendon Press.
- Vendetti, J. E., & Garland, R. (2019). Species name formation for zoologists: a pragmatic approach. *Journal of Natural History*, 53(47–48), 2999–3018.
<https://doi.org/10.1080/00222933.2020.1754482>
- Vignes-Lebbe, R. (2023). FAIR principles and TDWG standards: The case of morphological description of taxa and specimens. *Biodiversity Information Science and Standards*, 7, Art. e111859. <https://doi.org/10.3897/biss.7.111859>
- Walker, F. O. (1850). *List of the specimens of homopterous insects in the collection of the British Museum*. Edward Newman. <https://doi.org/10.5962/bhl.title.9063>
- Waller, J. (2022). Finding data gaps in the GBIF backbone taxonomy. *Biodiversity Information Science and Standards* 6, Art. e91312.
<https://doi.org/10.3897/biss.6.91312>

- Welter-Schultes, F., Görlich, A., & Lutze, A. (2016). Sherborn's Index Animalium: New names, systematic errors and availability of names in the light of modern nomenclature. *ZooKeys*, 550, 173–187. <https://doi.org/10.3897/zookeys.550.10041>
- Westwood J. O. (1838). Nouvelles diverses. Fulgores. *Annales de la Société Entomologique de France, Bulletin Entomologique*, 7. XXXVIII-XXXIX [in the supplement].
- Westwood J. O. (1839). On the family Fulgoridae, with a monograph of the genus Fulgora of Linnaeus. *Transactions of the Linnean Society of London, 2nd Series. Zoology*, 18(2), 133–153. <https://doi.org/10.1111/j.1095-8339.1838.tb00167.x>
- Whewell, W. (1840). *The philosophy of the inductive sciences founded upon their history*. John W. Parker, 1–660.
- White, M. J. (1992). Legal complexity and lawyers' benefit from litigation. *International Review of Law and Economics*, 12(3), 381–395. [https://doi.org/10.1016/0144-8188\(92\)90016-k](https://doi.org/10.1016/0144-8188(92)90016-k)
- Wieczorek, J., Bloom, D., Guralnick, R., Blum, S., Döring, M., De Giovanni, R., Robertson, T., & Vieglaiss, D. (2012). Darwin Core: An evolving community-developed biodiversity data standard. *Plos One*, 7(1), Art. e29715. <https://doi.org/10.1371/journal.pone.0029715>
- Wiley, E. O. (1981). *Phylogenetics: The theory and practice of phylogenetic systematics*. Wiley, 1–439.
- Witteveen, J. (2016). Suppressing synonymy with a homonym: The emergence of the nomenclatural type concept in nineteenth century natural history. *Journal of the History of Biology*, 49(1), 135–189. <https://doi.org/10.1007/s10739-015-9410-y>
- Woodbury, A. M. (1952). Ecological Taxonomy. *Science*, 115(2992), 3. <https://doi.org/10.1126/science.115.2992.3.s>
- World Spider Catalog. (n.d.). *Frequently Asked Questions*. Retrieved September 18, 2023, from <https://wsc.nmbe.ch/faq>
- Wüster, W., Thomson, S., O'Shea, M., & Kaiser, H. (2021). Confronting taxonomic vandalism in biology: conscientious community self-organization can preserve nomenclatural stability. *Biological Journal of the Linnean Society*, 133(3), 645–670. <https://doi.org/10.1093/biolinnean/blab009>
- Yoder, M., Dmitriev, D. A., Pereira, J. L., & Cigliano, M. M. (2017). Design and use of NOMEN, an ontology defining the rules of biological nomenclature. *Biodiversity Information Science and Standards*. <https://doi.org/10.3897/tdwgproceedings.1.20284>

Zamani, A., Fric, Z. F., Gante, H. F., Hopkins, T., Orfinger, A. B., Scherz, M. D., Bartoňová, A. S., & Pos, D. D. (2022). DNA barcodes on their own are not enough to describe a species. *Systematic Entomology*, 47(3), 385–389. <https://doi.org/10.1111>

Appendix 1: Data dictionary of *Lognom*

Table name Field name	Data type	Mandatory	Description
Name			
Id	PK	True	ID of the table
name_name	Varchar (100)	True	The name contains the scientific name of an animal organism. It's the central element of this table and of the software. There is no official limit to the length of a scientific name, but none known name is longer than a couple dozen letters.
name_availability	Boolean	True	This Boolean controls the availability status of the name. An unavailable name does not exist in the eyes of Nomenclature, but is still important to recover all information on a given taxon.
name_date	Integer	True	These fields are reserved for the date of creation of a name. It should be the year, minimum. By default, it is always the last possible date, so 2021 makes 12/31/2021, and 05/2021 makes 05/31/2021.
name_day	Integer	True	The day of publication
name_month	Integer	True	The month of publication
name_author	Varchar (100)	True	The field contains the author(s) of the name
name_registration	Boolean	True	This Boolean is true if the name is properly registered on <i>ZooBank</i>
name_formula	Boolean	True	This Boolean is true if the name is a zoological formula
name_external_homonymy	Boolean	True	This Boolean is on if the name has a homonym outside of Zoological nomenclature
name_binominality	Boolean	True	This Boolean controls that the name is correct under the rules of binominality
name_LAN	Boolean	True	This Boolean verifies that the name is present on a List of Available Name. However, in reality, the LANs are incomplete and unusable
name_intent_new_name	Boolean	True	This Boolean verifies that a name is introduced with the explicit intent to create a new name and also not as a synonym to another pre-existing name
name_disclaimer	Boolean	True	This Boolean controls that the name is not preceded by a disclaimer
name_description_diagnose	Varchar (200)	True	This field contains the taxonomic information accompanying the name.
name_validity	Boolean	True	This Boolean controls the validity status of the name. The valid name is the correct name to use for a taxon
name_synonymy	Boolean	True	This Boolean controls if a name is a synonym
name_synonymy_type	Varchar (50)	True	This field corresponds to the two possible types of synonymies, objective and

			subjective. These two values are the only possible values
name_homonymy	Boolean	True	This Boolean controls if a name is a homonym
name_near_homonymy	Boolean	True	This Boolean controls if a name is a near homonym -resembles or is the same name as a name in a close genus- (species series only)
name_synonym	Varchar (200)	True	This field controls the fact that the name was introduced as a synonym or conditionally
name_homonymy_type	Varchar (50)	True	This field corresponds to the two possible types of homonymies, primary and secondary. These two values are the only possible values
name_reversal_priority	Varchar (50)	False	This field corresponds to the status of a name regarding reversal of priority. Not always relevant
name_rp_number	Varchar (50)	False	This field attaches the reversal of precedence case number to the name if necessary. Not always relevant.
name_availability_suppression	Varchar (50)	False	This field corresponds to the status of a name availability when there is an intervention by the Commission. Not always relevant
name_as_number	Varchar (50)	False	This field attaches the case number to the name if necessary. Not always relevant
name_first_reviewer	Varchar (70)	False	This field corresponds to the action of the first reviser. Not always relevant
name_status	Varchar (50)	False	This field contains the status of the name in an advanced vocabulary that is present in the Zoonom thesaurus
name_status_URI	Text	False	A link to the status concept in the thesaurus
name_etymology	Text	False	The etymology of the name
name_replacement	Boolean	True	This Boolean controls if the name is a new replacement name
name_alternative_spelling	Text	False	List of alternative spellings of a name that do not have an independent existence, like lapsus calami, or original incorrect spelling
name_grammatical_case	Varchar (50)	False	This field corresponds to the grammatical case in the species series. Three possibilities: adjective or participle, genitive or indeclinable
name_combination	Boolean	True	This Boolean controls if a specific name is introduced with a combination
name_grammatical_gender	Varchar (30)	False	This field corresponds to the grammatical gender in the genus series. Three possibilities only: masculine, feminine and neuter
name_correct_form	Varchar (50)	False	This field contains the correct form for a family series name
name_root	Varchar (50)	False	This field contains the root of a family series name
name_rank_id	FK	True	This is the foreign key that links the name to the one rank it belongs to
name_type_name_id	FK	False	This is the foreign key that links a generic or familial name to the corresponding type

name_invalid_reason	Varchar (200)	True	This field contains the reasons explaining why the name is invalid
name_infrasubspecific_name_use	Boolean	True	This field corresponds to the use of an infrasubspecific name as a species or subspecies name before 1981
name_genus_type_invalid	Boolean	True	This Boolean controls if the type genus of a family name is considered invalid "The type genus is considered invalid by the author"
name_as_date	Integer	True	This field attaches the suppression case date to the name if necessary. Not always relevant.
name_explicit_type	Boolean	True	This Boolean controls if the type of the generic or familial name is explicit
name_work_id	FK	False	This is the foreign key that links the name to the article or book in which it was introduced
name_ergotaxon_id	FK	False	This is the foreign key that links the name to the ergotaxon on which the name applies
name_protaxon_id	FK	False	This is the foreign key that links the name to the protaxon on which the name applies
name_original_combination	Varchar (100)	False	The original combination of a species name
name_published	Boolean	True	This Boolean is true if the name is published or false if it is to be published
name_nominal_type_id	FK	False	This is the foreign key connected to the name table that links a generic or familial name to the corresponding type
name_rp_related_name_id	FK	False	This field corresponds to the name against which the given name has a special status. E.g., A-us can be declared invalid against B-us but not C-us. Not always relevant

