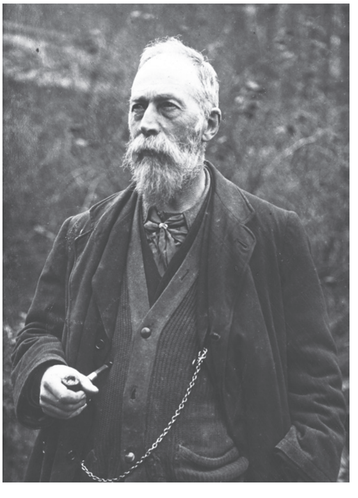
Anchoring Biodiversity Information: From Sherborn to the 21st century and beyond

Edited by

Ellinor Michel



PENSOFT

ZooKeys 550 (Special Issue)

Anchoring Biodiversity Information: From Sherborn to the 21st century and beyond

Edited by Ellinor Michel

First published 2016
ISBN 978-954-642-806-6 (paperback)

Pensoft Publishers
12 Prof. Georgi Zlatarski Street, 1700 Sofia, Bulgaria
Fax: +359-2-870-42-82
info@pensoft.net
www.pensoft.net

Printed in Bulgaria, January 2016

Contents

ZooKeys 550: 1–11 (2016)
doi: 10.3897/zookeys.550.7460
http://zookeys.pensoft.net

EDITORIAL

A peer-reviewed open-access journal
ZooKeys
Launched to accelerate biodiversity research

Anchoring Biodiversity Information: From Sherborn to the 21st century and beyond

Ellinor Michel[1]

1 *Department of Life Sciences, The Natural History Museum, Cromwell Road London SW7 5BD, UK*

Corresponding author: *Ellinor Michel* (e.michel@nhm.ac.uk)

Received 8 December 2015 | Accepted 8 December 2015 | Published 7 January 2016

http://zoobank.org/8EFBBB2D-BE60-43AA-A37D-23C9792CF075

Citation: Michel E (2016) Anchoring Biodiversity Information: From Sherborn to the 21st century and beyond. In: Michel E (Ed.) Anchoring Biodiversity Information: From Sherborn to the 21st century and beyond. ZooKeys 550: 1–11. doi: 10.3897/zookeys.550.7460

Charles Davies Sherborn provided the bibliographic foundation for current zoological nomenclature with his magnum opus *Index Animalium*. In the 43 years he spent working on this extraordinary resource, he anchored our understanding of animal diversity through the published scientific record. No work has equaled it and it is still in current, and critical, use.

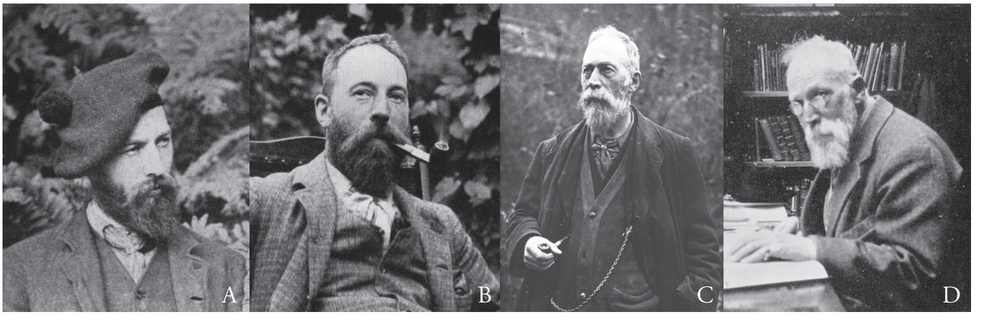


Figure 1. Charles Davies Sherborn aged 25 (**A**), 32 (**B**), 61 (**C**) and 72 (**D**).

Until now, Sherborn's contribution has been recognized and relied upon by professional taxonomists worldwide but he has escaped the celebration of his accomplishment that is his due. This changed on 28 October 2011, with a symposium held in his honour at the Natural History Museum (NHM), London, on the 150th year of his birth. The symposium was organized by the International Commission on Zoological Nomenclature (ICZN), in collaboration with the Society for the History of Natural History (SHNH). The full-day meeting included an international panel of experts on bibliography and biodiversity bioinformatics who linked a view of the past with an active debate on the future of these related fields. There were fifteen talks from distinguished speakers from around the world, and ten posters, including an exhibition of 'Sherborniana', or artifacts from Sherborn's tenure at the NHM. This volume expands on that meeting, with contributions from most of the presenters and selected additional contributors. The global and temporal reach of this event was extended through high quality recordings of all the talks, posters and discussion, including slides and poster downloads, through this site: http://backdoorbroadcasting.net/2011/10/anchoring-biodiversity-information-from-sherborn-to-the-21st-century-and-beyond/ and videos of all the talks through http://www.iczn.org/Sherborn.

The papers in this volume fall into three general areas. In the first section, seven papers present different facets of Sherborn as a man, scientist and bibliographer, and describe the historical context for taxonomic indexing from the 19th century to today. In the second section, five papers (with a major appendix) discuss current tools and innovations for bringing legacy information into the modern age. The final section, with three papers, tackles the future of biological nomenclature, including innovative publishing models and the changing tools and sociology needed for communicating taxonomy.

Because this volume is being produced as both a bound book and set of independent, Open Access papers free to download from the Web, there is a degree of overlap in some of the material covered. The papers need to be able to stand on their own, as well as to weave in to the whole overview of the accomplishments of this great man, his legacy and the roadmap for the future. In addition, because of the varied topics, the papers vary in style and length, some being more literary, some historical, some technical and some philosophical. Some are richly illustrated, others not at all. The only instructions to the authors were to attempt to reference each other's papers to the greatest practical degree, simulating the kind of cross-topic communication one might have by being present at a symposium. The papers were all peer reviewed - most had critical input from three independent specialists in the field. I hope this diversity of approaches, rigorous oversight and the cross-pollination make the volume stimulating to read as a whole.

Sherborn as a person, scientist and bibliographer, and his context

What kind of person takes on such a herculean task as did Sherborn? What was the source of his motivation? What were the related predecessors and descendants of his work? The first three papers address the whole of Sherborn's life, and historical context, with a focus on *Index Animalium*. While Sherborn's fame is based on *Index Animalium*, he undertook a number of other ambitious projects. Several of these also had lasting impacts on their respective fields. These are addressed in the next four papers in this section.

In 'Charles Davies Sherborn and the "Indexer's Club"' Neal Evenhuis provides personal and highly sympathetic insights into the incredible drive and bibliographic skills Sherborn had to harness in his effort to make an essentially universal index to all animal names. Evenhuis served as a Commissioner and President of the ICZN for many years, and is a self-described 'index-aholic' – he knows whereof he speaks in understanding Sherborn's motivation. As a plenary talk, and in this volume's cornerstone paper, Evenhuis sets the tone. His wry wit makes Sherborn's labours seem a natural endeavour, at least for those of an ambitious and altruistic mindset, motivated by the greatest of tasks, getting control over our information on the living world. Evenhuis also outlines forerunner and descendent projects of a similar nature.

Karolyn Shindler provides another, more personal portrait of Sherborn, adding a richness of experience of additional major projects in his life. Key among these was Sherborn's involvement with the archives of Sir Richard Owen, the great anatomist and founder of the British Museum (Natural History), now the Natural History Museum, London. She highlights the key phrase that Sherborn used for himself, but perhaps applies to most collections and bibliophilic workers: 'A Magpie with a Card Index Mind'. Although the historical facts that frame all the essays on Sherborn are the same, I found that the feeling for the man was quite different in each one. Shindler's piece brought me practically to tears with awe and appreciation for Sherborn's challenges and determination, but also feeling for his quirks. After reading it, I felt I had met the man himself.

Gordon McOuat's contribution provides a overview of the evolution of nomenclatural codes and controversies in the decades around Sherborn, bringing the history of science to life. His contribution has a number of key messages on the relationships between names (dubbing) and meanings (taxonomy), on the struggle between establishing nomenclature tied to rules (codes) or to specimens (the type concept and museum catalogues). These issues were intensely addressed in the early and mid-19th century and Sherborn's magnum opus played a foundational role in establishing the systems we now use for all biology, not just zoology. Nonetheless many taxonomists today continue to befuddle these relationships, often through lack of knowledge of the long history of the discussions.

C. Giles Miller delves in to the scientific starting point for Sherborn's indexing focus. Appropriately, this grew from taxonomic and collections work in the Natural History Museum on fossils, one of Sherborn's early research loves. However, Miller provides a telling comment, saying that while Sherborn's foraminiferan collections were respectable, they were not groundbreaking. In contrast, his bibliographic and indexing contributions changed the practice of micropalaeontology in the Natural History Museum, and thus the world. Moreover, his experience with this micropalaeo work set him on his future trajectory of ambitious indexing for all animal names and museum collections. It provided the focus for Sherborn to see his life's calling.

Figure 2. Sherborn's land and freshwater mollusc collections, with specimen labels in his own handwriting **A** Gastropods **B** Bivalves.

As an aside, it should be noted that Sherborn had several additional scientific and collections interests not dealt with in the papers in this volume. Notably, his background in malacology resulted in three drawers of collections made by him, held at the NHM, London (Fig. 2). There are also collections of coins and stamps, apparently held in the collections of the British Museum.

The final major work undertaken by Sherborn at the end of his life was a listing of natural history collections that was actually titled '*Where is the damned collection*?'. Clearly, Sherborn had a sense of humour! Michael A. Taylor gives a vibrant description of this work – how it came about, the business hurdles and social controversies around getting it published, and the rewards for subsequent museum collections work. It makes surprisingly compelling reading to learn about Sherborn's uncompromising statements such as his 'savage review of a foreign rival' that put sand in the gears of getting his own work published and given due credit. Nonetheless, he persisted. It was a work that, despite its apparent flaws, contributed to the development of collections research, and, like most of Sherborn's other contributions, is still in use.

Shifting to a taxon-specific focus, Edward Dickinson presents a detailed scrutiny of Sherborn's and Richmond's indexes in ornithology, a taxonomic best-case that illuminates problems that need attention in the larger whole of the corpus. He underscores that nomenclature is the un- (or under-) recognized foundation of taxonomy, thus calls on ornithologists to mobilize and collaborate to get the house in order for names of the approximately 10,000 bird species, as they are arguably the most public-facing of popular animal groups. This will require a level of attention to detail and collaboration that raises the game from previous ways ornithologists have worked.

F. Christian Thompson and Thomas Pape explain how research on the important (and beautiful!) megadiverse insect group Diptera has benefited from building an outstanding bibliographic index based on Sherborn's original work. As there are 160,000 currently recognized species of flies, with over 250,000 names, this group makes up a significant proportion of planetary diversity (an estimated 10%). *Systema Dipterorum* (http://www.diptera.org/) benefits from modern tools and additions that provide a resource of greater utility than even Sherborn could have imagined.

Current tools and innovations for bringing legacy information into the modern age

The Smithsonian Libraries have made Sherborn's *Index Animalium* freely accessible online (http://www.sil.si.edu/digitalcollections/indexanimalium/). Suzanne Pilsk, Martin Kalfatovic and Joel Richard describe how this was done, and how the transition from paper to bytes is the dawn of a new age for bibliographic information. They point out that traditional library metadata for a book title, which was sufficient to retrieve a physical book from the stacks, is now not fit-for-purpose for the vastly increased but distributed constituencies that modern libraries serve. Rising to the challenge, the

Smithsonian converted the 420,000 entries of *Index Animalium* to a detailed, fine-granularity bibliographic data set that can be used by researchers with greater speed and fidelity than print ever allowed.

A key concern for any data source is its error rate and identifying where the errors are concentrated. Francisco Welter-Schultes, Angela Görlich & Alexandra Lutze undertook a detailed study of error rates in Sherborn's magnum opus by comparing samples with their own large team project on original sources through the project AnimalBase (http://www.animalbase.org/). They found that Sherborn's error rate was remarkably low, at 1–2% or even less for individual entries. They point out that, 'this is low for any human endeavor, let alone one of such monumental scale requiring detailed work over many decades. It is all the more impressive when we realize that today we have comparable failure rates, despite having computer tools and teams of people to help with this kind of work.' However, there are areas where errors are concentrated such as names from chaotically organized original sources, from publications in languages that Sherborn did not speak, and from particular sources or taxa. Oddly, although Sherborn's error rate for molluscan names was higher than for insect names overall, his error rate for Fabricius, the largest source of insect names, was surprisingly high. The authors advise that *Index Animalium* not be used to determine correct authorship of a name, but original sources should be consulted because they see an unacceptable level of problems with rendition of authors. Welter-Schultes, et al., suggest that a 2–4% error rate is an intrinsic limit in manually compiled data of this kind at this scale; below that there are diminishing returns. They brought this point home in a criticism of the idea of Lists of Available Names (LANs, below) and as a caution to large-scale data input into projects like ZooBank (below).

In a paper that tackles technical issues, but with a good grounding in philosophical issues, Christopher Lyal describes the limitations of digitizing objects and information. His target is bringing legacy taxonomic literature into the digital sphere. Lyal underscores our current tendency to build forward from the past, using e-charged traditional methods to produce digital analogues of paper, rather than developing new tools that make the most of cybertechnology and assessment of future needs and opportunities. He explains how text mark-up with XML can open the door to allowing not only a viewable and searchable original text, but more powerfully, for subsets to be viewed, extracted and separately analyzed. This allows repurposing in a multitude of different contexts, extending the reach of the original publication, and creating new ways of using scholarly information. It also allows automatic population of large-scale information sources such as ZooBank or GBIF. Mark-up allows dynamic linking of new and extant information. In essence, it is what Sherborn was aiming to do with *Index Animalium*.

In a paper that should become required reading for all taxonomists, David Remsen takes the issues of taxonomic knowledge bases and systems of names to their philosophical foundations. Names are a handle or tag on larger sets of concepts (taxonomy) that can be fluid, however names also link to an objective standard, a type specimen, and have a single birthplace, or point of origination, in a publication. While dealing with unstable relations between these entities is a headache, Remsen explains with crisp language and clear diagrams how it can be clarified. 'Semiotics provides a general model for describing the relationship between taxon names and taxon concepts. It distinguishes syntactics, which governs relationships among names, from semantics, which represents the relations between those labels and the taxa to which they refer.' He places nomenclature in the context of a graphical triangle of reference, or semiotic triangle, as a model of how syntax and semantics are related to the objects they represent. The paper provides a critical link between lists of names, like *Index Animalium* and lists of species, which are the ultimate currency that interest most users.

With a more pragmatic perspective on issues of stability of scientific names of animals, Miguel Alonso-Zarazaga, Daphne Fautin and myself detail the requirements and opportunities for Lists of Available Names (LANs) to proceed through ICZN Article 79 to stabilize large taxonomic sections of nomenclature at once. Although it is not a light task to implement a LAN, a result is that 'nomenclatural archaeology' will find the footing pulled out from under it, thus increasing stability and transparency in scientific names of animals. Our short, succinct paper outlines the results of deliberations of several ICZN committees. It is supported by publication of a '*Manual for proposing a Part of the List of Available Names in Zoology*' by Miguel Alonso-Zarazaga, Philippe Bouchet, Richard Pyle, Nikita Kluge, Daphne Gail Fautin as an appendix to this volume. We hope these practical tools will result in well-documented, collaborative work by sectors of the taxonomic community to stabilize names that might otherwise create problems for information retrieval.

The future of biological nomenclature, including innovative publishing models and the changing sociology of science in taxonomy

The future for taxonomy is intimately bound with the future of publishing, and biological knowledge is on the cusp of a radical change in how it is delivered and archived. The atomization and automation of publications will make step changes in how information is used. Lyubomir Penev and 11 co-authors provide a very practical glimpse of what revolutionary e-tools look like, presenting a new work flow and publishing mechanism developed by this journal, *ZooKeys*. Their paper is a collaborative approach between four lead indexes of taxon names and nomenclatural acts, and thus achieves an additional objective of harmonizing practice across taxonomic disciplines. They point out how technical tools can radically change the landscape for the persistent, and previously intractable, controversies of registration and e-publication across all biological nomenclature.

In 'Surfacing the deep data of taxonomy' Rod Page observes that 'Names by themselves are of little value; it is the literature, specimens and data derived from those specimens that are the primary data of taxonomy. Yet much of this information remains hard to obtain (even discovering it exists can be challenging)'. His paper is a manifesto, targeted at a critical technical tool to achieve this goal – the form of the persistent digital link between units of information, the bibliographic identifiers. Because the revolution in digital information access has grown up through individual innovation and is facing a kind of free-market competition, not top-down infrastructural planning, we are currently in a situation where different projects have opted to use different kinds of identifiers (DOIs, or digital object identifiers, versus LSIDs, or Life Science Identifiers). Page makes a strong case that his preferred identifier system, DOIs, is the only one with the supporting features that allow complete, deep, linkage of all the primary taxonomic data. He suggests that the tracking features of DOIs allow them to potentially solve a huge problem for taxonomists in providing altmetrics that demonstrate the long and wide reach of their work. This can give taxonomists greater credit, countering the current skewed recognition based on journal impact factors. He also suggests that this bibliographic issue holds solutions to the problem of how to recognize 'dark taxa', those known from (usually molecular) data, but not recognized with a formal name. Page is known as a boat-rocking, take-no-prisoners provocateur; we are lucky that he has turned his sights on bibliography – follow his arguments to find where new disruptive technology will have a major constructive effect in taxonomy.

Richard Pyle's paper, based on his wrap-up plenary talk, makes a convincing case that, even in this time of major technological improvements for all taxonomic research tools, the greatest wholesale revolutionary change is the means by which we manage and communicate information. Names are at the nexus of that revolution. Pyle puts Sherborn's work at the center of the task of identifying and making order of our knowledge of biodiversity:

> 'the Linnaean nomenclatural system [is] a stable scaffold against which the ever-changing landscape of [taxonomic] species can be referenced. … In stark contrast to the dynamic, on-going, and seemingly endless debates about what a "species" is, the nomenclatural system used by taxonomists during the past two and a half centuries has been remarkably consistent, universal, and stable….. Whereas the majority of the nearly 4,400 species circumscriptions described by Linnaeus in his 1758 *Systema Naturae* bear very little resemblance to the species boundaries asserted by modern biologists, most of the scientific names he established are not only available under the current Code, but are in current use. …Although catalogs of species (e.g., Linnaeus 1758) may begin to lose their taxonomic relevance almost immediately after publication, the scientific names established within such catalogs retain their nomenclatural relevance indefinitely. Ultimately, this is why the career-long labors of Sherborn have retained their value well beyond his own life, up until today and continuing indefinitely into the future.'

Pyle exhorts that we are now responsible for the next iteration of this Linnaean enterprise in a new way. The new paradigm for all nomenclature projects is the Global Names Architecture (GNA), the dynamic index to interconnect and streamline the entire taxonomic enterprise through names. Pyle underscores the constructive collaboration of all the major taxonomic resources (e.g., GBIF, CoL, IPNI, EoL, ZooBank) to build a dynamic suite of web services to connect them all through the GNA. Together, this is 'the digital equivalent of the card catalogue of life – audacious task started by Linnaeus, dramatically extended by Sherborn'. Pyle convinces us that now we can now make another big leap in the content we cover, to encompass the entire living world using a consistent, interoperable information system that is accessible to all.

Concluding remarks

We are on the brink of a new and truly open taxonomy. The revolution has arrived through the development of technical tools that open up ways to atomize information and make it quickly findable, retrievable and recombinable. New ways of working and new results will result in a taxonomic 'Modern Synthesis'. Proactive collaboration will arise more fluidly between different systems with overlapping content. The philosophical underpinning of the mutual support between the flexibility of taxonomic interpretation and the stability of nomenclatural frameworks is becoming easier to define

Figure 3. The full panel of symposium speakers under the heading **WWSD? What would Sherborn Do?** From left to right: Suzanne Pilsk, Chris Lyal, Henning Scholz, Edward Dickinson, Neil Evenhuis, Daphne Fautin, Sandy Knapp, Lyubomir Penev, Rod Page, Chris Thompson, Chris Freeland, Gordon McOuat, (behind podium Richard Pyle, David Remsen).

through appropriate bibliographic tools. Similarly, past differences in how the major taxonomic codes deal with names are being decreased through use of shared technical tools and major infrastructures collaborating for information access. Registration of new names is on an active track for implementation in several taxonomic disciplines, some with a common framework. The authors of this volume have taken different approaches to the problems for animal names faced by Sherborn, but it adds up to a multifaceted and powerful approach for all biological nomenclatural issues.

At the end of the symposium that gave birth to this volume, we held a panel discussion under the banner 'WWSD? or 'What Would Sherborn Do?' With the contributions published here we now know much more about Sherborn as a man and scientist, about the long running nature of debates, about the current tools for making progress, and the bright future for the field. The answer is that Sherborn would have celebrated the new tools for this ambitious goal of linking all biological information through names, both machine and human readable. He would have understood its tremendous power for biodiversity science overall. And he would have knuckled down and got to work to make it happen.

Acknowledgements

The Society for the History of Natural History was pivotal in calling attention to Sherborn's sesquicentennial, and it was especially the support from Gina Lundy Douglas that brought the symposium to life. The quality of the papers was greatly improved by input from a large group of outstanding peer reviewers. These included most of the authors of the papers (who kindly reviewed their colleague's contributions) and also Donat Agosti*, Max Barclay, Daniel Bickel, Patrice Bouchard, Philippe Bouchet, Terrence Catapano, Ron Cleevley, Gene Coen, Paul Martyn Cooper*, Florian Ehrensperger, Neal Evenhuis*, Niko Franz, Stephen Gaimari, Graham Higley*, Kenneth Johnson, Alan Kabat, Martin Kalfatovic, Ross Mounce*, Roderick Page, Thomas Pape, Lyubomir Penev, Suzanne Pilsk, Ken Polzin, Richard Pyle, David Remsen*, Elliot Shubert, Frank Steinheimer, Michael A. Taylor, Michael P. Taylor, Jonathan Todd*, John Whittaker, Francisco Welter-Schultes, Deborah Wright, Nigel Robinson and Jeremy Young. A few, indicated with "*", provided many reviews or supporting illustrations that required considerable additional work, so deserve extra thanks.

Maureen MacGlashan gave an inspired lead into the world of professional indexing, where Bill Johncocks not only a professional job of indexing the book on biology's greatest indexer, but was unfailing in supportive comments on the wider interest of the work.

The NHM Photo Unit and Libraries and Archives are thanked for their professional imaging services and permission to use material for these papers. As Sherborn did his work in the NHM Library for his whole career, their interest and support is especially appreciated.

Thanks are due again to Backdoor Broadcasting, who provided professional recording and stable long term archiving of the talks, and slide presentations. I recommend listening to the talks as a supplement to the papers.

The staff of ZooKeys have provided a patient, supportive environment, and have executed their part of the publishing process with incredible efficiency and attention to detail. Lubyo Penev's support for Open Access for all the papers, and publishing as both as a book and individual units, is truly in the spirit of a new future for taxonomy.

The event was organised and sponsored by the ICZN (Int'l Commission on Zoological Nomenclature) and the Society for the History of Natural History, with significant sponsorship support from the Linnean Society, BHL-Europe (Biodiversity Heritage Library-Europe), Pensoft Publishers (ZooKeys), The NHM – Natural History Museum, and ViBRANT – Virtual Biodiversity.

ZooKeys 550: 13–32 (2016)
doi: 10.3897/zookeys.550.9697
http://zookeys.pensoft.net

RESEARCH ARTICLE

A peer-reviewed open-access journal
ZooKeys
Launched to accelerate biodiversity research

Charles Davies Sherborn and the "Indexer's Club"

Neal L. Evenhuis[1]

1 *Bishop Museum, 1525 Bernice Street, Honolulu, Hawaii 96817, USA*

Corresponding author: *Neal L. Evenhuis* (neale@bishopmuseum.org)

Academic editor: *E. Michel* | Received 26 March 2015 | Accepted 26 March 2015 | Published 7 January 2016

http://zoobank.org/796D9451-2C15-4687-BB20-684C2959AEEA

Citation: Evenhuis NL (2016) Charles Davies Sherborn and the "Indexer's Club". In: Michel E (Ed.) Anchoring Biodiversity Information: From Sherborn to the 21st century and beyond. ZooKeys 550: 13–32. doi: 10.3897/zookeys.550.9697

"*Nomina si pereunt, perit & cognitio rerum.*"
[If the names are lost, the knowledge also disappears.]
– C. Linnaeus, 1775, *Bigae Insectorum*, p. 305

Abstract

The first few words of the title of this symposium are "Anchoring Biodiversity Information". In order to properly anchor anything for a long-lasting future, a solid foundation needs to have been laid. For the zoological portion of biodiversity information, that firm foundation is best exemplified in the works of Charles Davies Sherborn. This man, like others of his ilk, was intimately focused on indexing names. This incredible focus was a life-long passion for him and culminated in his 9500-page *Index Animalium* of over 400,000 names of animals. This *Index* represents not only one of the most prodigious efforts in publication by a single man and the single most important reference to names in zoology, but a permanent legacy to the efforts of an indexer that proved to be an inspiration to many.

Keywords

Indexing, bioinformatics, zoology, biography

Introduction

Before we go into the life and work of Sherborn, a bit of an introduction needs to be made as to just who indexers are, and what makes them index. I call this group of individuals the "*Indexer's Club*". It is a unique gathering of like minds that for some reason have found comfort in essentially making order out of disorder for large groups of things. There are professional indexing societies in Australia and New Zealand, Britain, Canada, China, Germany, the Netherlands, and Southern Africa and an international quarterly journal, The Indexer, which covers a wide range of indexing-related matter. Technically, "indexes" and "lists" are two different things, but for convenience in this paper, I am lumping the two into "indexing" *sensu lato*.

In finding a way to make order out of an otherwise chaotic array of things, indexing is not necessarily making a long list of names in alphabetical order. It can be as simple as making a shopping list, a list of chores, a list of synonyms by category, a list of phone numbers, a list of birth-stones by month, or maybe even a list of past lovers (in chronological order, of course) or it can be a very complex and onerous task involving large numbers of unsorted items. We are all indexers in that we have made some of the lists just mentioned. Some lists may have been out of simple curiosity ("based on the data from the exams of my students that I have graded, I wonder who is at the top of the list in my class"); others may have been made because it helped us in some way ("based on the data from the exams of my students I have graded, I wonder if I may get a promotion"). Whether the user is us or others, simply said, indexers facilitate the various users of data to expedite their work by forming an ordered methodology to find what is being sought.

However, *bona fide* members of the "Indexer's Club" as defined herein – the ones that spend many years making lists of large groups of things – are not born that way but have, through experience with making a first list, found a unique form of satisfaction in making order out of something. It may not be as much the result of the efforts as the actual work of making order that is addicting or satisfying. Sherborn was one of these who found immense satisfaction from making lists of things, despite the incredible time and effort it often took (Fig. 1).

An indexer can naturally have a strong proclivity toward making lists but in some cases this obsession or addition may have come from an unhealthy or stressful background. Such was the case of one of the best-known list makers, the polymath Peter Mark Roget (1779–1869) (Fig. 2). Although his contemporaries thought he would be best known for his 2-volume, *Bridgewater treatise* on the physiology of plants and animals (Roget 1834), the work that would instead put his name into the vocabulary of millions upon millions and that would be referred to more than most dictionaries by undergraduate students worldwide was his *Thesaurus,* the work that is synonymous with synonyms. It is one of the most famous of all reference works, but few know of the life of its author or what drove him to list-making.

Like all indexers, Roget longed to put order into his world. Unfortunately, Roget's world as a child had his father die prematurely, his beloved uncle commit suicide in

Figure 1. Charles Davies Sherborn. Self caricature after he had finished the last entry to the *Index Animalium* (from Norman, 1944). The handwritten quip "I dunno wahriar" is a stunned and worn-out Sherborn saying "I dunno where I are".

Figure 2. Peter Mark Roget (from Wikimedia Commons).

his presence, and his maternal grandmother and mother each suffer from an unidentified mental illness. To escape this dark and dreary world, Roget comforted himself with words, and started making lists at the age of eight. One of his first lists was of his relatives and family and the dates of their deaths (Kendall 2008). From that morbid beginning, he went on to initially make lists of the things he found around him such as plants, animals, cloud types, and minerals, all with their Latin equivalents (Latin being a favorite subject of his, even at a very early age). He found solace in list-making and he didn't stop. Throughout his life, he kept files of all the words and their synonyms that he could find. Incredibly, it wasn't until he was 73 years of age before the first edition of *Roget's Thesaurus of the English language* appeared on bookshelves (Roget 1852). His indexing was thus a life-long but immensely personal obsession.

But to stereotype an indexer as being like Roget can be dangerously wrong. Not all come from the extremely depressing and stressful background of Roget. Yes, Sherborn could be easily defined as a "workaholic", involved often in a number of concurrent projects. Norman (1944) gave examples of Sherborn at various times during the production of his *Index Animalium* also working on other indexes such as the one of all the genera and species in Linnaeus's *Systema naturae* (Sherborn 1899). But he also did bibliographies and associated dating research, biographies and obituaries, synthesized all the known natural history collections into a single resource (Sherborn 1940), compiled his own family's genealogy [progress on work and enquiries for further material noted in Sherborn (1898) and final publication in Sherborn (1901)], and various and sundry other lists, all while keeping up with his hobby of collecting Byzantine bronze coins. Yet, despite his constantly keeping himself busy with projects, and in striking contrast to the dreary surroundings of Roget's life, Sherborn's life was much more on the "normal" side.

A brief vignette into the formative years of Charles Davies Sherborn

Charles Davies Sherborn, was born on 30 June 1861 in Gunter Grove, Chelsea (near central London but considered a rather rural area at the time with large open fields), and was baptized at St. Luke's Church, well known at the time as having been the venue for the marriage of Charles Dickens 25 years earlier. His father was Charles William Sherborn (1831–1912) [Charles Davies Sherborn always signed himself as "C. Davies Sherborn" possibly to disambiguate himself from his father, also a Charles], an etcher and engraver, especially known for his book plates, and his mother was Hannah Sherborn (née Simpson) (1829–1922). Charles was the eldest of five children (one having died in infancy).

Sherborn wrote a biography of his father (Sherborn 1912) and his description of him, also quoted in Norman (1944), echoes many of the same qualities that were said to have been of the younger Sherborn himself:

"My father was a robust person, about five feet nine inches high, easy-tempered and easy going, though intolerant of bores and politics, and strongly Protestant in his religious convictions. He went about little in Society, disliking formalities, and rarely entertained anyone at his own home." (Norman 1944: 12).

In the only place he described himself, the genealogy of his family, Sherborn simply stated:

"Educated by Miss Elizabeth Rye and at St. Mark's College School, Chelsea; was in business from 1876–84, when he went to Switzerland and Germany, afterwards devoting himself to the bibliography of the zoological and geological sciences" (Sherborn 1901: 142).

His early education was unremarkable but, after examinations, he did obtain from the South Kensington Museum (now the Victoria and Albert Museum) a certificate in geology which afforded him a life ticket to the library of that Museum. Geology being his favorite subject, this ticket to the library was heaven-sent and undoubtedly opened the door for his unquenchable thirst for knowledge. Having used the ticket often helped influence his philosophy toward education in that he felt that students should not be given facts, but should instead learn where to be able to find them (Norman 1944: 18). This self-professed credo of the essential tool in learning being the use of finding aids may have helped lead him toward his passion for indexing, as eventually Sherborn made his own finding aids for others to help empower them and expedite the search process, especially for things related to zoology.

Financial misfortune of his father's business forced Sherborn to abandon his education at the age of 14 and he soon found a job at the bookseller's and stationer's shop of Frederick William Stockley (1872–1948). Sherborn immersed himself in his work and soon became familiar with every aspect of the book trade. His duties included tending to the shop, cataloguing the stock, and traversing the streets of London collecting the day's orders of books, the last duty of which allowed him to find good bargains for his own personal book collecting. It is without a doubt that both his life ticket to the Museum library and his 6-year experience with the bookseller trade were to be linchpins in his future expertise with bibliography, dating research, and indexing. Although he was brought up with the financial hardships of his father's business, Sherborn himself was prudent with money throughout his life and after he passed away in 1944, probate records have his effects listed in the amount of £11619 (Anonymous 2010), which is equivalent to over US$500,000 in 2014 currency.

Stockley's bookselling business eventually went into bankruptcy in 1901 but before this, in 1883, Sherborn had left his employment in the bookselling trade to take on a few other odd jobs. The following year, Sherborn became employed by then-retired geologist and paleontologist Thomas Rupert Jones (1819–1911), and Charles's professional career had now been set on course. This association with Jones ultimately led to a visit to the new natural history museum at South Kensington and meeting the many scientists in the Geology Department. Jones had employed Sherborn to help illustrate and finish some monographic works on Foraminifera. It was not long after his initial work with forams that Sherborn realized a good bibliography and index were essential to better understanding and study of them. His work on the foram bibliography began around 1886 and was published a few years later (Sherborn 1888).

His acquaintances made at South Kensington led to Sherborn being employed by the British Museum (Natural History) around this same time. He was initially contracted in the Geology Department to mount specimens, but he quickly became involved in indexing and bibliographic work.

After his bibliography of the Foraminifera came his index to its genera and species (Sherborn 1893–1896). In discussing the foram bibliography and index to his biographer J.R. Norman, Sherborn gave a quote that encapsulated his life-long obsession:

"I suppose that I must have a card-index mind, because the preparation of my *Bibliography* and *Index* of that group (which my friends considered to be incredibly dull) gave me a lot of pleasure." (Norman 1944: 51).

Thus, with that first index, Sherborn was bitten by the indexing bug and never looked back. He was addicted. His life's path had been chosen.

With his "card-index mind" in full gear, working on the bibliography and concurrent assembling of the card-index for the Foraminifera gave Sherborn an idea. He felt that what could be done for the forams could also be done for all of zoology: an *Index Animalium* that would give a complete listing of every genus and species name, author, and accurate date of publication. Whether or not he understood the immensity of the task, this work would, in essence, captivate much of his time for the next 43 years.

A selected list of members of the Indexer's Club

During those next 43 years, and even a few years afterwards, Sherborn was involved in three types of indexing: bibliographies, ascertaining correct dates of publication (i.e., putting publications in proper chronological order), and nomenclators (lists of names). Many of his predecessors in this Indexer's Club who were involved in various types of indexing may well have been potential inspirations for him. Others in this Club may well have in turn been inspired by Sherborn in their own work.

Bibliographies make order of writings that otherwise would be scattered citations and it is one of the first forms of indexing. The Greek librarian Kallimachos (310–240 BC) "invented" the library catalog and was the first bibliographer (Blum 1991); and he is most famous for having made a bibliography of all the holdings of the Alexandrian Library. This listing was undoubtedly useful for users of the library, but Kallimachos unfortunately did not make a back-up copy and, when the Alexandrian Library was destroyed some 200 years after his death, the bibliography by Kallimachos was also destroyed, thus we will never know all that was in that library. Despite that unfortunate loss, the discipline of making bibliographies continues to this day and derives from the work of this man.

Louis Agassiz (1807–1873) (Fig. 3), a geologist and paleontologist by profession, became well known for two large series of works. One was his 4-volume bibliography, *Bibliographica zoologiae et geologiae* in which he was assisted by H.E. Strickland and W. Jardine (Agassiz and Strickland 1848, 1850, 1852; Agassiz et al. 1854). The other was his *Nomenclator zoologicus*, a list of animal generic names in a series of 12 fascicles pub-

Figure 3. Louis Agassiz (from Wikimedia Commons).

lished from 1842 to 1846; a summary volume of all the generic names was published in 1846 as *Nomenclatoris zoologici index universalis* (Agassiz 1846).

Another bibliographer who had conducted his research prior to Sherborn's working on his *Index* was ornithologist Elliott Coues (surname pronounced "cows", not "coos") (1842–1899) (Fig. 4). Coues worked at the Smithsonian and produced a four-part bibliography of ornithology (Coues 1878, 1879, 1880a, 1880b). His bibliographic work was excellent and showed the labor necessary to provide users with an ordered set of articles. However, it was a quote by him that I feel is worthy of repeating here. I quoted only a portion of it in my own bibliography of Diptera books (Evenhuis 1997) but the full quote should have been reproduced there as it typifies what can easily happen to bibliographers like Sherborn, myself, and possibly others once we start making bibliographies:

Figure 4. Elliott Coues (from Wikimedia Commons).

"... bibliography is a necessary nuisance and horrible drudgery that no mere drudge could perform. It takes sort of an inspired idiot to be a good bibliographer and his inspiration is as dangerous a gift as the appetite of the gambler or dipsomaniac – it grows with what it feeds upon, and finally possesses its victim like any other invincible vice." (Coues 1897: 39).

Coues was indeed addicted to bibliography and his fervent devotion to his work showed him to also do other types of indexing, such as also producing various checklists of North American birds. Some of these checklists were simple lists of common names and scientific names, but others came with classical language etymologies and sometimes even delved into proper orthoepy [correct pronunciation] (Coues 1882). Although it is a much-debated specialty, we need more works that research proper orthoepy in biology, Jaeger (1960) being one of the last major ones in that discipline.

Another predecessor of Sherborn who compiled both bibliographies and nomenclators was Samuel Hubbard Scudder (1837–1911) (Fig. 5). Scudder was primarily an entomologist and, like Agassiz, was interested in paleontology. While he was assistant librarian in charge of the catalogue at Harvard College [he resigned this post on 1 December 1882 to become the first editor for some new-fangled journal called *Science*], Scudder compiled a serialized bibliography. However, he restricted it to the

Figure 5. Samuel Hubbard Scudder (from Wikimedia Commons).

literature of fossil insects (Scudder 1880–1882). A later revised and annotated edition was published in Scudder (1890). But before he had even finished his first iteration of that bibliography in 1882, he had decided to assemble the names from the previous genus-group name nomenclators of Agassiz (e.g., 1846) and Marschall (1873) and bring the list up-to-date. By employing a team of colleagues worldwide who gave him lists of names in their discipline, he was able to provide the most complete list of genus-group names in zoology at that time in two parts (Scudder 1882, 1884). Scudder had indicated in his prefatory narrative that there would be periodic updates of this list but they never appeared. This absence of updates strained the patience of Scudder's colleagues at the British Museum (Natural History). And since many of the curatorial staff there were already compiling names of animals each year for the *Zoological Record* and had these names at hand, C.O. Waterhouse, with the assistance of David Sharp, eventually provided an update to Scudder's nomenclator to bring the generic names up to the year 1900 (Waterhouse 1902).

C.O. Waterhouse's (1902) nomenclator and a list of bird names a few years earlier by another Waterhouse (Waterhouse 1889) were no doubt inspirations to Sherborn and probably assisted him in his production of the *Index Animalium*. But the single person who without a doubt had the most influence on Sherborn's idea for an *Index*

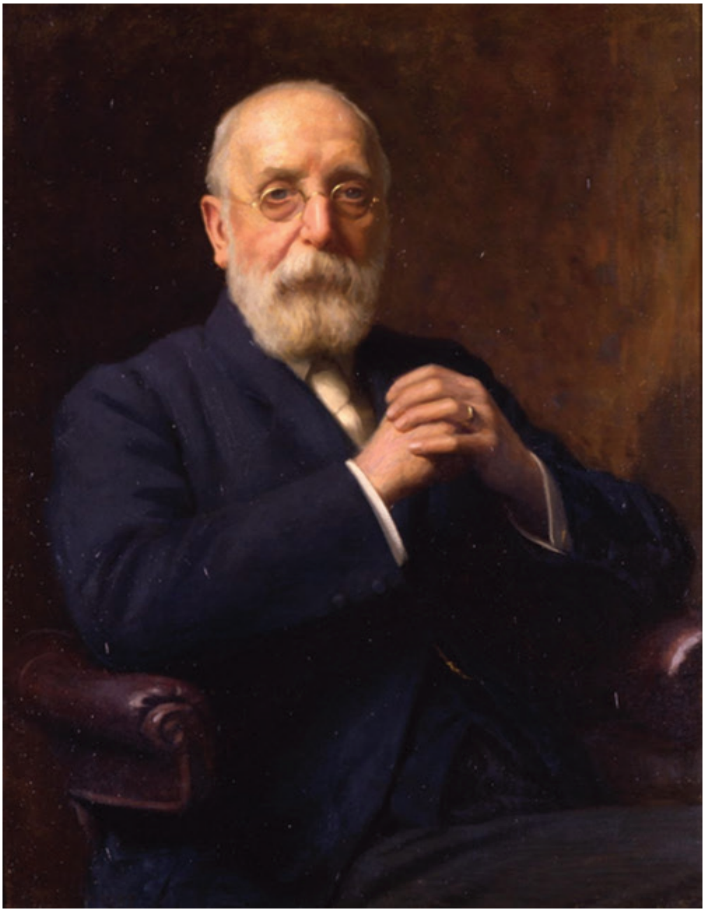


Figure 6. Benjamin Daydon Jackson (courtesy of the Zoological Society of London).