Work

id	PK	True	ID of the table
work_name	Varchar (800)	True	The title of the work, or a way to label it if necessary
work_date	Integer	True	The year of publication of the work.
work_copy	Boolean	True	This Boolean controls the fact that the work on CD in an edition containing simultaneously obtainable copies
work_registration	Boolean	True	This Boolean controls the fact that the work has been registered on <i>ZooBank</i>
work_obtainability	Boolean	True	This Boolean control the fact that the work is obtainable
work_date_statement	Boolean	True	This Boolean control the fact that the work states the date of publication in the work itself
work_disclaimer	Boolean	True	This Boolean is on if the work contains a disclaimer
work_permanent_record	Boolean	True	This Boolean controls the fact that the work was issued for the purpose of providing a public and permanent scientific record
work_statement	Boolean	True	This Boolean controls the presence of the distribution statement.
work_binominality	Varchar (200)	True	This Boolean is on if the work is binominal

work_archodeixy	Varchar (100)	True	This field lists the different possibilities of interventions by the Commission
work_archodeixy_number	Varchar (50)	True	This field attaches the case/Opinion number to the name if necessary. Not always relevant.
work_media	Varchar (100)	True	This field controls the type of media on which the work was published
work_availability	Boolean	True	This Boolean is on if the work is available
work_bibliography_id	FK	False	This FK links the work to its exact bibliography
Protaxon			
id	PK	True	ID of the table
taxon_working_description	Text	False	The taxon description, if needed
taxon_nature	Varchar (100)	True	This field contains a list of values that reflects the taxonomic nature of the element contained in the taxon
Ergotaxon			
id	PK	True	ID of the table
ergotaxon_work_id	Varchar (100)	True	The taxon work ID is a shortcut for the user to be able to organize his/her work. It can be a vernacular name or a theoretical naming (Taxon_34 or Subfamily_2)
ergotaxon_working_description	Text	False	The taxon description, if needed
Type Name			
Id	PK	True	ID of the table
type_name_name	Varchar (100)	True	The name contains the scientific name of an animal organism. In this case, it is the name of a scientific name used as a type for another
type_name_date	Integer	True	This field is reserved for the date of publication of a name. Only the year is of interest
type_name_author	Varchar (100)	True	This is the fields that contains the name to the author(s) who created the name
type_name_availability	Boolean	True	This Boolean controls the availability status of the name
type_name_suppression	Boolean	True	This Boolean controls if the genus name was suppressed
type_name_explicit	Boolean	True	This Boolean controls if the type name is explicit
Bibliography			
id	PK	True	ID of the table
bibliography_title	Text	False	Here is title of the relevant document where the nomenclatural act is contained.
bibliography_author	Text	False	The author(s) of the relevant document. They can be different from the author of the name.
bibliography_reference	Text	False	The whole bibliographical reference of the relevant document where the nomenclatural act is done.
bibliography_DOI	Text	False	The DOI of the publication
Onymophoront			
id	PK	True	ID of the table
onymophoront_number	Varchar (100)	True	The specimen number (in a collection) is an alphanumeric ID used in the relevant

			collection database. It is the preferred way to identify a specimen
onymophoront_institution	Varchar (300)	True	The institution in which the specimen(s) are stored
onymophoront_accessibility	Boolean	True	This Boolean controls if the onymophoront is accessible (not lost or missing)
onymophoront_label	Boolean	True	This Boolean controls if the onymophoront is labeled as a type
onymophoront_identification	Boolean	True	This Boolean controls if the onymophoront is identifiable
onymophoront_deposit_collection	Boolean	True	This Boolean controls if the onymophoront is explicitly said to be deposited in a collection
onymophoront_explicit_allocation	Boolean	True	This Boolean controls if the onymophoront is explicitly allocated to a name in the original publication
onymophoront_type_locality	Varchar (1000)	False	The type locality or onymotope of the onymophoront
Rank			
id	PK	True	ID of the table
rank_name	Varchar (100)	True	This field contains the rank name. It is the central element of this table.
rank_ending	Varchar (10)	False	The name contains the rank mandatory ending. Only 5 ranks have a mandatory ending, which are -oidea, -idae, -inae, -ini and -ina for ranks in between the Superfamily and the subtribe
rank_series_id	FK	True	This is the foreign key that links the rank to the one series (group) it belongs to.
Series			
id	PK	True	ID of the table
series_name	PK	True	This field contains the series name. It is the only element of this table.
Profile			
id	PK	True	ID of the table
profile_title	Varchar (30)	False	This field contains the title of the user (Dr., Pr. etc...)
profile_institution	Varchar (150)	True	This field contains the institution of the user
profile_user_id	FK	True	This field contains the corresponding user
User (Django generated)			
id	PK	True	ID of the table
password	Varchar (128)	True	This field contains the password of the user
last_login	datetime	False	This field contains the last time a user logged in
username	Varchar (150)	True	This field contains the username
last_name	Varchar (150)	True	This field contains the user's last name
email	Varchar (254)	True	This field contains the user's email
date_joined	datetime	True	This field contains the date when a user signed up
first_name	Varchar (150)	True	This field contains the user's first name

Nomfile			
id	PK	True	ID of the table
nomfile_name	Varchar (100)	True	This field contains the name of the file
nomfile_privacy	Boolean	True	This field controls the privacy of the file. True if public.
nomfile_user_id	FK	True	This field contains the user who created the file
Taxomen			
id	PK	True	ID of the table
taxomen_name	FK	True	The foreign key to the Name table
taxomen_onymophoront	FK	True	The foreign key to the Onymophoront table
NamexNomfile			
id	PK	True	ID of the table
namexnomfile_name	FK	True	The foreign key to the Name table
namexnomfile_nomfile	FK	True	The foreign key to the nomenclatural file table
WorkxNomfile			
id	PK	True	ID of the table
workxnomfile_work	FK	True	The foreign key to the Work table
workxnomfile_nomfile	FK	True	The foreign key to the nomenclatural file table
ErgotaxonxNomfile			
id	PK	True	ID of the table
ergotaxonxnomfile_ergotaxon	FK	True	The foreign key to the Ergotaxon table
ergotaxonxnomfile_nomfile	FK	True	The foreign key to the nomenclatural file table

Appendix 2: List of the Articles of the Code and their treatment in *Lognom*

Article Number	Status in <i>Lognom</i>	Treatment position	Comment
1.1.	N/A		Definition
1.1.1	Mentioned in text	All PublishName, All Availability	
1.2.1	Mentioned in text	HomePage	
1.2.2.	Implicit throughout the software		
1.3.1.	Stored in the database	All PublishName, All Availability	
1.3.2.	Stored in the database	All PublishName, All Availability	
1.3.3.	Stored in the database	All PublishName, All Availability	
1.3.4.	Redundant with Art. 45.6.4.1.		
1.3.5.	Stored in the database	SpeciesNameAvailability	
1.3.6.	Stored in the database	All PublishName, All Availability	
1.3.7.	Stored in the database	All Availability	
1.4.	Implicit throughout the software		
2.1.	Redundant with Art. 10.5.		
2.2.	Mentioned in text	On homonymy	
3.1.	To be added eventually as default names in the database		
3.2.	Stored in the database	All Availability	
4.1.	Stored in the database	All PublishName, All Availability	
4.2.	Implicit throughout the software		
5.1.	Stored in the database	All PublishName, All Availability	
5.1.1.	Redundant with Art. 11.4.		
5.1.2	Stored in the database	All PublishName, All Availability	
5.2.	Stored in the database	All PublishName, All Availability	
5.3.	Mentioned in text	On names, in All PublishName, All Availability	
6.1.	Implicit throughout the software		
6.2.	Missing		Might be added eventually

7.	N/A		Application note
8.1.	N/A		Application note
8.1.1.	Stored in the database	WorkAvailability	
8.1.2.	Stored in the database		
8.1.3.	Redundant with Art. 8.4.1.		
1.			
8.1.3.	Stored in the database	WorkAvailability, PublishWork	
2.			
8.2.	Stored in the database	WorkAvailability, PublishWork	
8.3.	Stored in the database	All PublishName, All Availability	
8.4.	Stored in the database	WorkAvailability, PublishWork	
8.4.1.	Stored in the database	WorkAvailability, PublishWork	
8.4.2.	Stored in the database	WorkAvailability, PublishWork	
8.4.2.	Stored in the database	WorkAvailability, PublishWork	
1.			
8.4.2.	Stored in the database	WorkAvailability, PublishWork	
2.			
8.5.1.	Stored in the database	WorkAvailability, PublishWork	
8.5.2.	Stored in the database	WorkAvailability, PublishWork	
8.5.3.	Stored in the database	WorkAvailability, PublishWork	
8.5.3.	Out of scope		Should be treated by <i>ZooBank</i>
1.			
8.5.3.	Out of scope		Should be treated by <i>ZooBank</i>
2.			
8.5.3.	Out of scope		Should be treated by <i>ZooBank</i>
3.			
8.6.	N/A		About the Commission or the structure of the Code
8.7.	Stored in the database	WorkAvailability	
8.7.1.	Mentioned in text		
8.8.	Out of scope		
9.1.	Stored in the database	WorkAvailability, PublishWork	
9.2.	Stored in the database	WorkAvailability, PublishWork	
9.3.	Stored in the database	WorkAvailability, PublishWork	
9.4.	Stored in the database	WorkAvailability, PublishWork	
9.5.	Stored in the database	WorkAvailability, PublishWork	
9.6.	Stored in the database	WorkAvailability, PublishWork	
9.7.	Stored in the database	WorkAvailability, PublishWork	
9.8.	Stored in the database	WorkAvailability, PublishWork	
9.9.	Stored in the database	WorkAvailability, PublishWork	
9.10.	Stored in the database	WorkAvailability, PublishWork	
9.11.	Stored in the database	WorkAvailability, PublishWork	
9.12.	Stored in the database	WorkAvailability, PublishWork	
10.1	Mentioned in text		

10.1.1.	Mentioned in text		
10.2.	Stored in the database	SpeciesNameAvailability	Implicit in PublishSpeciesName
10.3.	Missing		Might be added eventually
10.4.	Missing		Might be added eventually
10.5.	Out of scope		Needs to be referred to the ICN/ICNB
10.6.	Implicit throughout the software		
10.7.	Stored in the database	All PublishName, All Availability	
11.1.	Stored in the database	All PublishName, All Availability	
11.2.	Automatically corrected	ControlAlphabet	
11.3.	Cannot be verified		Relies on Semantics
11.4.	Stored in the database	All PublishName, All Availability	
11.4.1.	Stored in the database	All PublishName, All Availability	
11.4.2.	Stored in the database	All PublishName, All Availability	
11.4.3.	Stored in the database	WorkAvailability	
11.5.	Stored in the database	All PublishName, All Availability	
11.5.1.	Stored in the database	All PublishName, All Availability	
11.5.2.	Stored in the database	All PublishName, All Availability	
11.6.	Stored in the database	All PublishName, All Availability	
11.6.1.	Stored in the database	All PublishName, All Availability	
11.6.2.	Stored in the database	All PublishName, All Availability	
11.6.3.	Stored in the database	All PublishName, All Availability	
11.7.1.1.	Stored in the database	All PublishName, All Availability, ControlAlphabet	
11.7.1.2.	Mentioned in text	Family name pages	
11.7.1.3.	Automatically corrected	ControlAlphabet	

11.7. 1.4.	Stored in the database	FamilyNameAvailability, PublishFamilyName	
11.7. 1.5.	Stored in the database		
11.7. 2.	Missing		Might be added eventually
11.8.	Automatically corrected	ControlAlphabet	
11.8. 1.	Cannot be verified		Relies on Latin rules
11.9. 1.	Automatically corrected	ControlAlphabet	
11.9. 1.1.	Stored in the database	SpeciesNameAvailability, PublishSpeciesName	
11.9. 1.2.	Stored in the database	SpeciesNameAvailability, PublishSpeciesName	
11.9. 1.3.	Stored in the database	SpeciesNameAvailability, PublishSpeciesName	
11.9. 1.4.	Stored in the database	SpeciesNameAvailability, PublishSpeciesName	
11.9. 2.	Cannot be verified		Relies on Latin rules
11.9. 3.	Stored in the database	SpeciesNameAvailability, PublishSpeciesName	
11.9. 3.1.	Mentioned in text	On combination, SpeciesAvailability	
11.9. 3.2.	Mentioned in text	On combination, SpeciesAvailability	
11.9. 3.3.	Mentioned in text	On combination, SpeciesAvailability	
11.9. 3.4.	Mentioned in text	On combination, SpeciesAvailability	
11.9. 3.5.	Missing		Might be added eventually
11.9. 3.6.	Stored in the database	SpeciesNameAvailability	
11.9. 4.	Partially corrected automatically	ControlAlphabet	Relies on Latin rules
11.9. 5.	Cannot be verified		Relies on Latin rules
11.10 .	Missing		Might be added eventually
12.1.	Stored in the database	All Availability	
12.2.	Stored in the database	All Availability	
12.2. 1	Stored in the database	All Availability	
12.2. 2.	Stored in the database	All Availability	
12.2. 3.	Stored in the database	All Availability	

12.2.4.	Stored in the database	All Availability	
12.2.5.	Stored in the database	All Availability	
12.2.6.	Stored in the database	All Availability	Indirectly
12.2.7.	Stored in the database	All Availability	
12.2.8.	Stored in the database	All Availability	
12.3.	Stored in the database	All Availability	
13.1.	Stored in the database	All PublishName, All Availability	
13.1.1.	Stored in the database	All PublishName, All Availability	
13.1.2.	Stored in the database	All PublishName, All Availability	
13.1.3.	Stored in the database	All PublishName, All Availability	
13.2.	Stored in the database	FamilyNameAvailability, PublishFamilyName	
13.2.1.	Stored in the database	FamilyNameAvailability	
13.3.	Stored in the database	GenusNameAvailability, PublishGenusName	
13.3.1.	Missing		Might be added eventually
13.3.2.	Missing		Might be added eventually
13.3.3.	Missing		Might be added eventually
13.4.	Stored in the database	SpeciesNameAvailability, GenusNameAvailability, PublishSpeciesName, PublishGenusName	
13.5.	Stored in the database	FamilyNameAvailability, GenusNameAvailability, PublishFamilyName, PublishGenusName	
13.6.1.	Stored in the database	All PublishName, All Availability	
13.6.2.	Stored in the database	All PublishName, All Availability	
14.	Stored in the database	All PublishName, All Availability	
15.1.	Stored in the database	All PublishName, All Availability	
15.2.	Stored in the database	SpeciesNameAvailability, PublishSpeciesName	