Animalium was Benjamin Daydon Jackson (1846–1927) (Fig. 6), who was busy working on the first nomenclator of plants, the *Index Kewensis* (Jackson 1895). Jackson was already immersed in his index when Sherborn had begun his contract work at the British Museum. Jackson's list was not just of generic names but also of species names. It was the first list to bring together all the names known of plants at that time. If the plants could be done, why not the animals?

Work on the *Index Animalium*

Sherborn began his work organizing his *Index* in the late 1880s and in May 1890 sent a letter to *Nature* (Sherborn 1890) and a similar one to *Le Feuille des Jeunes Naturalistes* outlining his proposed project and requesting advice and feedback from readers. After receiving responses from colleagues throughout Europe, Sherborn began work in earnest on 1 July 1890. Yearly progress reports were given to the British Association for the Advancement of Science in the early going and after six year's of work, some 130,000 slips with names had been catalogued by Sherborn (Sherborn 1896a).

The methodology employed by Sherborn for his *Index* is exemplary of anyone who wishes to produce an accurate and complete list or database of names. He avoided perpetuating potential errors by others by not working from previous lists. Instead he examined each original publication, scanning each page and writing the binomials he found on two slips of paper: one for the alphabetical index by species; the other for the index by genus. Sherborn's methodology was painstakingly tedious but was the only way to ensure all names in a particular publication would be captured in his *Index*. Despite the rigor that went into this form of data entry, Sherborn's work is not without its errors and omissions. Poche (1939) gave a list of over 2,000 names that were missed by Sherborn, mostly malacological and subspecific or varietal names, but most were names already rejected by ICZN action in 1912, so are of little value to malacology. Welter-Schultes et al. (2016) gives further details on other errors in Sherborn's *Index*. Nevertheless, the incredibly small error rate for some 440,000 hand-written names from examination of nearly 28,000 books and articles over 43 years of work is indeed an astonishing feat that will probably never be duplicated.

Bibliographies

Both indexing and bibliographies were an important part of the work that occupied Sherborn for almost half of his life. His own words exemplify their need:

"The systematist requires certain tools for his work, of which not the least important are good bibliographies and indexes." (Norman 1944: 49).

We have already mentioned Sherborn's first bibliography on Foraminifera, but he obviously did not stop with that. Between 1888 and 1895 a "Bibliography of Malaya" appeared in serial form in the *Journal of the Straits Branch of the Royal Asiatic Society.* Other subsequent bibliographies included a list of natural science reference works (Sherborn 1894); a catalogue of the Linnaean Society Library (Sherborn 1896c); the writings of Gilbert White (Sherborn 1900); the conchological writings of Thomas Brown (Sherborn 1905); a bibliography of scientific literature relating to Egypt (Sherborn 1910, 1915); and a summary of natural history dating sources (Sherborn et al. 1936).

Dating works

Soon after beginning his work on compiling bibliographies, Sherborn realized the necessity for obtaining accurate dates of publication for the works he was listing. As Conrad (1853) stated in the prefatory sentences to his list of North American "Naiades":

"To render strict justice to every author according to date of publication, is not only the duty of the naturalist, but a necessity of science." (Conrad 1853: 243).

Sherborn's first article on dating (Sherborn 1891) was published soon after beginning his compilation of the first *Index Animalium* and an article pleading for the need for proper metadata on each publication of books was published a few years later (Sherborn 1896b). These studies on accurate dating were not only critical in resolving dating problems of some early zoological works that had never had been put into context with other works of that same period in time, but, in order to ascertain priority among taxonomic names, accurate dates of their original proposals were required. A number of dating articles of other early works soon followed his original 1891 paper and culminated in giving accurate dates in his bibliography that accompanied the first *Index Animalium* (Sherborn 1902).

Publication of the first *Index* did not stop Sherborn's work on dating since he needed to resolve further problems of dating for works that were to appear in his second *Index*. No fewer than 20 articles on dating or publication histories by Sherborn and various co-authors appeared between publication of the first *Index* and the publication of the last part of the second *Index*. As with his first *Index*, the bibliography of his second *Index* also was replete with proper dating.

Despite his *Index Animalium* being completed in 1933, Sherborn, being the consummate facilitator and indexer, did not stop being concerned with proper dates of publication and, with the assistance of bibliographer Francis James Griffin (1904–1990) and Kew Gardens librarian H.S. Marshall, published a synthesis of published sources that focused on bibliographical research and gave dates of publication for biological works. Their paper appeared as the first article in the first issue of the *Journal of the Society for the Bibliography of Natural History* (Sherborn et al. 1936), the society of which was founded by a group of fellow bibliographers who had an interest in seeing the results of bibliographic and dating research to be made public via a scientific journal. Not surprisingly, Sherborn was the Society's first president, typifying the zeal he had for this subject and also typifying his desire to further facilitate such information to fellow bibliographers and taxonomists worldwide. The formation of this society could be said to have been the beginnings of a formalized group of individuals who were interested in the discipline that would later be called "bioinformatics".

Influence on and future of bioinformatics

Sherborn did indeed construct a solid foundation for the future with his seminal works on bibliography, dating, and indexing. His works were followed by many others, either providing indexes, bibliographies, or catalogues, on small groups of organisms such as by family or country, or larger, more comprehensive studies. By way of a few examples, I will list a few of some of the more major works that have been produced since Sherborn's *Indexes* and inspired by his vision.

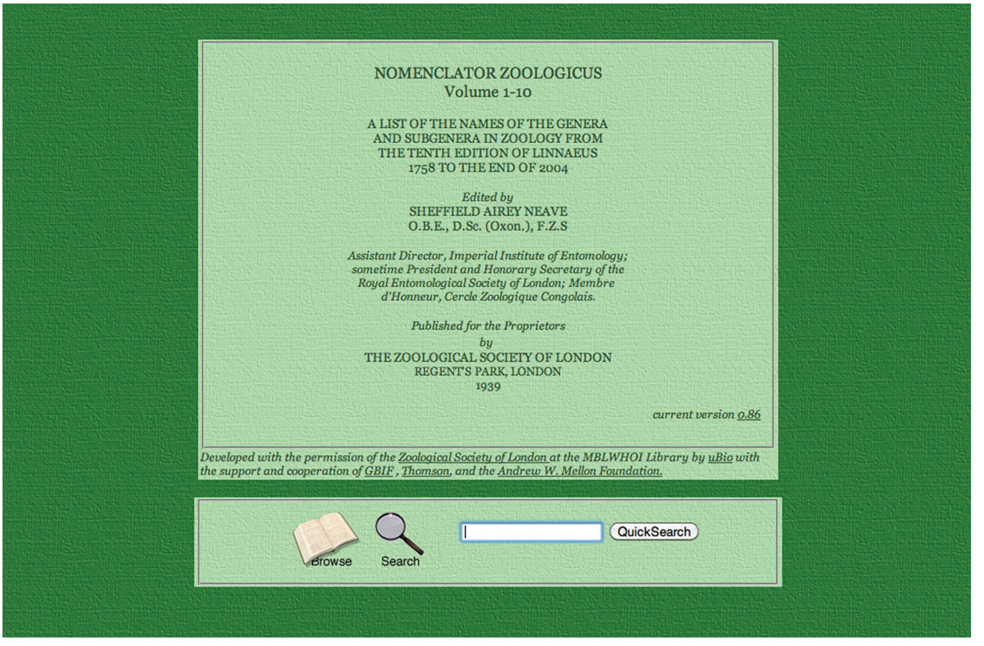


Figure 7. Screenshot of *Nomenclator Zoologicus* online.

Nomenclator Zoologicus

Sheffield Airey Neave (1879–1961), while working at the Imperial Bureau of Entomology, recognized the need for up-to-date information on all generic names in zoology and envisioned a nomenclator to index all of them. With initial funding from the Zoological Society of London, the original edition of his *Nomenclator Zoologicus* in four volumes (Neave 1939–1940) was published. Five supplement volumes appeared until 1994, but the absence of further supplements is not as devastating as originally thought by some. The hard-copy volumes, which used to be a major reference for any taxonomist wishing to describe a new genus and needing to check whether or not that name was used before, have all been digitized. With funding to a few dedicated staff from Thompson-Reuters, GBIF, and the Mellon Foundation, and partnering with the Zoological Society of London, over 340,000 genus-group names in the *Nomenclator Zoologicus* are now available online: http://uio.mbl.edu/NomenclatorZoologicus/ (Fig. 7).

Catalog of Fishes

In the 1980s, William N. "Bill" Eschmeyer (1939–), ichthyologist at the California Academy of Sciences, decided to organize all the taxonomic information on ichthyology. With initial partial funding from the National Science Foundation (and later technical support from the California Academy of Sciences), he began a task some thought impossible: to catalogue all the genera and species of fishes worldwide. Undaunted, his first volume on genus-group names appeared in 1990 (Eschmeyer and Bailey 1990) and eight years later his dream was fulfilled in the publication of his three-volume catalog of all genus- and species-group names of fishes and an associated bibliography (Eschmeyer 1998). He had followed the methodology of Sherborn in not only examining all the original literature, but made painstaking efforts to obtain accurate dates of publication for these works. But Bill was aware that hard copies of his work would not meet the needs of everyone, and the internet showed him the potential to reach workers who did not have the resources to purchase his volumes. He did not miss the opportunity and made available to the public all of the information he had compiled into a database over the years via a simple and user-friendly web interface [http://researcharchive.calacademy.org/research/Ichthyology/catalog/fishcatmain.asp]. The database is continually updated and is the best one-stop shopping for fish names and literature resources anywhere.

Systema Dipterorum

In 1984 at the XVII International Congress of Entomology in Hamburg, a group of dipterists working at the U.S. Department of Agriculture's Systematic Entomology Laboratory in Washington, DC proposed a plan to database all the names of two-winged flies (Diptera), which it turns out comprise a fairly large percentage of all animal names (15%). A great deal of interest was spurred from that presentation but, aside from a grant from GBIF in 2003, support through 4D4Life in 2009, the CoL Rotating Fund in 2010, and small yearly grants from the Schlinger Foundation during the last few years, meager funding over the years supports the time and staffing necessary to maintain and complete the project and funding ceased altogether in the last two years. This has not deterred F. Christian Thompson (1944–) from seeing this vision to fruition. The ensuing 25-some odd years since the announcement in 1984 saw Chris working diligently in the evenings in his home office to continually enter, update, and verify data in the *Systema Dipterorum* (SD). Thompson and Pape (2016) give further details on the history and status of SD, and as of October 2013, 4,653 family-group names (most catalogued in Sabrosky 1999), 23,437 genus-group names, and 198,258 species-group names in 32,900 published works have been entered into SD and are available for searching via a robust web interface at [www.diptera.org] (see also Pape and Thompson 2010). Because of the high standards put onto the methodology of data acquisition, entering, and vetting, and because data vetting follows the rigorous example set by Sherborn in his *Index Animalium* of examining every original paper, the SD is today one of the most accurate and complete databases of any megadiverse group of animals.

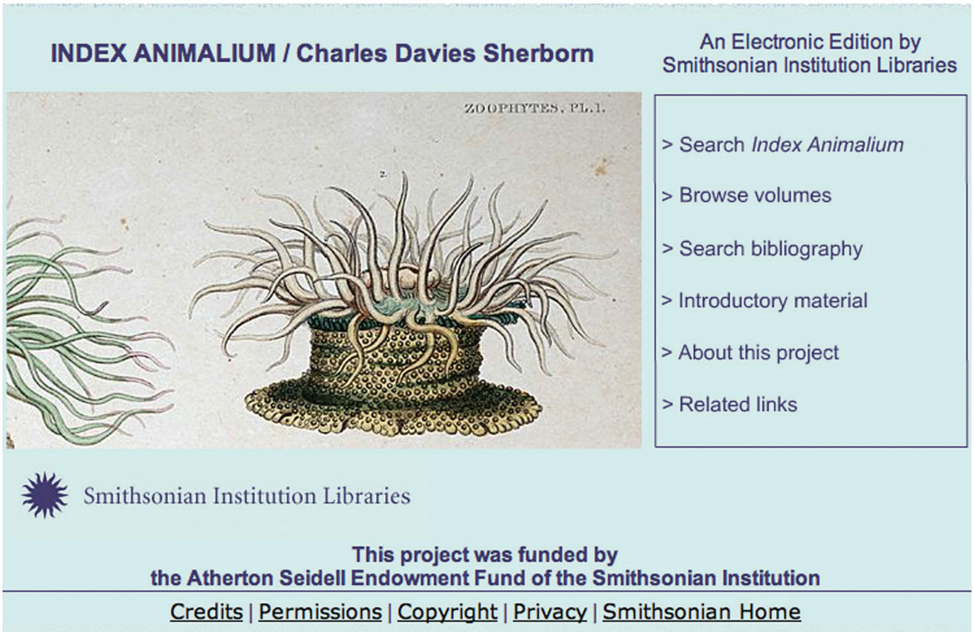


Figure 8. Screenshot of *Index Animalium* online.

Index Animalium Online

With exemplary foresight, in 2004 the Atherton Seidell Endowment Fund at the Smithsonian Institution brought the work of Sherborn into the 21st century by recognizing the importance of Sherborn's *Index Animalium* and making it available to as wide array of users as possible. It funded both the digitization of both editions, data parsing and re-keying, and design and implementation of the user interface on the web [http://www.sil.si.edu/digitalcollections/indexanimalium/TaxonomicNames/] (Fig. 8) allowing searching of every name that occurs in every part that was published from 1902 to 1933. Pilsk et al. (2016) give details on the labor that went into the digitization and parsing data on each page, with the goal of achieving a 99.995% accuracy rate in converting the OCR text.

The Future

It is the internet, and whatever iterations it evolves into, that is and will be the medium for making available the information we need on all aspects of cataloguing, nomenclature, bibliography, and dating. The final few papers in this volume (e.g., Penev et al. 2016; Page 2016; Pyle 2016) deal with how the names of animals and the information and metadata associated with them can be standardized and implemented for universal access. Although there is much yet to be done, we are making significant progress in serving up the information on biological names for future generations through immediately accessible electronic media.

Sherborn could never have dreamed that his small slips of paper with names handwritten on them would be replaced by 1s and 0s in binary form so that they could be transmitted electronically through an electronic medium that would have a viewing screen on everyone's desk or handheld device. But he can be comforted that his tireless work of 43 years in producing his *Index Animalium* has had a profound influence on what we do today to facilitate the research of others in studying biological taxa and the names associated with them.

Acknowledgements

In my indexing life I have been inspired largely by two individuals other than Sherborn: Chris Thompson and Bill Eschmeyer. The latter's work has been a goal of mine for Diptera genus-group names for many, many years and I can only dream to be able to achieve for Diptera what Bill has been able to do for fish. The former I am honored to call my friend and mentor, and all that I initially learned of bibliography, nomenclature, databasing, and finding accurate dates of publication I owe to him. Chris also introduced me to Sherborn and I have been in awe of him and his work ever since. I thank the thought provoking and sometimes almost cosmological discussions of the history and future of bioinformatics that I have had over the years with many colleagues, but especially Chris Thompson, Richard Pyle, and Thomas Pape. Thanks also to Alan Kabat who reviewed the manuscript and whose suggestions improved it. The Zoological Society kindly gave permission to reproduce the portrait of Benjamin Jackson. And special thanks to Ellinor Michel and the Society for History of Natural History for organizing the symposium honoring the 150th anniversary of Sherborn's birth and for honoring me with being selected to give this keynote address as the Ramsbottom Lecture for 2011.

References

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, nec non variis adnotationibus et emendationibus. Jent & Gassman, Soloduri, 393 pp.

Agassiz L, Strickland HE (1848) Bibliographia zoologiae et geologiae. A general catalogue of all books, tracts, and memoirs on zoology and geology by Prof. Louis Agassiz [corrected, enlarged, and edited by H.E. Strickland] I. Ray Society, London, 506 pp.

Agassiz L, Strickland HE (1850) Bibliographia zoologiae et geologiae. A general catalogue of all books, tracts, and memoirs on zoology and geology by Prof. Louis Agassiz [corrected, enlarged, and edited by H.E. Strickland] II. Ray Society, London, 492 pp.

Agassiz L, Strickland HE (1852) Bibliographia zoologiae et geologiae. A general catalogue of all books, tracts, and memoirs on zoology and geology by Prof. Louis Agassiz [corrected, enlarged, and edited by H.E. Strickland] III. Ray Society, London, 657 pp.

Agassiz L, Strickland HE, Jardine W (1854) Bibliographia zoologiae et geologiae. A general catalogue of all books, tracts, and memoirs on zoology and geology by Prof. Louis Agassiz [corrected, enlarged, and edited by H.E. Strickland] IV. Ray Society, London, 604 pp.

Anonymous (2010) England & Wales, National Probate Calendar (Index of Wills and Administrations), 1858–1966. Provo, Utah. http://ancestry.com [non-public database on-line]

Blum R (1991) Kallimachos – The Alexandrian Library and the origins of bibliography. University of Wisconsin Press, Madison, 282 pp.

Conrad TA (1853) A synopsis of the family of Naiades of North America, with notes, and a table of some of the genera and sub-genera of the family, according to their geographical distribution, and descriptions of genera and sub-genera. Proceedings of the Academy of Natural Sciences of Philadelphia 6: 243–269.

Coues E (1878) Bibliographical Appendix. List of faunal publications relating to north American ornithology. Miscellaneous Publication of the United States Geological Survey of the Territories 2: 567–784.

Coues E (1879) Second instalment of American ornithological bibliography. Bulletin of the United States Geological and Geographical Survey of the Territories 5(2): 239–330. doi: 10.5962/bhl.title.57781

Coues E (1880a) Third instalment of American ornithological bibliography. Bulletin of the United States Geological and Geographical Survey of the Territories 5(4): 521–1072.

Coues E (1880b) Fourth instalment of ornithological bibliography: being a list of faunal publications relating to British birds. Proceedings of the United States National Museum 2: 359–482. doi: 10.5479/si.00963801.101.359

Coues E (1882) The Coues check list of North American birds. Second edition, revised to date, and entirely rewritten, under direction of the author with a dictionary of the etymology, orthography, and orthoepy of the scientific names, the concordance of previous lists, and a catalogue of his ornithological publications. Estes & Lauriat, Boston, 165 pp.

Coues E (1897) Dr. Coues' column. The Osprey 2: 39–40.

Eschmeyer WN (1998) Catalog of fishes (3 volumes). California Academy of Sciences, San Francisco, 2905 pp.

Eschmeyer WN, Bailey RM (1990) Catalog of the genera of Recent fishes. California Academy of Sciences, San Francisco, 697 pp.

Evenhuis NL (1997) Litteratura Taxonomica Dipterorum (1758–1930) being a selected list of the books and prints of Diptera taxonomy from the beginning of Linnaean nomenclature to the end of the year 1930; containing information on the biographies, bibliographies, types, collections, and patronymic genera of the authors listed in this work; including detailed information on publication dates, original and subsequent editions, and other ancillary data concerning the publications listed herein. 2 vols. Backhuys Publishers, Leiden, 871 pp.

Jackson BD (1895) Index Kewensis. An enumeration of the genera and species of lowering plants from the time of Linnaeus to the year 1885 inclusive together with their author's names, the works in which they were first published, their native countries and their synonyms. 2 volumes. Clarendon Press, Oxford.

Jaeger EC (1960) The biologist's handbook of pronunciations. Thomas, Springfield, Illinois, 317 pp.

Kendall J (2008) The man who made lists. Love, death, madness, and the creation of Roget's thesaurus. G.P. Putnam's Sons, New York, 297 pp.

Marschall A (1873) Nomenclator zoologicus continens nomina systematica generum animalium tam viventium quam fossilium, secundum ordinem alphabeticum disposita. CR Societatis Zoologico-Botanicae, Vindobonae, 482 pp.

Neave SA (1939–1940) Nomenclator zoologicus. A list of the names of genera and subgenera in zoology from the tenth edition of Linnaeus 1758 to the end of 1935 (4 volumes). Zoological Society of London, London.

Norman JR (1944) Squire – Memories of Charles Davies Sherborn. GG Harrap & Co, London, 202 pp.

Page RDM (2016) Surfacing the deep data of taxonomy. In: Michel E (Ed.) Anchoring Biodiversity Information: From Sherborn to the 21st century and beyond. ZooKeys 550: 247–260. doi: 10.3897/zookeys.550.9293

Pape T, Thompson FC (2010) Systema Dipterorum. Version 1.0. Available at: http://www.diptera.org [accessed 6 May 2012]

Penev L, Paton A, Nicolson N, Kirk P, Pyle RL, Whitton R, Georgiev T, Barker C, Hopkins C, Robert V, Biserkov J, Stoev P (2016) A common registration-to-publication automated pipeline for nomenclatural acts for higher plants (International Plant Names Index, IPNI), fungi (Index Fungorum, MycoBank) and animals (ZooBank). In: Michel E (Ed.) Anchoring Biodiversity Information: From Sherborn to the 21st century and beyond. ZooKeys 550: 233–246. doi: 10.3897/zookeys.550.9551

Pilsk SC, Kalfatovic MR, Richard JM (2016) Unlocking *Index Animalium*: From paper slips to bytes and bits. In: Michel E (Ed.) Anchoring Biodiversity Information: From Sherborn to the 21st century and beyond. ZooKeys 550: 153–171. doi: 10.3897/zookeys.550.9673

Poche F (1939) Supplement zu C.D. Sherborns Index Animalium. Festschrift zum 60. Geburtstag von Professor Dr. Embrik Strand 5: 477–615.

Pyle RL (2016) Towards a Global Names Architecture: The future of indexing scientific names. In: Michel E (Ed.) Anchoring Biodiversity Information: From Sherborn to the 21st century and beyond. ZooKeys 550: 261–281. doi: 10.3897/zookeys.550.10009

Roget PM (1834) The Bridgewater treatises on the power wisdom and goodness of God manifested in the Creation (Treatise V). Animal and vegetable physiology considered with reference to natural theology (2 volumes). W. Pickering, London.

Roget PM (1852) Thesaurus of English words and phrases, classified and arranged so as to facilitate the expression of ideas, and assist in literary composition. Longman, Brown, Green & Longman's, London, 418 pp.

Sabrosky CW (1999) Family-group names in Diptera. An annotated catalog. Myia 10: 1–360.

Scudder SH (1880–1882) A bibliography of fossil insects. Harvard College Library Bulletin 15: 48–51; 16: 87–88 (1880); 17: 121–124; 18: 157–162; 19: 202–208; 20: 252–257 (1881); 21: 296–299; 22: 333–337; 23: 407–414 (1882).

Scudder SH (1882) Nomenclator zoologicus. An alphabetical list of all generic names that have been employed by naturalists for recent and fossil names from the earliest times to the close of the year 1879. I. Supplemental list of genera in zoology. List of generic names employed in zoology and paleontology to the close of the year 1879, chiefly supplemental to those catalogued by Agassiz and Marschall, or indexed in the Zoological Record. Bulletin of the United States National Museum 19: 1–376.

Scudder SH (1884) Nomenclator zoologicus. An alphabetical list of all generic names that have been employed by naturalists for recent and fossil names from the earliest times to the close of the year 1879. II. Universal index to genera in zoology. Complete list of generic names employed in zoology and paleontology to the close of the year 1879, as contained in the nomenclators of Agassiz, Marschall and Scudder, and in the Zoological Record. Bulletin of the United States National Museum 19: 1–340.

Scudder SH (1890) A classed and annotated bibliography of fossil insects. Bulletin of the United States Geological Survey 69: 7–101. doi: 10.1038/042054c0

Sherborn CD (1888) A bibliography of the Foraminifera, recent and fossil, from 1565–1888; with notes explanatory of some of the rare and little-known publications. Dulau & Co, London, 152 pp.

Sherborn CD (1890) Index Generum et Specierum Animalium. Nature 42: 54.

Sherborn CD (1891) Dates of the parts of P.S. Pallas's 'Icones Insect. p. Ross. Sibir.' and 'Nov. spec. Quadr. Glirium.' Annals and Magazine of Natural History (6) 7: 236. doi: 10.1080/00222939109460601

Sherborn CD (1893–1896) An index to the genera and species of the Foraminifera. Smithsonian Miscellaneous Collections 856 & 1031, 485 pp.

Sherborn CD (1894) Books of reference in the natural sciences. Natural Science 5: 115–127.

Sherborn CD (1896a) Explanation of the plan adopted for preparing an "Index Generum et Specierum Animalium". Proceedings of the Zoological Society of London 1897: 285–288.

Sherborn CD (1896b) The dating of books. Natural Science 9: 280.

Sherborn CD (1896c) Catalogue of the library of the Linnaean Society of London. New edition. Longmans, Green & Co., and Williams & Norgate, London, 727 pp.

Sherborn CD (1898) Charles Sherborn, engraver. Notes & Queries (9) 3: 11.

Sherborn CD (1899) An index to the generic and trivial names of animals, described by Linnaeus, the 10th and 12th editions of his "*Systema Naturae*". The Manchester Museum, Owens College. Museum Handbooks. Publication 25. Dulau, London, 108 pp.

Sherborn CD (1900) A bibliography of Gilbert White. In: Sharpe RB (Ed.) The natural history & antiquities of Selborne & a garden kalendar by the Reverend Gilbert White M.A. (Vol. II). ST Freemantle, London, 349–364.

Sherborn CD (1901) A history of the family of Sherborn. Mitchell & Hughes, London, 212 pp.

Sherborn CD (1902) Index animalium sive index nominum quae ab A.D. MDCCLVIII generibus et speciebus animalium imposita sunt. Sectio prima. A kalendis ianuariis, MDCCLVIII usque ad finem decembris, MDCCC. Typographico Academico, Cantabrigiae [= Cambridge], 1195 pp.

Sherborn CD (1905) The conchological writings of Captain Thomas Brown. Proceedings of the Malacological Society of London 6: 358–360.

Sherborn CD (1910) Bibliography of scientific and technical literature relating to Egypt, 1800–1900. Preliminary edition. National Printing Department, Cairo, 155 pp.

Sherborn CD (1912) A sketch of the life and work of Charles William Sherborn, painter-etcher. With a catalogue of his bookplates comp. by himself and George Heath Viner. Ellis, London, 109 pp.

Sherborn CD (1915) Bibliography of scientific and technical literature relating to Egypt, 1800–1900. Survey Department, Cairo, 310 pp.

Sherborn CD (1922–1932) Index animalium sive index nominum quae ab A.D. MDCCLVIII generibus et speciebus animalium imposita sunt. Sectio secunda. A kalendis ianuariis, MDCCCI usque ad finem decembris, MDCCCL. 33 parts. Longmans, Green & Co. & British Museum (Natural History), London.

Sherborn CD (1940) Where's the _____ collection? An account of the various natural history collections which have come under notice by the compiler. Cambridge university Press, Cambridge, UK, 148 pp.

Sherborn CD, Griffin, FJ, Marshall HJ (1936) A catalogue of papers concerning the dates of publication of natural history books. Journal of the Society for the Bibliography of Natural History 1: 1–30. doi: 10.3366/jsbnh.1936.1.1.1

Thompson CF, Pape T (2016) Sherborn's influence on *Systema Dipterorum*. In: Michel E (Ed.) Anchoring Biodiversity Information: From Sherborn to the 21st century and beyond. ZooKeys 550: 135–152. doi: 10.3897/zookeys.550.9447

Waterhouse CO (1902) Index zoologicus. An alphabetical list of names of genera and subgenera proposed for use in zoology as recorded in the "Zoological record" 1880–1900 together with other names not included in the "Nomenclator zoologicus" of S. H. Scudder. Gurney & Jackson, London, 421 pp.

Waterhouse FH (1889) Index generum avium. A list of the genera and subgenera of birds. RH Porter, London, 240 pp.

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. In: Michel E (Ed.) Anchoring Biodiversity Information: From Sherborn to the 21st century and beyond. ZooKeys 550: 173–183. doi: 10.3897/zookeys.550.10041

ZooKeys 550: 33–56 (2016)
doi: 10.3897/zookeys.550.9975
http://zookeys.pensoft.net

RESEARCH ARTICLE

A peer-reviewed open-access journal
ZooKeys
Launched to accelerate biodiversity research

A magpie with a card-index mind – Charles Davies Sherborn 1861–1942

Karolyn Shindler[1]

1 *Library Associate, Natural History Museum, London, United Kingdom*

Corresponding author: *Karolyn Shindler* (K.Shindler@nhm.ac.uk)

Academic editor: *Ellinor Michel* | Received 12 May 2015 | Accepted 13 May 2015 | Published 7 January 2016

http://zoobank.org/6A25367F-11B8-4B3F-AEDA-F431E167F036

Citation: Shindler K (2016) A magpie with a card-index mind – Charles Davies Sherborn 1861–1942. In: Michel E (Ed.) Anchoring Biodiversity Information: From Sherborn to the 21st century and beyond. ZooKeys 550: 33–56. doi: 10.3897/zookeys.550.9975

Abstract

Charles Davies Sherborn was geologist, indexer and bibliographer extraordinaire. He was fascinated by science from an early age and like so many Victorians, the young Sherborn was a passionate natural history collector and was obsessed with expanding his collection of land and freshwater shells. He later described himself as being a 'thorough magpie' and having 'a card-index mind', and these two traits coalesced in his monumental *Index Animalium*, the compilation of which occupied 43 years of his life. One of the first visitors through the doors of the Natural History Museum in South Kensington when it opened in 1881, Sherborn began work there seven years later as one of the small band of unofficial scientific workers, paid by the number of fossils he prepared. By the time of his death in 1942, Sherborn's corner in the Museum was the first port of call for generations of scientists seeking advice, information – or an invitation to one of his famous 'smoke and chat' parties.

In addition to his work on the *Index*, Sherborn is also responsible for rescuing from damp and probable destruction the huge archive of Sir Richard Owen, the great comparative anatomist and the prime mover behind the creation of the Natural History Museum, London. Without Sherborn, this invaluable resource of correspondence, manuscripts and books may well have been irretrievably ruined.

Charles Davies Sherborn's fascination with science began early. Like many small boys he collected rocks and fossils and was obsessed by expanding his collection of land and freshwater shells. Few boys, however have attempted to construct volcanoes in their gardens in west London, the consequent explosion resulting in a visit from the police. Sherborn's life was never going to be ordinary – even his first flat when he left home was above an undertaker's, with wood chippings from the coffins the fuel for his fire.

But it was his passion for collecting that triumphed – his 'magpie habits' as he called it – and was to result in his spectacular work of bibliography, the *Index Animalium*, a true labour of love (and shamefully little financial reward). His object was to provide zoologists with a complete list of all the generic and specific names that had been applied to animals from 1 January 1758, giving a reference to the book or journal in which it was first published, and the date of publication. It was to occupy 43 years of his life.

Two years after he had embarked on the *Index,* Sherborn received an extraordinary and challenging invitation. He was asked by the Reverend Richard Startin Owen to sort through and organise the papers of his grandfather, the great anatomist and creator of the Natural History Museum, Professor Richard Owen, who died in 1892. The archive was vast – and in a terrible state, heaped up, vulnerable to damp and rats and in urgent need of rescue. It was a massive task, involving tens of thousands of letters, manuscripts and books.

Simply one of these undertakings would have defeated most people, but for this self-educated, extraordinarily generous man with an encyclopedic brain, this was very heaven.

Charles Davies Sherborn was born in Chelsea on 30 June 1861, the eldest child of Charles William Sherborn, a renowned etcher and line-engraver, and his wife, Hannah. Sherborn was sent to school at the age of three, but his formal education ended abruptly when he was 14: his father's business misfortunes meant he had to leave school and earn a living. But he was already, like so many Victorians of all ages, a passionate natural history collector, particularly of rocks and fossils and was obsessed with expanding his collection of land and freshwater shells.

His first job was in an upmarket stationery and bookshop in Bond Street, which Sherborn later claimed laid the foundation for his expertise in bibliography (Cleevely 2004). He did not, though, give up on science. His next job was as a clerk in a tailor's near the Museum of Practical Geology in Jermyn Street in London.

His spare time was spent there, or reading in the library of the Victoria and Albert Museum – which he always called by its old name, the South Kensington Museum – while his weekends would be occupied in fieldwork. When the Natural History Museum opened at Easter 1881, Sherborn maintained he was one of the first half dozen visitors through its doors.

Two years later, he met the retired Professor of Geology at the Royal Military College, Sandhurst, Thomas Rupert Jones, who asked Sherborn for help with papers he was writing on microscopic fossils known as foraminifera (Miller 2016). By 1887 – all

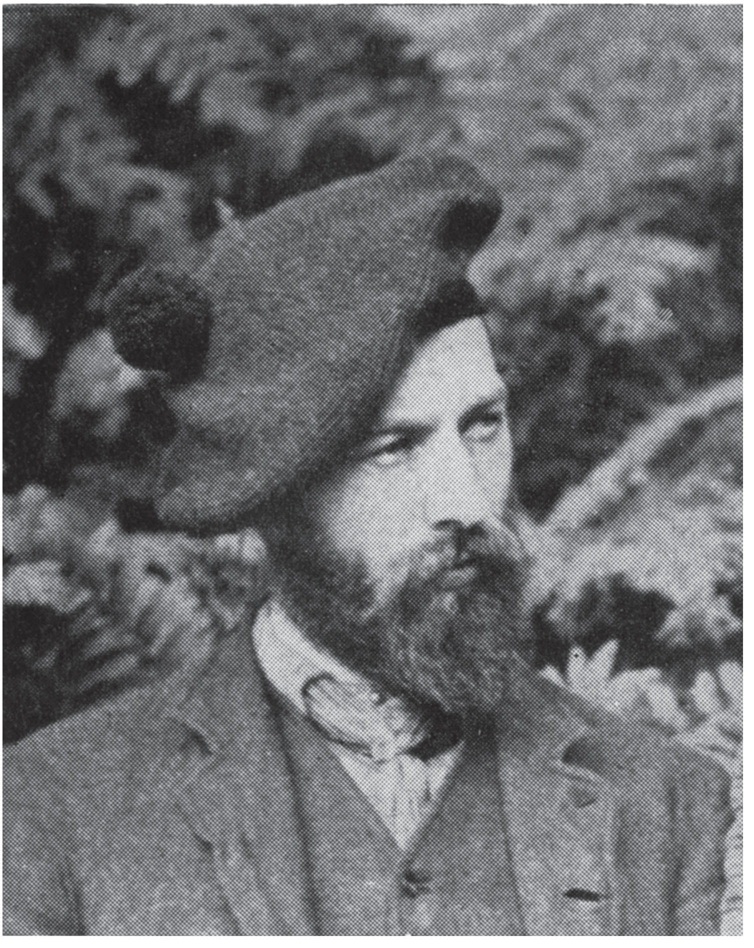


Figure 1. Sherborn at the age of about 25. This, according to his friend and biographer JR Norman, encapsulates 'Charles in the Geological Excursion Days'. (With permission of The Trustees of the Natural History Museum, London)

in his spare time – they had published three papers together, with Sherborn, who had considerable artistic talent, doing the drawings.

Prompted by the great number of journals they had had to consult, with Rupert Jones's encouragement Sherborn began to compile *A Bibliography of the Foraminifera* which was ready for publication in 1888. An American work on the same subject was published at this time, which Sherborn felt was so poor that he wrote a vitriolic criticism of it which was published in the journal *Nature* under the heading, 'An "Instructive" Bibliography of the Foraminifera'. According to Sherborn, it was 'absolutely untrustworthy', 'comparatively useless', with 'serious defects for which excuse must be difficult'. Furthermore, 'many of these errors and defects might have been avoided', he ended witheringly, 'had the compiler been used to public libraries' (Sherborn 1888). However justified Sherborn's criticism, the consequence for him was potentially

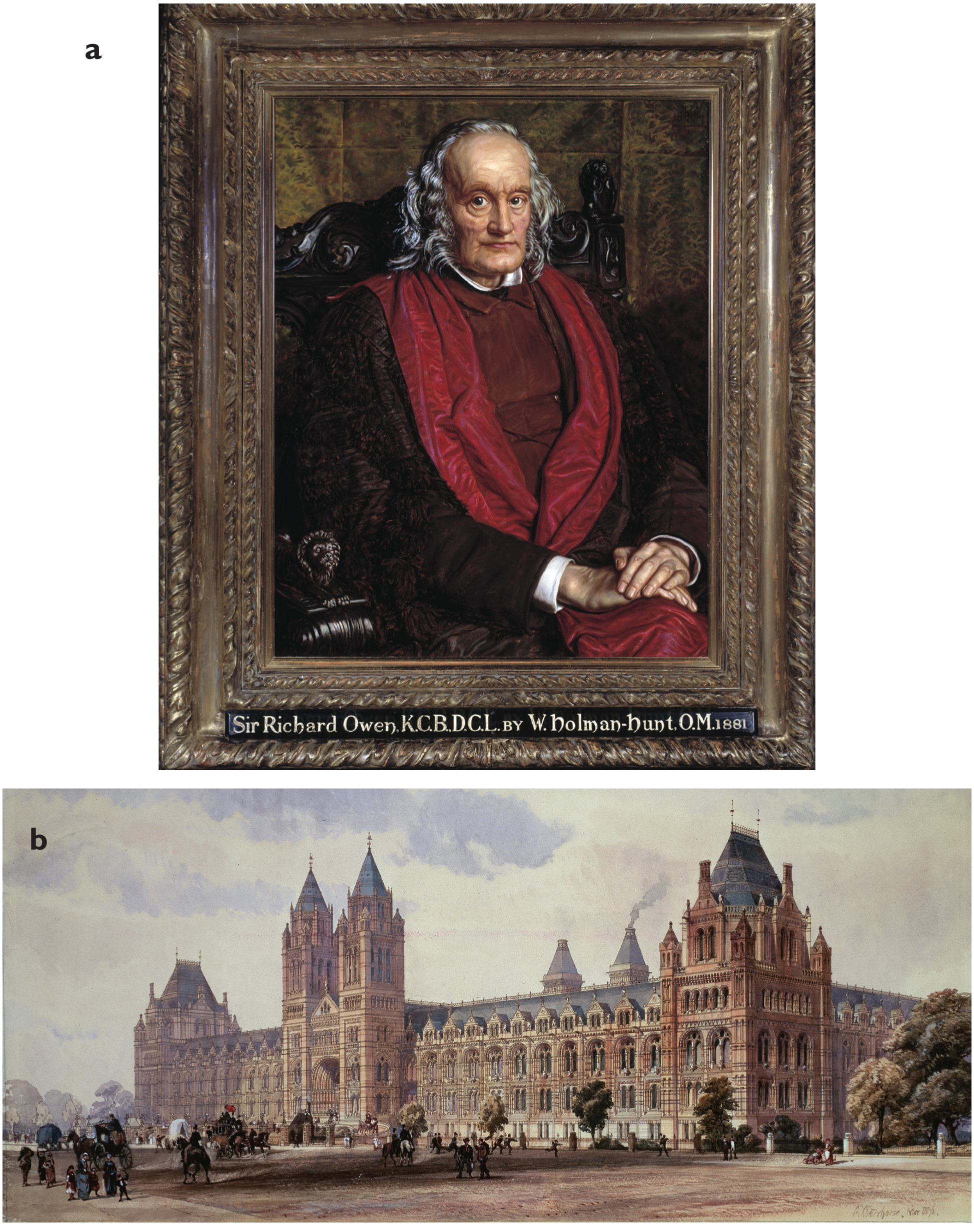


Figure 2. a Sir Richard Owen (1804–1892) in a portrait of 1881 by the Pre-Raphaelite painter William Holman Hunt (1827–1910). Owen's vision of a museum for natural history was realised that year with the opening of **b** the British Museum (Natural History) in South Kensington, which immediately became known as the Natural History Museum. This original illustration is by the Museum's architect, Alfred Waterhouse (1830–1905). During construction, the Treasury objected to the cost, and Waterhouse suggested it should be built in two stages, first the magnificent front, then later the back and two wings - one along Exhibition Road that you can just see in the drawing, and along Queen's Gate. With the front completed, the will to fund stage two vanished and the wings was never built. (With permission of The Trustees of the Natural History Museum, London)

Figure 3. The geologist and palaeontologist Thomas Rupert Jones (1819–1911) was Sherborn's scientific mentor and colleague. In 1864 he founded the *Geological Magazine* with Dr Henry Woodward of the British Museum's Geological Department.

disastrous. When he applied to the Royal Society for a grant for £100 for printing costs for his own *Bibliography*, it was refused on the grounds that he had written 'a savage criticism of a foreigner'. It was only the generous response of a publishing friend of his that ensured his work was published and, as Sherborn later wrote, 'wiped out completely the churlish action of the Royal Society'.

In 1888 Sherborn began part-time work in the Geology department at the Natural History Museum, preparing and cleaning fossils. He was paid according to the number of fossils he worked on. Quite literally, that changed his life. He loved the work, loved being in the world of natural history and, financially precarious though this was, he gave up his full-time employment – by this time he was working as a secretary at the Middlesex Hospital – and determined to try to earn a living through science. He became an unofficial scientific worker at the Museum, paid according to the amount of work he did. He never joined the staff.