15.2.1.	Redundant with Art. 45.6.4.		
16.1.	Stored in the database	All PublishName, All Availability	
16.2.	Stored in the database	FamilyNameAvailability, PublishFamilyName	
16.3.	Redundant with Art. 13.3.3. and Art. 13.3.2.		
16.4.1.	Stored in the database	SpeciesNameAvailability, PublishSpeciesName	
16.4.2.	Stored in the database	SpeciesNameAvailability, PublishSpeciesName	
17.1.	Implicit throughout the software		
17.2.	Implicit throughout the software		
17.3.	Implicit throughout the software		
18.	Implicit throughout the software		
19.1.	Missing		Might be added eventually
19.2.	Missing		Might be added eventually
19.3.	Missing		Might be added eventually
19.4.	Missing		Might be added eventually
20.	Stored in the database	FamilyNameAvailability, GenusNameAvailability, PublishFamilyName, PublishGenusName	
21.1.	Missing		Might be added eventually
21.2.	Missing		Might be added eventually
21.3.	Missing		Might be added eventually
21.3.1.	Missing		Might be added eventually
21.3.2.	Missing		Might be added eventually
21.4.	Missing		Might be added eventually
21.5.	Missing		Might be added eventually
21.6.	Missing		Might be added eventually

21.7.	Missing	Might be added eventually
21.7.1.	Missing	Might be added eventually
21.7.2.	Missing	Might be added eventually
21.8.1.	Missing	Might be added eventually
21.8.2.	Missing	Might be added eventually
21.8.3.	Missing	Might be added eventually
21.9.	Missing	Might be added eventually
22.	Implicit throughout the software	
23.1.	N/A	Definition
23.1.1.	Redundant with Art. 35.5. and Art. 40.	
23.1.2.	Redundant with Art. 55.3.1.1.	
23.1.3.	Redundant with Art. 20. and Art. 23.7.	
23.1.4.	Redundant with Art. 23.7.3.; 23.8.	
23.2.	N/A	Definition
23.3.	N/A	Application note
23.3.1.	Implicit throughout the Validity part of the software	
23.3.2.	Implicit throughout the Validity part of the software	
23.3.2.1.	Implicit throughout the Validity part of the software	
23.3.2.2.	Implicit throughout the Validity part of the software	
23.3.2.3.	Implicit throughout the Validity part of the software	
23.3.3.	Missing	Might be added eventually
23.3.4.	N/A	Application note
23.3.5.	N/A	Application note

23.3.6.	Implicit throughout the Validity part of the software		
23.3.7.	Missing		Might be added eventually
23.4.	Stored in the database	Validity	
23.4.1.	Redundant with Art. 55.; 39.		
23.5.	Missing		Might be added eventually
23.6.	Missing		Might be added eventually
23.7.1.	Missing		Might be added eventually
23.7.2.	Missing		Might be added eventually
23.7.3.	Missing		Might be added eventually
23.8.	Missing		Might be added eventually
23.9.1.	In a tool	NomenOblitum	
23.9.1.1.	In a tool	NomenOblitum	
23.9.1.2.	In a tool	NomenOblitum	
23.9.2.	In a tool	NomenOblitum	In text
23.9.3.	In a tool	NomenOblitum	In text
23.9.4.	Missing		Might be added eventually
23.9.5.	Missing		Might be added eventually
23.9.6.	In a tool	NomenOblitum	In text
23.10.	In a tool	NomenOblitum	In text
23.11.	In a tool	NomenOblitum	In text
23.12.	Missing		Might be added eventually
23.12.1.	Missing		Might be added eventually
23.12.2.	Missing		Might be added eventually
24.1.	Stored in the database	Validity	In part only, homonymy excluded

24.2.1.	N/A		Definition
24.2.2.	N/A		Application note
24.2.3.	Missing		Might be added eventually
24.2.4.	Missing		Might be added eventually
24.2.5.	Out of scope		
25.	N/A		Application note
26.	Cannot be verified		Relies on Latin and Greek rules
27.	Corrected automatically	ControlAlphabet	
28.	Implicit throughout the software		
29.1.	Stored in the database	FamilyNameAvailability, PublishFamilyName, ControlEnding	In part only, needs human interpretation
29.2.	Stored in the database	FamilyNameAvailability, PublishFamilyName, ControlEnding	
29.2.1.	Implicit throughout the software		
29.3.	N/A		Definition
29.3.1.	Cannot be verified		Relies on Latin rules
29.3.1.1.	Cannot be verified		Relies on Latin rules
29.3.2.	Cannot be verified		Relies on Latin rules
29.3.3.	Stored in the database	FamilyNameAvailability, PublishFamilyName	
29.4.	Absent		Might be added eventually
29.4.1.	Stored in the database	ControlEnding	
29.4.2.	Stored in the database	FamilyNameAvailability, PublishFamilyName	Might be added eventually
29.5.	Cannot be verified		Relies on the concept of Usage
29.6.	Mentioned in text	PublishFamilyName	
30.	N/A		Application note
30.1.1.	In a tool	GenusGender	In part only, relies on Latin rules
30.1.2.	In a tool	GenusGender	In part only, relies on Greek rules
30.1.3.	In a tool	GenusGender	In part only, relies on Latin rules

30.1.	Redundant with Art.		
4.1.	30.2.2.		
30.1.	In a tool	GenusGender	As text only, relies on Semantics
4.2.			
30.1.	Stored in the database, In a tool	GenusNameAvailability, PublishGenusName	In part only, needs human interpretation
4.3.			
30.1.	In a tool	GenusGender	In part only, needs human interpretation
4.4.			
30.1.	In a tool	GenusGender	As text only, relies on Semantics
4.5.			
30.2.	In a tool	GenusGender	In part only, relies on modern languages rules
1.			
30.2.	In a tool	GenusGender	
2.			
30.2.	In a tool	GenusGender	
3.			
30.2.	In a tool	GenusGender	
4.			
31.1.	Stored in the database	SpeciesNameAvailability, PublishSpeciesName	
31.1.	Cannot be verified		Relies on Latin rules
1.			
31.1.	Cannot be verified		Relies on Latin rules
2.			
31.1.	Out of scope		Might be added eventually
3.			
31.2.	Cannot be verified		Relies on Latin rules
31.2.	Cannot be verified		Relies on Semantics
1.			
31.2.	Cannot be verified		Relies on Semantics
2.			
31.2.	Cannot be verified		Relies on Semantics
3.			
32.1.	N/A		Definition
32.2.	N/A		Definition
32.2.	Absent		Might be added eventually
1.			
32.2.	Absent		Might be added eventually
2.			
32.3.	Implicit throughout the software		
32.4.	N/A		Definition

32.5.1.	Absent		Might be added eventually
32.5.1.1.	Absent		Might be added eventually
32.5.2.	Corrected automatically	ControlAlphabet	
32.5.2.1.	Corrected automatically	ControlAlphabet	
32.5.2.2.	Corrected automatically	ControlAlphabet	
32.5.2.3.	Corrected automatically	ControlAlphabet	
32.5.2.4.	Cannot be verified		Relies on Semantics
32.5.2.4.1.	Cannot be verified		Relies on Semantics
32.5.2.4.2.	Cannot be verified		Relies on Semantics
32.5.2.4.3.	Corrected automatically	ControlAlphabet	
32.5.2.4.4.	Cannot be verified		Relies on Semantics
32.5.2.5.	Corrected automatically		
32.5.2.6.	Corrected automatically	ControlAlphabet	
32.5.2.7.	Cannot be verified		Relies on Latin rules
32.5.3.1.	Corrected automatically	ControlEnding	
32.5.3.2.	Absent		Might be added eventually
32.5.3.3.	Absent		Might be added eventually
32.5.3.4.	Absent		Might be added eventually
33.1.	N/A		Definition
33.2.	Absent		Might be added eventually
33.2.1.	Absent		Might be added eventually
33.2.2.	N/A		Definition
33.2.3.	Absent		Might be added eventually
33.2.3.1.	Absent		Might be added eventually
33.3.	Absent		Might be added eventually

33.3.1.	Cannot be verified		Relies on the concept of Usage
33.4.	Absent		Might be added eventually
33.5.	Absent		Might be added eventually
34.1.	Corrected automatically	ControlEnding	
34.2.	Cannot be verified		Relies on Latin rules
34.2.1.	Absent		Might be added eventually
35.1.	N/A		Definition
35.2.	N/A		Definition
35.3.	N/A		Definition
35.4.	Absent		Might be added eventually
35.4.1.	Absent		Might be added eventually
35.4.2.	Absent		Might be added eventually
35.5.	Cannot be verified		Relies on the concept of Usage
36.1.	Implicit throughout the software		
36.2.	Implicit throughout the software		
37.1.	N/A		Definition
37.2.	N/A		Definition
38.	Redundant with Art. 39. and Art. 55.		
39.	Stored in the database	Validity	
40.1.	Implicit throughout the software		
40.2.	Cannot be verified		Relies on the concept of Usage
40.2.1.	Cannot be verified		Relies on the concept of Usage
41.	Redundant with Art. 65.2.		
42.1.	N/A		Definition
42.2.	Missing		Might be added eventually
42.2.1.	N/A		Definition
42.3.	N/A		Definition
42.3.1.	N/A		Definition
42.3.2.	N/A		Definition

42.4.	Redundant with Art. 10.3., Art. 10.4., Art. 11.8. and Art. 25. to 33.		
43.1.	N/A		Definition
43.2.	N/A		Definition
44.1.	N/A		Definition
44.2.	N/A		Definition
45.1.	N/A		Definition
45.2.	N/A		Application note
45.3.	N/A		Definition
45.4.	Redundant with Art. 11., Art. 19., Art. 20. and Art. 23.-34.		
45.5.	Stored in the database	SpeciesNameAvailability	In part only
45.5. 1.	Stored in the database	SpeciesNameAvailability	In part only
45.6. 1.	Stored in the database	SpeciesNameAvailability	
45.6. 2.	Stored in the database	SpeciesNameAvailability	
45.6. 3.	Stored in the database	SpeciesNameAvailability	
45.6. 4.	Stored in the database	SpeciesNameAvailability	
45.6. 4.1.	Stored in the database	SpeciesNameAvailability	
46.1.	N/A		Definition
46.2.	N/A		Definition
47.1.	N/A		Definition
47.2.	N/A		Definition
48.	N/A		Definition
49.	Missing		Might be added eventually
50.1.	N/A		Definition
50.1. 1	Missing		Might be added eventually
50.1. 2.	Missing		Might be added eventually
50.1. 3.	N/A		Application note
50.2.	Missing		Might be added eventually
50.3.	Missing		Might be added eventually
50.3. 1.	Implicit throughout the software		
50.3. 2.	Absent		Might be added eventually

50.4.	Missing		Might be added eventually
50.5.	Missing		Might be added eventually
50.6.	Redundant with Art. 24.		
50.7.	Missing		Might be added eventually
51.1.	Implicit throughout the software		
51.2.	Missing		Might be added eventually
51.2.1.	Out of scope		
51.3.	Missing		Might be added eventually
51.3.1.	Missing		Might be added eventually
51.3.2.	Missing		Might be added eventually
51.3.3.	Missing		Might be added eventually
52.1.	N/A		Definition
52.2.	Stored in the database	Validity	
52.3.	Missing		Might be added eventually
52.4.	Redundant with Art. 23.3.5., Art. 23.9.5., Art. 39., Art. 55. and Art. 60.		
52.5.	Redundant with Art. 54.4. and Art. 81.2.1.		
52.6.	Missing		Might be added eventually
52.7.	Implicit throughout the software		
53.1.	N/A		Definition
53.2.	N/A		Definition
53.3.	N/A		Definition
53.3.1.	N/A		Definition
54.1.	Missing		Might be added eventually
54.2.	Missing		Might be added eventually
54.3.	Missing		Might be added eventually
54.4.	Missing		Might be added eventually
55.1.	N/A		Application note
55.2.	Redundant with Art. 39.		

55.3.	Missing		Might be added eventually
55.3.1.	Missing		Might be added eventually
55.4.	Missing		Might be added eventually
55.5.	Missing		Might be added eventually
56.1.	N/A		Application note
56.2.	Missing		Might be added eventually
56.3.	Missing		Might be added eventually
57.1.	Mentioned in text	SpeciesNameAvailability, PublishSpeciesName	
57.2.	Missing		Might be added eventually
57.2.1.	Missing		Might be added eventually
57.2.2.	Missing		Might be added eventually
57.2.3.	Missing		Might be added eventually
57.3.	Missing		Might be added eventually
57.3.1.	Missing		Might be added eventually
57.3.2.	Missing		Might be added eventually
57.4.	Missing		Might be added eventually
57.5.	Missing		Might be added eventually
57.6.	In a tool		In part only
57.7.	Missing		Might be added eventually
57.8.	Missing		Might be added eventually
57.8.1.	Missing		Might be added eventually
57.8.2.	Missing		Might be added eventually
58.	Mentioned in text	ControlHomonymy	
58.1.	In a tool	ControlHomonymy	
58.2.	In a tool	ControlHomonymy	
58.3.	In a tool	ControlHomonymy	
58.4.	In a tool	ControlHomonymy	
58.5.	In a tool	ControlHomonymy	

58.6.	Absent		Needs human interpretation
58.7.	In a tool	ControlHomonymy	
58.8.	In a tool	ControlHomonymy	
58.9.	In a tool	ControlHomonymy	
58.10	In a tool	ControlHomonymy	
.			
58.11	In a tool	ControlHomonymy	
.			
58.12	Absent		Needs human interpretation
.			
58.13	In a tool	ControlHomonymy	
.			
58.14	In a tool	ControlHomonymy	In part only, needs human interpretation
.			
58.15	In a tool	ControlHomonymy	In part only, needs human interpretation
.			
59.1.	Stored in the database	Validity	
59.2.	Out of scope		Might be added eventually
59.3.	Out of scope		Might be added eventually
59.3.1.	Out of scope		Might be added eventually
59.4.	Out of scope		Might be added eventually
60.1.	Out of scope		Might be added eventually
60.2.	Implicit throughout the Validity part of the software		
60.2.1.	Implicit throughout the Validity part of the software		
60.3.	Implicit throughout the Validity part of the software		
61.1.	N/A		Definition
61.1.1.	N/A		Definition
61.1.2.	N/A		Definition
61.1.3.	N/A		Definition
61.2.	N/A		Definition
61.2.1.	Absent		Might be added eventually

61.2.2.	N/A		Definition
61.3.1.	Stored in the database, Implicit throughout the Validity part of the software	Validity	
61.3.2.	Missing		Might be added eventually
61.3.3.	Stored in the database	Validity	In part only
61.3.4.	Stored in the database	Validity	
61.4.	Missing		Might be added eventually
62.	N/A		Application note
63.	N/A		Definition
63.1.	N/A		Definition
64.	Missing		Might be added eventually
65.1.	N/A		Definition
65.2.	Missing		Might be added eventually
65.2.1.	Missing		Might be added eventually
65.2.2.	Missing		Might be added eventually
65.2.3.	Missing		Might be added eventually
66.	N/A		Application note
66.1.	Stored in the database	GenusNameAvailability, PublishGenusName	
67.1.	N/A		Definition
67.1.1.	N/A		Definition
67.1.2.	N/A		Definition
67.2.	Missing		Might be added eventually
67.2.1.	Missing		Might be added eventually
67.2.2.	Missing		Might be added eventually
67.2.3.	Missing		Might be added eventually
67.2.4.	Missing		Might be added eventually
67.2.5.	Missing		Might be added eventually

67.3.	Missing	Might be added eventually
67.3.1.	Missing	Might be added eventually
67.3.2.	Missing	Might be added eventually
67.4.	Missing	Might be added eventually
67.4.1.	Missing	Might be added eventually
67.5.	Missing	Might be added eventually
67.5.1.	Missing	Might be added eventually
67.5.2.	Missing	Might be added eventually
67.5.3.	Missing	Might be added eventually
67.6.	Missing	Might be added eventually
67.7.	Missing	Might be added eventually
67.8.	Missing	Might be added eventually
67.8.1.	Missing	Might be added eventually
67.9.	Missing	Might be added eventually
67.10.	Missing	Might be added eventually
67.11.	Missing	Might be added eventually
67.12.	Missing	Might be added eventually
67.12.1.	Missing	Might be added eventually
67.13.	Missing	Might be added eventually
67.13.1.	Missing	Might be added eventually
67.13.2.	Missing	Might be added eventually
67.14.	Missing	Might be added eventually
68.1.	Missing	Might be added eventually
68.2.	Missing	Might be added eventually