Figure 4. Sherborn aged 32. He was rarely seen without his beloved pipe.: (With permission of The Trustees of the Natural History Museum, London)

His passion at first was geology and palaeontology, but through his work with Rupert Jones, and then his collaboration with the palaeontologist Arthur Smith Woodward at the Natural History Museum, he found himself increasingly drawn to scientific bibliography. There was, he discovered, an overwhelming need for zoologists and palaeontologists to have available to them a complete index of all scientific names applied to animals, living and extinct, giving the exact date and place of publication. In the mid-18th century, the Swedish naturalist, Carl Linnaeus brought order to the chaos of natural history names with his binominal (often called binomial) solution: giving everything living – plants and animals - two Latin names, the genus or generic name, and the trivial or specific name – a descriptive one. Inspired by his own *Bibliography of the Foraminifera* and a work endowed by Charles Darwin, the *Index Kewensis*, which was devoted to plants, Sherborn put forward a plan for an *Index Animalium*. The 10th edition of Linnaeus's work, published in 1758 was regarded as definitive and that is why Sherborn begins his monumental work then. His original plan was to have ended the *Index* in 1899. He took advice from many scientists, and wrote to *Nature* to announce the project and that he would be starting work on 1 July, 1890 – which he duly did (Sherborn 1890). Endearingly, he noted later that his friends considered his rigorous, precise work 'incredibly dull', although 'it gave me a lot of pleasure' (see Evenhuis (2016), on 'the Indexer's Mind').

No one had ever attempted anything on this scale before, and, as he noted in 1896, 'The vastness of the record is appalling'. He went on to remark with unusual optimism that, 'given time all difficulties disappear', although whether he would have agreed with that sentiment 30 years later is another matter (Sherborn 1896). His first task was to work out how to tackle it. In an exercise book with a shiny reddish-brown cover, now a bit tattered, Sherborn painstakingly evolved his own rules. He noted problems as they arose – 'Question of Double-barrelled names!!' – and methodically worked out through the pages ways they might be resolved. 'How far', he asked himself, 'can one accept authors who use one, two or three words as a specific term?' The issue of sub genus occupies a number of pages and clearly caused him considerable trouble, 'it is absolutely impossible', he fumed, 'to get unanimity among authors, & often the author has not the dimmest idea what he himself means!' (Sherborn Collection nd)

When he embarked on this project, he was also working on labelling and registering type specimens in the Geological Society and preparing fossils at the NHM during the day – both of which gave him an income. The index he worked on only at night, taking the books he was working on home to his flat.

Yet in one year, he had worked through every page of some 500 volumes and indexed 40,000 names. 'References', he wrote, 'are taken from one book at a time – i.e., a

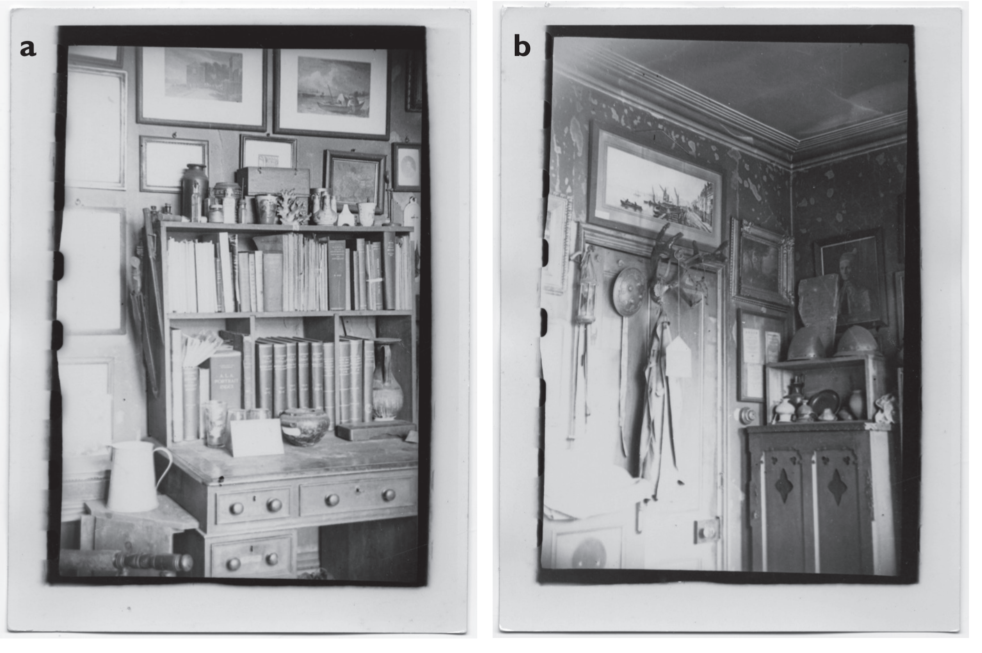


Figure 5. Sherborn's house at 49 Peterborough Road, London SW6 was a wonderful magpie's nest, filled with his eclectic and wide-ranging collections. These photos were probably taken after his death in 1942. (With permission of The Trustees of the Natural History Museum, London)

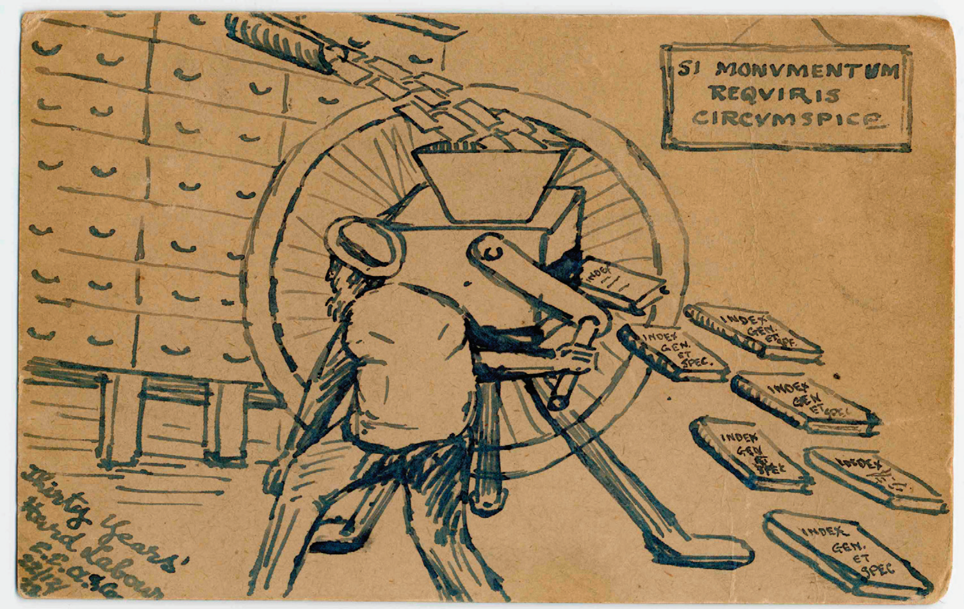


Figure 6. This splendid cartoon is of CDS operating a machine which turns his index cards into books. Hanging on the wall is the famous Latin inscription from Sir Christopher Wren's tomb in St Paul's Cathedral (which of course he designed): '*Si monumentum requiris circumspice*' - If you seek his monument, look around you. In the bottom left corner is 'Thirty Years Hard Labour, E.E.A.'. The cartoon is a postcard on which is written on the other side, 'Best of Good wishes for Christmas and the New Year, from all at Evaeria. 22.xii.23'. This is the name of the house of Major Ernest Edward Austen, DSO, FZS(1867–1938), assistant keeper of the NHM's Entomology Department and an authority on tsetse flies. (With permission of The Trustees of the Natural History Museum, London)

book is gone through from cover to cover – every genus and species, and every change of genus, being systematically recorded; thus completely disposing of that particular book, and ensuring the almost absolute certainty of every reference being taken'. (Sherborn 1896).

The statistics of this work are staggering. The 11 volumes, totalling more than 9000 pages, contain about 440,000 names, extracted by Sherborn from thousands of academic books and journals, in many different languages, each naming newly discovered species.

Each name, with the book or journal in which it was first published and the date, he recorded in black lead pencil on a small slip of paper, 127mm by 63mm (5 inches by $2^{1/2}$ inches) and then duplicated. Carbon-paper – blue or green – was soon used to make the duplicates, 'both methods having proved to be quite indelible', he wrote (Sherborn 1896). By 1916 there were more than one million slips. It took him one month to edit 10,000 and one hour to number 1500 of them.

From the outset, the *Index Animalium* project was hampered by lack of funds. Although the NHM gave him his own space in the museum's library, he was reliant on grants from scientific institutions. These were barely of subsistence level. In 1892,

Figure 7. This is just one of the drawers of Sherborn's index cards. It is now preserved in the NHM's Library. Sherborn recorded in pencil each name, with the book or journal in which it was first published and the date, on a small slip of paper, 127mm by 63mm, and then duplicated it. It took him one month to edit 10,000 and one hour to number 1,500 of them. By 1916 there were more than one million slips. It took 3 years to put them into rough alphabetical order. These cards were for the years 1850–1899, and were never published. The volume of material for those years was so great it would have demanded a team of workers to complete the task, not solely CDS, and that was simply unaffordable. (With permission of The Trustees of the Natural History Museum, London)

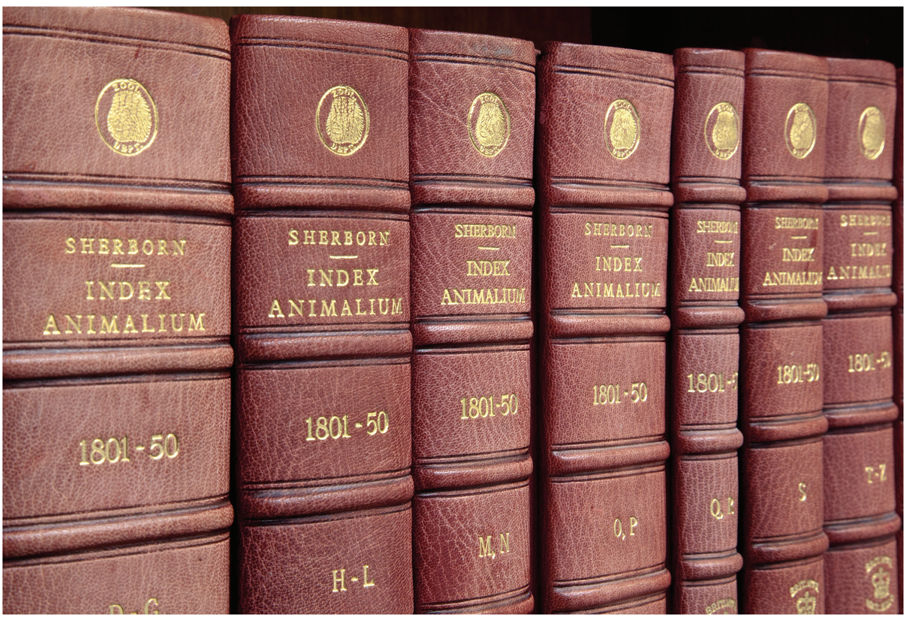


Figure 8. Some of the 11 volumes of Sherborn's *Index Animalium,* held in the NHM's Library. They total more than 9000 pages, and contain about 440,000 names. (With permission of The Trustees of the Natural History Museum, London)

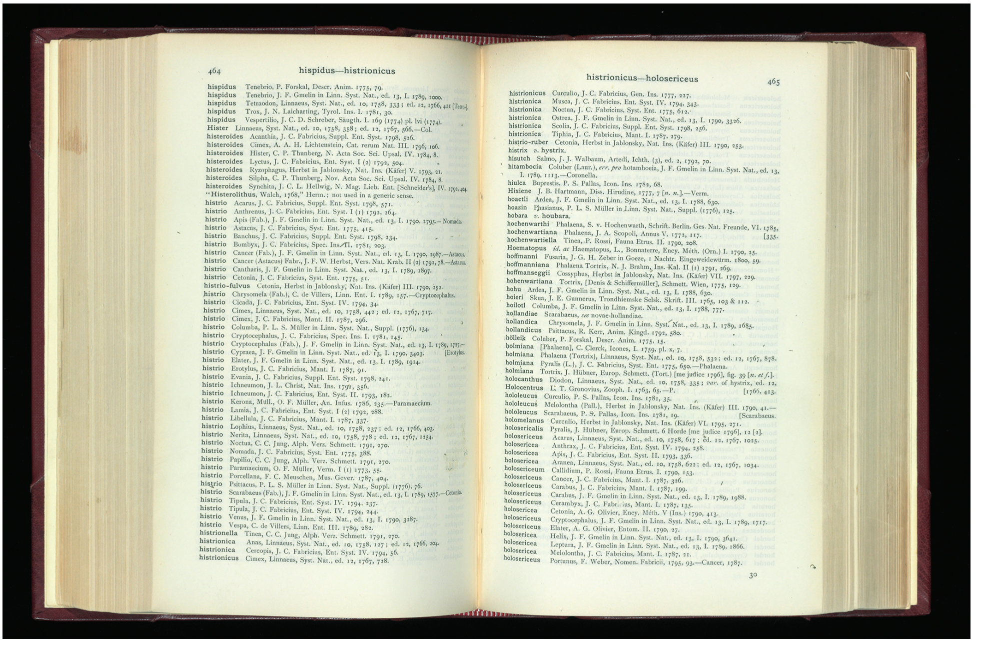


Figure 9. Pages from the first part of *Index Animalium*, 1758–1800. (With permission of The Trustees of the Natural History Museum, London)

the British Association for the Advancement of Science which to start with provided most of the funds, set up a committee to protect Sherborn from the added burden of administering the money. Year after year the committee wrote to various bodies pleading for funds so Sherborn could employ 'even a boy to do the sorting, alphabetical arrangement and numbering of slips', but the extra money never came (BAAS 1896-1912). Everything had to be done by Sherborn alone. It took three years just to put the million or so slips into rough alphabetical order.

In all those 43 years of labour, Sherborn received a total of just £5415 – or an average of £126 a year – in grants. It was not until 1912 that the Natural History Museum finally assumed responsibility for Sherborn – something for which his friends had been angling for some time. In 1909 the Keeper of Zoology, Sidney Harmer, raised the question with colleagues, but was discouraged by the museum's Secretary, Charles Edward Fagan, who thought the museum's finances were such that 'the moment is not a propitious one' (Harmer 1909). It was to be another three years before the Trustees awarded him an annual grant. In 1912 it was £100 a year. By 1931 that had risen to £250, where it remained until his death.

The first volume of the *Index* was published in 1902 – it took about 20 months to print the 1200 pages – and covers the years 1758–1800. Although twelve years had elapsed since he started work, the first volume was the easy bit, though he did not know it at the time, as he wrote in December 1902 to a friend, Ernest Hartert – a curator at Walther Rothschild's museum at Tring: 'I am so glad it is out, it has cost

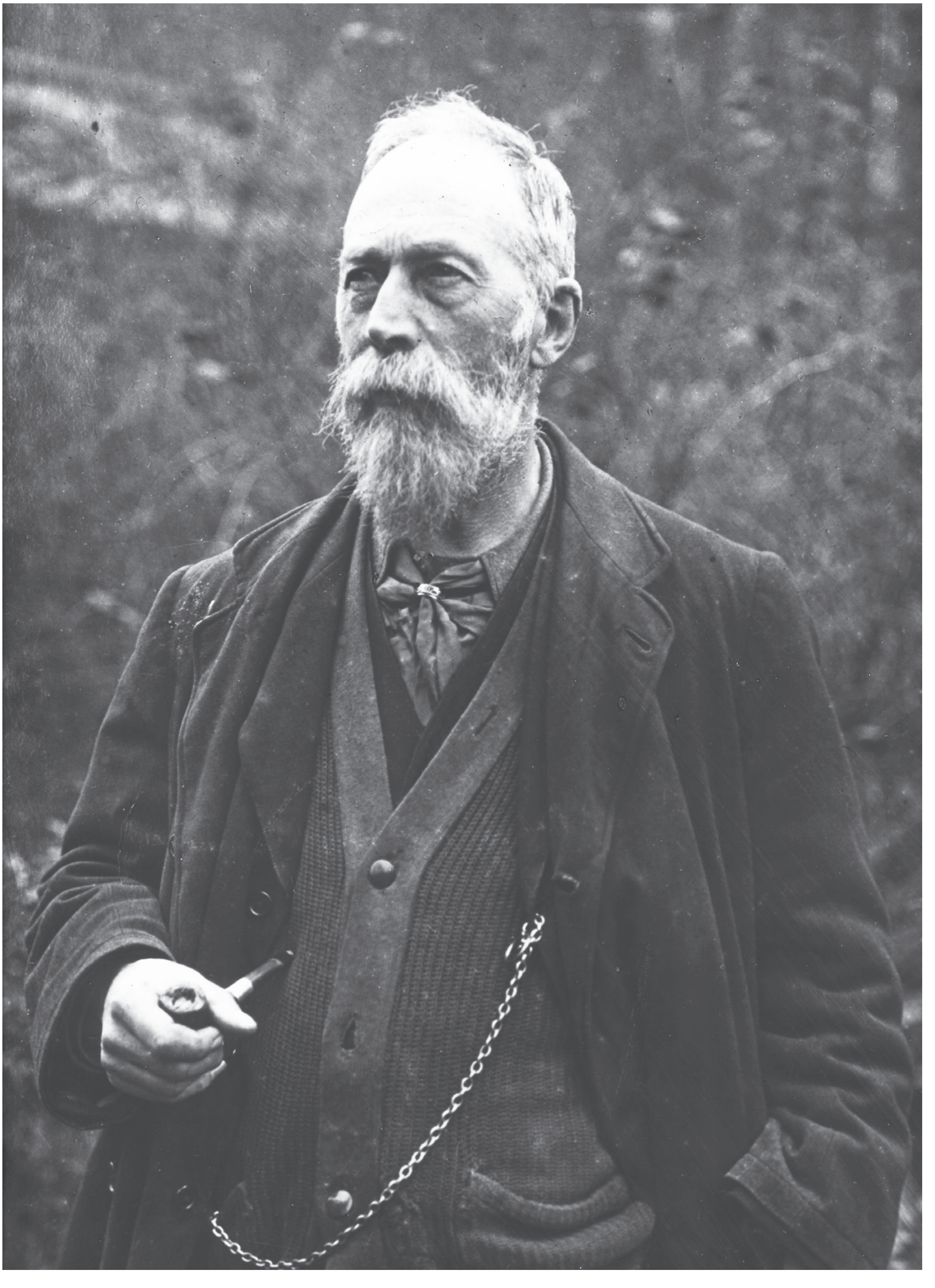


Figure 10. Sherborn at the age of 61. He was frugal and cared little for his appearance. He felt the cold keenly, and would keep warm by inserting a sort of apron of felt beneath his many layers of clothes. In extremis, he would also use a newspaper. Instead of a necktie, he wore a piece of folded red or black material, held together with an old gold ring. The black material, his biographer JR Norman discovered, was cut from an old umbrella. (With permission of The Trustees of the Natural History Museum, London)

me much work & ill health, but I feel like a giant refreshed now I have something to SHOW'. (Sherborn 1902b).

So severe was his illness, however, that he refers to it in the introduction to the first volume, recording that, 'an unfortunate breakdown in health, which has frequently recurred, laid me aside for three years, and thus the actual time spent on the manuscript has amounted to eight years" (Sherborn 1902). He was left with raging headaches and eye problems.

His ill health, however, was only partly due to his labours on the *Index*. Just two years after he began work on it, he received an extraordinary invitation. He was asked to collaborate on writing the biography of the great comparative anatomist, creator of the Natural History Museum and its first Superintendent, Sir Richard Owen. The invitation was from Owen's grandson, the Reverend Richard Startin Owen.

Sherborn had first met Sir Richard Owen in October 1889, when he was invited to dine at his home in Sheen Lodge, Richmond Park. Owen was then 85 and in poor health. As a memento of the occasion, Sherborn kept a photograph of Owen, inscribed on the back 'a memory of an interesting occasion dined with Richard Owen, Oct 27th 1889. C. Davies Sherborn.' (Sherborn 1887–1942).

If somewhere Sherborn wrote more than that, sadly it does not appear to have survived. Presumably at the dinner he would have met Owen's grandson, the Reverend R S Owen, and he may have been invited into Owen's over-flowing, book-lined study. He may even have been given a preliminary glimpse of the 60 years worth of manuscripts, correspondence and books that lay neglected and unsorted.

Sir Richard Owen's career had begun in 1827 when he became assistant conservator at the Hunterian Museum at the Royal College of Surgeons. After his retirement

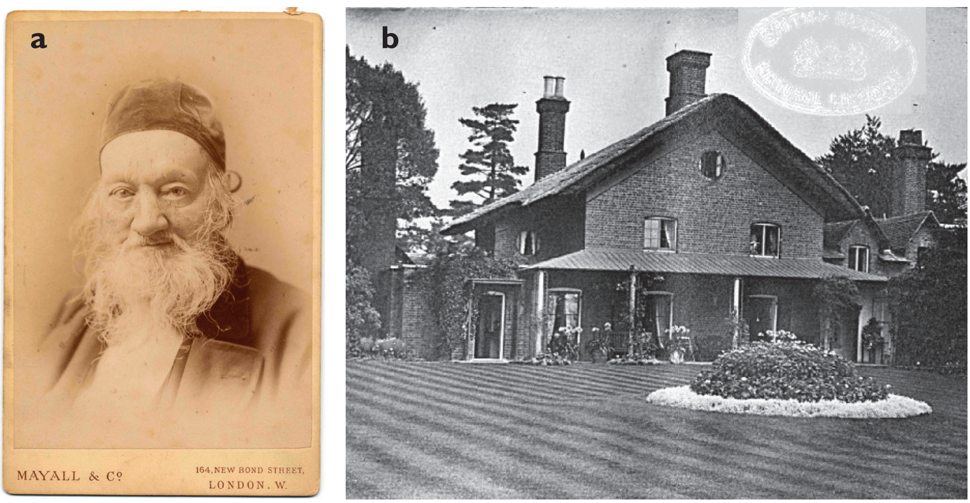


Figure 11. a Sir Richard Owen gave this photograph to Sherborn. CDS has written on the reverse: 'a memory of an interesting occasion dined with Richard Owen, Oct 27th 1889 C. Davies Sherborn'. The dinner took place at **b** Sheen Lodge, Sir Richard's grace-and-favour home in Richmond Park. (With permission of The Trustees of the Natural History Museum, London)

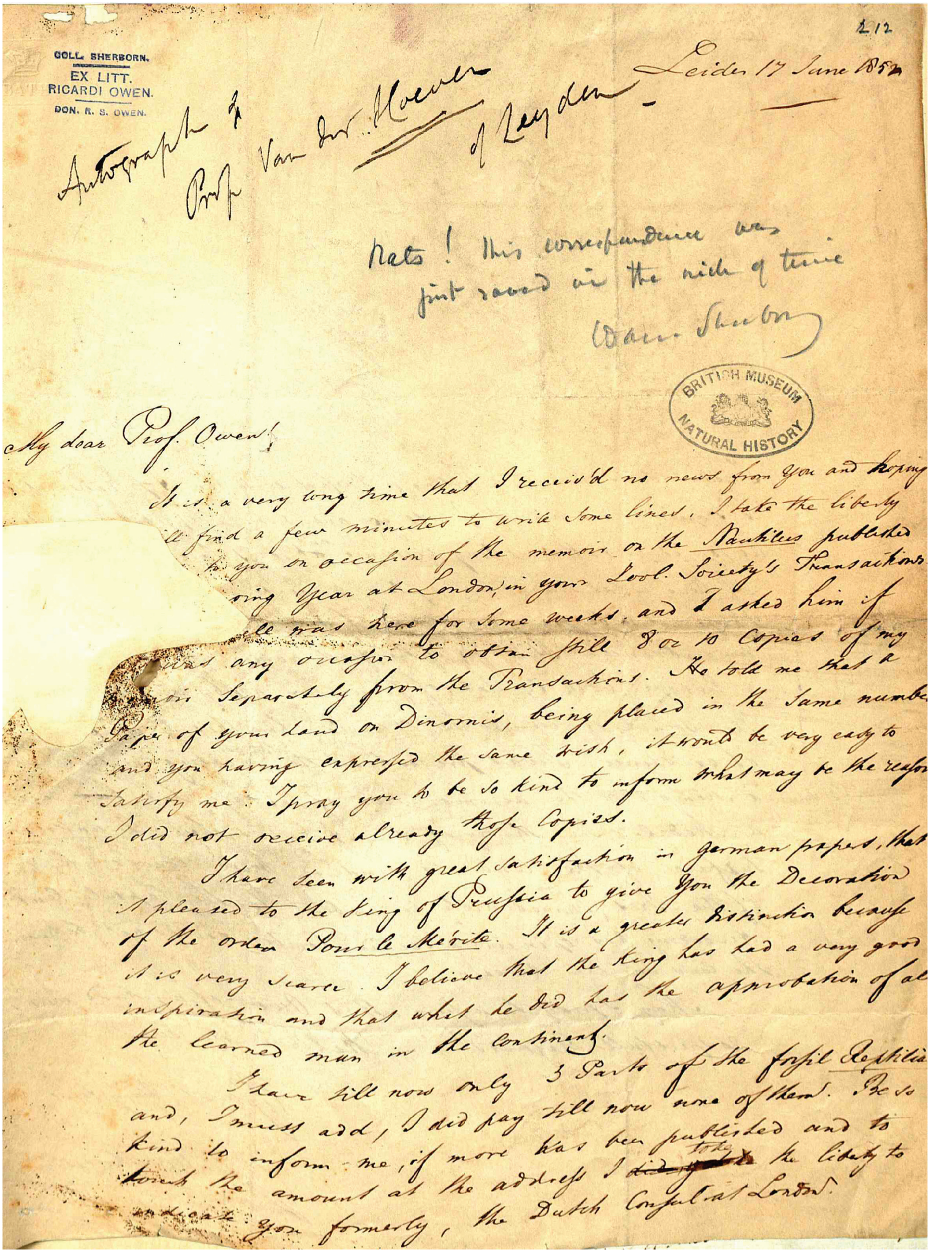
56 years later, as Superintendent of the new Natural History Museum, the contents of his office – including two cartloads of papers – followed him home to his grace-and-favour residence, Sheen Lodge. This mass of personal and professional correspondence, manuscripts and books led Owen to exclaim, 'I am compelled to part with a Gardener and to turn his cottage into their receiving house'.

What was to become of it all became an increasingly pressing problem as Owen's health declined. His beloved wife Caroline was dead, his only son William had apparently committed suicide in 1886, and none of his seven grandchildren, who lived with him at Sheen Lodge, appeared to have any interest in science, nor had they any idea of what in this huge mass of paper might be of importance or indeed, value. Unlike Charles Darwin's friends and family who brilliantly preserved and archived his papers, understanding full well Darwin's importance, there was no one with the inclination or knowledge to do this for Owen.

It was at some point after the dinner in 1889, that Sherborn was asked by Reverend RS (Richard Startin) Owen to sort through his grandfather's books and organise their sale. As Sherborn commented in his own autobiographical notes, Owen – or the family – had already sold a good number, 'so there remained only some ten years accumulation'. Sherborn also spent many hours with Owen (Norman 1944). Since his son William's death in 1886, Owen had become increasingly withdrawn. His beard grew, his hair, as photographs show, was straggly and unkempt. His 'great glittering eyes' that Thomas Carlyle had remarked upon, were huge in his gaunt face. His appearance frightened his younger grandchildren, but to Sherborn he reminisced and gossiped about 'all the great men of his youth'.

It was possibly not that surprising then, that in August 1892 – when Owen was so ill he could not speak or swallow – Sherborn received the invitation which, to say the least, clearly thrilled him. On 20 August 1892, he wrote to the librarian of the Natural History Museum, Bernard Barham Woodward, 'Some men are born great... and some have greatness thrust upon them... and I therefore entrust to you as a special friend the fact that I was summoned on Tuesday morning to East Sheen and asked to collaborate with the Rev R O on the Life and Letters of Richard Owen'. You can almost hear the glee in his voice. He urged Woodward to be discreet with the news. 'All is yet a secret, as by no means must it get to the Press until matters are settled.' There is no correspondence to show how well Sherborn had come to know the Owens, but he added, 'I am very proud that I should have been chosen by the family after all these years; and not only chosen but thanked again and again for all my kindness when first I went down there'. Spending hours talking to the frail professor was presumably what they had in mind, but as Sherborn noted, his conversations with Owen would be of 'great value' for the biography (Norman 1944).

When he wrote to BB Woodward, Sherborn was already hard at work sorting Owen's papers, and had nearly disentangled the correspondence from the manuscripts. He had, so far, found around 10,000 letters. Given the scientific and historical importance of the material, the condition in which the papers were kept is staggering. Sherborn found the papers in urgent need of preserving and sorting. They were 'in a cow-shed, exposed to rats and rain',



COLL. SHERBORN.
EX LITT.
RICARDI OWEN.
DON. R. S. OWEN.

212

Autograph of Prof Van der Hoeven of Leyden – Leiden 17 June 185[illegible]

Rats! this correspondence was just saved in the nick of time
Owen Sherborn

BRITISH MUSEUM NATURAL HISTORY

My dear Prof. Owen!

It is a very long time that I receiv'd no news from you and hoping
[illegible] find a few minutes to write some lines, I take the liberty
[illegible] to you on occasion of the memoir on the Nautilus published
[illegible] Year at London, in your Zool. Society's Transactions
[illegible] was here for some weeks, and I asked him if
[illegible] any occasion to obtain still 8 or 10 Copies of my
[illegible] separately from the Transactions. He told me that a
paper of yours on Dinornis, being placed in the same numbe[r]
and you having expressed the same wish, it would be very easy to
satisfy me. I pray you to be so kind to inform what may be the reason
I did not receive already those Copies.

I have seen with great satisfaction in german papers, that
it pleased to the King of Prussia to give you the Decoration
of the order Pour le Mérite. It is a greater distinction because
it is very scarce. I believe that the King has had a very good
inspiration and that what he did has the approbation of al[l]
the learned men in the Continent.

I have till now only 3 Parts of the fossil Reptilia
and, I must add, I did pay till now none of them. Be so
kind to inform me, if more has been published and to
touch the amount at the address I take the liberty to
indicate you formerly, the Dutch Consul at London.

Figure 12. One of the many letters from the Owen correspondence that Sherborn found in urgent need of preserving and sorting. On it he has written 'Rats! This correspondence was just saved in the nick of time'. (With permission of The Trustees of the Natural History Museum, London)

and this was no exaggeration (Owen Collection, OC62 General Library, NHM). The manuscripts were piled four metres high, while the correspondence filled many packing cases. But for Sherborn, who described himself as having a card-index mind, this was very heaven.

He told BB Woodward that so important was the Owen biography that, 'I must husband all my time and strength now, for it is a giant's task set before me, and this must cap, not sink below, my other works'.

Just two weeks after Sherborn's ecstatic letter to BB Woodward, on 3 September 1892 the Reverend Owen wrote to Sherborn with news that must have wounded him deeply: without further discussion he was withdrawing the invitation to collaborate on the biography. What appears to have changed Richard S Owen's mind was his consultation with some of his grand-father's old friends, including Sir Richard Owen's successor as Director of the Museum, Sir William Flower. They had concluded that as the biography was not to be a narrative of Owen's scientific life, but rather a record of his private life, 'the narrative will be drawn principally from the joint diaries kept by Sir R and his wife. I find that these are of so purely a family & private nature as to compel me to do all the work of extracting & compiling myself', the Reverend Owen told Sherborn. Sherborn was still to have a role, however. The Reverend Owen told him that as far as sorting the correspondence was concerned, there was no man 'better fitted for it than yourself', and he also wanted Sherborn to 'undertake to revise any mention of Sir R's scientific work so as to preclude the danger of error'. He would be paid and his work would be acknowledged in the preface (Owen RS 1892).

Sherborn agreed. As the scientific papers were not to be used in the biography, he dispersed many of them. The medical papers went to the Royal College of Surgeons and others went to the Geological Society, which had published a number of Owen's reports. John Marr, society secretary, wrote to Sherborn to thank him for the papers, 'The Librarian has made a selection of those which we do not possess, & the others are being returned to you, in accordance with your request'.

In 1894, the biography was published. The Reverend Owen in the preface thanked Sherborn 'for lending me much assistance throughout', and 'for carefully examining Sir Richard's correspondence' (Owen RS 1894). He then, Sherborn wrote, 'gave me the lot, and it filled a four-wheeler to take home. The manuscripts', Sherborn added devastatingly, 'were distributed to those interested all over the world.' As Sherborn did not compile a list of these, it is impossible to know now what was irretrievably lost. Three years later however, Sherborn discovered that R S Owen had not in fact given him 'the lot.' He had removed a large part of the collection and only gave them to Sherborn in 1897 as he was about to go to New Zealand for six months. 'Can you come round tomorrow morning,' he wrote to Sherborn on 6 October, 'and cart off the letters? ... Would you mind giving them houseroom while I am away, & also the prints of Hunter which I fear are being treated in rather a reckless fashion.' (Norman 1944)

The accommodation that Sherborn was able to provide for them at such short notice was scarcely less reckless. 'The cupboard full of letters & papers in my back room, the manuscript of Hunter in the safe, the papers in the cupboard in Smith Woodward's room at the Museum & the Diplomas of Sir R Owen there are held in trust for the Rev Richard Owen & are to be kept until he asks for them. The scientific letters he has promised to me & I intend them for the nation to be preserved at the Natural History Museum under BB Woodward's care as the Scientific Correspondence of Richard Owen.'

But even that still did not account for the whole archive. Sherborn himself, late in his life, noted briefly that, ‘about 100 letters kept back by his grandson, who sold them at Maggs [the antiquarian booksellers] in 1916. His own letters to his wife and sister[s] kept by his grandson.’ (Norman 1944) Sequences of correspondence were separated. I have read a letter in the British Library, only to find its reply in the Natural History Museum and the reply to that in the Royal College of Surgeons.

Letters from the famous were of course sold. Charles Kingsley, Alfred, Lord Tennyson, Charles Dickens – all were among Owen’s friends, but just about all their correspondence has vanished. Also missing are letters that refer to Owen’s disagreements with Charles Darwin and Thomas Huxley. It is extraordinary to believe in all that correspondence that it was scarcely referred to, yet there is hardly anything. Compare that with the Darwin correspondence where the Darwinites’ antipathy to Owen flows remorselessly in letter after letter. Apart from passing references, virtually all there is in the Owen correspondence is one letter in which Owen criticises Huxley’s ‘base and mendacious nature’ and on this Sherborn has written, ‘This is the only letter I remember in which Owen severely criticized an opponent’.

The scale of Sherborn’s task in sorting through this material was huge. Pages had become separated and had to be re-united, authors – and signatures – identified. He annotated many of the letters. On one from the Director of the National Portrait Gallery, George Scharf in 1889, to Owen’s daughter-in-law expressing his great satisfaction that he had ‘secured’ a portrait of ‘my dear old friend’ Sir Richard, Sherborn has scrawled scathingly, ‘This is nonsense, he first refused it owing to some silly regulation and was told that if they didn’t want it, it should be offered to the RCS. Then he jumped at it.’ (Scharf 1889) Before the first of 64 letters written to Owen by the Marchioness of Hastings, Sherborn has noted her brief biographical details. On a letter that Owen wrote to his wife Caroline concerning their grace and favour house, Sherborn has written, ‘Regarding house at Kew’ and then, worryingly, ‘keep’.

And this of course was one of Sherborn’s main difficulties. Confronted with so many thousands of letters, Sherborn edited rigorously, deciding what was worth keeping and, regrettably for subsequent researchers, what he deemed was not. When he presented the material to the Museum in 1908, Sherborn outlined to the librarian BB Woodward the parameters he had used for his selection. I have to admit to finding this horribly painful to read: ‘I have been carefully through the collection with a special knowledge of the history of science and of the collections of the British Museum (Natural History) and have destroyed several thousand letters of no value. Richard Owen kept everything and the great bulk of those destroyed were letters from tradesmen and similar unimportant persons from our point of view.’ Sherborn was of course a scientist and bibliographer, not an historian. He continued, ‘This collection is of infinite value to the British Museum, for hundreds of them refer to specimens actually in the various departments of Geology and Zoology … and will be a mine of information on general and bibliographic questions.’ (Norman 1944) In this of course, Sherborn is absolutely right. Owen’s papers are invaluable, in identifying and giving context to specimens in the collections.

In rescuing Owen's papers, Sherborn is owed a huge debt of gratitude. It was an extraordinary achievement, on top of all his other work, to bring order to this chaos and turn the mountain of paper into the invaluable collection it is today. Of course there are regrets about some of his decisions in editing and selling, but thanks to him, an immensely important archive still survives. He too was its beneficiary. Owen's manuscripts, Sherborn wrote, 'have been of the greatest service to me during the years I have held them in answering queries as to the date of publications, the movements of men, and other matters in connection with my *Index Animalium*.'

He needed all the help he could get. Such was the explosion in scientific literature in the first half of the 19th century, that while it had taken him 12 years to complete the first volume, 1758–1800, it was to take Sherborn another 21 years, until 1933, and 10 volumes to complete the Index just to 1850. That was still fifty years short of his original goal of completing the century. When Sherborn was asked if he would add 1850–1900, he replied that the museum could not expect to find another man 'with his knowledge and capacity content to work for an honorarium'. Furthermore, as the scientific literature since 1850 was so enormous, it would require a team of workers whose combined salaries would be something like £2000 a year – and that was simply unaffordable. He actually compiled a 'Time Sheet', which, he wrote, would 'be wanted should the Trustees continue the work after 1850. CDS'. It is neatly laid out and states that: 'It takes 1 month to edit 10,000 slips (c.3 feet?)... 1 month to look into queries

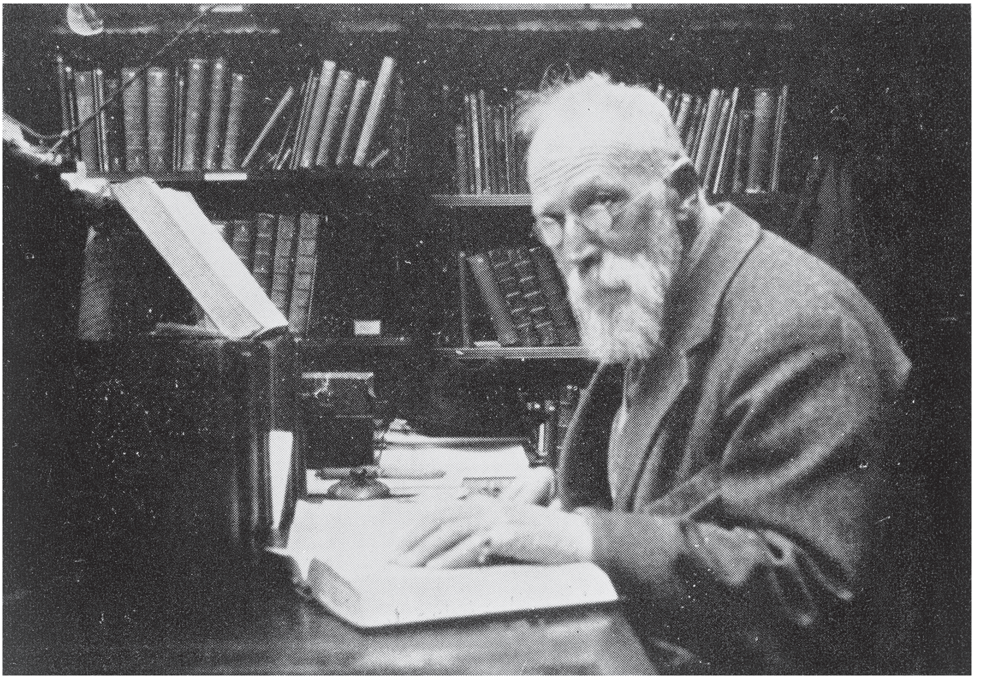


Figure 13. Sherborn at his desk in the Library of the NHM was the first port of call for generations of scientists seeking advice, information – or an invitation to one of his famous 'smoke and chat' parties. (With permission of The Trustees of the Natural History Museum, London)

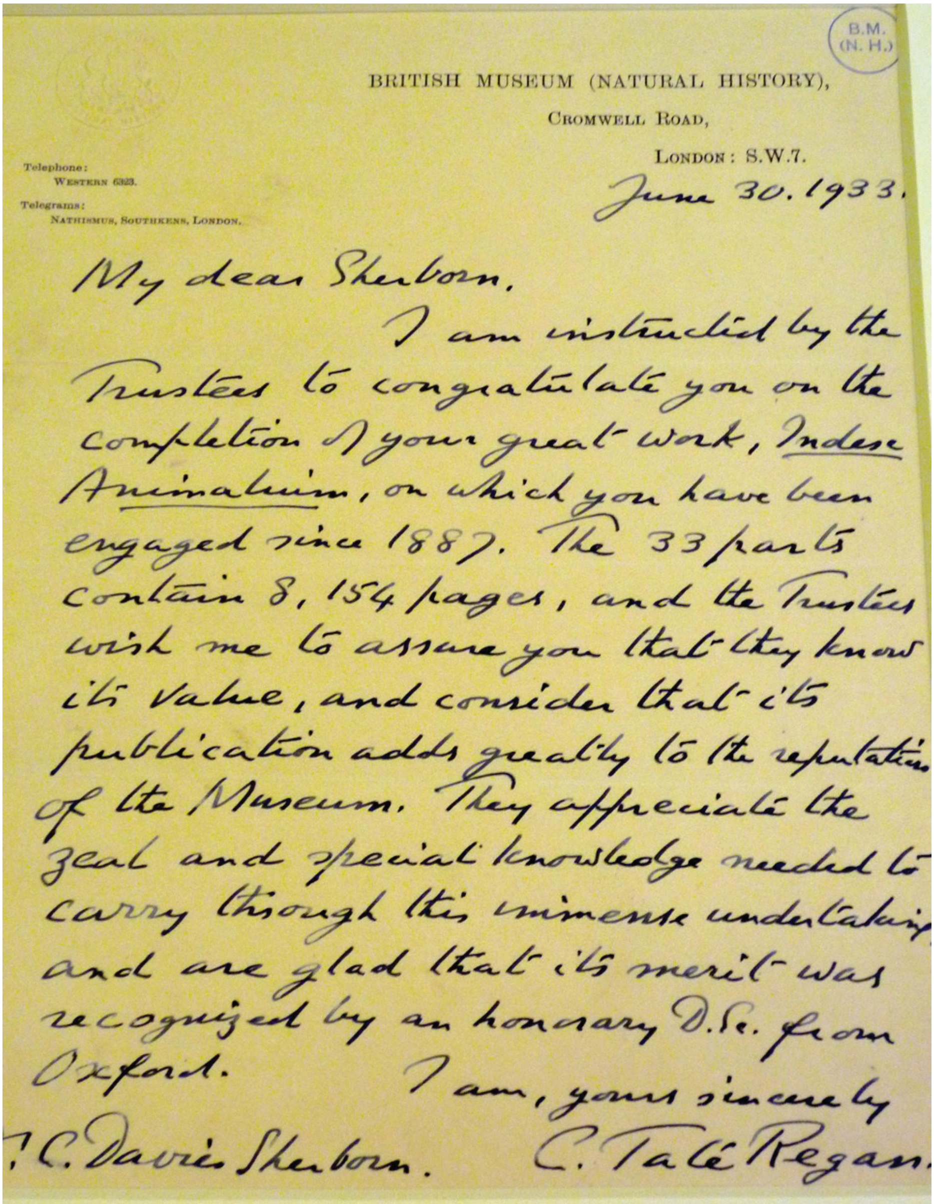
of same...1 hour to number 1500 slips...18 years to amass the ms. 1801-1850...5 years to print A-L...' and so on. (Sherborn 1924)

The *Index* was received with the gratitude and admiration Sherborn so richly deserved. The compliments came from all over the world, but this, from Dr Bashford Dean of the American Museum of Natural History in 1923, sums up the general response. In a letter to the Director, he wrote: 'I am very glad indeed to have this monumental work in my library, and I congratulate the South Kensington Museum [sic] with all my heart at completing the next stage of a magnum opus…such a work as Sherborn's is a labor of love of the greatest magnitude, and I feel sure that the 'index' field covering the whole of the animal kingdom will be a boon to zoologists the world over for all time'. [Dean 1923]

Apart from 5000 entries made 'by various friends abroad', every entry, Sherborn wrote, 'has been recorded from the original, arranged, sorted, checked, and passed for press by myself'. This was a true labour of love – for shamefully little financial reward. Sherborn compiled it single-handedly, and it took him 43 years, from 1890 to 1933, publishing each section as it was completed. His error rate was surprising low (Welter-Schultes et al. 2016) and the work formed the foundation for other similar projects with more focused taxonomic reach (e.g. Dickinson 2016). His reward was an honorary doctorate from Oxford, which gave him enormous pleasure – though to the end of his life he regretted that his mother had died before it was awarded to him. He also received the congratulations of the Trustees of the Natural History Museum, for which he had laboured for most of his life, though he never became a member of staff or received a regular salary or pension.

In addition to his work on the *Index*, he was president or fellow of various learned societies, wrote nearly 200 books and papers, including important contributions on microfossils (Miller 2016) and catalogues of natural history collections (Taylor 2016). He also catalogued the collections and library of the Geological Society of London. His interests were eclectic and wide-ranging and he was an avid collector of books, pictures and all kinds of antiquities. In a small notebook he would glue snippets of interesting stories or facts cut from newspapers, and jot down notes on pretty much anything that caught his eye. These ranged from the history of taxation, cures for the common cold, the derivation of symbols, and the Admiralty rules for the proportions of the Union Jack, to the best way to clean white marble – common washing soda and soap powder, or so he wrote. He was an enthusiastic theatre-goer, often seeing two or three productions a week. Almost inevitably, he collected theatre memorabilia, filling large volumes with reviews, posters, programmes and photographs of the theatrical stars of the age (Sherborn 1887–1942).

How he managed with so little income is extraordinary, but he was frugal in the extreme and cared little for his appearance. He felt the cold keenly, and would keep warm by inserting a sort of apron of felt beneath his clothes. In extremis, he would also use a newspaper. Instead of a necktie he wore a folded bit of red or black material drawn through an old gold ring. After his death, his friend and biographer John Roxborough Norman discovered that the black fabric had been cut from an old umbrella (Norman 1944).



B.M. (N. H.)

BRITISH MUSEUM (NATURAL HISTORY),
CROMWELL ROAD,
LONDON: S.W.7.

Telephone:
WESTERN 6323.
Telegrams:
NATHISMUS, SOUTHKENS, LONDON.

June 30. 1933.

My dear Sherborn,

I am instructed by the Trustees to congratulate you on the completion of your great work, Index Animalium, on which you have been engaged since 1887. The 33 parts contain 8,154 pages, and the Trustees wish me to assure you that they know its value, and consider that its publication adds greatly to the reputation of the Museum. They appreciate the zeal and special knowledge needed to carry through this immense undertaking, and are glad that its merit was recognized by an honorary D.Sc. from Oxford.

I am, yours sincerely

C. Tate Regan.

C. Davies Sherborn.

Figure 14. Letter from the Museum's Director, Charles Tate Regan, on behalf of the Trustees, congratulating Sherborn on the completion of his 'great work' in 1933. Although he laboured for most of his life on behalf of the Museum, Sherborn never became a member of staff or received a regular salary or pension. (With permission of The Trustees of the Natural History Museum, London)

By the time of his death in 1942, Sherborn had become a vital pillar of the Natural History Museum. His corner in the museum library was the first port of call for generations of scientists – anyone who wanted to know anything went to Sherborn for information, advice, or to consult the ever-expanding boxes of *Index Animalium* slips, which from the outset provided an invaluable resource to the museum's staff. (Harmer 1921). He was known by colleagues as 'Squire': JR Norman records that his family had links to the old manorial title of Squire of the Fawns and Cock Bell at Bedfont. To his junior colleagues he was Sherb, while to family and his closest friends he was Sherby or Charlie.

In the course of his work, Sherborn came across many volumes he thought should be in the Natural History Museum's libraries. If he failed to persuade the museum to buy them, he bought them himself – and then either sold them to the museum for what he had paid for them, or simply donated them. One work, he has noted, was 'Bought by CDS after refusal by Zool. Dept & presented 1930.' Between 1891 and 1939 he acquired some 1600 volumes for the museum.

Sherborn's generosity was not just to the museum's libraries. His friends were beneficiaries of his great kindness and his collections. 'No man was ever more generous', recalled Francis Griffin of the Society for the Bibliography of Natural History (Sherborn was its first president), 'in handing over treasures to his friends'. Griffin himself had received an armoured breastplate from Sherborn with the injunction to wear it under his jacket - 'Much the easiest way of carrying it you know!' (Griffin 1953)

As a young man new to the museum, Sherborn had been hugely appreciative of the 'at homes' given by Henry Woodward, Keeper of Geology. He carried that on with his own institution – his famous 'smoke and chat' parties. Sherborn believed strongly in the importance of bringing people together, and that it was the responsibility of senior colleagues to entertain their juniors and visitors in their homes. Today he would have been a sought-after networker. 'A cup of coffee and a few biscuits all round', was what he offered, together with 'a good mix of chaps'. He invited not just scientific staff, but all grades in the museum, as well as foreign visitors and non-scientific friends.

These were informal, all male events. His biographer and friend, JR Norman quotes him as saying, 'There's no need to make it a social affair, with a lot of women and boiled shirts', and with that attitude, it is perhaps not surprising that Sherborn acquired a reputation as a misogynist. Norman, however believed it may well have been a pose, but it was one Sherborn pursued with some vigour, even writing, 'I never liked women and never chose them as companions'. He also deplored the 'amazing standard of ignorance in the so-called educated woman', complained they did not read the great masterpieces and 'that is why the bulk of women are so extraordinarily uninteresting'.

Disparaging and sexist as these remarks are, the reality of his relationships with women seems very different. I knew from my research into the life of the palaeontologist Dorothea Bate who was associated with the museum for more than fifty years, how generous and kind he had been to her, and so he seemed to be generally with his female acquaintances.

Indeed, women were among his closest friends and correspondents and, as he revealed to Norman, as a young man he had actually once contemplated marriage. 'I once esteemed a woman very highly and engaged to marry her,' he told Norman, 'she was highly intelligent and well read'. However, his career came first, he realised he had no prospect of keeping her in comfort, and after ten years they parted, 'and she wisely married another'. Whether his subsequent attitude was to disguise a profound hurt is

Figure 15. Sherborn at the feet of three unknown friends. (With permission of The Trustees of the Natural History Museum, London)

impossible now to know, but Norman notes his views on marriage were largely cynical. When Norman himself married, Sherborn sent him this:

> 'Dear JR,
> Cheerio!
> You have my entire sympathy.
>
> Yours ever,
>
> C. Davies Sherborn
>
> Still raining!' (Norman 1944)

One of his greatest friends and confidantes was Agnes Arber, to whom he wrote constantly. She also happened to be a botanist, philosopher and Fellow of the Royal Society. Sherborn professed to be uninterested in children, yet his papers reveal how witty and charming he was to the offspring of his numerous friends. To Mrs Arber's daughter, Muriel, who had been unwell, he sent a sketch of himself in 1922 when she was nearly nine, and wrote:

'Dear Muriel,

This is me. All you would see of me if you found me in the field. I am having a long rest in a few days.

I hope you are quite well again. It is no use being ill, it is such a waste of time.

Your friend,

C. Davies Sherborn

The sketch shows Sherborn, invisible apart from a large hat and puffs of smoke from his ubiquitous pipe, lying in grass under a tree with rabbits hopping happily around. (Sherborn 1922). Muriel, incidentally, was to become a respected geologist and teacher. (Robinson 2007).

The last years of Sherborn's life were beset by ill-health, although he still went daily to the museum. A tremor in his hands made writing agonizingly difficult and he suffered

Figure 16. 'Talking to the Brer Rabbits'. Sherborn sent this drawing of himself in 1922 to the 'nearly nine' -year-old Muriel Arber who had been unwell. She was the daughter of his great friend, the botanist, philosopher and Fellow of the Royal Society, Mrs Agnes Arber. All that can be seen of Sherborn is a large hat and the smoke from his ubiquitous pipe. 'I hope you are quite well again,' he wrote to Muriel, 'It is no use being ill, it is such a waste of time. Your friend, C. Davies Sherborn'. (With permission of The Trustees of the Natural History Museum, London)

increasingly from severe colds, though that did not persuade him to give up his beloved pipe. By 1940, as he wrote to a friend, the effect of the blitz meant he had 'nothing to do at the Museum now, so stay put and read and smoke'. His visits to the Museum were reduced to Fridays only, and his housekeeper told Norman how lonely he was. In December 1941 he wrote, 'Shall be alone Xmas Day, five sardines for dinner and a good pipe after'. His last visit to the museum was on Friday, 19 June 1942, when he met his friends afterwards at Lyons teashop in South Kensington as they had done for so many years. The following Monday as he was running a bath, he had a heart attack. He rallied a little when the doctor came, but he died that afternoon. Sherborn was cremated at Golders Green crematorium in north London the following Friday, at the exact time, Norman noted, that he would have been meeting his friends at Lyons teashop.

Three months later, a letter arrived for him from an old friend, the Reverend CR Bower. Unaware that Sherborn had died, he wrote that he had been thinking a lot about him and had intended writing, but had been too busy. He asked his old friend to drop him a line: 'we must not after all these years lose touch. Your friendship is one of my most treasured possessions. My wife,' he ended, 'sends her love'. (Bower 1942)

Charles Davies Sherborn's contribution to zoology, bibliography and the collections of the Natural History Museum is unique. His rescue of Sir Richard Owen's archive alone was of outstanding importance, but combined with his great memorial, the *Index Animalium*, this dedicated, idiosyncratic, self-taught man with an encyclopedic brain deserves to be celebrated and acclaimed with gratitude. Dr Dean of the AMNH could not have put it better in his letter to the NHM's Director, Dr Sidney Harmer: 'I think Sherborn is a marvel, and if he were not so devoted and patriotic, I tell you frankly that I would have stolen him bodily long since'.

References

BAAS (1896–1912) Sherborn Minute Book of the British Association Committee for the compilation of an Index Animalium, General Library, Natural History Museum, MSS Brit.

Bower CR (1942) Letter from CR Bower to Sherborn 16 Sept 1942, General Library, Natural History Museum, Sherborn Collection, MSS SHE B, Box 2.

Cleevely RJ (2004) Sherborn, Charles Davies. Oxford Dictionary of National Biography, OUP.

Dean B (1923) Letter to SF Harmer 15 Oct 1923, Archives, Natural History Museum, London, DF 1004/cp 778 Publications Index Animalium.

Dickinson EC (2016) Reinforcing the foundations of ornithological nomenclature: Filling the gaps in Sherborn's and Richmond's historical legacy of bibliographic exploration. In: Michel E (Ed.) Anchoring Biodiversity Information: From Sherborn to the 21st century and beyond. ZooKeys 550: 107–134. doi: 10.3897/zookeys.550.10170

Evenhuis NL (2016) Charles Davies Sherborn and the "Indexer's Club". In: Michel E (Ed.) Anchoring Biodiversity Information: From Sherborn to the 21st century and beyond. ZooKeys 550: 13–32. doi: 10.3897/zookeys.550.9697

Griffin FJ (1953) 'Charles Davies Sherborn (1861-1942): some personal recollections. Jnl of the Soc for the Bibliography of Natural History, 3, pt 1: 1–4. doi: 10.3366/jsbnh.1953.3.1.np

Harmer SF (1909) Letter to Arthur Smith Woodward 6 October, Archives NHM DF 106/32.

Harmer SF (1921) Preface. In: Sherborn CD (Ed.) *Index Animalium* 1801-1850, A-B.

Miller CG (2016) Sherborn's foraminiferal studies and their influence on the collections at the Natural History Museum, London. In: Michel E (Ed.) Anchoring Biodiversity Information: From Sherborn to the 21st century and beyond. ZooKeys 550: 71–81. doi: 10.3897/zookeys.550.9863

Norman JR (1944) Squire, Memories of Charles Davies Sherborn, London.

Owen RS (1892) Letter to Charles Davies Sherborn 3 Sept 1892, Vol VII Science British Library add. MSS 42581 f247

Owen RS (1894) The Life of Richard Owen. Cambridge University Press, London, 409 pp.

Robinson E (2007) The influential Muriel Arber: a personal reflection. In: Burek CV, Higgs B. The role of women in the history of geology, Geological Society Special Publication 281: 287–293.

Scharf G (1889) Letter to Mrs Owen. OC62 General Library, NHM.

Sherborn CD (1890–1940) 'Slips used in the compilation of *Index Animalium*, General Library, Natural History Museum, MSS SHE A.

Sherborn CD (1887–1942) 'The Sherborn Collection: notebooks, address books, pamphlets, ephemera, photographs, geological maps and other items', General Library, Natural History Museum, MSS SHE B.

Sherborn CD (nd) 'Working Notes on Index Animalium' folder 8, The Sherborn Collection: notebooks etc, General Library, Natural History Museum, MSS SHE B, Box 2.

Sherborn CD (1888) 'An "instructive" bibliography of the foraminifera'. Nature 37: 583–584. doi: 10.1038/037583a0

Sherborn CD (1890) 'Index Generum et Specierum Animalium'. Nature 42: 54. doi: 10.1038/042054c0

Sherborn CD (1896) Explanation of the Plan adopted for preparing an *Index Generum et Specierum Animalium*. Proceedings of the Zoological Society of London 1896: 610–614.

Sherborn CD (1902) *Index Animalium*, 1, CUP.

Sherborn CD (1902b) Letter to E Hartert 15.12.02 NHM Archives, Tring Museum Correspondence and papers.

Sherborn CD (1922) Letter to Muriel Arber 1 July 1922, NHM General Library, Sherborn Collection MSS SHE B, Box 3.

Sherborn CD (1924) Note on the state of the ms of the Index 1924, NHM General Library, MSS SHE A.

Taylor MA (2016) '*Where is the damned collection?*' Charles Davies Sherborn's listing of named natural science collections and its successors. In: Michel E (Ed.) Anchoring Biodiversity Information: From Sherborn to the 21st century and beyond. ZooKeys 550: 83–106. doi: 10.3897/zookeys.550.10073

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. In: Michel E (Ed.) Anchoring Biodiversity Information: From Sherborn to the 21st century and beyond. ZooKeys 550: 173–183. doi: 10.3897/zookeys.550.10041

ZooKeys 550: 57–69 (2016)
doi: 10.3897/zookeys.550.7399
http://zookeys.pensoft.net

RESEARCH ARTICLE

A peer-reviewed open-access journal
ZooKeys
Launched to accelerate biodiversity research

Naming and Necessity: Sherborn's Context in the 19th Century

Gordon McOuat[1]

1 *University of King's College, Halifax, NS, CANADA*

Corresponding author: *Gordon McOuat* (gmcouat@dal.ca)

Academic editor: *E. Michel* | Received 3 December 2015 | Accepted 3 December 2015 | Published 7 January 2016

http://zoobank.org/697452CA-8410-42C8-ADB3-13E425BEEC55

Citation: McOuat G (2016) Naming and Necessity: Sherborn's Context in the 19th Century. In: Michel E (Ed.) Anchoring Biodiversity Information: From Sherborn to the 21st century and beyond. ZooKeys 550: 57–69. doi: 10.3897/zookeys.550.7399

Abstract

By the late 19th Century, storms plaguing early Victorian systematics and nomenclature seemed to have abated. Vociferous disputes over radical renaming, the world-shaking clash of all-encompassing procrustean systems, struggles over centres of authority, and the issues of language and meaning had now been settled by the institution of a stable imperial museum and its catalogues, a set of rules for the naming of zoological objects, and a new professional class of zoologists. Yet, for all that tranquillity, the disputes simmered below the surface, re-emerging as bitter struggles over synonyms, trinomials, the subspecies category, the looming issues of the philosophy of scientific language, and the aggressive new American style of field biology – all pressed in upon the received practice of naming and classifying organisms and the threat of anarchy. In the midst rose an index. This paper will explore the context of CD Sherborn's *Index Animalium* and those looming problems and issues which a laborious and comprehensive "index of nature" was meant to solve.

Editor's note

This paper is a transcription of the talk presented by Professor McOuat in the symposium *Anchoring Biodiversity Information: from Sherborn to the 21st century and beyond*, 28 October 2011, Natural History Museum, London. It is an exciting read about an important topic for this volume – it sets the historical and philosophical context for Sherborn's contribution to nomenclature and taxonomy clearly and vibrantly. It has a number of key messages on the relationships between names (dubbing) and meanings (taxonomy), on the struggle between establishing nomenclature tied to rules (codes) or to specimens (the type concept and museum catalogues). These issues were intensely addressed in the early and mid 19th century and Sherborn's magnum opus played a foundational role in establishing the systems we now use for all biology, not just zoology. Nonetheless many taxonomists today continue to befuddle these relationships, often through lack of knowledge of the long history of the discussions. I felt it was critical that this history is included in this volume, because it adds a different and necessary perspective on Sherborn's context and influence. Although we were not successful in getting Gordon McOuat to send his written text for the volume, I have decided to publish this as a transcript, with minor edits for flow and a few images for expanded context, as the talk is in the public domain and its presentation was fully funded by the symposium organisers. The paper should thus be read as a transcript only.

Talk and slides

http://backdoorbroadcasting.net/2011/10/gordon-mcouat-sherborn%E2%80%99s-context-cataloguing-nature/

Early Victorian recognition of the value of names

Although he worked in the late 19th Century, Sherborn's context starts with the very earliest groundwork for modern taxonomy, systematics and nomenclatural practice in the early 19th Century. This time included the origins of well-known disputes, of ruckuses in early Victorian biology, some of which are still with us today. Understanding these origins helps understand the issues in Victorian times and today.

Early Victorians knew the value of names, often couching the discussion in monetised terms. Sir William Kirby, in his Foundational Address of the Zoological Club of the Linnean Society, 1823, expressed the value that a name brings:

Nomina si pereunt, perit et cognitio rerum

"Names are the foundation of knowledge: and unless they have a 'a name' as well as a 'local habitation' with us, the zoological treasures that we so highly prize might almost as well have been left to perish in their native deserts or forests, as have grown mouldy in our drawers or repositories. But when once an animal subject is named and described, it becomes a possession for ever, and the value of every individual specimen of it, even in a mercantile view, is enhanced."

This is matched by the words of the radical anatomist, Robert Grant in his presentation to the Parliamentary Commission on the Affairs of the British Museum in 1835–1836:

'An object may not be the value of a farthing until it is identified and properly named. Its value may be raised to 30, 40 or 50 guineas once it is named, even though it has not gained an ounce.'

Both Kirby and Grant expressed the value of names at a time when there was turmoil in the process of giving names, and there was a process being born to establish stability and an anchor. There was a radical new club in the Linnean Society of London that harboured those who aimed to break the hold of Linnaeus over systematics and meaning. They aimed to introduce new ideas, imported, for example from France, to break the hold of the Linnaean world system. This is where Kirby made his presentation. Robert Grant, as a radical who called for the overthrow of all received systems, was himself a Lamarckian, an evolutionist and political radical before Darwin. These are presented in 'reform-bill Britain' where democratic forces threatened Tory privileges, much as the Occupy movement attempted at St Paul's, or as we see in the current challenges to the existing political and economic systems.

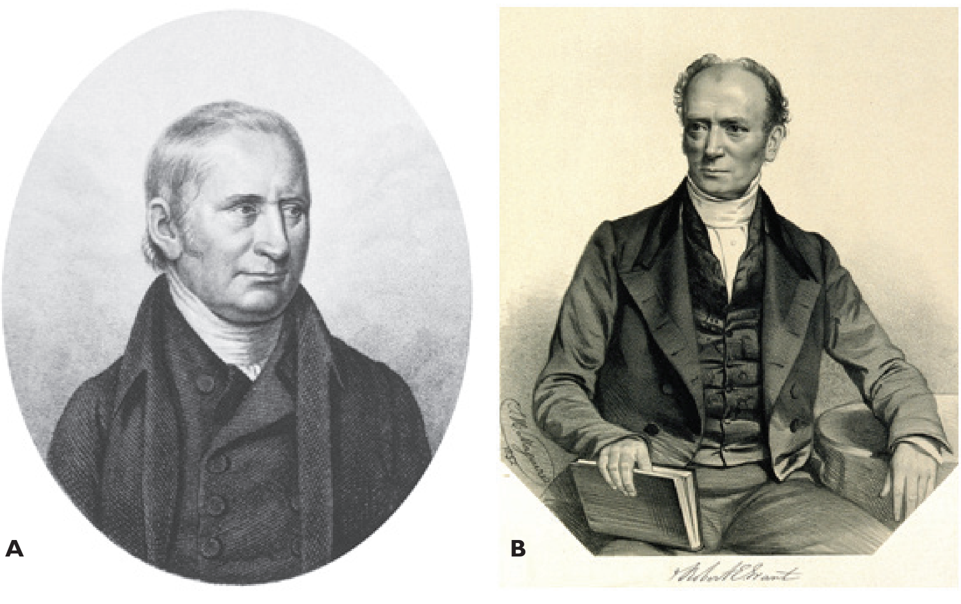


Figure 1. A Sir William Kirby, date and age uncertain **B** Robert Grant in 1852, aged 59.

The Stricklandian Code – the first attempt at an international code governing language in any science.

Any conference on nomenclature – its problems and its history – must harken back to the pioneering document, the founding creed of nomenclatural rules, the Code of Zoological Nomenclature drafted in 1842 by Hugh Edwin Strickland (1811–1853) under the patronage of the British Association for the Advancement of Science.

Strickland's committee was a veritable who's who of British natural history: John Stevens Henslow, Jennings, William Ogleby, JO Westwood, Richard Owen, Charles Darwin, William Yarl, WE Shuckard and GR Waterhouse. The committee convened its meetings in Darwin's house, as he still lived in London at the time. Here is an early draft of Strickland's rules with Darwin and Ogilby's comments on what should be changed and what should be kept (courtesy of Cambridge University Library): I cannot over emphasise the importance of these rules as a founding document. They are the first attempt at an international code governing language in any science. Any modern code, whether botanical or zoological, can trace its direct ancestry to this code. Many of the structures of modern codes, and many might say some of the problems, and zoological nomenclature in particular, can be traced directly to this code and its rules.

There are some important peculiarities of this document and its inheritance. We should unpack it a bit and give some grounding for Sherborn and his monumental project. The Stricklandian Code starts with a series of paragraphs with a very detailed account of the philosophy of language.

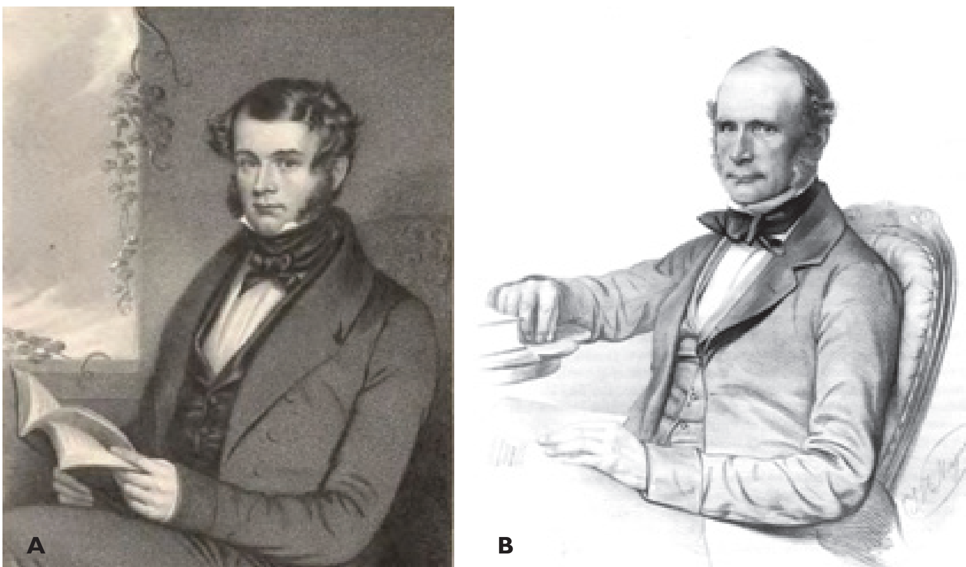
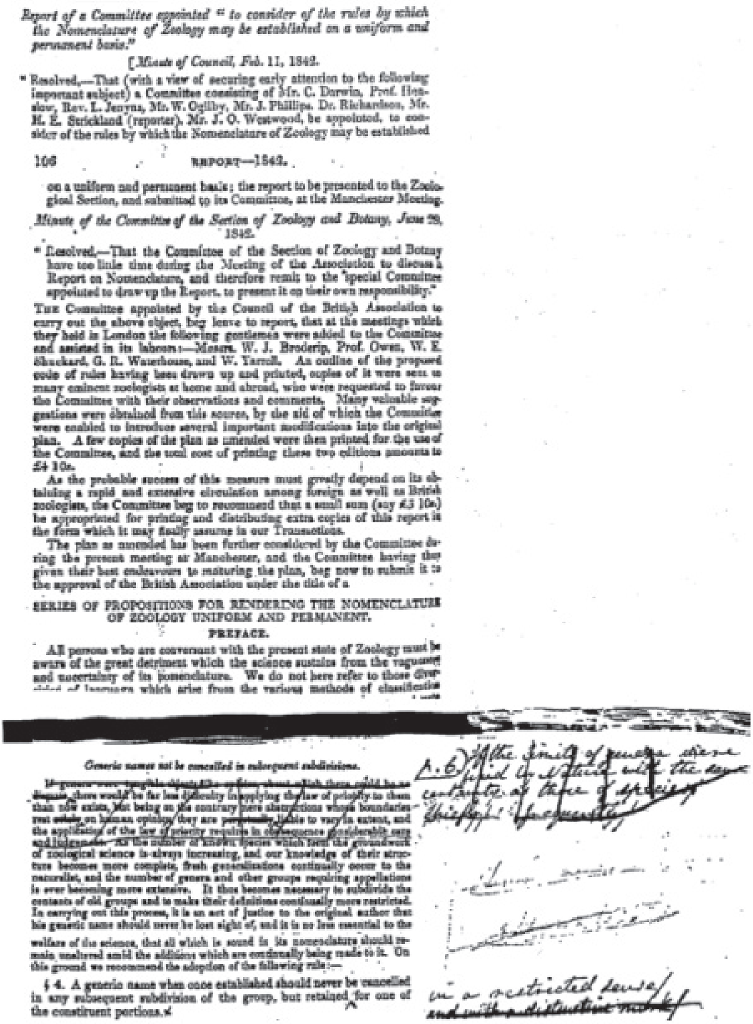


Figure 2. Hugh Edwin Strickland **A** age 26 **B** aged 42, when he died.



Report of a Committee appointed "to consider of the rules by which the Nomenclature of Zoology may be established on a uniform and permanent basis."

[*Minute of Council, Feb.* 11, 1842.

"Resolved,—That (with a view of securing early attention to the following important subject) a Committee consisting of Mr. C. Darwin, Prof. Henslow, Rev. L. Jenyns, Mr. W. Ogilby, Mr. J. Phillips, Dr. Richardson, Mr. H. E. Strickland (reporter), Mr. J. O. Westwood, be appointed, to consider of the rules by which the Nomenclature of Zoology may be established

106 REPORT—1842.

on a uniform and permanent basis; the report to be presented to the Zoological Section, and submitted to its Committee, at the Manchester Meeting.

Minute of the Committee of the Section of Zoology and Botany, June 28, 1842.

"Resolved,—That the Committee of the Section of Zoology and Botany have too little time during the Meeting of the Association to discuss a Report on Nomenclature, and therefore remit to the special Committee appointed to draw up the Report, to present it on their own responsibility."

THE Committee appointed by the Council of the British Association to carry out the above object, beg leave to report, that at the meetings which they held in London the following gentlemen were added to the Committee and assisted in its labours:—Messrs. W. J. Broderip, Prof. Owen, W. E. Shuckard, G. R. Waterhouse, and W. Yarrell. An outline of the proposed code of rules having been drawn up and printed, copies of it were sent to many eminent zoologists at home and abroad, who were requested to favour the Committee with their observations and comments. Many valuable suggestions were obtained from this source, by the aid of which the Committee were enabled to introduce several important modifications into the original plan. A few copies of the plan as amended were then printed for the use of the Committee, and the total cost of printing these two editions amounts to £4 10s.

As the probable success of this measure must greatly depend on its obtaining a rapid and extensive circulation among foreign as well as British zoologists, the Committee beg to recommend that a small sum (say £3 10s.) be appropriated for printing and distributing extra copies of this report in the form which it may finally assume in our Transactions.

The plan as amended has been further considered by the Committee during the present meeting at Manchester, and the Committee having thus given their best endeavours to maturing the plan, beg now to submit it to the approval of the British Association under the title of a

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

PREFACE.

All persons who are conversant with the present state of Zoology must be aware of the great detriment which the science sustains from the vagueness and uncertainty of its nomenclature. We do not here refer to those diver[illegible] which arise from the various methods of classificatio[illegible]

Generic names not to be cancelled in subsequent subdivisions.

~~If genera were tangible objects, the confines about which there could be no dispute,~~ there would be far less difficulty in applying the law of priority to them than now exists, but being on the contrary mere abstractions whose boundaries rest ~~solely~~ on human opinion, they are ~~perpetually~~ liable to vary in extent, and the application of the law of priority requires in consequence considerable care ~~and judgment~~. As the number of known species which form the groundwork of zoological science is always increasing, and our knowledge of their structure becomes more complete, fresh generalizations continually occur to the naturalist, and the number of genera and other groups requiring appellations is ever becoming more extensive. It thus becomes necessary to subdivide the contents of old groups and to make their definitions continually more restricted. In carrying out this process, it is an act of justice to the original author that his generic name should never be lost sight of, and it is no less essential to the welfare of the science, that all which is sound in its nomenclature should remain unaltered amid the additions which are continually being made to it. On this ground we recommend the adoption of the following rule:—

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

A. 6. ~~If the limits of genera were fixed in Nature with the same certainty as those of species~~ [illegible]

in a restricted sense/ ~~and with a distinctive mark~~

Figure 3. Early draft of the Stricklandian Code with handwritten comments by Darwin and Ogilby.

Language and meaning – dubbing not definitions

The system of naming and reference were in contention in Britain at this very time; followers of William Whewell had entirely different understanding of how things were named from the followers of John Locke. The Stricklandian Committee held a Lockian view of the meaning of meaning, as so remarkably espoused in these paragraphs.

Strickland himself had written numerously and voluminously on the notion of language, on the meaning and use of names. Strickland wrote:

'Words are only conventional signs. This should be enough to check those who are constantly trying to subvert the language of zoology. Names do not capture essences; they are not definitions. So how do they get authority and reference? By first dubbing.

RULES

OF

ZOOLOGICAL NOMENCLATURE.

PART I.

RULES FOR RECTIFYING THE PRESENT NOMENCLATURE.

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

In proposing a measure for the establishment of a permanent and universal zoological nomenclature, it must be premised that we refer solely to the Latin or systematic language of zoology. We have nothing to do with vernacular appellations. One great cause of the neglect and corruption which prevails in the scientific nomenclature of zoology, has been the frequent and often exclusive use of vernacular names in lieu of the Latin binomial designations, which form the only legitimate language of systematic zoology. Let us then endeavour to render perfect the Latin or Linnaean method of nomenclature, which, being far removed from the scope of national vanities and modern antipathies, holds out the only hope of introducing into zoology that grand desideratum, an universal language.

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

It being admitted on all hands that words are only the conventional signs of ideas, it is evident that language can only attain its end effectually by being permanently established and generally recognized. This consideration ought, it would seem, to have checked those who are continually attempting to subvert the established language of zoology by substituting terms of their own coinage. But, forgetting the true nature of language, they persist in confounding the *name* of a species or group with its *definition*; and because the former often falls short of the fulness of expression found in the latter, they cancel it without hesitation, and introduce some new term which appears to them more characteristic, but which is utterly unknown to the science, and is therefore devoid of all authority*. If these persons were to object to such names of men as *Long*, *Little*, *Armstrong*, *Golightly*, &c., in cases where they fail to apply to the individuals who bear them, or should complain of the names *Gough*, *Lawrence*, or *Harvey*, that they were devoid of meaning, and should hence propose to change them for more characteristic appellations, they would not act more unphilosophically or inconsiderately than they do in the case before us; for, in truth, it matters not in the least by what conventional sound we agree to designate an individual object, provided the sign to be employed be stamped with such an authority as will suffice to make it pass current. Now in zoology no one person can subsequently claim an authority equal to that possessed by the person who is the first to define a new genus or describe a new species; and hence it is that the name originally given, even though it may be inferior in point of elegance or expressiveness to those subsequently proposed, ought as a general principle to be permanently retained. To this consideration we ought to add the injustice of erasing the name originally selected by the person to whose labours we owe our first knowledge of the object; and we should reflect how much the permission of such a practice opens a door to obscure pretenders for dragging themselves into notice at the expense of original observers. Neither can an author be permitted to alter a name which he himself has once published, except in accordance with fixed and equitable laws. It is well observed by Decandolle, "L'auteur même qui a le premier établi un nom n'a pas plus qu'un autre le droit de le changer pour simple cause d'impropriété. La priorité en effet est un terme fixe, positif, qui n'admet rien, ni d'arbitraire, ni de partial."

* Linnaeus says on this subject, "Abstinendum ab hac innovatione quae nunquam cessaret, quin indies aptiora detegerentur ad infinitum."

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

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

Having laid down this principle, we must next inquire into the limitations which are found necessary in carrying it into practice.

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

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

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

Figure 4. Stricklandian Code – discussion of the philosophy of language.

Not by accurately capturing any meaning or sense, but rather that very first dubbing. These rules are about dubbing, and about disciplining that dubbing.'

This is a remarkable start for a set of rules on zoological nomenclature. It is only understandable in the face of the radical attempts, all through the early 19th century, and also today, to radically alter the words or names of things to match their place in scientific place and practice; to have names capture the reform and meaning of science.

Strickland had cut his teeth on fighting such radical attempts, and they were legion in the early 19th century, to entirely reform the whole system zoology and to adjust the names of things to match that reform.

Strickland's biggest enemy was Neville Wood, who was a popular writer, ornithologist, and eventually one of the leaders of alternative medicine in the late 19th century. Neville Wood would write 'It is essential for the improvement of ornithological science that names be frequently altered, for when a new system is proposed – and there are few who would advocate the Linnaean system now – new names must necessarily be introduced.'

New systems abounded. Anyone who is an historian of early 19th century natural history knows there were bifurcating systems and quinarian systems and Cuvierians… all of them associating the new names that they were establishing to match their system. But for Strickland, names are arbitrary – they are dubbings that hold on to that reference irrespective of the meanings in the systems to which they belong. But, asked Strickland, if there is a first dubbing, where is it to occur? Somewhat contravening his own philosophy, Strickland gives an arbitrary date of the 12th edition of Linnaeus' *Systema Naturae* where he thinks we find the solidification of binomial (binominal) nomenclature. It is from that moment that the dubbing of names should begin. This is where Strickland introduces the law (now principle) of priority. The very first rule states "the name originally given by the founder of a group or the describer of a species should be permanently retained."

Here the rules are giving rules for procedure and not for construction (or meanings) of the names themselves. The rest of the Code outlines where such descriptions can be found: published in certain received authoritative journals and books, and not in the popular press. All this was aimed at preventing amateurs from forming new names willy-nilly, removing the anchor and changing the very nature of zoological discourse. Thus, Strickland kept the issue of the meaning of names at bay. But notice how this brings up the issue of priority and genealogy.

The overarching priority of priority

The emergence of the Stricklandian Code was not without its own controversy. The 'British Association for the Advancement of Science Rules' were not actually passed by that organisation. They were cleverly inserted by Strickland into the report of 1842, but they were not actually adopted by the BAAS because of enormous opposition to the first rules.

The strongest opposition came from John Edward Gray (1800–1875), the chief Keeper of Zoology at the British Museum (which was still in Bloomsbury), who kept the rules from being approved by the BAAS. He was adamant that the Stricklandian rules should not be established to control the nature of discourse in natural history. Why? Because he was, at the same time, establishing a different source for authority on naming and discourse for natural history. He was working on his own solution to systematic and nomenclatural anarchy, his own material anchor to the biodiversity problem. For Gray, the British Museum catalogues of types would establish names and reference and be the site of authority. Not some regulatory rules, but real concrete catalogues and type specimens that would solidify the names.

Interestingly this huge fight between rules and museums continued in to the middle of the 19th century. Gray used the anarchy that seemed to exist in zoology as a way to lobby aristocratic trustees of the British Museum to publish the catalogues and establish them as the worldwide authorities of types and thus species. These began publication in the 1840s.

The huge fight had its short-term resolution, in a certain sense, in Darwin. His monograph on barnacles was the first to explicitly use the new Stricklandian rules, but also the first to use the new Gray catalogues. It was Darwin who tried to create a resolution of the rules from the committee of which he was a member, and the catalogues of types. This was an uneasy compromise that has not been a complete success – the controversy ran through the late 19th century.

Figure 5. John Edward Gray, the chief Keeper of Zoology at the British Museum (Bloomsbury).