68.2.1.	Missing	Might be added eventually
68.2.2.	Missing	Might be added eventually
68.3.	Missing	Might be added eventually
68.3.1.	Missing	Might be added eventually
68.4.	Missing	Might be added eventually
68.5.	Missing	Might be added eventually
68.6.	Missing	Might be added eventually
69.	Missing	Might be added eventually
69.1.	Missing	Might be added eventually
69.1.1.	Missing	Might be added eventually
69.1.2.	Missing	Might be added eventually
69.2.	Missing	Might be added eventually
69.2.1.	Missing	Might be added eventually
69.2.2.	Missing	Might be added eventually
69.2.3.	Missing	Might be added eventually
69.2.4.	Missing	Might be added eventually
69.3.	Missing	Might be added eventually
69.4.	Missing	Might be added eventually
70.1.	Missing	Might be added eventually
70.1.1.	Missing	Might be added eventually
70.1.2.	Missing	Might be added eventually
70.2.	Missing	Might be added eventually
70.3.	Missing	Might be added eventually
70.3.1.	Missing	Might be added eventually

70.3.2.	Missing		Might be added eventually
70.4.	Missing		Might be added eventually
70.4.1.	Missing		Might be added eventually
70.4.2.	Missing		Might be added eventually
71.	N/A		Application note
72.1.	N/A		Definition
72.1.1.	N/A		Definition
72.1.2.	N/A		Definition
72.1.3.	N/A		Definition
72.2.	Missing		Might be added eventually
72.3.	Stored in the database	SpeciesNameAvailability, PublishSpeciesName	
72.4.1.	N/A		Definition
72.4.1.1.	Missing		Might be added eventually
72.4.2.	Missing		Might be added eventually
72.4.3.	Missing		Might be added eventually
72.4.4.	Missing		Might be added eventually
72.4.5.	Missing		Might be added eventually
72.4.6.	Missing		Might be added eventually
72.4.7.	Missing		Might be added eventually
72.5.	Missing		Might be added eventually
72.5.1.	Missing		Might be added eventually
72.5.2.	Missing		Might be added eventually
72.5.3.	Missing		Might be added eventually
72.5.4.	Missing		Might be added eventually
72.5.5.	Missing		Might be added eventually

72.5.6.	Missing	Might be added eventually
72.6.	Missing	Might be added eventually
72.7.	Missing	Might be added eventually
72.8.	Missing	Might be added eventually
72.9	Missing	Might be added eventually
72.10	Missing	Might be added eventually
73.1.1.	N/A	Definition
73.1.1.	N/A	Definition
73.1.2.	Missing	Might be added eventually
73.1.3.	N/A	Definition
73.1.4.	N/A	Definition
73.1.5.	Missing	Might be added eventually
73.2.1.	N/A	Definition
73.2.1.	N/A	Definition
73.2.1.1.	N/A	Definition
73.2.2.	N/A	Definition
73.2.3.	N/A	Definition
73.3.1.	N/A	Definition
73.3.1.	N/A	Definition
73.3.2.	Missing	Might be added eventually
74.1.1.	Missing	Might be added eventually
74.1.2.	Missing	Might be added eventually
74.1.3.	Missing	Might be added eventually
74.2.	Missing	Might be added eventually
74.3.	Missing	Might be added eventually

74.4.	Missing	Might be added eventually
74.5.	Missing	Might be added eventually
74.6.	Missing	Might be added eventually
74.6.1.	Missing	Might be added eventually
74.6.1.1.	Missing	Might be added eventually
74.6.1.2.	Missing	Might be added eventually
74.7.	Missing	Might be added eventually
74.7.1.	Missing	Might be added eventually
74.7.2.	Missing	Might be added eventually
74.7.3.	Missing	Might be added eventually
75.1.	N/A	Definition
75.2.	N/A	Definition
75.3.	Missing	Might be added eventually
75.3.1.	Missing	Might be added eventually
75.3.2.	Missing	Might be added eventually
75.3.3.	Missing	Might be added eventually
75.3.4.	Missing	Might be added eventually
75.3.5.	Missing	Might be added eventually
75.3.6.	Missing	Might be added eventually
75.3.7.	Missing	Might be added eventually
75.4.	Missing	Might be added eventually
75.4.1.	Missing	Might be added eventually
75.5.	Missing	Might be added eventually
75.6.	Missing	Might be added eventually
75.7.	Missing	Might be added eventually

75.8.	Missing	Might be added eventually
76.1.	N/A	Definition
76.1.1.	N/A	Definition
76.2.	N/A	Definition
76.3.	N/A	Definition
77.1.	N/A	About the Commission or the structure of the Code
77.2.	N/A	About the Commission or the structure of the Code
77.3.	N/A	About the Commission or the structure of the Code
77.3.1.	N/A	About the Commission or the structure of the Code
77.3.2.	N/A	About the Commission or the structure of the Code
77.3.3.	N/A	About the Commission or the structure of the Code
77.3.4.	N/A	About the Commission or the structure of the Code
77.3.5.	N/A	About the Commission or the structure of the Code
77.4.	N/A	About the Commission or the structure of the Code
77.5.	N/A	About the Commission or the structure of the Code
78.1.	N/A	About the Commission or

		the structure of the Code
78.2.	N/A	About the Commission or the structure of the Code
78.2. 1.	N/A	About the Commission or the structure of the Code
78.2. 2.	N/A	About the Commission or the structure of the Code
78.2. 3.	N/A	About the Commission or the structure of the Code
78.2. 4.	N/A	About the Commission or the structure of the Code
78.3.	N/A	About the Commission or the structure of the Code
78.3. 1.	N/A	About the Commission or the structure of the Code
78.3. 2.	N/A	About the Commission or the structure of the Code
78.3. 3.	N/A	About the Commission or the structure of the Code
78.4.	N/A	About the Commission or the structure of the Code
78.4. 1.	N/A	About the Commission or the structure of the Code
78.4. 2.	N/A	About the Commission or the structure of the Code

78.4.3.	N/A	About the Commission or the structure of the Code
78.4.4.	N/A	About the Commission or the structure of the Code
78.4.5.	N/A	About the Commission or the structure of the Code
79.1.	Out of scope	About the LANs
79.1.1.	Out of scope	About the LANs
79.1.2.	Out of scope	About the LANs
79.1.3.	Out of scope	About the LANs
79.1.4.	Out of scope	About the LANs
79.1.5.	Out of scope	About the LANs
79.2.1.	Out of scope	About the LANs
79.2.2.	Out of scope	About the LANs
79.2.2.1.	Out of scope	About the LANs
79.2.2.2.	Out of scope	About the LANs
79.2.2.3.	Out of scope	About the LANs
79.2.2.4.	Out of scope	About the LANs
79.2.2.5.	Out of scope	About the LANs
79.2.2.6.	Out of scope	About the LANs
79.3.	Out of scope	About the LANs
79.3.1.	Out of scope	About the LANs
79.3.2.	Out of scope	About the LANs
79.3.3.	Out of scope	About the LANs
79.4.	Out of scope	About the LANs
79.4.1.	Out of scope	About the LANs

79.4.2.	Out of scope	About the LANs
79.4.3.	Out of scope	About the LANs
79.5.	Out of scope	About the LANs
79.5.1.	Out of scope	About the LANs
79.5.2.	Out of scope	About the LANs
79.6.	Out of scope	About the LANs
80.1.	N/A	About the Commission or the structure of the Code
80.2.	N/A	About the Commission or the structure of the Code
80.2.1.	N/A	About the Commission or the structure of the Code
80.2.2.	N/A	About the Commission or the structure of the Code
80.3.	N/A	About the Commission or the structure of the Code
80.4.	N/A	About the Commission or the structure of the Code
80.5.	N/A	About the Commission or the structure of the Code
80.6.	N/A	Definition
80.6.1.	Out of scope	About the OL
80.6.2.	Out of scope	About the OL
80.6.3.	Out of scope	About the OL
80.6.4.	Out of scope	About the OL
80.6.5.	Out of scope	About the OL