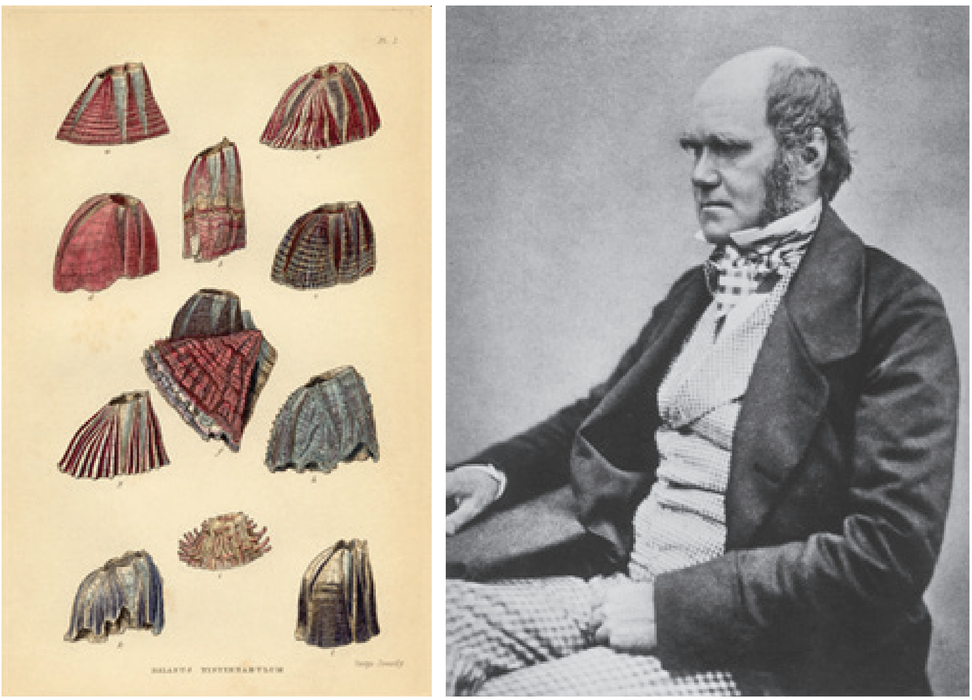


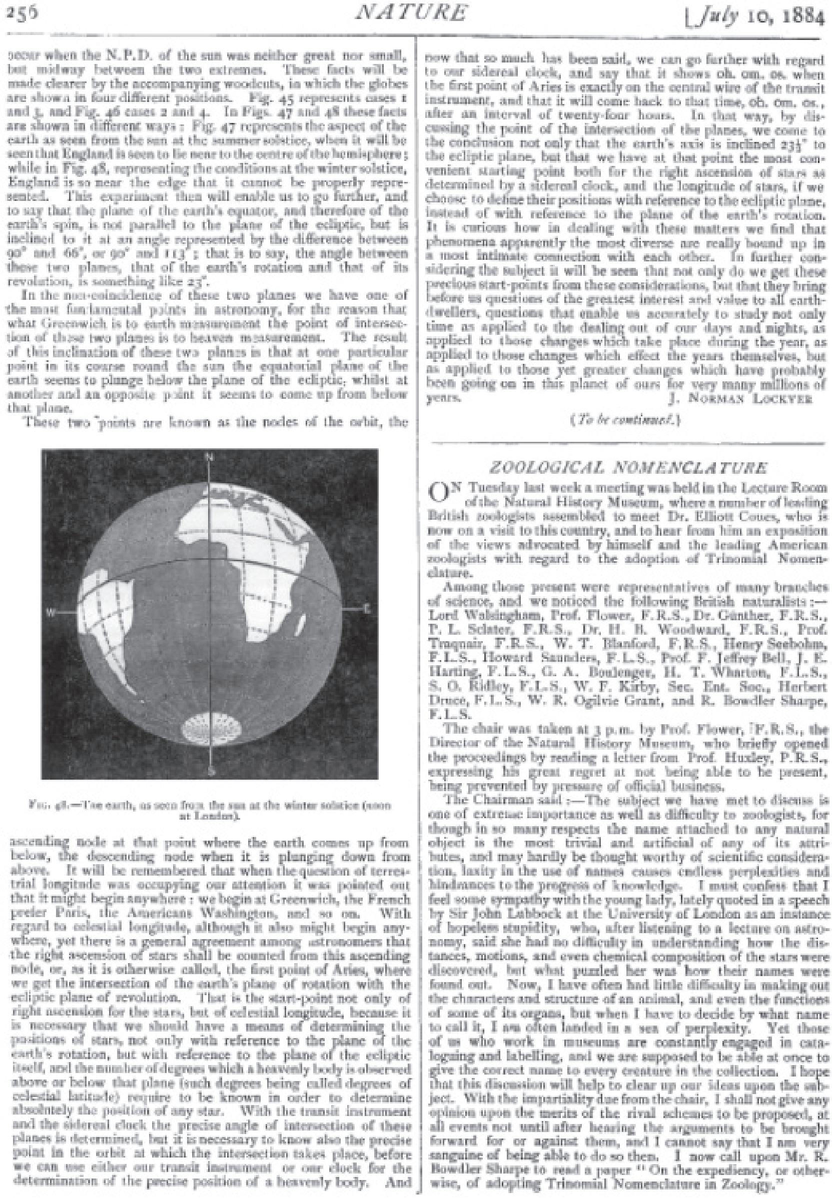
Figure 6. Charles Darwin and a plate from his work on barnacles.

Figure 7. A Elliot Coues **B** Joel A. Allen

It was then in the new Natural History Museum in South Kensington where the next steps were taken in the great nomenclatural debate, and the seeds were sown for Sherborn's great project. A great but controversial American zoologist, Elliot Coues (1842–1899) (Evenhuis this volume; Dickinson, this volume) happened to be visiting spiritualist sites throughout Europe. He and his partner in crime, Joel A. Allen (1838–1921), arrive at the Natural History Museum and advocated a new American way of doing field-based zoology, specifically ornithology, instead of the stodgy museum-based biology of the Old World. They set out an ornithological set of rules for nomenclature, which was supported by the American Ornithological Union.

The AOU rules were basically grounded on the Stricklandian rules except for one striking addition – the introduction of subspecies names, based on geographical distribution. Organisms would now be identified by a trinomial that would include the geographical location. All three parts would comprise the organism's name. For the British this was an utter travesty from ignorant Americans that promised a return to anarchy, to use the phrasing from William Flower, the director of the Natural History Museum's words. For the British, this was clearly mixing up, negating, the original Lockian perspective. It mixed up naming and meaning, violating all that had been achieved in establishing a system-free nomenclatural authority.

Thus, on July 1st, 1884 Coues presented his new system of trinomial nomenclature in a meeting in the new Natural History Museum London. Every British zoologist



256 *NATURE* [*July* 10, 1884

occur when the N.P.D. of the sun was neither great nor small, but midway between the two extremes. These facts will be made clearer by the accompanying woodcuts, in which the globes are shown in four different positions. Fig. 45 represents cases 1 and 3, and Fig. 46 cases 2 and 4. In Figs. 47 and 48 these facts are shown in different ways: Fig. 47 represents the aspect of the earth as seen from the sun at the summer solstice, when it will be seen that England is seen to lie near to the centre of the hemisphere; while in Fig. 48, representing the conditions at the winter solstice, England is so near the edge that it cannot be properly represented. This experiment then will enable us to go further, and to say that the plane of the earth's equator, and therefore of the earth's spin, is not parallel to the plane of the ecliptic, but is inclined to it at an angle represented by the difference between 90° and 66°, or 90° and 113°; that is to say, the angle between these two planes, that of the earth's rotation and that of its revolution, is something like 23°.

In the non-coincidence of these two planes we have one of the most fundamental points in astronomy, for the reason that what Greenwich is to earth measurement the point of intersection of these two planes is to heaven measurement. The result of this inclination of these two planes is that at one particular point in its course round the sun the equatorial plane of the earth seems to plunge below the plane of the ecliptic, whilst at another and an opposite point it seems to come up from below that plane.

These two points are known as the nodes of the orbit, the

FIG. 48.—The earth, as seen from the sun at the winter solstice (noon at London).

ascending node at that point where the earth comes up from below, the descending node when it is plunging down from above. It will be remembered that when the question of terrestrial longitude was occupying our attention it was pointed out that it might begin anywhere: we begin at Greenwich, the French prefer Paris, the Americans Washington, and so on. With regard to celestial longitude, although it also might begin anywhere, yet there is a general agreement among astronomers that the right ascension of stars shall be counted from this ascending node, or, as it is otherwise called, the first point of Aries, where we get the intersection of the earth's plane of rotation with the ecliptic plane of revolution. That is the start-point not only of right ascension for the stars, but of celestial longitude, because it is necessary that we should have a means of determining the positions of stars, not only with reference to the plane of the earth's rotation, but with reference to the plane of the ecliptic itself, and the number of degrees which a heavenly body is observed above or below that plane (such degrees being called degrees of celestial latitude) require to be known in order to determine absolutely the position of any star. With the transit instrument and the sidereal clock the precise angle of intersection of these planes is determined, but it is necessary to know also the precise point in the orbit at which the intersection takes place, before we can use either our transit instrument or our clock for the determination of the precise position of a heavenly body. And now that so much has been said, we can go further with regard to our sidereal clock, and say that it shows 0h. 0m. 0s. when the first point of Aries is exactly on the central wire of the transit instrument, and that it will come back to that time, 0h. 0m. 0s., after an interval of twenty-four hours. In that way, by discussing the point of the intersection of the planes, we come to the conclusion not only that the earth's axis is inclined 23½° to the ecliptic plane, but that we have at that point the most convenient starting point both for the right ascension of stars as determined by a sidereal clock, and the longitude of stars, if we choose to define their positions with reference to the ecliptic plane, instead of with reference to the plane of the earth's rotation. It is curious how in dealing with these matters we find that phenomena apparently the most diverse are really bound up in a most intimate connection with each other. In further considering the subject it will be seen that not only do we get these precious start-points from these considerations, but that they bring before us questions of the greatest interest and value to all earth-dwellers, questions that enable us accurately to study not only time as applied to the dealing out of our days and nights, as applied to those changes which take place during the year, as applied to those changes which effect the years themselves, but as applied to those yet greater changes which have probably been going on in this planet of ours for very many millions of years.

J. NORMAN LOCKYER

(*To be continued.*)

ZOOLOGICAL NOMENCLATURE

ON Tuesday last week a meeting was held in the Lecture Room of the Natural History Museum, where a number of leading British zoologists assembled to meet Dr. Elliott Coues, who is now on a visit to this country, and to hear from him an exposition of the views advocated by himself and the leading American zoologists with regard to the adoption of Trinomial Nomenclature.

Among those present were representatives of many branches of science, and we noticed the following British naturalists:—Lord Walsingham, Prof. Flower, F.R.S., Dr. Günther, F.R.S., P. L. Sclater, F.R.S., Dr. H. B. Woodward, F.R.S., Prof. Traquair, F.R.S., W. T. Blanford, F.R.S., Henry Seebohm, F.L.S., Howard Saunders, F.L.S., Prof. F. Jeffrey Bell, J. E. Harting, F.L.S., G. A. Boulenger, H. T. Wharton, F.L.S., S. O. Ridley, F.L.S., W. F. Kirby, Sec. Ent. Soc., Herbert Druce, F.L.S., W. R. Ogilvie Grant, and R. Bowdler Sharpe, F.L.S.

The chair was taken at 3 p.m. by Prof. Flower, F.R.S., the Director of the Natural History Museum, who briefly opened the proceedings by reading a letter from Prof. Huxley, P.R.S., expressing his great regret at not being able to be present, being prevented by pressure of official business.

The Chairman said:—The subject we have met to discuss is one of extreme importance as well as difficulty to zoologists, for though in so many respects the name attached to any natural object is the most trivial and artificial of any of its attributes, and may hardly be thought worthy of scientific consideration, laxity in the use of names causes endless perplexities and hindrances to the progress of knowledge. I must confess that I feel some sympathy with the young lady, lately quoted in a speech by Sir John Lubbock at the University of London as an instance of hopeless stupidity, who, after listening to a lecture on astronomy, said she had no difficulty in understanding how the distances, motions, and even chemical composition of the stars were discovered, but what puzzled her was how their names were found out. Now, I have often had little difficulty in making out the characters and structure of an animal, and even the functions of some of its organs, but when I have to decide by what name to call it, I am often landed in a sea of perplexity. Yet those of us who work in museums are constantly engaged in cataloguing and labelling, and we are supposed to be able at once to give the correct name to every creature in the collection. I hope that this discussion will help to clear up our ideas upon the subject. With the impartiality due from the chair, I shall not give any opinion upon the merits of the rival schemes to be proposed, at all events not until after hearing the arguments to be brought forward for or against them, and I cannot say that I am very sanguine of being able to do so then. I now call upon Mr. R. Bowdler Sharpe to read a paper "On the expediency, or otherwise, of adopting Trinomial Nomenclature in Zoology."

Figure 8. Report of the meeting discussing trinomial nomenclature.

of note was there – Schlater, Bolter, Guenther, Sharp. Huxley sent a note saying he couldn't attend but give'em hell. All were there to give the upstart Americans, Coues and Allen, a piece of their minds and defend their rules and their museum. The verbatim report from *Nature* makes interesting reading from a philosophical standpoint, as all the debates from the early 19th century are rehearsed in 1884 (Fig. 8). In fact, these arguments about the meaning of language, of dubbing, of authority, are rehearsed again and again subsequently, and perhaps still through the ICZN. Fears of anarchy are continually raised if there were to be a rejigging the 'meaning of meaning' for all of zoological nomenclature.

The British scientists argued that, by identifying location, trinomials were giving meaning within the name itself. This was liable to abuse, and would destabilise the system of authority so deeply established by rules and by the museum. Coues attempted to fight his corner but to no avail. The meeting was raucous and Coues was sent limping. The debate lasted long past that Tuesday 1st July, carrying on for the rest of the year in the press and journals.

Enter Sherborn and the *Index*

Recently employed by the geologist Thomas Rupert Jones in the British Museum (who might well have attended the raucous discussions on rules and meaning in names), was the 23 year-old Charles Davies Sherborn. He had already shown a predilection for indexing. We saw that Elliot Coues had tried to provoke him by saying only an inspired idiot could perform such a work. With the inspiration of Flower, Guenther, Slater and others, Sherborn published the announcement for the project of his great work in the May 1890 issue of *Nature* (Fig. 9). As stated in the announcement, the index was to be built on binomials (binominals) alone. He would constantly write that if something was a trinomial, it was not a name. The list was alphabetised by species, not genus. And the philosophical rule of priority, of first dubbing, now set to be from the 12th edition of Linnaeus, was to apply. The index became a deciding foundation to the problematic first presented by an attempt to anchor zoological discourse in a philosophy of language. It was a method of grounding and dubbing. It wore its origins and philosophical genealogy proudly.

As Sherborn stated later in life, in a 1933 private letter to Vaughn, the head of the Scripps Institute in the United States,

"After all this work, there are only two rules that are any good: First - Priority, which dates from 1st January 1758, and Second, that the first trivial is the type. If the generic diagnosis does not agree, then so much the worse for the genus, and it must be revised, unless the type is specifically mentioned. Them's my sentiments."

Sherborn continues,

"The International Zoological Committee is of little value as it meets only once in five years and then talks, but decides nothing. What we want is a Mussolini who can decide. Not a congress or a conference or such body who merely argue and make suggestions. I regard the first trivial name in a genus as the type unless it is otherwise fixed."

Images are in the public domain through Wikipedia and Wikimedia Commons unless otherwise noted.

NATURE [May 15, 1890

"Index Generum et Specierum Animalium."

Naturalists have long needed a reference book to the names of genera and species. Such a want has already been partially supplied by Agassiz, Bronn, Morris, Marschall, Scudder, Waterhouse, and others—only Bronn and Morris having attempted palæontological species—but no one book including references to all names given to living and fossil animals has yet been attempted. Botanists, more fortunate, will soon possess Daydon Jackson's index to flowering plants. The idea has therefore suggested itself to me to begin at the end of June next, such an "Index Generum et Specierum Animalium," taking the following rules for guidance:—

(1) The earliest reference is to date from the twelfth edition of Linnæus, 1766.

(2) The last reference to close with December 31, 1899.

(3) The names of genera and species to be given in a single alphabetical sequence, and accompanied by a reference to the original source.

(4) The names of species of each genus to be also quoted in alphabetical order under that genus.

(5) No attempt at synonymy to be given; but, to assist reference, the various genera in which a species has from time to time been placed, to be indicated under that species.

(6) Pre-Linnæan names to be quoted as founded by the author first using them after 1766:—*e.g. Echinocorys*, Leske, 1778 (*ex* Klein, 1734). Should a pre-Linnæan species or genus have been re-named after 1766, before the post-Linnæan use of that pre-Linnæan name, the new name is to stand. [References will be given to Artedi, Brisson, and Scopoli, in accordance with British Association rules.]

Among the many offers of assistance, that of Prof. Flower, F.R.S., Dr. Günther, F.R.S., and Dr. Henry Woodward, F.R.S., who have promised the necessary space for the storage of the MS. in the Natural History Museum, is most valuable, as it practically ensures safety from fire, and renders the MS. easily accessible to those wishing to consult it while still imperfect.

The contribution of inaugural addresses, theses, or other publications difficult to obtain, would be of great assistance; and, after use, such pamphlets would be handed over to the library at the Museum.

Any suggestions for the improvement of this plan, before the commencement of the undertaking, would be gladly received and carefully considered.

Appended is a rough outline of the scheme:—

[cordatus -a, -um]
—— Amphidetus (Penn.) Düb. and Koren, Zool. Bid. 285 1844
[*v.* Echinus]
—— Amphidotus (Penn.) E. Forbes, Brit. Starf. 190, fig. 1841
[*v.* Echinus]
—— Echinocardium (Penn.) J. E. Gray, Cat. R. Ech. 43 1855
[*v.* Echinus]
—— Echinus, Pennant, Brit. Zool. iv. 58, xxxiv. 2, xxxvi. 2 1777
[*v.* also Amphidetus, Amphidotus, Echinocardium, Spatangus]
—— Spatangus (Penn.) Flem. Brit. Anim. 480 ... 1828
[*v.* Echinus]
Cordia, Stål, Hem. Afric. iv. 78 ... ... Hem. 1866
[albilaterata, peragrans.]
Cordienia, A. Rouault, B. S. géol. France, v. 207.. Gast. 1848
[biaritziana, iberica, palensis, pyrenaica, all *nom. nud.*]

Charles Davies Sherborn.

540 King's Road, London, S.W.

Figure 9. Announcement of Sherborn's plans for his hugely ambitious project.

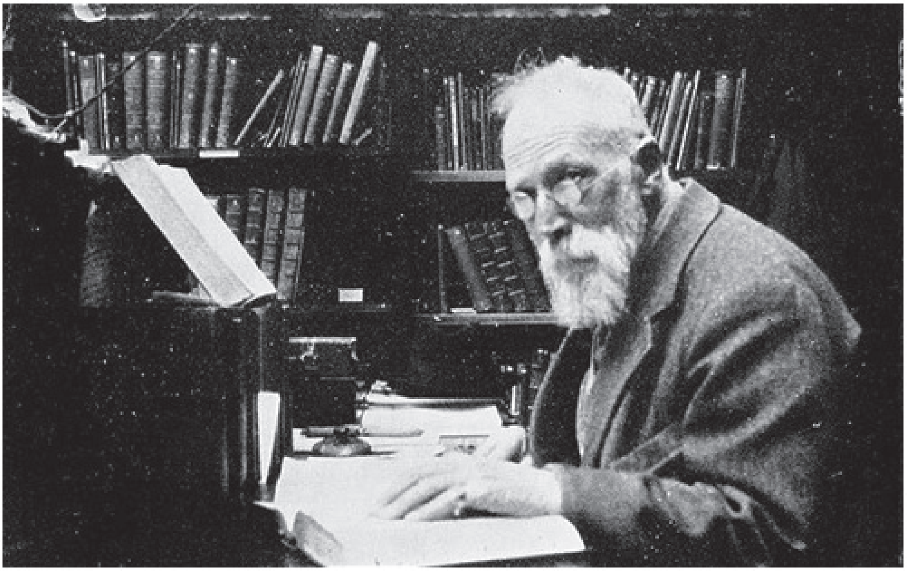


Figure 10. An iconographic picture of Sherborn in later years – staged, but revealing and taken at about the time of the final quote.

ZooKeys 550: 71–81 (2016)
doi: 10.3897/zookeys.550.9863
http://zookeys.pensoft.net

RESEARCH ARTICLE

A peer-reviewed open-access journal
ZooKeys
Launched to accelerate biodiversity research

Sherborn's foraminiferal studies and their influence on the collections at the Natural History Museum, London

C. Giles Miller[1]

1 *Department of Earth Sciences, The Natural History Museum, London. SW7 5BD*

Corresponding author: *C. Giles Miller* (G.Miller@nhm.ac.uk)

Academic editor: *E. Michel* | Received 23 April 2015 | Accepted 24 April 2015 | Published 7 January 2016

http://zoobank.org/C7769BF0-324A-4BC6-B637-F82374078CE6

Citation: Miller CG (2016) Sherborn's foraminiferal studies and their influence on the collections at the Natural History Museum, London. In: Michel E (Ed.) Anchoring Biodiversity Information: From Sherborn to the 21st century and beyond. ZooKeys 550: 71–81. doi: 10.3897/zookeys.550.9863

Abstract

Sherborn's work on the Foraminifera clearly provided the initial spark to compile the major indexes for which he is famous. Contact and help from famous early micropalaeontologists such as T. Rupert Jones and Fortescue William Millett led Sherborn to produce his *Bibliography of Foraminifera* and subsequently a two-part *Index of Foraminiferal Genera and Species*. Edward Heron-Allen, whose mentor was Millett, was subsequently inspired by the bibliography to attempt to acquire every publication listed. This remarkable collection of literature was donated to the British Museum (Natural History) in 1926 along with the foraminiferal collections Heron-Allen had mainly purchased from early micropalaeontologists. This donation forms the backbone of the current NHM micropalaeontological collections. The NHM collections contain a relatively small amount of foraminiferal material published by Sherborn from the London Clay, Kimmeridge Clay and Speeton Clay. Another smaller collection reflects his longer-term interest in the British Chalk following regular fieldwork with A. W. Rowe. Other collections relating to Sherborn's early published work, particularly with T. R. Jones, are not present in the collections but these collections may have been sold or deposited elsewhere by his co-workers.

Keywords

Foraminifera, C. D. Sherborn, Fortescue William Millett, T. Rupert Jones, Edward Heron-Allen, A. W. Rowe, C. P. Chatwin, T. H. Withers, Natural History Museum, Chalk, London Clay, Kimmeridge Clay

Introduction

Whilst Sherborn is best known for his *Index Animalium*, his scientific career began with work on Foraminifera and Ostracoda. Foraminifera are single celled organisms that secrete amazingly diverse microscopic shells or tests of mainly calcite. Occasionally they use available ocean bottom sediment to create their tests. Ostracoda are microscopic bivalved crustaceans common in most aquatic environments and found through most of the fossil record. This paper aims to provide details of Sherborn's relationships to early micropalaeontological workers, to summarise his work on the Foraminifera and to investigate how this is reflected by the collections currently held at the Natural History Museum, London. For convenience in this paper, the title "Natural History Museum" is used throughout even where the original name of the institution was The British Museum (Natural History).

The Natural History Museum foraminiferal collections

Roughly half of the Natural History Museum's microfossil collection of approximately 550,000 slides represent examples of foraminifera with the remainder including ostracods, palynomorphs, calcareous nannofossils, radiolarians and conodonts. The museum's micropalaeontological collection is built around the donation in 1926 of a remarkable collection of foraminiferal books and slides assembled by Edward Heron-Allen mainly during the early 20th century and subsequently much added to (Hodgkinson 1989, Whittaker 2013). These collections contain fossil and mainly Recent Foraminifera and were originally part of the Zoology Department collections. They were subsequently moved to The Geology Department, part of which later became The Palaeontology Department that later formed part of the current Earth Science Department. A full history of the NHM micropalaeontological collections and their custodians has yet to be written but some historical details of Heron-Allen's collection of Foraminifera and its acquisition have been provided by Hodgkinson and Whittaker (2003) and other papers such as Adams et al. (1980) give an overview of all type material deposited by 1980. Details of most of the type and figured part of the collection can be found on-line on the museum's web site (http://data.nhm.ac.uk/).

Sherborn and T. Rupert Jones

T. Rupert Jones (1819–1911) was a London surgeon who became interested in palaeontology and particularly Foraminifera and Ostracoda. He was later Professor of Geology at the Royal Military College in Sandhurst, Fellow of the Geological Society and was elected Fellow of the Royal Society in 1872 (Siveter and Lord 2004). Evenhuis (2016), Shindler (2016) and Siveter and Lord (2004) describe how in retirement, T. R. Jones employed Sherborn to help illustrate and complete some works on the

Foraminifera leading Sherborn to The Department of Geology at the new Natural History Museum at South Kensington where he made contact with several members of staff in The Geology Department. Sherborn claims that he was one of the first dozen visitors through the doors at the new museum when it opened in 1881 and shortly afterwards was employed by The Geology Department to mount specimens (Shindler 2016). Sherborn and T. R. Jones jointly published three papers; two on Foraminifera (Jones and Sherborn 1886, 1887) and another a large monograph on the Ostracoda (Jones and Sherborn 1889).

Sherborn and Millett

Fortescue William Millett (1833–1915) was one of the leading micropalaeontologists of the 19th Century working mainly on Recent Foraminifera (Hodgkinson 2006, Hart et al. 2011). Although he was born in Cornwall, SW England and retired there, he spent over 30 years in London where he was one of the founder members of the Quekett Microscopical Club in 1865. This is presumably where he came into contact with Sherborn although this is not recorded. Millett had an encyclopaedic knowledge of the Foraminifera and a very good library (Hodgkinson 2006). Sherborn's acknowledgement of Millett in the 1893 *Index of Genera and Species* is stated in a prefatory note dated October 1893 Millett's 'knowledge of literature of the subject is remarkable and peculiar' (Sherborn 1893, un-numbered page prior to p. 1). Sherborn wrote a short obituary of Millett (Sherborn 1915), which was expanded by Hodgkinson (2006) to provide an in-depth biography of Millett including listing Millett's 13 papers on the Foraminifera.

Documents rescued by Edward Heron-Allen from Millett's house after his death suggest that Millett had a strong connection with T. R. Jones who often passed him collections to study (Hodgkinson 2006) and both Sherborn and Millett are acknowledged in T. R Jones's introduction to the foraminiferal part of the Crag Monograph of East Anglia (Jones 1895).

Sherborn and Heron-Allen

We can only assume that Edward Heron-Allen (1861–1943) met Sherborn because Heron-Allen was also greatly influenced by his mentor Millett and was also a member of the Quekett Microscopical Club. The polymathic Heron-Allen had become fascinated by the Foraminifera aged 14 but only started serious study of them relatively late in life. He made it an aim to acquire all of the early works on the Foraminifera listed in Sherborn's 1888 *Bibliography of Foraminifera* (Jones 2005). Several copies of this 1888 *Bibliography* are present in the Heron-Allen Library at the NHM. Heron-Allen had his personal copy rebound and annotated its margins with references to books in his personal library that he later donated to the British Museum (Natural

Figure 1. C. D. Sherborn bookplate acquired by Heron-Allen and pasted into the first couple of pages of Heron-Allen's copy of Sherborn (1888b). Heron-Allen wrote 'The original figure *Globigerina bulloides* is 23.5cm in height, is quite accurately reproduced.'

History). Attached into the front of the same book is C. D. Sherborn's bookplate (Fig. 1) designed and engraved by Sherborn's father C. W. Sherborn that illustrates a bust of Shakespeare and a profile portrait of Darwin (Jones 2005). Heron-Allen, a prolific gatherer of related materials, often pasted articles or letters or annotated them but it is not clear how he acquired this particular item. Next to the book plate he left the following handwritten note:

'Symbolic book plate of the author designed and engraved by his father C. W. Sherborn, the most notable book plate engraver of the XIXth century. NB the miniature reproduction of Plate 77 of H. B. Brady's Report on the Foraminifera of the Challenger Expedition, London 1884. The original figure *Globigerina bulloides* is 23.5cm in height, is quite accurately reproduced.'

The annotations in Heron-Allen's copy of the 1888 index suggest that Heron-Allen came close to his aim of acquiring all the references listed in the bibliography as about 80 per cent are accounted for. Whether Heron-Allen and Sherborn ever met is not known for certain. What is certain is that the index guided Heron-Allen to accumulate the amazing collection of foraminiferal books and references that forms the backbone of the NHM Micropalaeontology Sectional Library that now bears the name 'The Heron-Allen Library'. The library, augmented with Heron-Allen's annotations and attached documentation is a unique and unrivalled resource for anyone wishing to study the Foraminifera (Whittaker 2013).

Sherborn's research on the Foraminifera

Shindler (2016) and Evenhuis (2016) provide a more detailed history of the production of the bibliography that was started in 1886 and published in 1888. While pre-

Figure 2. Handwritten note with the collections referring to slides donated by C. D. Sherborn from excavation to widen the Cannon Street Rail Bridge. Representative illustration of Lias Foraminifera from Northamptonshire from Crick and Sherborn (1891, part of fig. 1)

paring the *Bibliography* he also published on the foraminifera from the London Clay of Piccadilly, London (Sherborn and Chapman 1886) and the Jurassic of England (Jones and Sherborn 1886). He continued his collaboration with T. R. Jones with a publication on the variability in cristellarian Foraminifera (Jones and Sherborn 1887) and their collaboration culminated with the publication of a monograph on the Tertiary Entomostraca (ostracods) of England (Jones and Sherborn 1889). Just before the publication of the *Bibliography* (Sherborn 1888b) he published a savage review of the American Anthony Woodward's foraminiferal bibliography (Sherborn 1888a) and this almost compromised the publication of his own bibliography (Shindler 2016, Taylor 2016). He also published a short note on the foraminiferan *Webbina irregularis* (d'Orb) from the Oxford Clay at Weymouth (Sherborn 1888c), by way of a comment on the collection of R. Formby Esquire of Bath. An additional note on the Foraminifera of the London Clay from the Drainage Works in Piccadilly, London (Sherborn and Chapman 1889) was published in the same year as a paper on the London Clay from Sheppey (Chapman and Sherborn 1889). Two papers were written in collaboration with H. W. Burrows (Burrows et al. 1890) on the Foraminifera of the Red Chalk of Yorkshire, Norfolk and Lincolnshire and shortly afterwards on the London Clay from Cannon Street rail bridge (Sherborn and Burrows 1891, Fig. 2).

There followed two publications with Walter Drawbridge Crick (1857–1903) on the Liassic Foraminifera from Northamptonshire (Crick and Sherborn 1891, 1892). To accompany his 1888 *Bibliography of the Foraminifera*, Sherborn then compiled his *Index of Foraminiferal Genera and Species* in two parts (Sherborn 1893, 1896). At around the same time, he published on ostracods from the Gault at Folkestone (Chapman and Sherborn 1893) remarking that these had been recovered during the

Figure 3. A. Sherborn (left) and A. W. Rowe (right) doing field work on an outcrop of chalk reproduced from Hart and Bailey (2013) with permission from Natural History Museum. B. Sherborn (right) and C. P. Chatwin (left) during fieldwork on the British Chalk.

production of a monograph on the Foraminifera. Sherborn was a lifelong friend and field companion of A. W. Rowe (1858–1926; Fig. 3A) who published a string of papers on the zonation of the English Chalk including one in 1930 that was finished after his death by Sherborn and T. H. Withers (1883–1953) (Gale and Cleevely 1989, Hart and Bailey 2013). This paper includes listings of foraminiferal species but not illustrations or discussions on Foraminifera. In C. S. Carter's Presidential address to the Lincolnshire Naturalists' Union for 1928 (Carter 1929) he states that Rowe and Sherborn worked closely together on the Chalk of Lincolnshire, assisted by local amateurs. Gale and Cleevely (1989) provide details of the relationship between Rowe and Sherborn including examples of work Sherborn was responsible for and various anecdotes that suggest that Rowe used Sherborn as a dogsbody. Sherborn is also known to have carried out fieldwork on the Chalk with Charles P. Chatwin (1887–1971, Fig. 3B) who published on the Foraminifera of the Chalk from Oxfordshire and Berkshire with T. H. Withers (Chatwin and Withers 1908). This published collection is held at The Natural History Museum (NHMUK PM P 8762-8786; P 8711-8714). Chatwin was at the time an attendant at the museum and went on to become Librarian at the Geological Society (1913–19), lecturer in Palaeontology at the University of Liverpool (1919–20) and worked at the Geological Survey from 1920 to 1941 (Andrew Morrison, pers. com.).

Figure 4. Examples of Sherborn slides from the Natural History Museum collections.

Sherborn's foraminiferal collections at The Natural History Museum

The collections are relatively modest compared to the number of Sherborn's publications. It may be that the collections were deposited elsewhere or sold if Sherborn was not the first author. The registers show that Sherborn sold material to the British Museum in 1886 and later donated material in 1890. We know that many of T. R. Jones's collections were sold after his death (Siveter and Lord 2004). Collections not present include material from the London Clay of Sheppey (Chapman and Sherborn 1889) and the material from the Northamptonshire Lias published by Crick et al. (1891, 1892). Examples of Sherborn slides in The Natural History Museum collection are shown (Fig. 4) in the hope that similar slides might be recognised in other collections.

Four discrete collections remain at the Natural History Museum:

1 21 slides from an excavation of the London Clay at Piccadilly (NHMUK PM P 3669-3726) were purchased from Sherborn in October 1886 and represent the material relating to Sherborn and Chapman (1886).

2 18 slides from the works to widen the Cannon Street Rail Bridge (NHMUK PM P 4370, 9722-9738) were presented by Sherborn in 1890 and relate to the publication Sherborn and Burrows (1891).

3 25 slides from several Kimmeridgian, Jurassic sites at Roslyn Pit, Ely, Cambridgeshire, Gillingham Brick Works, Dorset and from near the cemetery at Mere, Wiltshire (NHMUK PM P 42004-42119, 42180-42193). 2 other slides (NHMUK PM P 33316-33317) are from Ely and marked as collected and presented by Sherborn c. 1899. They were published by Ovey (1938).

4 14 slides from the Red Chalk at Hunstanton and from the Speeton Clay are unregistered but the slide labels indicate that they were prepared by the Rev. G. Bailey. They may relate in part to the publication of Burrows et al. (1890).

Concluding statements

Sherborn's foraminiferal collections at the NHM are relatively modest in size and some key collections that he published on are not present. His collections and work on the Foraminifera cannot be considered to be particularly ground breaking. In contrast, the production of the Bibliography and Indexes had a profound effect on Edward Heron-Allen whose subsequent donation of literature and collection forms the backbone of the current Natural History Museum micropalaeontology collection. Publication of the foraminiferal bibliography and indexes also had a profound effect on Sherborn who went on to publish his *Index Animalium*. It seems that later in his life Sherborn continued to encourage workers such as A. W. Rowe, T. H. Withers and C. P. Chatwin to work on foraminifera from the British Chalk, some collections of which are also housed at the Natural History Museum.

Footnote

The collections at The Natural History Museum continue to be influenced by a modern day C. D. Sherborn. Dr John Williams has been compiling an index of all papers relating to Palaeopalynology. His index of currently stands at 25,502 items, all cross referenced by a vast card index (Riding et al. 2012).

Acknowledgements

This contribution has been greatly improved following reviews by Michael A. Taylor (University of Leicester; http://orcid.org/0000-0002-1495-8215) and Jeremy Young (University College, London). John Whittaker (Scientific Associate, Natural History Museum), who also reviewed this paper, is thanked for his help and encouragement to undertake this short study. Haydon Bailey (Network Stratigraphic, Potters Bar) provided information about A. W. Rowe as did Chris Wood (formerly of The British Geological Survey) and Prof Andy Gale (University of Portsmouth) helped source Figure 3A which is reproduced from the Natural History Museum Archives with permission. The dates of birth and death cited in this paper are extracted from Cleevely (1983). Details of Charles P. Chatwin were provided by Andrew Morrison (Archivist, British Geological Survey).

References

Adams CG, Harrison CA, Hodgkinson RL (1980) Some primary type specimens of Foraminifera in the British Museum (Natural History). Micropalaeontology 26(1): 1–16. http://micropal.geoscienceworld.org/content/26/1/1.full.pdf+html

Brady HB (1884) Report on the Challenger foraminifera dredged by H. M. S. Challenger during the years 1873–1976. Report on the Scientific Results of the Voyage of H. M. S. Challenger (Zoology Series) 9: 1–814.

Burrows HW, Sherborn CD, Bailey G (1890) The Foraminifera of the Red Chalk of Yorkshire, Norfolk and Lincolnshire. Journal of the Royal Microscopical Society 1890(5): 549–566. doi: 10.1111/j.1365-2818.1890.tb00798.x

Carter CS (1929) Presidential address to the Lincolnshire Naturalists' Union – The White Chalk of Lincolnshire. Lincolnshire Naturalists Union Transactions for 1928: 45–69.

Chapman F, Sherborn CD (1889) Foraminifera from the London Clay at Sheppey. Geological Magazine 3(4): 498–499. doi: 10.1017/S001675680018954X

Chapman F, Sherborn CD (1893) On the Ostracoda of the Gault at Folkestone. Geological Magazine 3(10): 345–349. doi: 10.1017/S0016756800170645

Chatwin CP, Withers TH (1908) The zones of the Chalk in the Thames Valley between Goring and Shiplake. Proceedings of the Geologists' Association 20(5): 390–419. doi: 10.1016/S0016-7878(08)80289-1

Cleevely RJ (1983) World Palaeontological Collections. British Museum (Natural History), Mansell Publishing Limited, 365 pp.

Crick WD, Sherborn CD (1891) On some Liassic Foraminifera from Northamptonshire. Journal of the Northamptonshire Natural History Society 6: 208–213.

Crick WD, Sherborn CD (1892) On some Liassic Foraminifera from Northamptonshire Part II – The Leda-Ovum Beds of the Upper Lias. Journal of the Northamptonshire Natural History Society 7: 67–73.

Evenhuis NL (2016) Charles Davies Sherborn and the "Indexer's Club". In: Michel E (Ed.) Anchoring Biodiversity Information: From Sherborn to the 21st century and beyond. ZooKeys 550: 13–32. doi: 10.3897/zookeys.550.9697

Gale AS, Cleevely RJ (1989) Arthur Rowe and the Zones of the White Chalk of the English coast. Proceedings of the Geologists' Association 100(4): 419–431. doi: 10.1016/S0016-7878(89)80018-5

Hart MB, Bailey HW (2013) Key figures from the history of research on the Foraminifera of the Chalk Group in the UK. In: Bowden AJ, Gregory FJ, Henderson AS (Eds) Landmarks in foraminiferal micropalaeontology. Micropalaeontological Society Special Publication, The Geological Society, 85–102.

Hart MB, Boulton SJ, Hart AB, Leighton AD (2011) Fortescue William Millett (1833–1915): The man and his legacy in South-West England. Geoscience in South West England 12: 295–303.

Hodgkinson RL (1989) The Heron-Allen and Earland Type slide collection of Foraminifera in the British Museum (Natural History). Journal of Micropalaeontology 8(2): 149–156. doi: 10.1144/jm.8.2.149

Hodgkinson RL (2006) The life and works of Fortescue William Millett (1833–1915), foraminiferologist. Journal of Micropalaeontology 25: 165–174. doi: 10.1144/jm.25.2.165

Hodgkinson RL, Whittaker JE (2003) The Beauty and Mystery of the Foraminifera. The scientific research of Edward Heron-Allen FRS. Opusculum IV, The Heron Allen Society, 38 pp.

Jones C (2005) Some treasures of the Heron-Allen Library at the Natural History Museum, London. In: Whittaker JE (Ed.) Edward Heron-Allen FRS: Scientist. Proceedings of the 4th Heron-Allen Symposium, 2004. Opusculum VII, The Heron Allen Society, c/o W Sussex Registry Office, Chichester, 32–50.

Jones TR (1895) A monograph of the Foraminifera of the Crag Part II (1895). Monograph of The Palaeontographical Society, 73–210.

Jones TR, Sherborn CD (1886) On the Microzoa found in some Jurassic Rocks of England. Geological Magazine 3(3): 271–274. doi: 10.1017/S0016756800145030

Jones TR, Sherborn CD (1887) Remarks on the Foraminifera, with especial reference to their variability of form, illustrated by the Cristellarians. Journal of the Royal Microscopical Society 1887(4): 545–557. doi: 10.1111/j.1365-2818.1887.tb01579.x

Jones TR, Sherborn CD (1889) A supplementary monograph of the Tertiary Entomostraca of England. Monograph of The Palaeontographical Society of London 42: 1–68.

Ovey CD (1938) Difficulties in establishing relationships in Foraminifera. Proceedings of the Geologists' Association 49(2): 160–170. doi: 10.1016/S0016-7878(38)80015-8

Riding JB, Pound MJ, Hill TCB, Stukins S, Feist-Burkhardt S (2012) The John Williams Index of Palaeopalynology. Palynology 36(2): 224–233. doi: 10.1080/01916122.2012.682512

Rowe AW (1930) The Zones of the White Chalk of Lincolnshire. The Naturalist December 1929(875): 411–439. [Written posthumously by C. D. Sherborn and T. H. Withers]

Sherborn CD (1888a) An "Instructive" Bibliography of the Foraminifera. Nature 37: 583–584. doi: 10.1038/037583a0

Sherborn CD (1888b) A bibliography of the Foraminifera fossil and recent from 1565–1888: with explanatory notes of some of the rare and little known publications. Dulau & Co, London, 152 pp. doi: 10.5962/bhl.title.20095

Sherborn CD (1888c) Note on *Webbina irregularis* (d'Orb) from the Oxford Clay at Weymouth. Proceedings of the Bath Natural History and Antiquarian Field Club 6(3): 332.

Sherborn CD (1893) An Index of the Genera and Species of the Foraminifera. Part I A to Non. Smithsonian Miscellaneous Collections 856.

Sherborn CD (1896) An Index of the Genera and Species of the Foraminifera. Part II Non to Z. Smithsonian Miscellaneous Collections 1031.

Sherborn CD (1915) Fortescue William Millett. Geological Magazine 6(2): 288. doi: 10.1017/S0016756800177945

Sherborn CD, Burrows HW (1891) Report on microscopical examination of some samples of London Clay from the excavations for the widening of Cannon Street Railway Bridge, London, 1887. Proceedings of the Geologists' Association 12: 4–7. doi: 10.1016/S0016-7878(91)80022-0

Sherborn CD, Chapman F (1886) On some Microzoa from the London Clay exposed in the Drainage Works, Piccadilly, London, 1885. Journal of the Royal Microscopical Society 2(6): 737–763. doi: 10.1111/j.1365-2818.1886.tb02072.x

Sherborn CD, Chapman F (1889) Additional note on the Foraminifera of the London Clay exposed in the Drainage Works, Piccadilly, London in 1885. Journal of the Royal Microscopical Society 1889(4): 483–488.

Shindler K (2016) A magpie with a card-index mind – Charles Davies Sherborn 1861–1942. In: Michel E (Ed.) Anchoring Biodiversity Information: From Sherborn to the 21st century and beyond. ZooKeys 550: 33–56. doi: 10.3897/zookeys.550.9975

Siveter DJ, Lord AR (2004) Thomas Rupert Jones: pioneering studies on ostracods. In: Siveter DJ, Lord AR (Eds) In the footsteps of T. R. Jones: Lower Palaeozoic of Shropshire and post-Palaeozoic of Avon, Dorset and Kent. Field Guide for the Thirteenth International Symposium on Ostracoda, Chatham 1997. British Micropalaeontological Society Field Guide, 10: 7–10

Taylor MA (2016) '*Where is the damned collection?*' Charles Davies Sherborn's listing of named natural science collections and its successors. In: Michel E (Ed.) Anchoring Biodiversity Information: From Sherborn to the 21st century and beyond. ZooKeys 550: 83–106. doi: 10.3897/zookeys.550.10073

Whittaker JE (2013) Edward Heron-Allen F. R. S. (1861–1943). In: Bowden AJ, Gregory FJ, Henderson AS (Eds) Landmarks in foraminiferal micropalaeontology. Micropalaeontological Society Special Publication, The Geological Society, 47–58.

ZooKeys 550: 83–106 (2016)
doi: 10.3897/zookeys.550.10073
http://zookeys.pensoft.net

RESEARCH ARTICLE

A peer-reviewed open-access journal
ZooKeys
Launched to accelerate biodiversity research

'*Where is the damned collection*?' Charles Davies Sherborn's listing of named natural science collections and its successors

Michael A. Taylor[1]

1 *School of Museum Studies, University of Leicester & Department of Natural Sciences, National Museums Scotland, Chambers St., Edinburgh EH1 1JF, Scotland (http://orcid.org/0000-0002-1495-8215)*

Corresponding author: *Michael A. Taylor* (mat22@le.ac.uk)

Academic editor: *E. Michel* | Received 28 May 2015 | Accepted 28 May 2015 | Published 7 January 2016

http://zoobank.org/109EE95D-9C64-47F4-9AF6-E77789B511B3

Citation: Taylor MA (2016) '*Where is the damned collection*?' Charles Davies Sherborn's listing of named natural science collections and its successors. In: Michel E (Ed.) Anchoring Biodiversity Information: From Sherborn to the 21st century and beyond. ZooKeys 550: 83–106. doi: 10.3897/zookeys.550.10073

Abstract

C. D. Sherborn published in 1940, under the imprint of Cambridge University Press but at his own expense, *Where is the – Collection?* This idiosyncratic listing of named natural science collections, and their fates, was useful, but incomplete, and uneven in its accuracy. It is argued that those defects were inevitable, given Sherborn's age and wartime conditions, and that what might seem one of Sherborn's less impressive works was in fact a pioneering work highly influential in stimulating the production of successor works now much used in curation, and in systematic and descriptive biology and palaeontology. The book also contributed to the development of collections research in the natural sciences, and the history of collections and of museums.

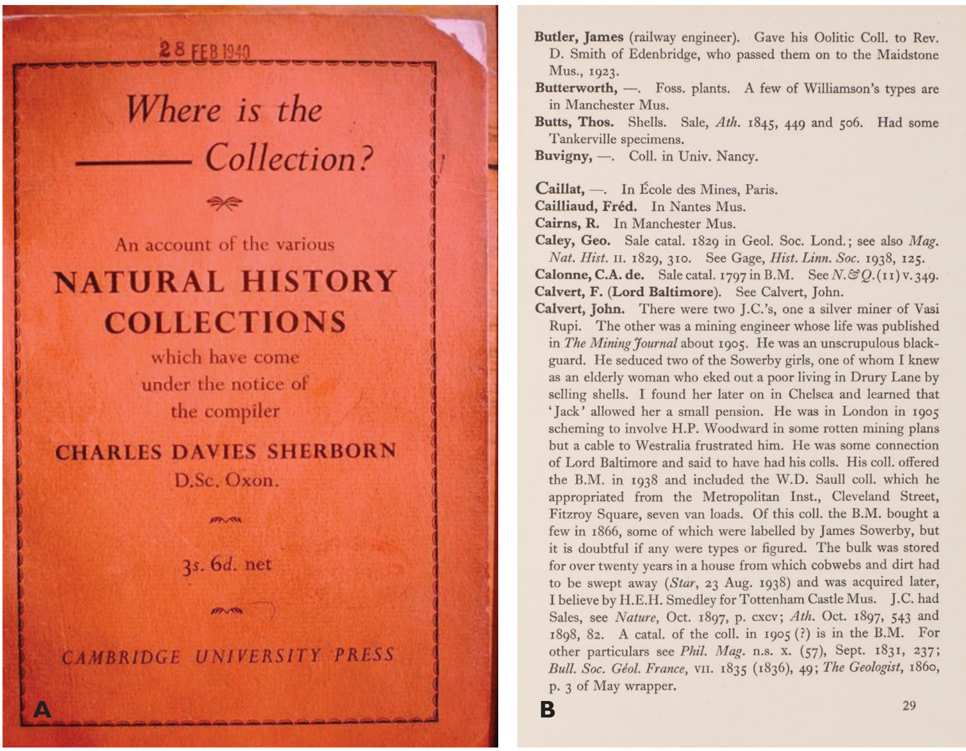
Keywords

Charles Davies Sherborn, collections, geology, biology, taxonomy, museums

Introduction

Charles Davies Sherborn (1861–1942) was a geologist and above all a scientific bibliographer (Anonymous 1942b, [Hinton] 1942, Norman 1942, 1944, Griffin 1953, Cleevely 2004, Dickinson 2016, Evenhuis 2016, Shindler 2016, Welter-Schultes et al. 2016). His last significant publication was a small book called *Where is the – Collection?* (Fig. 1a, b; Sherborn 1940). This paper describes the book's genesis and content, and assesses its significance at the time, the value of its contained information, and its importance as a precedent and nucleus for systematics, curation and collections-historical research in the natural sciences.

Sherborn did not explicitly give his reasons for writing the book. It is evident from his introduction that the aim was to help researchers, and especially systematists, locate named collections, and thereby particular specimens: the important point is that the collections were named. The dash in the title is usually taken as standing for the name of the relevant collection, but Sherborn once privately called his book "*Where is the damned Collection?*" (Norman 1944, p. 81), and one reviewer commented that "The difficulty of discovering the resting place of some important specimen [...] doubtless justifies the 'blue-pencilled' word which the author may or may not have hinted at in the title" (Ritchie 1940, p. 80; "blue-pencilled" here means censored as an expletive).



Butler, James (railway engineer). Gave his Oolitic Coll. to Rev. D. Smith of Edenbridge, who passed them on to the Maidstone Mus., 1923.
Butterworth, —. Foss. plants. A few of Williamson's types are in Manchester Mus.
Butts, Thos. Shells. Sale, *Ath.* 1845, 449 and 506. Had some Tankerville specimens.
Buvigny, —. Coll. in Univ. Nancy.

Caillat, —. In École des Mines, Paris.
Cailliaud, Fréd. In Nantes Mus.
Cairns, R. In Manchester Mus.
Caley, Geo. Sale catal. 1829 in Geol. Soc. Lond.; see also *Mag. Nat. Hist.* II. 1829, 310. See Gage, *Hist. Linn. Soc.* 1938, 125.
Calonne, C.A. de. Sale catal. 1797 in B.M. See *N.&Q.* (11) V. 349.
Calvert, F. (Lord Baltimore). See Calvert, John.
Calvert, John. There were two J.C.'s, one a silver miner of Vasi Rupi. The other was a mining engineer whose life was published in *The Mining Journal* about 1905. He was an unscrupulous blackguard. He seduced two of the Sowerby girls, one of whom I knew as an elderly woman who eked out a poor living in Drury Lane by selling shells. I found her later on in Chelsea and learned that 'Jack' allowed her a small pension. He was in London in 1905 scheming to involve H.P. Woodward in some rotten mining plans but a cable to Westralia frustrated him. He was some connection of Lord Baltimore and said to have had his colls. His coll. offered the B.M. in 1938 and included the W.D. Saull coll. which he appropriated from the Metropolitan Inst., Cleveland Street, Fitzroy Square, seven van loads. Of this coll. the B.M. bought a few in 1866, some of which were labelled by James Sowerby, but it is doubtful if any were types or figured. The bulk was stored for over twenty years in a house from which cobwebs and dirt had to be swept away (*Star*, 23 Aug. 1938) and was acquired later, I believe by H.E.H. Smedley for Tottenham Castle Mus. J.C. had Sales, see *Nature*, Oct. 1897, p. cxcv; *Ath.* Oct. 1897, 543 and 1898, 82. A catal. of the coll. in 1905 (?) is in the B.M. For other particulars see *Phil. Mag.* n.s. x. (57), Sept. 1831, 237; *Bull. Soc. Géol. France*, VII. 1835 (1836), 49; *The Geologist*, 1860, p. 3 of May wrapper.

29

Figure 1. A The cover of *Where is the – Collection?* **B** A sample page (p. 29) from *Where is the – Collection?*, including the entry for John Calvert. The Sowerby women are thought to be the daughters of G. B. Sowerby I (1788–1854) (R. J. Cleevely, pers. comm. 2014).

Sherborn's book was, strictly speaking, not the first listing of collections. Cleevely (1983, p. 9) records sporadic lists of collections published as early as 1812, and notes the presence of collections location data in a listing of geologists in the *Fossilium Catalogus* series (Lambrecht et al. 1938). Sherborn himself noted (p. 5) a prior listing of fossil insect collections, and the listing of some British collections in his own *Catalogue of British fossil Vertebrata* (Woodward and Sherborn 1890).

One of the most important early general works on fossils was Sowerby's *Mineral Conchology* (Cleevely 1983, pp. 7, 9, 11, 14–16). Sherborn (1935) published a paper that listed all the collectors mentioned by the Sowerbys in this work (1812–1846); he cited the references that helped to identify the 237 collector / collections listed. This was a particularly interesting meta-analysis as 28 of those collectors were women, and often very significant participants, such as Etheldred Benett (1776–1845), contrary to the impression one sometimes gets from the secondary and popular literature of today that Mary Anning was almost the only female collector in this period. Certainly Sherborn's listing of the collectors who provided material for the Sowerbys must have been an important preliminary stage of compiling *Where is the – Collection*.

However, in its wide scope, *Where is the – Collection?* was for decades unique as a practical reference which listed such information on named natural sciences collections and their fates as he had come across in his decades of work at the British Museum. Sherborn's interests meant that the emphasis was on palaeontological and malacological collections, mainly in Britain, with a sprinkling of other categories such as mineralogy, ornithology, and botany, and manuscripts. Sherborn also commented on collections which had been destroyed, for instance by fire or flood.

Methods

In this paper, for space reasons, and because they feature strongly in Sherborn's book, I use palaeontological collections as my main examples, but in fact similar developments occurred across the entire field of natural science collections. Sherborn's book was a listing of named collections rather than an institutional directory, so I here use "collection" in the sense of a collection of specimens made by a named person or body, rather than the holding institution as a whole. Admittedly this definition is still ambiguous; for instance, it includes both field and cabinet collectors (cf. Torrens 2006, Lucas and Lucas 2014). Sherborn did not attempt to produce a listing of institutions either directly, or indirectly by indexing, and I therefore do not cover lists of institutions in detail (but do refer to them when relevant). However, Sherborn did include some institutional collections, especially when they had been transferred and dispersed amongst other institutions: effectively, they became collections under the name of the original institutions. The modern equivalents of 1940 values are determined using the Bank of England inflation calculator (URL: http://www.bankofengland.co.uk/education/pages/inflation/calculator/flash/default.aspx, accessed 23 January 2014).

References, archives and repositories: where only pagination is given in a reference, Sherborn's book (1940) is intended. "British Museum", in the usual shorthand of Sherborn's time, here denotes the British Museum (Natural History), London, now the Natural History Museum. Repository abbreviations: BMNH, British Museum (Natural History), now NHM; CUL, Cambridge University Library, West Road, Cambridge CB3 9DR, England; Cambridge University Press, University Printing House, Shaftesbury Road, Cambridge CB2 8BS, England; NHM, Natural History Museum, Cromwell Road, London SW7 5BD, England.

Origin and content

In a guide to sources for collections research in the first *Newsletter* of the new Geological Curators Group (GCG), Hugh Torrens called Sherborn's book "the only primary source on collections known to me", and described it admirably (Torrens 1974, pp. 12–13):

> [The book has] 149 pages but every other one is blank to allow annotation. [...] It is scarce only 500 copies having been printed. This is an account of the various Natural History Collections which Sherborn came across between 1880–1939. It is not exhaustive or always accurate but contains an immense amount of information. Furthermore it is often fascinating reading. [...] His biography by J. R. Norman [1944] [...] is also equally entertaining reading. His primary interests were geological and palaeontological so there is a useful [for GCG members] bias towards these collections in his book. [...] it had amazingly to be published at his own expense.

Sherborn said that the book contained "facts accumulated over sixty years in answer to inquiries", and that its "original MS" had "been on my table at the British Museum (Natural History) and of daily use to the Staff or others" (p. [5]). Norman (1944, 80–81) added that "much of his material was collected from old sale catalogues, biographies, obituary notices, museum guides" and the like. No doubt much of Sherborn's information came as a by-product to his work on *Index Animalium*, but Sherborn plainly carried out additional research, as shown by his file of MSS notes and clippings (still in the BMNH libraries in the 1970s, R. J. Cleevely 1983, and pers. comm. 2014). Examples are Sherborn's inquiries for John Phillips's fossils (see below), and his searches through journals such as *Gentleman's Magazine* and *Notes and Queries*, of which he bought 143 volumes for the purpose (Norman 1944, p. 81).

Sherborn completed his literature searches around March 1939, and in due course finished his manuscript and sent it to the Museums Association. He later reported the disappointing results to a friend in a letter of 27 December 1939:

> ... that astute body [the Museums Association] hummed over it for two months, and this tho' I offered to pay for it, that I sent for the MS. back, [and] sent it on to the Cambridge [University] Press [...] (Norman 1944, p. 81).

Sherborn already had an excellent relationship with the Press, who reportedly called him the "best editor" with whom they had ever worked (Norman 1944, p. 79). On 22 November 1939 he wrote to them, evidently as part of an ongoing discussion (CUP archive, CUL UA Pr.A.S.429):

> By the by, you might say if you would undertake the publishing, I to pay cost of printing, to keep say fifty copies and give you the remainder if you pay me say 1/6 on all sold copies. This is only a suggestion as I shall want some publisher on the T[itle]. P[age]. and would rather you than anyone.

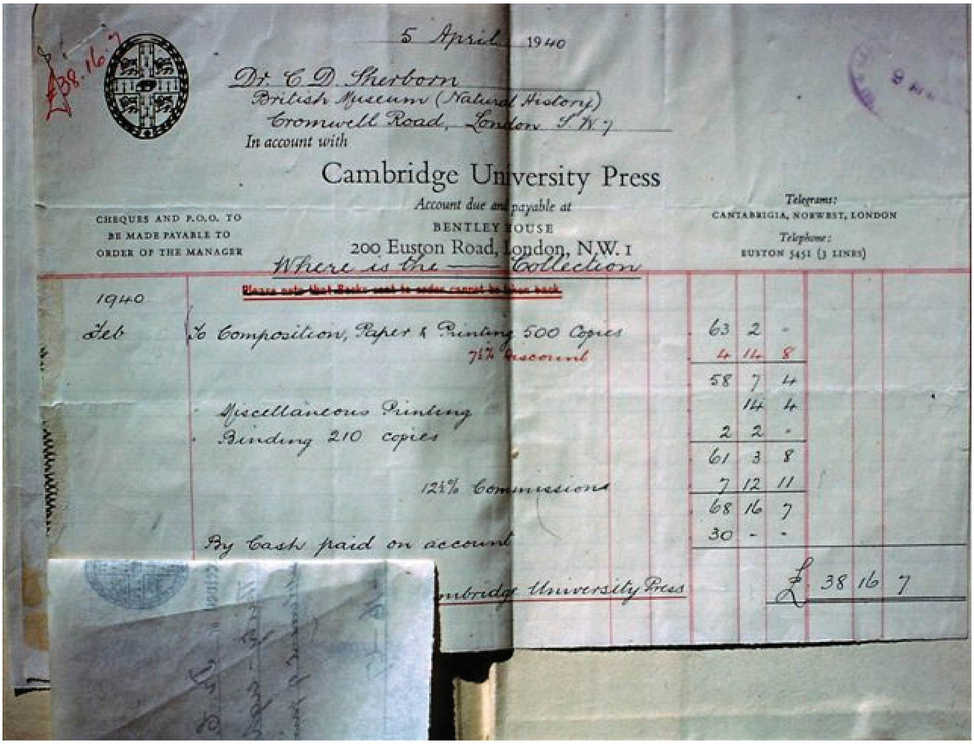
Sherborn suggested a price of 3/6 or 4/6 (3/6, three shillings and sixpence in pre-decimal United Kingdom currency, is nominally equivalent to 17.5p today but then worth much more). The Press Syndicate decided at its meeting of 8 December 1939 to "undertake the publication of his proposed catalogue of Natural History References, on commission" (CUL UA Pr.V.82, Syndicate Minutes for 1935–1939). The standard 'Memorandum of Agreement', i.e. a printed contract for printing and publishing the book at his expense, survives in the Contract Archive at CUP (K. Thompson, Brand Protection Officer, CUP, pers. comm. 2014), bearing Sherborn's MSS annotations. Sherborn evidently returned it with a covering letter of 12 December (CUL UA Pr.A.S.429). Amongst other matters, he confirmed a print run of 500 copies of which he was to have 50, and suggested that the *Times Literary Supplement* and Nature were "the only papers likely to be of advertising value of such a book, but I leave it to you". He specified binding in paper: "I cant afford the cloth. Rest of cash available for you when asked for, do not increase it more than you can help for this is a bit of an effort on my part."

Sherborn soon reported to his friend in the letter of 27 December 1939 cited above:

> [...] Cambridge Press [...] accepted my terms, set it up at once, in ten days the whole proofs went back to Cambridge, and it will be printed and ready by mid-January. Cost me £70, sells at 3/6, 500 copies. So that's that. (Norman 1944, p. 81).

He would need to sell 80% of all copies to recoup his £70, even ignoring other costs (which apparently included 12.5% commission to CUP specified in the Memorandum). The risk was not trivial as £70 was equivalent to almost £4000 in 2012 values. But to call it 'a bit of an effort' surely reflected his habitual economy rather than actual poverty, as he was relatively well off (Norman 1944, but see Shindler 2016). After Sherborn's death, his estate would be valued at £11,619; his executors later sold the copyright and all rights in the stock of the book to CUP for £5 5s (Anonymous 1942a; *England & Wales, National Probate Calendar (Index of Wills and Administrations), 1858–1966*, probate granted at Llandudno, 5 September 1942; receipt attached to Memorandum of Agreement in CUP Contract Archive).

The Press Syndicate Minutes for 2 February 1940 report that the agreement with Sherborn was 'completed', whatever that meant (CUL UA Pr.V.82). The printing and



£38.16.7
5 April 1940
Dr. C. D. Sherborn
British Museum (Natural History)
Cromwell Road, London S.W.7
In account with
Cambridge University Press
Account due and payable at
BENTLEY HOUSE
200 Euston Road, London, N.W. 1
CHEQUES AND P.O.O. TO BE MADE PAYABLE TO ORDER OF THE MANAGER
Telegrams: CANTABRIGIA, NORWEST, LONDON
Telephone: EUSTON 5451 (3 LINES)
Where is the — Collection

1940					
Feb	To Composition, Paper & Printing 500 Copies	63	2	-	
	7½% discount	4	14	8	
		58	7	4	
	Miscellaneous Printing		14	4	
	Binding 210 copies	2	2	-	
		61	3	8	
	12½% Commission	7	12	11	
		68	16	7	
	By Cash paid on account	30	-	-	
	Cambridge University Press	£ 38	16	7	

Figure 2. The final bill from Cambridge University Press, tipped into a copy of *Where is the – Collection?* in the NHM Library. Photo courtesy R. J. Cleevely, NHM.

binding were in any case done in time for the final bill, dated 5 April 1940, which came to just under the expected £70 (Fig. 2). The book was out in time to be reviewed in the 20 July 1940 issue of *Nature* (Ritchie 1940). An initial search of the CUP archives and of sales catalogues for the period has not turned up any record of an official publication date (R. Grooms, CUP Archivist, CUL, pers. comm. 2014), but this may simply reflect the book's anomalous nature, the lack of advertising, and the wartime conditions.

Sherborn's letter of 27 December gives the impression that he withdrew his book from the Museums Association because of the Association's dilatoriness, but he does not give any reason for this delay. It is possible that the Association had reservations about the book itself, especially if it had the book assessed by the same person who later reviewed it harshly for the Association's *Museums Journal* (quoted below, "C. M.", 1940). Another possibility is that the Museums Association feared that some entries were defamatory. Sherborn was seemingly inclined to unrestrained criticism at times. In 1888 the Royal Society of London refused him support for his bibliography of foraminiferans because of his savage review of a foreign rival (Miller 2016, Shindler 2016). In a 1905 letter to Arthur Smith Woodward, then Keeper of Palaeontology, Sherborn described the British Museum's Chalk echinoderms as "disgraceful material" (NHM Archives DF 100/39/256; P. M. Cooper, pers. comm. 2014). Sherborn was just as forthright in his little book, noting for instance that Professor W. J. Sollas (1849–1936) “destroyed the *Phascolotherium* jaw” of an exceptionally rare Jurassic mammal in the James Parker collection at Oxford (p. 105). The University Museum’s curators have never forgiven Sollas, who was trying to study the jaw’s internal structure through serial sectioning, a crude and inherently destructive predecessor of computed tomography (Vincent 1994, pp. 28–29, 39).

As part of the standard agreement, Sherborn had to indemnify Cambridge University Press for any libel or copyright claims, and in his letter of 12 December (CUL UA Pr.A.S.429) he said, “[...] please read items Groom and Calvert. All parties are long since dead and my remarks are historically valuable and should stand if possible.” Charles Ottley Groom (1839–1894), an impostor who called himself the Prince of Mantua and Monferrat, went by the Scottish lairdly title of Napier of Merchiston (Davenport-Hines 2004). Sherborn described him as a “notorious rogue and thief, tried to kill Thomas Davies [presumably the geologist (1837–1912)] by dropping a boulder upon him from a high ladder in Tennant’s shop in the Strand” (p. 63). And Sherborn’s account of John Calvert (1814–1897), fraudster, traveller, self-proclaimed mining expert, and mineral collector, was the “most notorious entry in his otherwise genial catalogue of collectors” (Fig. 1), stimulating later research by several historians, according to Cooper (2006, pp. 86–87). In fact, Sherborn seems to have attributed to John some of the doings of Albert (1872–1946), John’s also unscrupulous and then still alive grandson - or son: the Calverts were never too clear about this (Birman 1979, Rothwell 2010, Cooper 2006, pp. 85–105). The Press took the precaution of sending the original proofs of those two entries (but nothing else) to Field Roscoe & Co., a London firm of solicitors, on 1 January 1940, and the firm replied the next day, saying that “there is no doubt that the passages in question are defamatory”. But there was no problem if Calvert and Groom were dead, so long as a small change was made to avoid giving the inadvertent imputation of dubious behaviour on the part of one W. G. Ball, who had been selling Calvert’s material on behalf of another firm of solicitors lumbered with it in lieu of a bad debt (CUL, CUP archive, UA Pr.AS.429; Cooper 2006, p. 101). Sherborn was evidently willing to alter his text, for Mr Ball does not appear in the book.

Sherborn’s comments on some museums obviously did not worry the Press, even though English libel law allows corporations to sue. But they might have created a sticking point for the Museums Association, because some of its institutional members were mentioned unflatteringly. He referred to “the way this local museum [Liverpool Museum] has treated types”, Elgin Museum as “a dump of useful stuff uncared for”, and Wilson’s insects “in Perth Mus[eum]. in ‘shocking state’”, and even the Stebbing collection in the British Museum itself where “most of the spirit had evaporated and specimens were practically useless” (pp. 11, 49, 111, 127, 145; his Perth seems to be that in Scotland rather than Australia, from the admittedly incomplete match of other ‘Perth’ entries with Stace et al. 1987).

However, it seems just as likely, if not more so, that the Museums Association’s real problem with the book lay in its timing. Time must have pressed grievously on

Sherborn while he sought to publish this last work of any substance. He was, from 1934, unable to work for long periods, and was becoming increasingly aged and unwell, suffering significant deterioration in 1938, and an episode of poor health in the winter of 1938–1939. He now had to cope with the outbreak of a war whose likelihood he had professed not to take seriously (Norman 1944). The British Museum's staff and facilities were already being diverted to wartime priorities, and Sherborn would know very well from his Great War experiences how severe such disruption could become. If Sherborn and the Cambridge University Press were as efficient as they seem to have been, then the manuscript was presumably with the Museums Association during September and October 1939, give or take a few weeks either way: in other words, the period of final mobilisation, the declaration of war on 3 September, and the first few weeks of war. The Museums Association would have been hugely distracted by the problems which the war posed for its members and itself, and Sherborn's book would have seemed a very low priority. Perhaps the Association never even got as far as actually considering Sherborn's proposal.

In hindsight Sherborn was wise to take the initiative by abandoning the Museums Association, and pushing through the book's rapid publication elsewhere. His sister died in January 1940, he developed heart disease at the end of 1940, and his last years were a time of increasing wartime disruption at both home and the British Museum, especially after the destructive air raids from September 1940 and the closure of the libraries in 1942, the year of his own death (Norman 1944).

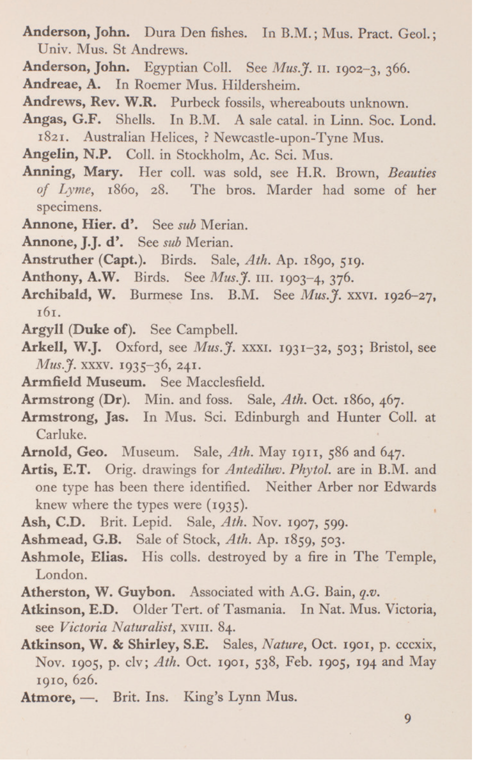
Assessment

Sherborn's book was uneven, with the biases in subject content already noted. It was organized only by collector, without any indexing by holding institution. It was inadequately edited. The brevity of its sometimes cryptic entries, with inconsistent names and abbreviations for the Royal Scottish Museum, for instance, annoyed the *Nature* reviewer (Ritchie 1940) – surely James Ritchie (1882–1958), Professor of Zoology at the University of Edinburgh, and previously Keeper of Natural History at that same museum. Doughty (1984, p. 160) described the book as "sometimes a little obscure, veiled [...] in the pedant's sophist[r]y and waggishness". I am more inclined to ascribe this to the book's origin as a collection of notes which acted as personal memory-joggers. The entry for Street Museum in Somerset (now the Alfred Gillett Trust) (p. 129) says in part "Nothing of value except the *Ichthyosaurus* (E. I. White 1934). Wallis of Bristol had a pick of specimens and books". Sherborn obviously knew what this meant, but the reader needs some background knowledge to conclude that, presumably, Errol White (1901–1985), palaeontologist at the British Museum, had made comments to Sherborn about a visit in 1934, and that Dr F. S. Wallis (d. 1979), Director of Bristol City Museum, had helped with a partial dispersal. Sometimes the reader is left tantalised. "Weeks – (formerly Cox). Had the mechanical spider" (p. 141) actually refers to a popular automaton in the museum of mechanical curiosities in Tichborne Street, London, ca. 1803–1835, assembled by a person named Weeks and stemming, at least in part, from the 18th century collection of James Cox (c. 1723–1800) (Coleman et al. 1902, Altick 1978, Smith 2008). And despite recent studies (Hodgkinson 2006, Miller 2016), one is left in the dark as to why the foram worker Fortescue W. Millett (1833–1915) "kept his rare books in the W. C. under the seat" (p. 97). Did Sherborn mean that Millett had an ultra-superior throne carefully integrated into the room's wood panelling, with convenient bookshelves designed in? Or was this a triple pun on rare, rear (as in backside), and rears (as in English "public school" slang for lavatories)?

The book, as Sherborn himself admitted, was "not exhaustive; that were too much to expect and almost an impossibility" (p. [5]). Nor is the book particularly reliable in detail (Torrens 1974, pp. 12–13, Cleevely 1983, p. 9). Sherborn stated (p. 107) that some of the fossil collection of John Phillips (1800–1874) was stolen and dumped in the River Thames, but this is now known to be an exaggeration of another author's canard - though he did take the trouble to check the Blackfriars Bridge engineers' records (Torrens 1975b, Edmonds 1977, Cleevely 1983, p. 231, Nikolaeva and Morgan 2010). The collection of the Wernerian geologist and mining engineer Thomas Weaver (1773–1855), "used to form hard core of a urinal at Bewdley" (p. 141), seems in fact to have been the unwanted residue after sales and donations to museums (Torrens 2004).

There is some evidence that Sherborn simply decided to stop work and go to print, rather than delay any longer, even if it meant cutting corners. He and his friend W. D. Lang (1878–1966) both cited a relatively unusual source for Mary Anning (1799–1847), a Lyme Regis guidebook (p. 9, Fig. 3 here; Brown [1857], Lang 1950, Taylor and Torrens 2014). Yet Sherborn's entry completely fails to mention her numerous and important Liassic vertebrates in the British Museum. It is true that she did not amass much of a personal collection, being a commercial collector who sold her finds, but this cannot be the reason as the same omission recurs in the entry for Thomas Hawkins (1810–1889) (p. 67, Fig. 4 here)(Torrens 1995, Evans 2010). This suggests that Sherborn did not systematically collate his manuscript with the official history of the British Museum's collections (British Museum (Natural History) 1904–1912), or have it read over by colleagues such as Lang and the palaeoherpetologist W. E. Swinton (1900–1994). Perhaps this was because of wartime disruption. However, the *Museums Journal* reviewer noted other examples of Sherborn's failure to collate information from other publications, even ones which Sherborn had cited ("C. M." 1940).

Wartime conditions surely meant that the book received fewer reviews and notices than it might otherwise have had. Even the *Journal of the Society for the Bibliography of Natural History*, co-founded by Sherborn, did not print one till 1943 (Anon. 1943). Reviewers generally noted the book's incompleteness and, to some extent, unreliability, while focussing on their own areas of expertise. The *Quarterly Review of Biology* reviewed it in 1940 (Anon. 1940), and the ichthyologist George S. Myers (1905–1985) of Stanford University assessed it, sympathetically but briefly, in *Copeia* (Myers 1941). Sherborn's friend Frederick Chapman (1864–1943), of the National Museum, Melbourne, discussed mainly Australian collections in another sympathetic review in *Victorian Naturalist* (Chapman 1942). A hatchet job in the *Museums Journal* was perhaps



Anderson, John. Dura Den fishes. In B.M.; Mus. Pract. Geol.; Univ. Mus. St Andrews.
Anderson, John. Egyptian Coll. See *Mus.J.* II. 1902–3, 366.
Andreae, A. In Roemer Mus. Hildersheim.
Andrews, Rev. W.R. Purbeck fossils, whereabouts unknown.
Angas, G.F. Shells. In B.M. A sale catal. in Linn. Soc. Lond. 1821. Australian Helices, ? Newcastle-upon-Tyne Mus.
Angelin, N.P. Coll. in Stockholm, Ac. Sci. Mus.
Anning, Mary. Her coll. was sold, see H.R. Brown, *Beauties of Lyme*, 1860, 28. The bros. Marder had some of her specimens.
Annone, Hier. d'. See *sub* Merian.
Annone, J.J. d'. See *sub* Merian.
Anstruther (Capt.). Birds. Sale, *Ath.* Ap. 1890, 519.
Anthony, A.W. Birds. See *Mus.J.* III. 1903–4, 376.
Archibald, W. Burmese Ins. B.M. See *Mus.J.* XXVI. 1926–27, 161.
Argyll (Duke of). See Campbell.
Arkell, W.J. Oxford, see *Mus.J.* XXXI. 1931–32, 503; Bristol, see *Mus.J.* XXXV. 1935–36, 241.
Armfield Museum. See Macclesfield.
Armstrong (Dr). Min. and foss. Sale, *Ath.* Oct. 1860, 467.
Armstrong, Jas. In Mus. Sci. Edinburgh and Hunter Coll. at Carluke.
Arnold, Geo. Museum. Sale, *Ath.* May 1911, 586 and 647.
Artis, E.T. Orig. drawings for *Antediluv. Phytol.* are in B.M. and one type has been there identified. Neither Arber nor Edwards knew where the types were (1935).
Ash, C.D. Brit. Lepid. Sale, *Ath.* Nov. 1907, 599.
Ashmead, G.B. Sale of Stock, *Ath.* Ap. 1859, 503.
Ashmole, Elias. His colls. destroyed by a fire in The Temple, London.
Atherston, W. Guybon. Associated with A.G. Bain, *q.v.*
Atkinson, E.D. Older Tert. of Tasmania. In Nat. Mus. Victoria, see *Victoria Naturalist*, XVIII. 84.
Atkinson, W. & Shirley, S.E. Sales, *Nature*, Oct. 1901, p. cccxix, Nov. 1905, p. clv; *Ath.* Oct. 1901, 538, Feb. 1905, 194 and May 1910, 626.
Atmore, —. Brit. Ins. King's Lynn Mus.

9

Figure 3. A sample page (p. 9), including the entry for Mary Anning. This misses her specimens in the British Museum (Natural History).

written by Claude Morley (1874–1951), an entomologist linked to Ipswich Museum ("C. M." 1940, p. 73): "the book cannot be described as one which the enquirer may consult with a reasonable probability of finding the whereabouts of any particular collection, even a large one". There was some truth in this; in *Nature*, Ritchie (1940) noted the omission of important collections in his own Royal Scottish Museum, including that of Hugh Miller (1802–1856) – though, despite Ritchie's comments, Sherborn was right to mention Miller fossils in Cromarty (Waterston 1954). The CUP archive contains an album with reviews and notices pasted in, presumably recording those known to the Press (CUL, Cambridge University Press newspaper clippings S-1940). Apart from those already mentioned, it contains pieces from *Extraits de la Revue des Questions Scientifiques* (no details, in English); *Science Newsletter*, 17 August 1940; *Biological Abstracts*, Vol. 15, No. 5, 1941; *Ciencia*, Vol. 1, No. 2, 1941 (in Spanish); and *Mexican Society of Natural History*, No. 4, Vol. 1, 1941 (in Spanish).

Where is the – Collection? might, at first sight, seem an anticlimactic end to Sherborn's career, and the least impressive of his works especially when compared to his 11-volume *Index Animalium*. It was, of course, a work of its time. Given Sherborn's age and the war, he had to publish what he had when he did, or not at all. A separate

Hartley Institute. See Southampton.
Hartmann, J.D.W. Shells in St. Gallen Mus.
Haslar Hospital. All coll. came to B.M. in 1855.
Hastings, Barbara (Lady). Tert. Hampshire and Isle of Wight. Bought by B.M. 1855. Sale, *Ath.* 1855, 539; Catal. in Geol. Soc. Lond.
Haviland, Thos. Portrait by Gainsborough in Fitzwilliam, Cambridge, has two Ammonites on table. Mentioned in Pennant's Catal. p. 339 as of Bath. No further information.
Hawke, Wm. Lepid. Sale, *Ath.* July 1899, 50; *Nature*, July 1899, p. xciii.
Hawkins, H.H. Eggs. Sale, *Ath.* March 1898, 327, May 1902, 642.
Hawkins, H.J. Eggs. Sale, *Nature*, March 1898, p. cxlvi.
Hawkins, Herbert S. Birds and Eggs. Sale, *Ath.* June 1867, 775.
Hawkins, John. Min. Sale, *Nature*, Ap. 1905, p. ccxxiv.
Hawkins, S.J. In own hands.
Hawkins, Thos. Foss. Rept. Sale, *Ath.* 1844, 585; *Lond. Geol. Journ.* 1846, 16.
Haworth, A.H. Coll. See *Mag. Nat. Hist.* VII. 1834, 192. Warburton had many of H.'s types.
Hayes, Paddy. Local Sheppey collector about 1840. See *Life of Shrubsole* in Gen. Liby. B.M. 1927.
Headley, W. (or Hedley). Bridlington Crag Shells. Passed into G.W. Lamplugh's Coll.; mixed up and not now separable, 1938.
Heath, T. Cave remains, Owens Coll.
Hébert, E. In Sorbonne, Paris.
Hedin, Sven. Colls. See *Mus.J.* II. 1902–3, 193.
Heer, Oswald. Much of his coll. went to van Breda of Haarlem.
Helps, J.A. Ins. Sale, *Ath.* May 1906, 627.
Hemming, — (of Brighton). Ins. Sale, *Ath.* March 1857, 295.
Hennah, — (of Newport). Shells. Sale, *Ath.* Ap. 1885, 427.
Hennah, Rich. Corals in Geol. Soc. Lond. were lent to W. Lonsdale and mostly returned to the Soc. and placed 'in a heap' in a drawer in top room. A few types identified in 1934 and now in B.M.
Henslow, J.S. Sale, *Ath.* May 1861, 680, July 1861, 35. Catal. in Geol. Soc. Lond.
Herbst, J.F.W. Crabs in Berlin.

A 67

HAWKINS, Thomas, FGS
(1810–15 Oct. 1889)
Lived at Sharpham Park, near Glastonbury. Collected skeletons of Fossil Reptilia from Lower Lias of Somerset and Dorset and was the first to begin to extract them with some care and method (see *Proc. Dorset Nat. Hist. Archaeol. Soc.*, 1938: 157–9). De la Beche commented on his wholesale collecting: 'Why, you carry away whole quarries!' Hawkins made several fossil reptile collections and figured his more important specimens in: *Memoirs on Ichthyosauri* (1834) and *The Book of the Great Sea-Dragons* (1840). A collection of reptiles was sold at J. C. Stevens on 25 July, 1844 (see *Athenaeum*, 1844: 585).

Q. Jl. Geol. Soc., Lond., **46**, 1890: Proc.: 48; see extract concerning Hawkins' collection in Investigation of British Museum Parliamentary Report published in *Mag. Nat. Hist.*, N.S., **4**, 1840: (Appendix to No. 39) pp. 11–44; see account of collection of ichthyosaur in Mantell, *Petrifications and their Teachings* (1851): 379.

Cambridge, Sedgwick Museum. Presented Liassic Reptilia in 1856.
London, BM(NH). Purchased collection of specimens figured in his two books in 1834 (see Mantell, *Lond. Geol. Jl.*, **1846:** 16, and Mantell, 1851: 490). Further collection purchased in 1840.
Oxford, Univ. Museum. Also acquired some Fossil Reptilia.