80.7.	Out of scope		About the OI
80.7.1.	Out of scope		About the OI
80.7.2.	Out of scope		About the OI
80.8.	Out of scope		About the OI
80.9.	N/A		About the Commission or the structure of the Code
81.1.	Stored in the database	All Availability, Validity	
81.2.1.	Stored in the database	All Availability	
81.2.2.	Stored in the database	All Availability	
81.2.3.	Stored in the database		In part only, homonymy excluded
81.2.3.1.	Implicit throughout a function	Validity	In part only
81.2.3.2.	Implicit throughout a function	Validity	Might be added eventually explicitly
81.2.4.	N/A		About the Commission or the structure of the Code
82.1.	Absent		Relies on the concept of Usage
82.2.	N/A		About the Commission or the structure of the Code
83.	N/A		About the Commission or the structure of the Code
84.	N/A		About the Commission or the structure of the Code
84.1.	N/A		About the Commission or the structure of the Code
84.2.	N/A		About the Commission or the structure of the Code

85.	N/A	About the Commission or the structure of the Code
86.1.	N/A	About the Commission or the structure of the Code
86.1.1.	N/A	About the Commission or the structure of the Code
86.1.2.	N/A	About the Commission or the structure of the Code
86.2.	N/A	About the Commission or the structure of the Code
86.3.	N/A	About the Commission or the structure of the Code
87.	N/A	About the Commission or the structure of the Code
88.	Implicit throughout the software	
89.1.	Implicit throughout the software	
89.1.1.	N/A	About the Commission or the structure of the Code
89.2.	Implicit throughout the software	
90.	N/A	About the Commission or the structure of the Code

Appendix 3: Test of *Lognom*

English version

1) In 1952, Prof. J. Raffe described a taxon in the rank race named *Zirrafa bicolora* race *leucochloris* Raffe, 1952. He did not associate a type specimen with it.

In 1962, Dr. O. Truche described a taxon of species rank named *Zirrafa leucochloris* Truche, 1962.

In 1975, Prof. A. Rableux considered *Zirrafa bicolora* var. *leucochloris* Raffe, 1952 to be a senior homonym of *Zirrafa leucochloris* and proposed a replacement for Dr. Truche's name. All other conditions of availability being considered met, what is the status of *Zirrafa bicolora* race *leucochloris* Raffe, 1952? What's the name's rank?

2) Prof. A. Régner found an old book by T. Génère in her collections, in which an unknown name appeared, *Aranea cyanoventris* Génère, 1852. As she reads, she realizes that *A. cyanoventris* refers to the same species as *A. purpurooculis* Lamy & Gale, 1975, which is a widely used name.

Knowing the following:

The last time the name *Aranea cyanoventris* was used, according to Prof. A. Régner's research, was in T. Génère's book, where the species is described for the first time.

A. purpurooculis Lamy & Gale, 1975 was used in 2023.

A. purpurooculis Lamy & Gale, 1975 was used by 9 different authors.

A. purpurooculis Lamy & Gale, 1975 is cited by the World Spider Catalogue.

A. purpurooculis Lamy & Gale, 1975 was used as valid in 15 works before 2010, in 12 works between 2010 and 2012 and 3 times after 2012.

Can A. Régner declare *Aranea cyanoventris* Génère, 1852 as a *nomen oblitum*? If not, why?

Cite the relevant articles.

3) T. Tard spent his childhood asleep in ancient language class. Now that he has to determine the grammatical genders of the following three genus names, he's got his work cut out for him. Can you help him?

Argyrops, derived from Greek, introduced with the specific noun *sinensis* (indeclinable), with no indication of grammatical gender, but with the following etymology "from ancient Greek, ἄργυρος (argyros) for silver and ὄψ (ops) for eye".

Parma, derived from Italian, introduced with the species *reggiano*. *reggiano* is indicated as masculine, *Parma* is a feminine noun in Italian.

Propopon, which contains no indication other than that the author wrote that the name was inspired by the sound his son made when playing near him. *Propopon* is introduced with the species *P. ponpon*, whose grammatical gender is unknown.

4) Are the names *Flamingo monakus* and *Flamingo monacus*, both derived from the word for monk, to be considered homonyms? What about *Placebo sojicaum* and *Placebo sojicarum*?

5) If the work (modified from the original) presented on the following page were only published online, in PDF format, would it be available? If not, please cite the relevant articles.

Hint: the work is correctly binominal and there is no relevant information to determine availability after the first page. [The work given is a *Zootaxa* article where the date of publication was removed]

6) Here's the text accompanying the description of a new fictional species. You are a reviewer for this article. What advice would you give the author to ensure that his article follows all the rules and recommendations of the Code? Cite the relevant articles.

Rana îris Author sp. nov.

Definition: *R. îris* is recognizable by its bright colored back and big blue eyes.

Diagnosis: *R. îris* differs from *R. mettatica* by the fact that the latter has black eyes and duller colors on its back.

R. îris has no spine on its back, unlike *R. cyanops*.

Etymology: Noun in apposition, in the singular feminine, after the Greek goddess of the rainbow, Îris.

Type specimens: MNHN RA-8526-8756, MNHN RA-8526-8757.

Both specimens are deposited in the collections of the MNHN, Paris.

Type locality: Parc naturel régional de Camargue, France

7) Which of the following three names for the same family rank taxon is the valid one? All three names are available.

Family Paonidae Brutus, 1852; published in "Oiseaux d'Europe et d'Asie mineur", page 152

Tribe Lanini Pavmov, 1831; published in "Ceremonial Birds of Britain", page 89

Subfamily Muripainae Pavmov, 1831; published in "Ceremonial Birds of Britain", page 98

If you're solving this exercise on the software, log in to the account with "Tester23" as a login and "validity123" as a password.

In French

1) En 1952, le Pr. J. Raffe décrit un taxon au rang race nommée *Zirrafa bicolora* race *leucochloris* Raffe, 1952. Il ne lui associe pas de spécimen-type.

En 1962 le Dr. O. Truche décrit un taxon au rang espèce nommé *Zirrafa leucochloris* Truche, 1972.

En 1975 le Pr. A. Rableux considère que *Zirrafa bicolora* var. *leucochloris* Raffe, 1952 est un homonyme sénior de *Zirrafa leucochloris* et propose un nom de remplacement pour le nom du Dr Truche. Toutes autres conditions de disponibilité étant considérées comme remplies, quel est le statut de *Zirrafa bicolora* race *leucochloris* Raffe, 1952 ? À quel rang est-il appliqué ?

2) La Pr. A. Régner trouve un vieux bouquin de T. Génère dans ses collections, dans lequel un nom inconnu au bataillon apparaît, *Aranea cyanoventris* Génère, 1852. En lisant, elle comprend que *A. cyanoventris* se rapporte à la même espèce que *A. purpurooculis* Lamy & Gale, 1975, qui est un nom largement utilisé. Sachant les éléments suivants :

La dernière fois que le nom *Aranea cyanoventris* a été utilisé d'après les recherches du Pr. A. Régner, est dans le livre de T. Génère où l'espèce est décrite pour la première fois.

A. purpurooculis Lamy & Gale, 1975 a été utilisé en 2023.

A. purpurooculis Lamy & Gale, 1975 a été utilisé par 9 auteurs différents.

A. purpurooculis Lamy & Gale, 1975 est cité par le World Spider Catalogue.

A. purpurooculis Lamy & Gale, 1975 a été utilisé comme valide dans 15 travaux avant 2010, dans 12 travaux entre 2010 et 2012, et 3 fois après 2012.

A. Régner peut-elle déclarer *Aranea cyanoventris* Génère, 1852 comme *nomen oblitum*? Si non, pourquoi ? Citez les articles correspondants.

3) T. Tard a passé son enfance à dormir en cours de langue ancienne. Maintenant qu'il doit déterminer les genres grammaticaux des trois noms de genre suivants, il a bien du souci.

Pouvez-vous l'aider ?

Argyrops, dérivé du grec, introduit avec le nom spécifique *sinensis* (indéclinable), sans indication de genre grammatical, mais avec l'étymologie suivante « du grec ancien, ἄργυρος (*argyros*) pour argent et ὄψ (*ops*) pour œil ».

Parma, dérivé de l'italien, introduit avec l'espèce reggiano. *reggiano* est indiqué comme masculin, *Parma* est féminin en italien.

Propopon, qui ne contient aucune indication, si ce n'est que l'auteur a écrit que ce nom lui a été inspiré par le son que son fils faisait en jouant près de lui. *Propopon* est introduit avec l'espèce *P. ponpon* dont le genre grammatical est inconnu.

4) Les noms *Flamingo monakus* et *Flamingo monacus* dérivés tout deux du mot moine sont-ils considérés des homonymes ? Qu'en est-il de *Placebo sojicaum* et *Placebo sojicarum* ?

5) Si le travail (modifié depuis l'original) présenté à la page suivante était uniquement publié en ligne, sous format PDF, serait-il disponible ? Si non, citez les articles correspondants.

Indication: le travail est correctement binominal et il n'y a aucun élément pertinent pour déterminer la disponibilité après la première page.

6) Voilà le texte accompagnant la description d'une nouvelle espèce fictive. Vous êtes reviewer pour cet article. Quelle(s) indication(s) et conseils donneriez-vous à l'auteur pour faire en sorte que son article suive toutes les règles et recommandations du Code ? Citez les articles correspondants.

Rana îris Author sp. nov.

Definition: *R. îris* is recognizable by its bright colored back and big blue eyes.

Diagnosis: *R. îris* differs from *R. metallica* by the fact that the latter has black eyes, and duller colors on its back. *R. îris* has no spine on its back, unlike *R. cyanops*.

Etymology: Noun in apposition, in the singular feminine, after the Greek goddess of the rainbow, Îris.

Type specimens: MNHN RA-8526-8756, MNHN RA-8526-8757.

Both specimens are deposited in the collections of the MNHN, Paris.

Type locality: Parc naturel régional de Camargue, France

7) Parmi les trois noms suivants s'appliquant au même taxon de rang famille, lequel est le nom valide ? Les trois noms sont disponibles.

Famille Paonidae Brutus, 1852 ; publiée dans « Oiseaux d'Europe et d'Asie mineur », page 152 Tribu Lanini Pavmov, 1831 ; publiée dans « Ceremonial Birds of Britain », page 89

Sous-famille Muripainae Pavmov, 1831 ; publiée dans « Ceremonial Birds of Britain », page 98

Si vous faites cet exercice sur le logiciel, connectez-vous au compte ayant pour pseudo « Tester23 » et « validity123 » comme mot de passe.

Résumé

La nomenclature est la branche de la taxonomie responsable de la création et de la gestion des noms scientifiques attribués aux catégories d'êtres vivants. Elle assure la continuité de la transmission de toutes sortes de données et de connaissances accumulées sur les taxons. Les zoologistes se réfèrent pour cela au Code International de Nomenclature Zoologique. Le Code contient les règles qui permettent de comprendre et d'appliquer correctement cette discipline. La nomenclature s'est complexifiée au fil des siècles, pour s'adapter aux changements scientifiques et technologiques. Aujourd'hui, la nomenclature joue, par l'intermédiaire des noms scientifiques, un rôle crucial dans les bases de données de biodiversité. Mais ces dernières présentent cependant des limitations en termes de modèles de données et de l'utilisation de celles-ci pour des travaux nomenclaturaux. Cette dimension de la nomenclature est explorée dans la thèse.

Si les données nomenclaturales sont présentes en grande partie dans les bases de données en ligne, la discipline elle-même est restée relativement imperméable à l'informatisation. Une analyse approfondie du Code a permis d'établir une liste d'objets formels et des propriétés les accompagnant nécessaires pour modéliser les règles de cette discipline. En effet, la structure des règles du Code se veut logique et non-ambigüe, et est donc idéale pour une traduction de ces dernières en série d'algorithmes.

Cette hypothèse a mené à la création d'une application web appelée *Lognom*, pour 'logiciel de nomenclature'. *Lognom* est un logiciel basé sur des algorithmes aidant à la prise de décision en matière de nomenclature zoologique. L'application ne repose pas sur des bases de données préexistantes, mais fournit une réponse en fonction des données entrées par l'utilisateur. Ce logiciel vise à soutenir les taxonomistes dans leur gestion de la nomenclature au quotidien, en déterminant si un nom ou un travail est disponible, si les règles orthographiques ont été correctement appliquées et si toutes les exigences précédant la publication d'un nouveau nom ou d'un nouveau travail ont été respectées. *Lognom* permet également à l'utilisateur d'établir quel nom est le nom valide parmi plusieurs candidats préenregistrés, et la liste des synonymes qui résulte de cette détermination. Il comprend également quelques outils pour répondre à des questions diverses de nomenclature, telle que la détermination du genre grammatical d'un nom de genre.

Toutes les règles du Code de nomenclature zoologique n'ont cependant pas pu être intégrées à l'application. Certaines règles reposent sur une interprétation sémantique qu'il est très complexe d'automatiser. De plus, même s'il contient quelques contrôles, *Lognom* est très sensible à la qualité des données fournies par ses utilisateurs. Une proposition de classification des règles est fournie, afin de mieux cerner les forces et les faiblesses du Code quant à sa possible informatisation exhaustive, ainsi que des recommandations quant à l'optimisation de son caractère logique et non-ambigu. De même, diverses questions en rapport avec la nomenclature et ses applications informatiques sont explorées, et une brève analyse expliquant les difficultés sociales liées à l'amélioration de ces règles est évoquée. Il existe une multitude d'applications futures possibles pour les algorithmes développés lors de cette thèse. Ces perspectives incluent la possibilité de travailler collaborativement sur des projets concernant la nomenclature d'un même groupe taxonomique. Ceci pourrait conduire à établir des listes dynamiques de noms. Une modification de ces algorithmes pourrait également simuler l'impact d'une modification des règles du Code sur les noms et les travaux existants. Sur le long terme, un outil tel que *Lognom* pourrait conduire à la possibilité de modéliser la nomenclature dans sa totalité, et d'ouvrir la porte à une gestion plus efficace et plus coordonnée de cette discipline pluricentenaire.

Abstract

Nomenclature is the discipline of taxonomy responsible for creating and managing the scientific names assigned to categories of living beings. It ensures continuity in the transmission of all kinds of accumulated data and knowledge about taxa. To this end, zoologists refer to the International Code of Zoological Nomenclature. The Code contains the rules for understanding and correctly applying this discipline.

Nomenclature has become increasingly complex over the centuries, to keep pace with the evolution of scientific and technological knowledge. It currently plays, through scientific names, a crucial role in biodiversity databases. However, databases have their limitations in terms of structure when it comes to nomenclatural endeavors. The role of nomenclature in databases is explored in the thesis.

While nomenclatural data is largely present in online databases, the discipline itself has remained relatively impervious to computerization. An in-depth analysis of the Code enabled the creation of a list of formal objects and their properties, which are needed to model the rules of this discipline. Moreover, the structure of the Code's rules is intended to be logical and unambiguous, which makes it ideal for translating into a series of algorithms.

This hypothesis led to the creation of a web application called *Lognom*. *Lognom* is an algorithm-based software that supports decision-making in zoological nomenclature. The application does not rely on pre-existing databases, but provides an answer based on data entered by the user. The software aims to support taxonomists in their day-to-day nomenclature management, by determining whether a name or work is available and whether spelling rules have been correctly applied. It can also verify whether all requirements preceding the publication of a new name or work have been met. Additionally, *Lognom* allows the user to establish which name is the valid name among several candidates, and the list of synonyms that results from this decision. It also includes several tools for answering various nomenclatural questions, such as the determination of the grammatical gender of a genus name.

However, it has not been possible to integrate all the rules of the International Code of Zoological Nomenclature into the application. Some rules are based on semantic interpretation, which is very complex to automate. Moreover, *Lognom* is highly sensitive to the quality of the data supplied by its users, even if it does provide a few controls.

A proposed classification of the Code's rules is included, to better identify the strengths and weaknesses of the Code in terms of its possible complete computerization. Recommendations for the optimization of its logical and unambiguous character are also mentioned. Similarly, various issues relating to nomenclature and its computer applications are explored, as well as a brief analysis of the social difficulties that might impede the improvement of these rules.

There are still many possible future applications for the algorithms developed for *Lognom*. These include the possibility of working collaboratively on projects concerning the nomenclature of a given taxonomic group. This could lead to the creation of dynamic lists of names: Furthermore, the algorithms should be able to simulate the impact of changes in the rules of the Code on existing names and works. In the long term, a tool such as *Lognom* could enable the possibility of modeling nomenclature in its entirety, opening the door to more efficient and coordinated management of this centuries-old discipline.