B

Figure 4. Differing treatments of Thomas Hawkins. **A** The original Sherborn entry in *Where is the – Collection?* (p. 67) **B** The much more extensive entry in *World Palaeontological Collections* (Cleevely 1983, p. 147).

issue is that for Sherborn to do much better would have involved the organization of a systematic questionnaire, well beyond the energy and resources of a single elderly worker (Dickinson 2016, Evenhuis 2016, Shindler 2016, Welter-Schultes et al. 2016). Moreover, such a questionnaire would have been pointless even if Sherborn could obtain major institutional support. Too many potential target museums were disrupted during the war, even if they did not end up being targets of another kind. Such collections research is, in any case, decentralised by its nature, dealing with collections, documents and archives in many places: far beyond Sherborn on his own in London. It would not help that collections research is naturally more chaotic by nature than Sherborn's more familiar bibliographical-taxonomic work. A single collection can end up in many places thanks to the vagaries of the owner's swaps, gifts, sales, and bequests, and then of the holding institutions. For instance, Sherborn (p. 97) failed to note that a significant part of the fossil collection of Charles Moore (1815–1881) at the Royal Literary and Scientific Institution, Bath, had been transferred to Taunton by the honorary curator Rev. H. H. Winwood (1830–1920) (Copp et al. 2000). (Sherborn also stated that the collection at Bath was being cared for by Winwood in 1925: "not an easy task for a man dead 4 years", dryly noted Torrens 1975a, p. 113).

Some of Sherborn's information, such as the story of Groom and Davies, plainly came unattributably from colleagues, probably losing precision and introducing error along the way, but with a core of truth, as is the way of oral history. This is perhaps how he knew that A. M. B. Anderson of Brighton was in fact a later alias for Alexander Montagu Browne (1837–1923), curator of the New Walk Museum, Leicester, and a major figure in the history of British taxidermy (p. 7). Rather disappointingly, however, Sherborn failed to confirm the oral tradition amongst successive Leicester curators (including J. A. Cooper and M. Evans, pers. comm. 2014, and MAT) that Montagu Browne was sacked for running a brothel round the corner from the museum; the actual, or at least official, reason was a disagreement with the museum committee over his curatorial training scheme, and perhaps also the museum's modernisation (McCann 1981, Morris 2010, pp. 339–342). Nevertheless the entry reminds us that Sherborn's book remains a worthwhile source today, so it is unfortunate that neither the book, nor his biography by Norman (1944), are fully available on the internet today. Might not one of the annotated copies of Sherborn's book at NHM be made available on Biodiversity Heritage Library?

Despite its problems, Sherborn's book was the only one of its kind, and a great deal better than nothing. Most importantly of all, Sherborn and some (but not all) of his contemporaries appreciated that his book was simply a starting point, an initial stage towards something better, as implied by its publication with every other page left blank. I now turn to the issue of its long–term influence.

Sherborn's successors: collections research

A key reason for the rise of the specialist Geological Curators Group (GCG) in Britain and Ireland in the 1970s was the realisation that much needed to be done to improve the quality of museum work in geology (Doughty 1999, Knell 2002). Much of the Group's attention was devoted to issues of collection care and usage, and specimen conservation.

Survey work was done to find which institutions housed geological material, and the state of these collections and their usage. This work led to publications listing these institutions and analysing the resulting data, notably the classic "State and Status" survey conducted by Phil Doughty (1937–2013) (Doughty 1981, 1999, and more recently Nudds 1999 and Fothergill 2005). Although not collections listings in the Sherborn sense, they often gathered useful information of this kind. Other important examples more globally are Glenister et al. (1977), Prieur (1980), and Webby (1989, 1992).

Under the influence and example of such workers as Hugh Torrens, GCG encouraged research on the history of collections, for it was realised that this had to be understood before a collection could be properly curated and used (Doughty 1984, 1992, 1999, Knell 2002). Such work by Group members and others elucidated, amongst other things, the fates and present locations of collections, and effectively followed on from Sherborn. Some of this research was published as books, such as that which Andrews (1982) wrote specifically to locate certain fossil fish specimens published by Louis Agassiz (1807–1873). But a significant proportion appeared in the *Newsletter of the Geological Curators Group*, latterly titled *The Geological Curator*. This was, and remains, a collective work in progress, with the "Lost and Found" column providing for collections inquiries and for short pieces on new information that does not justify a whole article. This corpus is now available on www.geocurator.org.

Parallel developments took place for biological collections under the aegis of the Biology Curators Group, with its own journals such as the *Biology Curators Group Newsletter*. The Group is now part of the Natural Science Collections Association (NatSCA; an increasing proportion of the older publications are accessible on www.natsca.org).

Sherborn's successors: collections reference books

Ron Cleevely of the British Museum became interested in gathering information on collections in the early 1970s, with the intent of producing a new revision of Sherborn's book, stemming originally from the need to locate type material to support the work of Leslie R. Cox (1897–1965) for the *Treatise of Invertebrate Palaeontology*, and using the data in an annotated copy of Sherborn's book in the Fossil Mollusca Section. Cleevely developed the book using links with the Society for the Bibliography of Natural History, and with the then new Geological Curators' Group, including survey data from Doughty's 'State and Status' work and an earlier survey by Douglas Bassett of the National Museum of Wales in 1966–1967 (Anon. 1972, Torrens 1974, Cleevely 1977, 1983, especially introductory essays, Hancock and Pettitt 1981, Pettitt and Hancock 1981, R. J. Cleevely, pers. comm. 2014). Fortunately the Museum's management recognised the value of this project and Cleevely was able to spend official time on it. He had originally simply intended a more modern version of Sherborn's effort, and, like it, inexpensive with alternate blank pages. However, its formal adoption by the Museum, and a management decision, led to its publication as a markedly more substantial and more expensive project.

Cleevely's book *World Palaeontological Collections* provided far more detail than Sherborn, and on many more collections (Cleevely 1983). It systematically incorporated references to collectors' obituaries and biographies, museum catalogues, and other useful sources (Figure 4). As a result, it is also a very useful biographical reference. It is also much better organised, with indexation by institution and not just collector. Cleevely's book is not a strict equivalent to Sherborn's, as he had to restrict his main scope to fossils for reasons of project size (the unused information, mainly on zoological collections, remains on file at NHM). But he did not rigidly exclude minerals, molluscs and other non-fossils if they were relevant, as in the case of a multidisciplinary collector. In origin, spirit and at least partly in coverage, Cleevely's book is the most direct successor to Sherborn's. It remains very valuable today, and Doughty (1999) regarded it as notably "worthy of revision".

Cleevely's work was preceded, and has been followed by, books listing collections in other natural sciences. Peter Dance's classic history of shell collecting listed scientifically important collections of Recent shells as an appendix (1966; the 1986 edition appears to have the same appendices though the main text is different). This listing is now largely superseded by Kabat and Boss (1992). Dance (1966: 275, 1986: 201) specifically cited Sherborn's book as an "extremely useful" predecessor, greatly helping specialists locate important collections. Dance described his own list as containing a fraction of those collections that had existed, and he usually omitted fossil molluscs, but he recorded collections that had been destroyed during the Second World War.

Sherborn's book listed auction sales, and another direct successor is therefore *Natural history auctions 1700-1972: register of sales in the British Isles* (Chalmers-Hunt et al. 1976). This remains a valuable reference today. "[A]mong bibliographical aids to [its] compilation [...] first and foremost" was Sherborn's book, both in itself and in an extensively annotated copy in the British Museum (Natural History) (Chalmers-Hunt et al. 1976, pp. ix–x).

Knowing about collections is not just of research value. Area Museum Councils, now mostly abolished in the United Kingdom, were non-governmental public agencies which provided support for, and directed resources to, museums not otherwise funded by central government. During the 1980s, several Area Museum Councils set up advisory schemes to support museums with "orphaned" geological collections, using specialist curatorial and conservation staff, sometimes from larger local museums. This work was in direct response to the depressing results of the GCG's "State and Status" survey of collections (Doughty 1981, 1999, Taylor 1987, Knell and Taylor 1990). (The same issues, and parallel developments, occurred for biological collections.) An "orphaned" geological collection is one in a museum without specialist geological or natural sciences staff. The persons managing the museum cannot make decisions about the collection, let alone spend resources on it, if they know nothing about it. During the middle decades of the 20th century it was common for a member of staff from a larger institution to remove the scientifically interesting material from an orphan collection, sometimes abandoning or even dumping the remainder. Advisory schemes provided an informed alternative (Knell 1986, Gill and Knell 1988, Knell and Taylor 1990). Non-geological management might still find it hard to understand the scientific importance of their collection, even with specialist advice, but they would have no trouble appreciating its value for public display, and would also grasp the concept of the wider historical and local significance of a collection. However, they needed information and advice to fulfil the collection's potential. The advisors' reports, often drawing upon collections research publications such as Cleevely's book, helped justify expenditure on those collections' preservation and use. They also raised the regard in which the collections were held, and encouraged the assignment of permanent and temporary staff (e. g. Taylor 1987, Torrens and Taylor 1990, Copp et al. 2000). All those gave geological collections a greater chance to survive in a world where once they had been discarded with impunity.

Sherborn's successors: the Collections Research Units, FENSCORE, and other online sources

Where is the – Collection? was specifically recognised as a direct predecessor to the Collections Research Units which were organised in the UK during the 1970s and 1980s (Pettitt and Hancock 1981, p. 73). These units stemmed from the push for collections information embodied by the Geological Curators Group and the Biology Curators Group. These Groups' joint conference with the Systematics Association in Liverpool in 1977, on the "Function of local natural history collections", led directly to the first scheme, in north-west England, and then to others elsewhere (Hancock and Pettitt 1981, Pettitt and Hancock 1981, Davis and Brewer 1986, Hartley et al. 1987, Stace et al. 1987, especially vi-x, Bateman and McKenna 1993, Museum Documentation Association 1993, Walley 1993, Doughty 1999, Hancock and Hounsome 2010). The Units were usually based in major museums, using the support and regional structure of the Area Museum Councils, and often with additional aid from the government job creation schemes of the time, and from funding bodies such as the Wolfson Foundation. They gathered information on geological and biological collections, mostly in museums and other institutions such as schools and universities, but sometimes held by individuals. Under the influence, in particular, of Bill Pettitt (1937–2009) of Manchester Museum (Hancock and Hounsome 2010), those projects were seen as suitable for computerised data handling. Such modern techniques were also seen as raising the perceived status of natural science curators and helping the survival of their specialist positions. The processed output typically summarised the collector, content, and source localities for each collection, in thick volumes supplemented sometimes by microfiches for the bulk of the detail, as in the Scottish volume (Stace et al. 1987). However, before the entire United Kingdom was covered, these books came to be complemented by an online computer database under the aegis of FENSCORE (Federation for Natural Sciences Collections Recording), founded in 1981 but now dormant. The database is hosted by the University of Manchester (www.fenscore.man.ac.uk). It is understood to contain the data from all regions, including those (such as the South West) for which no book was ever published. It can be searched in different ways from the books. This is a valuable resource, which Doughty (1999) reckoned had basic information on over 95% of the natural sciences contents of museums in the United Kingdom. It does not, however, include the Natural History Museum collections (unless mentioned in some other entry as an "Associated Name").

This collections research work also fed biological and geological site and locality data into the new county or regional environmental records centres, often based in museums. This work was valuable in itself. It was also useful in gathering political support for those museums which were seen to be responding to the new environmental concerns, and also to be playing their part in job creation schemes at a time when unemployment was a major concern (Ely 1994).

The future

There seems little immediate prospect of future updates to Sherborn's successors, or at least those dealing with collections in the United Kingdom. One reason must be the pressure on museum staffing levels, combined with the structural changes within museum organizations which have led to a disproportionate reduction in specialist curatorial staffing over the last two or three decades. All this, combined with the elimination of some Area Museum Councils, inevitably discourages joint curatorial projects, whether between the museums of an area, or by the members of a specialist curatorial group in their own time. Maybe the existing databases are simply seen as sufficiently satisfactory that the further work needed for completion and updating is hard to justify against other pressures and priorities. Perhaps, also, collections research is no longer novel and fashionable, and has to some extent been displaced by newer initiatives relating to such things as social inclusion, health and wellbeing, and communities. New databases seem more likely to be at the specimen rather than collections level, be intended for taxonomic use, and be accessible online. At least initially, too, they seem likely to be at the level of the individual institution, such as the PalaeoSaurus database operated by the British Geological Survey (BGS: http://www.bgs.ac.uk/palaeosaurus/). However, the obvious need for cross-institutional platforms is leading to joint initiatives, if so far still specimen-based ones, such as the JISC-funded and BGS-led *GB3D types online project*, a database of British type fossils, with high-resolution images, stereo-anaglyphs and three-dimensional digital scans (http://www.3d-fossils.ac.uk/home.html). So perhaps we will see the fruition of the early hopes of the Collections Research Units workers for a union catalogue of type specimens (Hancock and Pettitt 1981, Pettitt and Hancock 1981).

Confidentiality has always had to be taken into account (e. g. Bateman and McKenna 1993), but a new problem arises because of legislation (at least in the UK) concerning the confidentiality of personal data on computer databases. This can cause problems where the original collector's name is part of the data sought by the inquirer. The BGS have had to consider this issue for their databases (M. Howe, pers. comm. 2014). PalaeoSaurus compromises by omitting the donor/collector name from the online display, but one can still search by using the collector name; and more recently donors have been asked to give permission for their names to be put online. In my view, there seems a strong argument for the default position to be the routine publication of names, with them being taken down if requested. The names of collectors and donors can be critical for research, and requests not to publish names are rare or non-existent, while names were routinely published in print in the days when museums still produced full annual reports.

As far as the field as a whole is concerned, one obvious way forward would be a regularly updated digital version of Cleevely's book, and its equivalents for other fields, perhaps online and presumably incorporating information from FENSCORE. Until then, it seems likely that as far as overall databases are concerned, we will have to rely on Sherborn's first-generation successors, not forgetting Sherborn himself, and (for the UK) FENSCORE, with internet and literature searches to catch more recent publications. FENSCORE, at least, might perhaps be modernised by converting the data into a modern system of data management, which could be updated directly by curators allowed password access (G. Hancock, pers. comm. 2014). This reminds us of the increasing importance of on-line sources, of which an example is the web publication *2,400 Years of Malacology* by Eugene V. Coan, Alan R. Kabat and Richard E. Petit (http://www.malacological.org/2004_malacology.html). It lists papers about malacologists, such as biographies, bibliographies, and lists of taxa and their present status, often noting the present repositories of relevant type specimens. Most importantly, like other on-line resources, this can be relatively easily updated, as happens near the beginning of each calendar year (E. Coan, pers. comm. 2014).

Some museums also contain historical accounts and other information on their websites, but those sites have a primary role in marketing, education and public presentation, and are liable to radical modification thanks to marketing-driven changes. It is prudent to keep such academic information in an explicitly permanent area, perhaps best of all in a completely separate formal repository.

Conclusions

Guides such as Sherborn's are needed more than ever, with the great increase in our knowledge of collections and their fates, and their usage in research and education. See, for instance, the essays by Cleevely (1983) and Doughty (1984, 1992), and the comments above on orphaned collections. Here are, briefly, a few further case studies.

A biologist or palaeontologist may only be concerned with individual specimens of a single taxon, and which institution holds them. But to find those specimens needs a knowledge of collections, the intermediate level between specimen and museum, and also how to use evidence such as specimen labels and catalogues. Such work led to the location of the lost holotype of the ammonite *Ammonites defossus* Simpson, 1843, at the Sedgwick Museum, Cambridge, informing a decision of the International Commission of Zoological Nomenclature (Torrens 1979, Forbes 1980, Brunton et al. 1985).

It can be important to find the institutions holding a named collection. A researcher on the Wealden fossil reptiles of the dinosaur pioneer Gideon Mantell (1790–1852) could find it valuable to know the museums to which his collection was partly dispersed by the British Museum in the late 1880s (Cleevely and Chapman 1992, p. 354).

There are other reasons to be aware of collections as entities in their own right. The documentation of collections, in the widest sense, includes diaries, field notes and correspondence. When a collection is split between museums, one institution is likely to end up holding data relevant to specimens in another institution. An example is the Alfred Leeds (1847–1917) collection of Middle Jurassic fossil vertebrates from Peterborough, England, divided between museums in different countries (Liston and Noè 2004, Araújo et al. 2008, Noè et al. 2010). A knowledge of the collection in question can suggest other important issues; for instance, a researcher using the Jurassic marine reptiles collected by Thomas Hawkins (1810–1889) needs to know that these contain a number of deceptively fabricated composites (Lomax and Massare 2012).

Finally, the creation and use of collections is a major subject of research in its own right, which addresses important questions in the sociology and history of science, and in wider Western culture. A good example is the work of Simon Knell (2000) on early 19th century English geology, which was inspired by Hugh Torrens's biographical studies of an underclass of practical men and women. Torrens (1995) has argued that historians must remember that making a collection can itself be a major contribution to a field of study, even if the collector produces no publications (see also Doughty 1992, pp. 518–519). Such a person was Mary Anning (1799–1847), commercial collector of Lyme Regis. Remarkably, she has attracted more biographical attention than almost all British or Irish geologists (Oldroyd 2013). This admittedly arises partly because of her story as a poor working class woman in a romantic Regency resort, but the excitement of fossil collecting is an important element in her appeal: hence the stream of popular Anningian books and articles, and museum activities such as those at the Natural History Museum, London, and Lyme Regis Museum. All this is, of course, based in part on Torrens's research (1995), reminding us that collections research has an important role in public education and recreation. Anning is admittedly an ironic example, for we would probably know more about her if Sherborn had not dispersed, and, one presumes, also partly destroyed much of her personal archive as valueless for scientific research. This came about because her papers had been handed to Richard Owen (1804–1892), latterly Director of the British Museum (Natural History), and thereby passed to Sherborn who was given the huge and problematical task of dealing with Owen's papers (Gruber and Thackray 1992, Torrens 1995). This explains what happened when the American palaeontologist George Gaylord Simpson (1902–1984), visiting the British Museum in 1926–1927 to work on Mesozoic mammals, was befriended by Sherborn. Sherborn gave him "some treasures, an ms & autograph letter of Owen's, [and] a sheaf of notes in Clift's hand on the famous 'Missourium'" – William Clift (1775–1849) being Owen's father-in-law and predecessor as Conservator at the Hunterian Museum (Laporte 1987, p. 62). This attitude of Sherborn's must have contributed to the historiographic problems which today beset any writer trying to make sense of what has been written about Anning while paying due respect to elementary accuracy at any level (Torrens 1995, Taylor and Torrens 2014 and refs therein).

A knowledge of collections is, in short, useful for curation and research, and in developing the managerial and political will to support those collections and their museums. But this requires the underpinning of a corpus of organised information about the collections, and this is what Sherborn pioneered, as *Nature*'s reviewer instantly realised (Ritchie 1940, p. 80):

> [The book's] deficiencies can be put right in time; the chief concern is that Dr. Sherborn's vast knowledge and painstaking labour have created a foundation upon which a complete *Catalogus Thesaurorum* [i.e. Catalogue of Collections] may be erected, and which in the meantime will be invaluable for reference.

For its defects, Sherborn's book was more than useful enough to show the value of such works, while its inadequacies repeatedly reminded the user that something better was not only possible, but must be done. The seed which he planted did indeed take root and grow. How it will develop in the future is, perhaps, another matter.

Acknowledgements

Simon Knell, like me a peripatetic advisory curator during the 1980s, kindly discussed the questions raised by Sherborn's often overlooked work. Our mutual reflections make it timely to acknowledge the debt that our generation of curators owes to Hugh Torrens, Ron Cleevely and the late Phil Doughty, and the other founders of the Geological Curators Group, and their fellow pioneers in the Biology Curators Group. Their work led natural science curators to rethink our relationship to our collections, with a great impact on the values which we sought to pursue in our profession.

I am most grateful to the referees, Ron Cleevely, Eugene Coan, Paul Martyn Cooper and Alan R. Kabat for their helpful reviews and information. I also thank John Cooper, Mark Evans, David Gelsthorpe, Geoff Hancock, Mike Howe, Andrew Kitchener, Mark Shaw, and Hugh Torrens for discussion and information. I thank Rosalind Grooms, Cambridge University Press Archivist, Cambridge University Library, for searching for and providing copies of archival material in CUL, and Katherine Thompson, Brand Protection Officer, CUP, for locating the publication contract and providing a copy. The Cambridge University Library and Cambridge University Press are thanked for permission to cite archival material. I am grateful for the support of the libraries of the University of Leicester and National Museums Scotland and to the Natural History Museum Image Library for copies of books and documents illustrating this paper.

References

Altick RD (1978) The shows of London. Belknap Press, Cambridge, Massachusetts, 568 pp.

Andrews SM (1982) The discovery of fossil fishes in Scotland up to 1845 with checklists of Agassiz's figured specimens. Royal Scottish Museum, Edinburgh, 87 pp.

Anonymous (1940) [Review, Where is the – Collection?] Quarterly Review of Biology 15: 468–469.

Anonymous (1942a) Bequests. Museums Journal 42: 224.

Anonymous (1942b) Dr. C.D. Sherborn – The "Index Animalium". The Times 1942: 7. [24 June]

Anonymous (1943) Where is the – collection? Journal of the Society for the Bibliography of Natural History 1: 479. doi: 10.3366/jsbnh.1943.1.12.479

Anonymous (1972) Revision of Where is the – collection by C. Davies Sherborn, Ds.C., Cambridge U.P., 1940. Proceedings of the Malacological Society of London 40: 236.

Araújo R, Smith AS, Liston JJ (2008) The Alfred Leeds fossil vertebrate collection of the National Museum of Ireland-Natural History. Irish Journal of Earth Sciences 26: 17–32. doi: 10.3318/ijes.2008.26.17, http://www.jstor.org/stable/20788277

Bateman J, McKenna G (Eds) (1993) Register of natural science collections in south east Britain. Guide to the project and retrieval of data. South Eastern Collections Research Unit and Area Museum Service for South East England. [no place of publication or pagination given]

Birman W (1979) Calvert, Albert Frederick (1872–1946). Australian dictionary of biography. http://adb.anu.edu.au/biography/calvert-albert-frederick-5469/text9293 [accessed 10 January 2014]

Blake JF (1902) Lists of the types and figured specimens recognised by C.D. Sherborn in the collections of the Geological Society of London. Geological Society, London, 100 pp.

British Museum (Natural History) (1904–1912) The history of the collections contained in the Natural History Departments of the British Museum. Trustees of the British Museum (Natural History), London, 782 pp.

Brown HR [1857] The beauties of Lyme Regis, Charmouth, the Land-slip, and their vicinities; topographically and historically considered. Dunster, Lyme Regis, 34 pp.

Brunton CHC, Besterman TP, Cooper JA (Eds) (1985) Guidelines for the curation of geological materials. Geological Society Miscellaneous Paper 17. http://www.geocurator.org/pubs/Guidelines.pdf

Chalmers-Hunt JM (1976) Natural history auctions 1700–1972: a register of sales in the British Isles. Sotheby Parke Bernet, London, 189 pp.

Chapman F (1942) Scientific collections: their romance and tragedy. Victorian Naturalist 58: 160–162.

Cleevely RJ (1974) The Sowerbys, the Mineral Conchology, and their fossil collection. Journal of the Society for the Bibliography of Natural History 6: 418–481. doi: 10.3366/jsbnh.1974.6.6.418

Cleevely RJ (1977) The realities of a rainbow – towards an index of natural history collections. Museums Journal 77(3): 129.

Cleevely RJ (1983) World palaeontological collections. British Museum (Natural History) and Mansell, London, 365 pp.

Cleevely RJ (2004) Sherborn, Charles Davies (1861–1942). Oxford dictionary of national biography. http://www.oxforddnb.com/view/article/39240 [accessed 26 March 2014]

Cleevely RJ, Chapman SD (1992) The accumulation and disposal of Gideon Mantell's fossil collections and their role in the history of British palaeontology. Archives of Natural History 19: 307–364. doi: 10.3366/anh.1992.19.3.307

Coleman EH, Baldock GY, Newton EE (1902) Weeks's Museum. Notes and Queries (9)9: 97. doi: 10.1093/nq/s9-IX.214.97-c

Cooper MP (2006) Robbing the sparry garniture. A 200 year history of British mineral dealers 1750–1950. Mineralogical Record, Tucson, Arizona, 358 pp.

Copp CJT, Taylor MA, Thackray JC (2000) Charles Moore (1814–1881), Somerset geologist. Proceedings of the Somerset Archaeological and Natural History Society 140: 1–36.

Dance SP (1966) Shell collecting: an illustrated history. Faber and Faber, London, 344 pp.

Dance SP (1986) A history of shell collecting. Brill, Leiden, 265 pp.

Davenport-Hines R (2004) Groom, Charles Ottley (1839–1894). Oxford dictionary of national biography. http://www.oxforddnb.com/view/article/54058 [accessed 8 January 2014]

Davis P, Brewer C (1986) A catalogue of natural science collections in north-east England with biographical notes on the collectors. North of England Museum Service, 333 pp. [no place of publication given]

Dickinson EC (2016) Reinforcing the foundations of ornithological nomenclature: Filling the gaps in Sherborn's and Richmond's historical legacy of bibliographic exploration. In: Michel E (Ed.) Anchoring Biodiversity Information: From Sherborn to the 21st century and beyond. ZooKeys 550: 107–134. doi: 10.3897/zookeys.550.10170

Doughty PS (1981) The state and status of geology in United Kingdom museums. Geological Society Miscellaneous Paper 13: 1–118.

Doughty PS (1984) Research: geological collections. In: Thompson JMA (Ed.) Manual of Curatorship. Butterworths and the Museums Association, London, 156–163.

Doughty PS (1992) Researching geological collections. In: Thompson JMA (Ed.) Manual of Curatorship. Butterworths and the Museums Association, London, 513–524.

Doughty PS (1999) Museums then and now: collection developments in museum geology since 1981. The Geological Curator 7: 3–10. http://www.geocurator.org/arch/Curator/Vol7No1.pdf

Edmonds JM (1977) The legend of John Phillips's 'lost fossil collection'. Journal of the Society for the Bibliography of Natural History 8: 169–175. doi: 10.3366/jsbnh.1977.8.2.169

Ely W (1994) Natural history museums and biological recording. In: Stansfield G, Mathias J, Reid G (Eds) Manual of natural history curatorship. Her Majesty's Stationery Office, London, 167–189.

Evans M (2010) The roles played by museums, collections and collectors in the early history of reptile palaeontology. Geological Society, London, Special Publications 343: 5–29. doi: 10.1144/SP343.2

Evenhuis NL (2016) Charles Davies Sherborn and the "Indexer's Club". In: Michel E (Ed.) Anchoring Biodiversity Information: From Sherborn to the 21st century and beyond. ZooKeys 550: 13–32. doi: 10.3897/zookeys.550.9697

Forbes CL (1980) Credit where it is due. Newsletter of the Geological Curators Group 2: 404. http://www.geocurator.org/arch/Curator/Vol2No7.pdf

Fothergill H (2005) The state and status of geological collections in United Kingdom museums: 2001. The Geological Curator 8: 53–136, 199–200. http://www.geocurator.org/arch/Curator/Vol8No3.pdf, http://www.geocurator.org/arch/Curator/Vol8No4.pdf

Gill MAV, Knell SJ (1988) Tunbridge Wells Museum: geology and George Abbott (1844–1925). The Geological Curator 5: 3–16. http://www.geocurator.org/arch/Curator/Vol5No1.pdf

Glenister BF, Bolton TE, Broadhead TW, Eldredge N, Silberling NJ, Sweet WC, Waller TR (1976) Fossil invertebrates – collections in North American repositories. The Paleontological Society, Lawrence, Kansas, 67 pp.

Griffin FJ (1953) Charles Davies Sherborn (1861–1942): some personal recollections. Journal of the Society for the Bibliography of Natural History 3: 1–4. doi: 10.3366/jsbnh.1953.3.1.np

Gruber JW, Thackray JC (1992) Richard Owen commemoration. Three studies. Natural History Museum Publications, London, 181 pp. doi: 10.5962/bhl.title.87002

Hancock EG, Pettitt CW (Eds) (1981) Register of natural science collections in north west England. Manchester Museum, Manchester. [unpaginated]

Hancock EG, Hounsome MV (2010) Charles Arthur William "Bill" Pettitt (20 August 1937–26 March 2009): Zoological Curator at Manchester Museum. NatSCA News 19: 1–5. http://www.natsca.org/sites/default/files/publications-full/NatSCA%20News%20Issue%2019.pdf

Hartley MM, Norris A, Pettitt CW, Riley TH, Stier MA (Eds) (1987) Register of natural science collections in Yorkshire and Humberside. Area Museum and Art Gallery Service for Yorkshire and Humberside, 343 pp.

[Hinton] MAC (1942) [Obituary, C. D. Sherborn.] Museums Journal 42: 223. [as 'M. A. C. H.', and presumably Sherborn's BMNH colleague of that name]

Hodgkinson RL (2006) The life and works of Fortescue William Millett (1833–1915), foraminiferologist. Journal of Micropalaeontology 25: 165–174. doi: 10.1144/jm.25.2.165

Kabat AR, Boss KJ (1992) An indexed catalogue of publications on molluscan type specimens. Occasional Papers on Mollusks 5: 157–336; (1997) Addendum. 5: 337–370.

Knell SJ (1986) Some little known geology collections in the South East. The Geological Curator 4: 350–352. http://www.geocurator.org/arch/Curator/Vol4No6.pdf

Knell SJ (2000) The culture of English geology, 1815–1851. A science revealed through its collecting. Ashgate, Aldershot, 377 pp.

Knell SJ (2002) Collecting, conservation and conservatism: late twentieth century developments in the culture of British geology. Geological Society, London, Special Publications 192: 329–351. doi: 10.1144/GSL.SP.2002.192.01.16

Knell SJ, Taylor MA (1990) Museums on the rocks. Museums Journal 9: 23–25.

Lambrecht W, Quenstedt W, Quenstedt A (1938) Palaeontologi: Catalogus bio-bibliographicus. Fossilium Catalogus I: Animalia 72: 1–495. [reprinted 1978, Arno Press, New York]

Lang WD (1950) More about Mary Anning, including a newly-found letter. Proceedings of the Dorset Natural History and Archaeological Society 71: 184–188.

Laporte LF (1987) Simple curiosity; letters from George Gaylord Simpson to his family, 1921–1970. University of California Press, Berkeley, 340 pp.

Liston JJ, Noè LF (2004) The tail of the Jurassic fish *Leedsichthys problematicus* (Osteichthyes: Actinopterygii) collected by Alfred Nicholson Leeds – an example of the importance of historical records in palaeontology. Archives of Natural History 31: 236–252. doi: 10.3366/anh.2004.31.2.236

Lomax DR, Massare JA (2012) The first reported *Leptonectes* (Reptilia: Ichthyosauria) with associated embryos, from Somerset, England. Paludicola 8: 263–276.

Lucas AM, Lucas PJ (2014) Natural history 'collectors': exploring the ambiguities. Archives of natural history 41: 63–74. doi: 10.3366/anh.2014.0210

M---- C. [as 'C. M.'] (1940) Index to natural history collections [review of Sherborn 1940]. Museums Journal 40(5): 73.

McCann AG (1981) Montagu Browne (Alexander Montagu Browne-Anderson) 1842(?)-1923. Guild of Taxidermists Newsletter 7 (April 1981): 21–22.

Miller CG (2016) Sherborn's foraminiferal studies and their influence on the collections at the Natural History Museum, London. In: Michel E (Ed.) Anchoring Biodiversity Information:

From Sherborn to the 21st century and beyond. ZooKeys 550: 71–81. doi: 10.3897/zookeys.550.9863

Morris PA (2010) A history of taxidermy. Art, science and bad taste. MPM Publishing and P. A. Morris, Ascot, 396 pp.

Museum Documentation Association (1993) Register of natural science collections in south east Britain: guide to the project & retrieval of data by Museum Documentation Association. Museum Documentation Association, Duxford, unpaginated.

Myers GS (1941) [Review, Where is the – Collection?.] Copeia for 1941: 122.

Nikolaeva S, Morgan N (2010) Lost and found. Geology Today 26: 16–20. doi: 10.1111/j.1365-2451.2010.00739.x

Noè LF, Liston JJ, Chapman SD (2010) 'Old bones, dry subject': the dinosaurs and pterosaur collected by Alfred Nicholson Leeds of Peterborough, England. Geological Society, London, Special Publications 343: 49–77. doi: 10.1144/SP343.4

Norman JR (1942) Dr. C. D. Sherborn. Nature 150: 146–147. doi: 10.1038/150146a0

Norman JR (1944) Squire. Memories of Charles Davies Sherborn. Harrap, London, 202 pp.

Nudds JR (Ed.) (1999) Directory of British geological museums. Geological Society, Miscellaneous Paper 18: 1–141.

Oldroyd DR (2013) A preliminary analysis of biographies and autobiographies of geologists. Earth Sciences History 32(1): 7–12.

Pettitt CW, Hancock EG (1981) Natural Science Collections Research Units, their origin, aims and current status. Museums Journal 81: 73–74.

Prieur A (1980) Catalogue des collections paléontologiques françaises. Université Claude-Bernard, Lyon, 307 pp.

Ritchie J (1940) Where is the – Collection? [review of Sherborn 1940]. Nature 146: 80. doi: 10.1038/146080a0

Rothwell N (2010) The red highway. 2nd edition. Black, Collingwood, Australia.

Sherborn CD (1935) Sowerby's Mineral Conchology: annotated list of the collectors mentioned by the Sowerbys. Proceedings of the Malacological Society of London 21: 238–243.

Sherborn CD (1940) Where is the – Collection? An account of the various natural history collections which have come under the notice of the compiler Charles Davies Sherborn D.Sc. Oxon. between 1880 and 1939. Cambridge University Press, Cambridge, 149 pp.

Shindler K (2016) A magpie with a card-index mind – Charles Davies Sherborn 1861–1942. In: Michel E (Ed.) Anchoring Biodiversity Information: From Sherborn to the 21st century and beyond. ZooKeys 550: 33–56. doi: 10.3897/zookeys.550.9975

Smith R (2008) Cox, James (c.1723–1800). Oxford dictionary of national biography. http://www.oxforddnb.com/view/article/54567 [accessed 18 January 2014]

Stace HE, Pettitt CWA, Waterston CD (1987) Natural science collections in Scotland. National Museums of Scotland, Edinburgh, 373 pp.

Taylor MA (1987) A strategy to safeguard the geological collections of the smaller museum. The Geological Curator 4: 413–420. http://www.geocurator.org/arch/Curator/Vol4No7.pdf

Taylor MA, Torrens HS (2014) An account of Mary Anning (1799–1847), fossil collector of Lyme Regis, Dorset, England, published by Henry Rowland Brown (1837–1921) in the second edition (1859) of Beauties of Lyme Regis. Proceedings of the Dorset Natural History and Archaeological Society 135: 62–70.

Torrens HS (1974) Locating and identifying collections of palaeontological material. Newsletter of the Geological Curators Group 1: 12–17. http://www.geocurator.org/arch/Curator/Vol1No1.pdf

Torrens HS (1975a) The Bath geological collections. Newsletter of the Geological Curators Group 1: 88–124. http://www.geocurator.org/arch/Curator/Vol1No3.pdf

Torrens HS (1975b) Yorkshire Museum collectors. Newsletter of the Geological Curators Group 1: 153–154. http://www.geocurator.org/arch/Curator/Vol1No3.pdf

Torrens HS (1979) Detection at the Sedgwick: an illustration of the importance of data retention. Newsletter of the Geological Curators Group 2: 333–340. http://www.geocurator.org/arch/Curator/Vol2No6.pdf

Torrens HS (1995) Mary Anning (1799–1847) of Lyme; 'the greatest fossilist the world ever knew'. British Journal for the History of Science 28: 257–284. doi: 10.1017/S0007087400033161

Torrens HS (2004) Weaver, Thomas (1773–1855). Oxford dictionary of national biography. http://www.oxforddnb.com/view/article/28915 [accessed 8 January 2014]

Torrens HS (2006) Notes on 'The Amateur' in the development of British geology. Proceedings of the Geologists' Association 117: 1–8. doi: 10.1016/S0016-7878(06)80027-1

Torrens HS, Taylor MA (1990) Geological collectors and museums in Cheltenham 1810–1988: a case history and its lessons. The Geological Curator 5: 175–213. http://www.geocurator.org/arch/Curator/Vol5No5.pdf

Vincent EA (1994) Geology and mineralogy at Oxford 1860–1986. History and reminiscence. Published by the author, Oxford, 245 pp.

Walley GP (1993) All Midlands Collection Research Unit. Register of natural science collections in the Midlands of England. Nottingham City Museum for West Midlands Area Museum Service and East Midlands Area Museum Service, Nottingham, 320 pp.

Waterston CD (1954) The Hugh Miller Cottage, Cromarty. Museums Journal 53(10): 255–257.

Webby BD (Ed.) (1989) Fossil collections of the world: an international guide. International Palaeontological Association, Washington, D.C., 216 pp.

Webby BD (1992) Fossil collections of the world: supplement to an international guide. Lethaia 25: 119–122. doi: 10.1111/j.1502-3931.1992.tb01796.x

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. In: Michel E (Ed.) Anchoring Biodiversity Information: From Sherborn to the 21st century and beyond. ZooKeys 550: 173–183. doi: 10.3897/zookeys.550.10041

Woodward AS, Sherborn CD (1890) A catalogue of British fossil Vertebrata. Dulau, London, 396 pp. doi: 10.5962/bhl.title.62185

ZooKeys 550: 107–134 (2016)
doi: 10.3897/zookeys.550.10170
http://zookeys.pensoft.net

RESEARCH ARTICLE

A peer-reviewed open-access journal
ZooKeys
Launched to accelerate biodiversity research

Reinforcing the foundations of ornithological nomenclature: Filling the gaps in Sherborn's and Richmond's historical legacy of bibliographic exploration

Edward C. Dickinson[1]

1 *Flat 3, 19 Bolsover Road, Eastbourne, East Sussex, BN20 7JG, U.K.*

Corresponding author: *Edward C. Dickinson* (edward@asiaorn.org)

Academic editor: *E. Michel* | Received 19 June 2015 | Accepted 19 June 2015 | Published 7 January 2016

http://zoobank.org/CAB6BB96-6DAA-443C-A0EC-359D1F2FFAF7

Citation: Dickinson EC (2016) Reinforcing the foundations of ornithological nomenclature: Filling the gaps in Sherborn's and Richmond's historical legacy of bibliographic exploration. In: Michel E (Ed.) Anchoring Biodiversity Information: From Sherborn to the 21st century and beyond. ZooKeys 550: 107–134. doi: 10.3897/zookeys.550.10170

Abstract
Due to its public popularity, ornithology has a huge corpus of scientific publication for a relatively small number of species. Although there are global checklists of currently recognised taxa, there has been only limited, mainly individual, effort to build a nomenclatural database that the science of ornithology deserves. This is especially true in relation to concise synonymies. With the arrival of *ZooBank* and the *Biodiversity Heritage Library*, the time has come to develop synonymies and to add fuller bibliographic detail to databases. The preparation for both began at the start of the 20th century with extensive work by Sherborn and Richmond. I discuss their legacy, offer notes on significant work since then, and provide suggestions for what remains to be done. To make solid the foundations for ornithological nomenclature and taxonomy, especially for synonymies, ornithologists will need to collaborate much more and contribute to the digital infrastructure.

Keywords
Ornithology, stability, priority, ICZN Code, dates of publication, ZooBank, verification, LANs, family-group names, genus-group names, species-group names, taxon-sampling, synonymy, objective synonyms, subjective synonyms

Introduction

As an old and popular science, ornithology has a very substantial literary foundation. Linnaeus (1758, 1766) provided the starting point, and the elaborated rules for scientific nomenclature meant that the content of the ornithological literature can be organised and is capable of providing detailed histories of our understanding of each avian taxon. Such histories are easiest to compile if names have not changed. However names do change, especially when a species is re-interpreted to belong to a different genus with the addition of new data or changing taxonomic perspectives. The Linnaean binomial (binominal) system provides for such changes in our understanding of relationships first, by allowing for new genus-group names to be introduced and second, in the context of the required combination of two names – the genus-group name and the species-group name – by maintaining the species-group name when transferred, subject only to gender agreement. Linnaean nomenclature is rooted in Latin and Greek, although it has been enriched by the acceptance of names from other sources. Because of these classical roots and the worldwide convention of acceptance of these rules, scientific names form the *lingua franca* of the world's zoologists. Ornithologists in Morocco, Japan or the United States of America understand the name *Eremophila alpestris* to refer a species of bird that is consistent across their language and geographic differences.

To be nomenclaturally 'available' in the technical sense, (i.e., validly published according to accepted nomenclatural rules) the name when first used must have been the first applied to the taxon after 1757 in binominal format (i.e. in a combination of genus-group name and species-group name). Thus, first use implies that the date it is introduced determines whether it gains priority and can be used. Other names may be available but date precedence will normally dictate which name should be chosen – the oldest name should be used according to the Principle of Priority in the *International Code of Zoological Nomenclature* (current edition, ICZN 1999; hereafter 'the Code'). This approach, set out as a declared Principle (or foundational rule) is thought to have been taken from early doctrine in patent law. Thus, the determinant evidence for each contending name includes the publication date along with the name of the author. The decision on priority may potentially rest on the very day of publication.

However, if a name is later used in a different combination due to assignment to a different genus, homonyms (identical names, in this case at the species-group level) must be resolved. Any other taxon found within the newly relevant genus that already bears the same specific or subspecific name can prevent the retention of the original species-group name. This can happen surprisingly often, as species-group names may refer to relatively common characteristics (e.g., *alba* for white, *atlantica* for distribution, etc.). To be retained, the transferred name must be older than any contending name. In any such case one of the two homonyms must be discarded and a replacement name found from the list of available synonyms if possible, or be freshly established if necessary. Making this evidence available is vital for understanding the logic behind historical name changes.

Sherborn (1861–1942) recognised the importance of this evidence for maintaining sense in the shifting meanings of the world's diversity. His creation of a card index, which, when sorted, became the *Index Animalium*, is remarkable both for the size of the task he set himself and for his years of application to that task for minimal reward (Evenhuis 2016, Shindler 2016, Taylor 2016, Welter-Schultes et al. 2016). The *Index Animalium*, which ran to almost 9600 pages, has been scanned and made available through the 'Smithsonian Libraries' website (http://www.sil.si.edu/digitalcollections/indexanimalium/ see Pilsk et al. 2016). In so far as ornithology is concerned this work provided a near-complete, and very largely accurate, dataset of bird names from 1758 to 1850. It should be emphasised that this index of old names provides the foundation, and thus is the most critical basis, for a stable modern nomenclature and links to past published scientific information. In compiling *Index Animalium* Sherborn worked first on the period 1758 to 1800 (Sherborn 1902) with the library resources of the Natural History Museum in London (NHM-London) at his disposal, doubtless working all the way through each volume in turn. He then tackled 1801 to 1850 in a 33-part work appearing from 1922 to 1935. His card index is held in the Rare Books Room at the NHM-London and remains of value as some cards reveal more than Sherborn included in his one-line entries in the *Index*.

Sherborn's work came to the attention of Charles Wallace Richmond (1868-1932) at the Division of Birds at the National Museum of Natural History (NMNH, a part

Figure 1. Charles Wallace Richmond as depicted in his obituary in *The Auk* by Stone (1933).

of the Smithsonian Institution Washington, D.C.) who had begun a card catalogue of bird names by 1896, while working closely with his mentor, Robert Ridgway, who was writing the *Birds of North and Middle America*. By then Richmond had been collecting such names since around 1889 (Stone 1933).

Richmond, whose card index related solely to birds, made it his business to build his content well beyond the 1850 date reached by Sherborn, and this task was taken up by those who followed him in the NMNH. He was, or became, just as interested in the importance of dates of publication as Sherborn and the two corresponded on this topic. Richmond was very determined in his search for avian genus-group names and his card index of these became the basis for four supplements to the *Index Generum Avium* of Waterhouse (1889) (Richmond 1902, 1908, 1917, 1927) – and these often included names that had escaped both Waterhouse and Sherborn.

The Richmond Index, published in microfiche form in 1992, sixty years after Richmond's death, was a relatively comprehensive reference system when Richmond last worked upon it. Just how comprehensive is unclear as Olson and Browning (1992) offered the caveat that "Richmond died in 1932 but his contributions to the Index had probably diminished well prior to that. Stone (1933) relates that Richmond's health began to fail about the onset of the First World War, after which his work on the Index was only desultory". Thus there may be significant gaps during a period from about 1914 to the 1930s. Over the sixty years between Stone's death and publication, there was a clear recognition of its utility by the Division of Birds, NMNH, which held the resource in its library. Smithsonian staff did much to add to it. However, again a caveat is in order from Olson and Browning (1992), who wrote "this was attended to with varying degrees of competence and dedication. More diligence has been applied in recent years, but there remains a period for which the *Index* is certain to be recognizably incomplete". Nonetheless, the list can reasonably claim to be more complete than any other for birds available today.

The Richmond Index includes both genus-group name cards and species-group name cards. These were microfilmed and appeared in microfiche form (Richmond 1992). More recently these cards have been scanned by Alan Peterson and made available on www.zoonomen.net. Richmond also used index cards to record his research into the dates of publications, both part-works and journals and to keep notes about authors. These were very much works-in-progress, and have not been published; they are housed in the library of the Division of Birds, NMNH.

Sadly Stone did not tell us precisely when Richmond began to correspond with Sherborn saying "through all these years and up to the time of his death Richmond maintained a correspondence ..." not only with Sherborn, but also "with Gregory Mathews and others interested in bibliographic research". But the start of their sharing of their findings surely cannot have been later than 1902 when the first part of the *Index Animalium* appeared. Letters kept by Sherborn are archived in the Palaeontology Library NHM-London, but these still need to be explored in depth.

Thus, when Sherborn died in 1942 the primary sources for avian names were the *Index Animalium* up to 1850 and, up to and beyond that date, the Richmond card index although it was not put in wider circulation for another 50 years. My aim with this chapter is to define the gap between what they left us and what we ought now to have, and to discuss the extent to which that gap has now been closed. To do this I need first to suggest what I believe we should have available to us. Then, I will record some of the major works in ornithology which have filled large parts of that gap. And, finally, I will offer a summary of what remains to be done.

The organised resources that the ornithological community needs:

A comprehensive set of all validly published avian scientific names complete with their authors, dates and citations. This dataset should consist of original combinations with original spellings, right or wrong. Where introduced with dual or multiple spellings, each such spelling should be in the dataset and information on the subsequent selection of a spelling as correct by a First Reviser should be located and a citation to that added [Art. 32.2.1, ICZN Code]. First Revisers have a special definition and value in nomenclature [defined in the ICZN glossary as "The first author to cite names (including different original spellings of the same name) or nomenclatural acts published on the same date and to select one of them to have precedence over the other(s)." and supported in Art. 24].

The structure of that dataset should link every name that first appeared on the same date in the same work thus allowing any date change to cause change in each linked record. Also relevant will be any published First Reviser action in which precedence of one work rather than another has been asserted because this affects the dates of publication of record for both publications and may affect more than just the specific case dealt with by the First Reviser (so the record for each such work should hold a notation of this kind).

Fully functional nomenclatural synonymies need to be organised. Not synonymies of the kind found in 19th century works like the *Catalogue of the birds in the British Museum* where the objective was to list every use of a name listing each of its combinations and spellings – although such synonymies have value in a different context. Instead, a nomenclatural synonymy needs to show the relationships between senior and junior names for the same taxon. In the context of genera any names of subgenera must be included as related subordinate names and in the case of species the subspecies and their synonyms must be included. At genus-group level, where phylogeneticists could immediately benefit, the broad genus would have in its synonymy all the subgenera and the synonyms that relate to the broad genus name. These should be qualified according to whether they may be objective synonyms – based on the same type species – or subjective synonyms which, in the right circumstances, taxonomists might choose to bring into use. At the species group level where, in principle, each name is based on a type specimen there are again objective synonyms (based on the same type specimen) but more often synonyms at this level are subjective. Due to taxonomic change these synonymies would need periodic review because subjective synonyms might well come into use. There are tools available today to help maintain such lists and some branches of zoology, such as icthyology and several sections of entomology, have established networks of scientists who have committed to help with such tasks. Ornithology, having been judged to be “now well-known” for over 50 years, is finding that it needs to catch up!

The *Code* requires that account be taken of other related issues. Thus the main dataset should include, or be linked to, full information on the approved changes to original spellings. This, with the possible exception of changes due to gender agreement (Art. 32.3) as they can be dynamic, implies: (i) corrected spellings as governed by Art. 32.5 (with retention and signalling of an incorrect original spelling because that is what a researcher may find when checking the original), (ii) justified emendations as governed by Art. 33, or mandatory changes as defined in Art. 33 and (iii) the need for a notation as to the chosen original spelling selected by a First Reviser (Art. 32.2.1) from two or more original spellings. Finally, notes must be added regarding any decisions made by the Commission that fix a spelling (see ICZN 1987, 2001).

ZooBank (http://zoobank.org), the registration platform of the International Commission on Zoological Nomenclature, was conceived not just to hold names but also nomenclatural acts – although progress towards accommodating such acts has so far been limited. The Code places central importance on such acts and implicitly requires that a collection be made of all such acts thus the establishment of *ZooBank* is the logical outcome of the Code and the shift to digital taxonomic tools. The extent to which any zoological discipline has compiled such lists is unclear but it is not easy to find any such lists made for ornithological cases. Centralisation makes sense, so *ZooBank* is the logical place for confirmation that Code-based requirements have been met and for showing the effects on the original name. The stability of spellings is partly dependent on access to this information, but so too is the avoidance of contradiction of any First Reviser action relating to a spelling choice. Special attention will be needed for recording First Reviser actions that give precedence to one work over another. These actions need association with the first works involved and the discovery of what other names, perhaps in a different discipline, may be affected. Anyone accrediting any zoological or botanical name to that work will need to do the underlying bibliographic research to do the job properly.

Even if the basic information is only partially complete, deposition in *ZooBank* makes any set of names accessible, i.e., immediately retrievable, for a global community of users. Thus importing the names from Sherborn’s *Index Animalium* makes sense although there would then remain the challenges of completing the dataset, verifying the names, and establishing which names are ‘available names’ in the Code-specified technical meaning. *ZooBank* will need to signal for each registered name whether it has been verified, and, where appropriate, to signal which published names have been found to not be ‘available’ in the sense of the Code.

The development of Lists of Available Names (LANs; see Art. 79 of the Code) would be assisted by such information access, which seems to be a necessary preliminary step. In this context see Alonso-Zarazaga et al. (2016).

Some zoological groups, whether taxon-rich or not, lack the mass of serious publication per taxon that is found in ornithology. No current ICZN Commissioner works extensively on ornithological taxonomy and it may be that the value of such a mass of literature is under-appreciated by commissioners not working in fields that are similarly rich in bibliographic history. A thorough understanding of the literature as it relates to any given taxonomic group gives a much stronger qualification for decisions on name availability for that group.

Proposed LANs need to provide an appendix of unavailable names, giving the reason each was decided to be unavailable (Alonso-Zarazaga et al. 2016). There are problems with the suppression of names, not only because they may confound past decisions on homonymy, but also because lack of use on its own is an insufficient reason for suppression – especially at a time when molecular studies are revealing sibling taxa to which some of our 'unused' names are being found to apply. Although a reservoir of synonyms should be valued and not considered a problem, it would be reasonable to suppress old names that have not been retained in synonymies. In particular it would be a mistake to suppress genus-group names that may well be available and valid for unrecognised subgenera. By contrast, issues of homonymy aside, names which are objective synonyms may merit suppression although, in this information age, it is hard to see any gain coming from full suppression as opposed to listing as unavailable with potential approval for restoration.

What help, since the work of Sherborn and Ridgway, has been provided for the building of synonymies?

Neither of these compilers left us fully comprehensive lists of avian names or sought to arrange synonymies at any rank. However, since then the work of Bock (1994) covering family-group names is available. Bock (1994: 13) acknowledged the help available to him from a card index of family-group names which Ernst Mayr had prepared about 1960. Bock's work, proposed during discussion of the 4th edition of the Code, was intended to pave the way for a List of Available Names for ornithology although unfortunately Bock's expectations of the Code were not all met in the final drafts of the relevant articles. Indeed until recently (Alonso-Zarazaga et al. 2016) the ICZN had not provided clear guidelines on format or procedure. Some well-considered suggestions were made after Bock's work appeared and all these elements will need to be taken into account when a submission is prepared for a LAN for avian family-group names, for which Bock's work provides a very helpful foundation.

At the level of genus-group names help is at hand from general zoological nomenclators. The well-known *Nomenclator Zoologicus* by Sheffield Airey Neave (1879–1961) was published in four volumes (1939–40), and has been complemented by supplements that have continued into the present century and is accessible online (http://uio.mbl.edu/NomenclatorZoologicus/). The less well known, but useful, *Nomenclator Animalium generum et subgenerum* by Franz Eilard Schulze (1840–1921) was issued in parts from 1926 to 1954. One or two ornithologists with a particular personal interest in genus-group names will facilitate the timely preparation of a synonymy of genus-group names if they can be recruited.

As for where names of "missing" new species and subspecies of birds may be found, we have a huge corpus of ornithological literature, and synonymies can be found in many of the more scientific works.

At the global level, for anyone seeking to develop a synonymy of avian names, two works provide the backbone. The first was the *Catalogue of the Birds in the British Museum* (1874 to 1898), totalling 27 volumes, compiled mostly by Richard Bowdler Sharpe (1847–1909). This took the starting point of binomial nomenclature as Linnaeus (1766 – the 12th edition of the *Systema Naturae*) and thus, unfortunately, excluded all earlier references. Acceptance of the 1758 10th edition of the *Systema Naturae* (Linnaeus, 1758) as the start point for zoological nomenclature dates from 1886 in America (Banks 2004), and in Europe apparently only from the *Règles* (ICZN 1905), but the *Catalogue* had been begun with the 12th edition as its start point and this was applied through to its completion. Obviously this eight-year difference is important, works such as those of Brisson (1760) and Pallas (1764) were excluded. A second problem with this resource is the spelling of scientific names. Sharpe and some or all of his co-workers believed names should comply with their understanding of the rules of Latin and Greek grammar and they made many corrections. In the majority of cases original spellings are faithfully reproduced in the citations in the helpful synonymies for each taxon discussed, which draw on most of the relevant books, but there are certainly some cases where the spellings here were also emended. These emendations seem to be the single major reason why ornithology has suffered from competing spellings. In spite of these two problems, the *Catalogue* contains a huge bedrock of knowledge. However, almost all the volumes appeared before Sherborn (1902). Thus, for example, the name *Strix barbata* Latham, 1790 (see Sherborn 1902: 108) was missed by Sharpe (1875) which has led to the name being almost invariably attributed to Pallas (1811). Sharpe's five volume *Handlist of the Genera and Species of Birds* (1899–1909) is a useful summary of the *Catalogue*.

In the early 20th century there was a general understanding of the value of such lists and of the need to keep them up to date. By the 1920s it was apparent on both sides of the Atlantic that much new information had been accumulated and Sclater (1924: [iii]) wrote:

> "The scheme for the publication of a systematic list of the Birds of the World, according to Zoogeographical Regions, had its origins in a proposal laid before a Committee of the British Ornithologists' Union in 1919, when a special committee was appointed to take the matter into consideration." "The Committee have [sic] held many meetings, and have been in communication with the Secretary of the American Ornithologists' Union, and an agreement has been reached in conjunction with that Union in the preparation of the Lists of the Birds of each Zoogeographical Region, the B.O.U. being responsible for those dealing with the Old World." This led to coverage of the Ethiopian Region (Sclater, 1924, 1930) and the Australasian region (Mathews, 1927, 1930 and supplements), but not to global lists for the bulk

of Asia, nor to a Palaearctic list. In the case of the latter, Ernst Hartert's *Die Vögel der paläarktischen Fauna* (1909–1934) was filling the gap and no work in English was started.

In America work had already begun on the *Catalogue of birds of the Americas and the adjacent Islands in the Field Museum of Natural History including all species and subspecies known to occur in North America, Mexico, Central America, South America, the West Indies, and islands of the Caribbean Sea, the Galapagos Archipelago, and other islands which may be included on account of their faunal affinities*. Of this 15 volumes appeared between 1918 and 1950, the authors being Charles Barney Cory (1857–1921), Carl (Charles) Eduard Hellmayr (1878-1944) and (Henry) Boardman Conover (1892-1950).

Although the above catalogue was still unfinished it gradually became evident that the coverage of Asia and the Palaearctic was either non-existent or dated. Perhaps it was the need to deal with these gaps, or just the obvious value of having all the birds of the world in one reference work in the English language, that led to the initiative to do just that. The new conception, and the second key source for material for avian synonymies, was the *Check-list of Birds of the World*, begun by James Lee Peters (1889-1952) in 1931. This 15-volume work emanating from Harvard University was completed (except for the sixteenth volume holding the General Index) by the publication of volume 11, in 1986, 34 years after the death of Peters. It soon became *the* standard work, but only volume one was ever updated. The completion of this after the death of Peters was due to the work of a variety of internationally-known ornithologists led and encouraged by Ernst Walter Mayr (1904–2005). This was a period when the perception was that many species were mere local variants and the merging of species, with little or no pre-publication of reasons, was common. A similar compression of genus-group names occurred; many became synonyms or hidden subgenera.

From the beginning it was Peters' intention that names listed in synonymy in the above mentioned *Handlist* and still considered synonyms would not be re-listed – in spite of the well-established use, by now, of subspecies. Over the years that intention was held to, but deliberate omissions of synonyms began due to other major works, but regional, not global ones.

In the Introduction to volume I of his *Check-list* Peters (1931: [v]) wrote "It is now nearly thirty-two years since the first volume of Sharpe's *Handlist* .... made its appearance. The five volumes comprising that work have long been the one and only standard catalogue available to ornithologists, and it is a pity that Sharpe's work could not have remained so, but the rapidity in the increase of ornithological knowledge has made it clear for a number of years that a new work along the same, or perhaps slightly more elaborate, lines was needed". One of the strengths of this new compendium was the listing of the type locality for each listed name. A few key points from that Introduction (p. vi) need reporting. "... it does *not* include a complete principal synonymy. Synonyms, both generic, specific and subspecific, are given only for genera, species and subspecies described since the publication of the first volume of Sharpe's *Handlist*".

However, Peters (op. cit.) also wrote “Synonyms not to be found in any of the volumes of the *Catalogue of Birds of the British Museum* are cited in full”, which was slightly at odds with his earlier sentence and implied that names prior to 1766 had all been accounted for and were given, which is not strictly correct. With the appearance in 1960 of volume IX of Peters’ *Check-list of Birds of the World* Ernst Mayr and James Cowan Greenway Jr. (1903-1989) were appointed editors; and in fact it was Mayr in particular who guided the project to its completion. Volume IX was the eighth to appear because volume VIII had been delegated to John Todd Zimmer (1889-1957) and he had died five years after Peters leaving that volume unfinished. In the Introduction to volume IX Mayr and Greenway (1960: vii) revised the limitations to the included synonymy, writing “The synonymy of Old World taxa includes all names proposed since Sharpe’s *Handlist*, or, where appropriate, Hartert’s *Vög. pal. Fauna*, while synonyms of New World taxa correctly cited in Hellmayr’s *Catalogue* are not repeated.” Users of the second edition of Peters *Check-list* volume I, which dates from 1979, should note from its Introduction (p. vi) that Mayr and Cottrell (1979) stated “Synonyms correctly listed in the first edition have been omitted ...”. Because of these and other limitations it is clear that to derive full synonymies other works, and not just those mentioned above, need to be consulted.

As work on Peters’ *Check-list* seemed to head towards completion, the Introduction to volume X (Mayr and Paynter 1964: v) claimed leadership throughout the zoological world saying of the *Check-list* “There is nothing like it in the world literature for any other kind of organism.” However, the last volume to appear (Vol. XI, 1986), apart from the index (vol. XV, 1987), came out 22 years later. Whether or not this work led the world it remains of very great use to ornithologists for its content, especially its inclusion of type localities.

The Richmond card index may have been reasonably complete in 1930, but after that adding new cards depended on his successors. While some of these are known to have shared his enthusiasm for this resource there will certainly have been periods when a lower priority was assigned to this task. The *Zoological Record* should help with completion, but its journal coverage has never been as complete as its compilers would have wished. Finally, a few private individuals have been assiduous in collecting new names and their information will need to be obtained. However, it seems probable that since the 1990s no ornithologist has been paid to maintain a list, despite the fact that new species and subspecies of birds continue to be found and named every year and new genus-group names proposed. Amateurs, however, have done their best to fill the gap and the ‘*Howard and Moore complete checklist of birds of the world*’ edited by Dickinson (2003) made full use of their gatherings even if sometimes only in footnotes.

In the period from 1851 onwards, i.e., after the *Index Animalium*, scientific zoological journals begin to multiply and then specialise and the earliest ornithological journals had their beginnings about that date (*Naumannia* in 1849, the *Journal für Ornithologie* in 1853 and *The Ibis* in 1859). The cataloguing of zoological journals arrived earlier.

The *Royal Society Catalogue of Scientific Papers* covered the years 1800 to 1863 (and in a second series covered up to 1900); but the output in zoology was so large that in

1864 the Zoological Society of London launched *The Zoological Record*, now commercially published by Thomson Reuters. This provides separate listings for the literature of each class of zoology. Almost certainly everything reported in these lists will have been indexed by Richmond and his successors. Some tens of journals from smaller countries, with limited facilities for the study of zoology, have at times neglected to provide the indexers with their works and the extent to which new names have been missed is not certainly known, but the number we lack is probably quite small, say less than 2 or 3%.

An effort similar to that of the Royal Society was made in Germany where the *Archiv für Naturgeschichte* provided quality information from 1839 onwards on the main zoological subjects covering both books and periodicals. The sections on birds were called *Bericht über die Leistung in der Naturgeschichte der Vögel während des Jahres ….* and the successive compilers were Andreas Wagner, Gustav Hartlaub, August von Pelzeln and Anton Reichenow. When these ceased towards the end of the 19th century Reichenow ensured that similar material appeared in the *Ornithologischen Monatsbericte* and indeed the primary journals (such as those mentioned above) that had arisen to serve ornithology all provided such information as they were able to collect and consider.

Peters (1934) was fulsome in his praise of the utility of the works of Sherborn and Richmond and also John Todd Zimmer (1889–1957) whose *Catalogue of the Edward E. Ayer Ornithological Library* appeared in 1926. Zimmer's bibliographic work was paralleled by research by Gregory Macalister Mathews (1876–1949) published in 1925 as two supplements to his *Birds of Australia*, and followed later by an important and more detailed but unfinished work by Robert Morrow Mengel (1921–1990) of which two parts were published (A–B in 1972 and C–D in 1983). The complex part works of John Gould (1804-1881) were very carefully documented by Gordon Sauer (1982), who, like Mengel, used the important Ralph Ellis Library of Ornithology at the University of Kansas. Studies such as these were encouraged by Sherborn as a founder member of the Society for the Bibliography of Natural History. This has now broadened its interests and been renamed the Society for the History of Natural History.

A recent account of the various sources likely to be needed by bibliographers working in ornithology, but also important to those working in other zoological disciplines, is to be found in Dickinson et al. (2011). The foundations for ornithological nomenclature ultimately comprise the many books and periodicals published that made possible these and later compilations.

What remains to be done?

In the 20th Century it began to be said that we knew all the birds, as the rate of discovery of new species suggested to the public that accumulation of ornithological knowledge overall had plateaued. Peters (1931: v) had referred to the "rapidity of the increase of ornithological knowledge" since Sharpe's *Catalogue*. But Zimmer and Mayr (1943) wrote "practically all the widespread species of the birds of the world have been discovered, whether they be rare or common. There still remain a number of tropical islands, mountain ranges, or isolated peaks on which additional new species will be discovered". Mayr (1957) wrote "... I doubt that more than 20 new species will be discovered in the next 10 years". However, with at least 35 new species discovered in the decade following that comment, Mayr later (1971) suggested that this rate of discovery might continue and indeed it has.

Ornithologists found that research into behaviour and ecology opened new frontiers, revealing new taxa, filling needs for conservation biology and supporting the interests of a growing community of bird-watchers who became more serious with increased leisure time and cheaper travel. This was also supported by the introduction of field guides with good colour plates and an increasing willingness by publishers to depict every species in colour (Moss 2009). Television further popularised birds and they are clearly among the star turns that support interest in, and fund-raising for, conservation.

More recently phylogenetic studies of birds have formed the vanguard of evolutionary biology. Stresemann (1959) lamented that 200 years since Linnaeus we still lacked a reliable phylogenetic framework for the relationships of birds. However, the discovery of the structure of DNA in 1953 created new opportunities for molecular approaches to phylogenetics, seized upon by ornithologists. Charles Sibley's comparative data on egg white proteins are recognised as game-changing for building molecular phylogenies. Sibley (1965) wrote "we have studied ... all 27 orders and ... 146 of the 170 families of living birds" and followed this with DNA-DNA hybridization work leading to the first avian checklist built on inferred phylogenetic relationships (Sibley and Monroe 1990). Molecular biology of birds remains at the forefront of both academic and applied research, with the extraction of DNA from blood, feathers, bone or tissue of live and dead birds now commonplace. We have been finding, first, that many perceived relationships inferred from anatomy and morphology have been oversimplified and, second, that we have misperceived the diversity and need more genera and families to sensibly structure avian diversity. Names help structure knowledge. The 170 families mentioned by Sibley contrasts with today's checklists of 100 non-passerine families plus 136 passerine families (Dickinson and Remsen 2013; Dickinson and Christidis 2014): a 39% increase in just under 50 years.

Currently, tools for disambiguation of bird taxonomy are developing that also include nomenclatural links. There is a close working relationship between the authors of the 2013-14 checklists (Dickinson and Remsen 2013; Dickinson and Christidis 2014) and Avibase (Lepage et al. 2014 and http://avibase.bsc-eoc.org/), a large taxonomic reconciliation, distribution and current names resource. The Avibase compiler is the database manager for the checklist work and has approval to insert checklist content into Avibase as soon as the checklist is actually published.

This is where the access to nomenclatural information requires an upgrade. Many of the genus-group names needed exist in synonymy but this is not immediately obvious due to the lack of detailed, organised synonymies – synonymies (which, so that date precedence can play its role, must give the authors and dates for the genus-group names mentioned) should clarify the type species and make clear which synonyms are objective, because they are based on the names of species already used for an earlier genus-group name, and which are subjective, and are listed where they are only because of a taxonomic judgment that the type-species concerned is satisfactorily placed within a broader genus. In fact, such synonymies would greatly improve the selection of species for taxon sampling because screening the type species of genus-names in synonymy allows more explicit results to be postulated.

Wilson (2005) set out a vision of rapid progress by the biological sciences with considerable expectations from molecular and cellular biology. In ornithology the number of publications based on molecular studies year is clearly growing. But, molecular studies are being increasingly complemented by evidence of behavioural traits, including song, that can be strongly suggestive of species limits, and indeed much recent change reflects anatomical and morphological evidence that had led to the recognition of extra genera that were swept into synonymy in the mid 20^{th} century.

Table 1. Recent published molecular studies in ornithology. Separate columns list papers of global relevance; those related to the "Old World" (including Australasia), and those related to the Americas based on the coverage of each study (source: Dickinson and Remsen 2013).

Year	Global	Old World	New World	Total
2001	23	18	37	78
2002	22	27	45	94
2003	29	27	49	105
2004	42	43	45	130
2005	28	44	56	128
2006	33	31	67	131
2007	32	62	69	163
2008	40	59	86	185
2009	44	48	68	160
2010	39	62	98	199
2011	50	61	100	211
2012	36	71	92	199

There is no detailed published synonymy of avian genus-group names of the kind explained above. We have some 10,000 species of birds and over the years since 1758 perhaps double that number of genus-group avian names has been proposed. As only about 2500 of these names are in current use, it seems that each genus name we employ must have about seven synonyms! Molecular biologists researching birds need a comprehensive synonymy of avian genus-group names; one where they can determine what names, within a broad genus, are found in its synonymy and how some of these names will dictate any subdivision of that genus. Of course it will sometimes be the case that a genus including two or more sections, which appear in the phylogenetic tree as clades, lacks a name in synonymy that is representative of and applicable to each clade. In such cases new generic names will be needed. However, if the genus or family under study is 'mapped out' in terms of its synonymy and the taxon sampling includes the type species of the genus-group names in synonymy, the taxonomic evaluation of the results will be making optimal use of nomenclatural structuring of knowledge. This will help reduce the need for corrections to changes proposed without full evaluation.

This serious lack of full synonymies is, I suggest, partly due to how few alpha-taxonomists are now paid to do such work, but there is also a generalised failure in ornithology to recognise this need and to collaborate internationally, and to use the revolutionary tools offered by digital information systems to create databases such as synonymies. The vision shown by Sherborn, Richmond, Peters, Mayr and others has not been sustained in a world where ornithology is organised without a sufficient consideration of prioritisation of resources and without central direction for international projects other than those in conservation. Ornithology is largely organised in national societies, with limited terms of elected office and those taking office at such levels are the willing and the available, and are rarely the long-term thinkers and the visionaries. Behaviourally such office-bearers are more like politicians: they have a shorter-term focus. An integrated approach to taxonomy such as that suggested by Padial et al. (2010) will simply not materialise without the creation of expert groups and determined delegation to them by such elected officials. This already often occurs at the national level but it is not visibly working at the international level.

Other zoological disciplines have managed data collection and organisation much better. Ichthyology for example has "Fishbase" (http://fishbase.org). It is interesting and relevant that this was largely an unfunded labour of love, with a key founder, Bill Eschmeyer, but that there was also a huge cooperative research community. The value of "Fishbase" to all fish workers, in all aspects of research and applied work, is widely agreed (Pyle 2015). There are some ornithologists who also perform such labours of love; however all, or almost all, are amateurs working individually. Unfortunately the ornithological community has not pulled together in the same way as the fish folk! Although the sheer volume of the bird literature is a challenge, it is certainly a surmountable one with distributed, collaborative effort. The bird community should make it a priority to produce a similar resource over an appropriate time period.

There is now some potential for collaboration. What was the Standing Committee on Ornithological Nomenclature, established by the International Ornithological Congress has, in 2015, morphed into the Working Group on Avian Nomenclature of the International Ornithological Union. The change has led to an increased number of members and to recognition that it must involve itself in the challenges in producing Lists of Available Names. What is unclear is how much can be achieved in a timely manner without financial support.

If there is a problem with a lack of a source of synonyms, now apparent at the generic level but certainly suffered at the species and subspecies level, this also affects names at the family level. Bock (1994: 159) mentioned his difficulties in examining some of the very early works in which he sought for the reasons for changes in usage of generic names (in the cases where such names provide the stems for family group names). This is a complex problem; most of us will not have total recall and not know where to find crucial evidence, in addition actually locating some older works and gaining access has been a problem. Since Bock's period of study much has changed and a growing body of such works is available through the Biodiversity Heritage Library (www.biodiversitylibrary.org). Our museum librarians, it seems, do have the vision needed and have identified ways to fund a tremendous resource – benefitting especially the smaller countries of the world with limited local library facilities and also the amateur working at home with limited resources.

The value of bringing together data that relates to particular taxonomic groups is increasingly recognised; scientific names and common names, in all languages, are being linked in "name-use catalogues" or indexes by organisations such as GBIF (the Global Biodiversity Information Facility – www.gbif.org). However, as identical names are used for taxonomic concepts that are not identical (e.g. for a broad species with perhaps 10 subspecies or for the nominate subset of that which follows a separation into two or more species) there is an increasing need for taxonomists to mediate the understanding of what such aggregations of data tell us. This is where the lead taken by Avibase is important because within Avibase each such concept is mapped so that the scope of each concept attached to one and the same name and can be seen clearly. Nine different taxonomic authorities can be compared within this resource, which contains 14 million records of about 10,000 species and 22,000 subspecies of birds.

Filling the gap – the efforts made so far

The compilation of institutional indexes or databases of avian names, such as the card index developed by Richmond at the Smithsonian Institution or a similar one for recent avian names maintained by the Department of Ornithology of the American Museum of Natural History, New York (AMNH) until the late 1990s, have probably all been discontinued due to pressures on personnel and a focus for computerisation on collection holdings and thus data-capture focussed on specimen registers.

However, such card indexes have proved their value. Drawing on the AMNH card index, and with extra inputs from Norbert Bahr, he and I developed a list of new avian names since the volumes of Peters's *Check-list*, which, by 2001, allowed the 2003 edition of the *Howard and Moore complete checklist of the birds of the world* (Dickinson 2003) to mention every name then known to have been proposed since Peters *Check-list*, to which it became a vital companion work. No other recent global ornithological checklist had previously set out to do this. All the subspecies names in that 2003 checklist were supplied to Alan Peterson, and he now has the best-organised database of ornithological names known to me (although it is incomplete). He makes this available on www.zoonomen.net.

Alongside his database, which is focussed on names in use, Peterson displays scans of all the cards of the principal card index developed by Richmond thus providing information on the both original citations of names in use and of synonyms that found their way into Richmond's card index.

Because of both the relative completeness of Peterson's data and the way he has it organised it, this data should be the basis for preliminary population of ZooBank with avian names. He already makes his content available to ITIS (the Integrated Taxonomic Information System – www.itis.gov) and the Encyclopedia of Life (www.eol.org). However, only in ZooBank is it seriously likely that eventually there will be carefully-structured validation of all these names; and also ZooBank is perhaps the most logical repository to promote to encourage ornithologists to strive to create for comprehensive coverage at an even higher level, for example by including family-group names. This is a collaborative task and it should be complementary and enabling for LANs (Lists of Available Names).

As in other zoological fields, there remain bibliographic problems that have not been resolved. Dickinson et al. (2011) listed and discussed 148 books containing new names in ornithology (and sometimes other fields) that have presented, and in some cases still present, problems of accurate dating. Although such cases were often explored by Sherborn, or other later authors, evidence unknown to them is still being found and changing our understanding of the correct dates to attribute to taxon names. Detective work in this field can be very rewarding. I explain briefly below four cases where recent studies have either caused dates to change (or to be retained – although known to be wrong).

The *Nouveau recueil de planches coloriées* of Conraad Temminck and Meiffren Laugier (1820-1839): Sherborn (1898) made us aware that of the 101 parts of this work the first twenty initially lacked texts and that in these cases because the plates gave only French vernacular names it was necessary to find the wrappers of each part to discover, and be sure of, the spelling of any new original name. Wrappers are the encasing parts produced separately but to be grouped for binding (sometimes obviously based on pagination, sometimes in a format instructed later). These often include specific publication information that is not included once the volumes are bound. Wrappers may include the only precise indications of publication dates and thus can be very useful in bibliographic detective work. Of these twenty wrappers he knew of just two. Over a century later Dickinson et al. (2011) knew of no more, but later the same year all the remaining wrappers began to surface and they were photographed and published (Lebossé and Bour 2011, Dickinson 2012). However, another problem plagued this work. Temminck had promised that each part would contain six plates but 600 plates, the full complement, required 101 parts to be published and which parts had fewer than six had not been determined so that numerous citations were potentially wrong. This riddle was resolved by Dickinson (2001), and proved by Temminck's handwritten list in the archives of the museum in Leiden, meaning that each plate is now associated with just one part and thus date of publication.

The monograph on pigeons and doves with text by Temminck and plates by Pauline Knip (née de Courcelles) in 1808–1811: although it was generally known that the artist had in some way made herself appear to be the mainspring of this work, exactly what happened and how this might affect dates of publication or citations was unclear. Here, evidence was available but it had not been compared and understood. The full details of what turned out to be a fraud were published by Dickinson et al. (2010) and several dates of publication had to be changed. It was concluded that Temmick

(1808–1810) was the sole author of the first 13 livraisons (one of the numbers of a book published in parts) under the title *Histoire naturelle générale des pigeons* and that the last two livraisons, under the title *Les Pigeons* must be credited to Knip and Temminck (1811) (the latter alone being responsible for any new names).

The *Catalogue of the birds of the Peninsular of India* by Thomas C. Jerdon: began as serialised parts in the *Madras Journal of Literature and Science* but the [first] Supplement, in part 30 of that journal, was long delayed and we found that Jerdon had had a 200 page catalogue privately printed in 1840 or 1841 which included the published parts and the delayed part, and that this was three or four years before its formal appearance in the journal in 1844. The dates of several new names from that supplement were thus advanced to 1841. See Dickinson et al. (2004).

The *Histoire naturelle des oiseaux de l'Amerique septentrionale* of Vieillot: was discussed by Richmond (1899) who mistakenly believed, and seems to have led Browning and Monroe (1991) to believe, that the first part was published on 1 December, 1807. In fact what Richmond saw was merely an announcement that appeared four weeks after that saying it "will be published" on that date. A later notice in July 1808 changes the story and said that it "will be published" on 1 September 1808. The date of each part that was suggested by Browning and Monroe was a merely a projection of a very rapid, and almost certainly over-optimistic, timetable based on a start date that was at least nine months out. Indeed, based on the delay of the first part, it is extremely unlikely that subsequent issues followed on schedule every month. But here we have no correct dates to use and for the moment the incorrect dates are retained to avoid premature changes to nomenclatural stability (Dickinson 2011).

Of the many periodicals also discussed by Dickinson et al. (2011) the two most complex problem cases appear to be the *Journal für Ornithologie* and the *Proceedings of the Academy of Natural Sciences, Philadelphia*. The dates of many of the early issues of both are doubtful, because some issues are known to have been delayed. A complete picture of which issues were delayed and which were published on time requires detailed painstaking work and may well still be far from resolution. Numerous other periodicals, with fewer new avian names in them, also await serious review.

Articles on subjects like these were quite plentiful in the early years of the *Journal of the Society for the Bibliography of Natural History* (now the *Archives of Natural History*), up to approximately 1980, by authors such as C.E. Cowan, F.J. Griffin, M. Guédès, F. Hemming, L.G. Higgins, M.E. Jahn, R.I. Johnson, W.L. McAtee, N.F. McMillan, H.S. Marshall, E.C. Nelson, J.H. Price, F.C. Sawyer, C.D. Sherborn, W.T. Stearn, J.C. Thackray, A.C. Townsend, A. Wheeler and P.J.P. Whitehead (names sourced from the list of authors in Nelson 2001), but unfortunately such detective work, ideal for retired zoologists with good bibliographic skills and an appetite for rigour, attracts few recruits these days. The decline in such publications between 1980 and 1989 is shown in Fig. 2.

However, there has been a resurgence of interest in the subject at the end of the millennium as zoologists realise that digital tools can make this kind of bibliographic foundation easier to build and more important than ever for the age of biodiversity bioinformatics. In this period undoubtedly the most influential author has been Neal

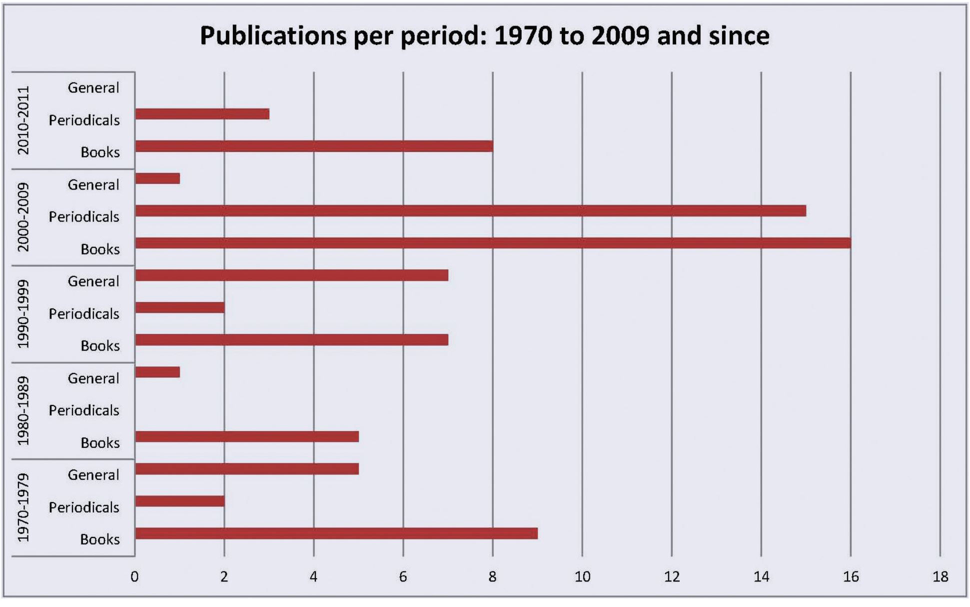


Figure 2. Bibliographic publications per decade, 1970 to late 2011, by subject. Treating books, periodicals and general works such as catalogues (source: references cited by Dickinson et al. 2011).

Evenhuis (e.g., Evenhuis 2003a, 2003b, 2008), but others such as Kraig Adler, Kees Rookmaaker and Florence Pieters have contributed, as has the present author. Lyal (2016) outlines how legacy zoological literature should be most effectively digitised and utilised.

Filling the gap – what remains to be done

Looking to the future and assuming that ZooBank becomes the long-term solution to finding all avian names, this will require its population, verification of such retrospective imported records and a programme to complete the entries of old names. This programme might well be started as part of the development of Lists of Available Names (LANs). I suggest first a LAN for avian family names, then for avian genus-group names and eventually for species-group names. This is being facilitated by the dissemination of full guidelines for the submission to the I.C.Z.N. of such lists, and for their consideration and potential adoption (Alonso-Zarazaga et al. 2016).

Bock's (1994: 121–123) reservations relating to the previous Code are still of relevance to these guidelines. In a nutshell, Bock argued that meeting the original expectations of the Commission required a wholly disproportionate time commitment. His six or more years of research led him to list 1400 names (although perhaps 15% of these need not be counted again as they reflect the operation of the Principle of

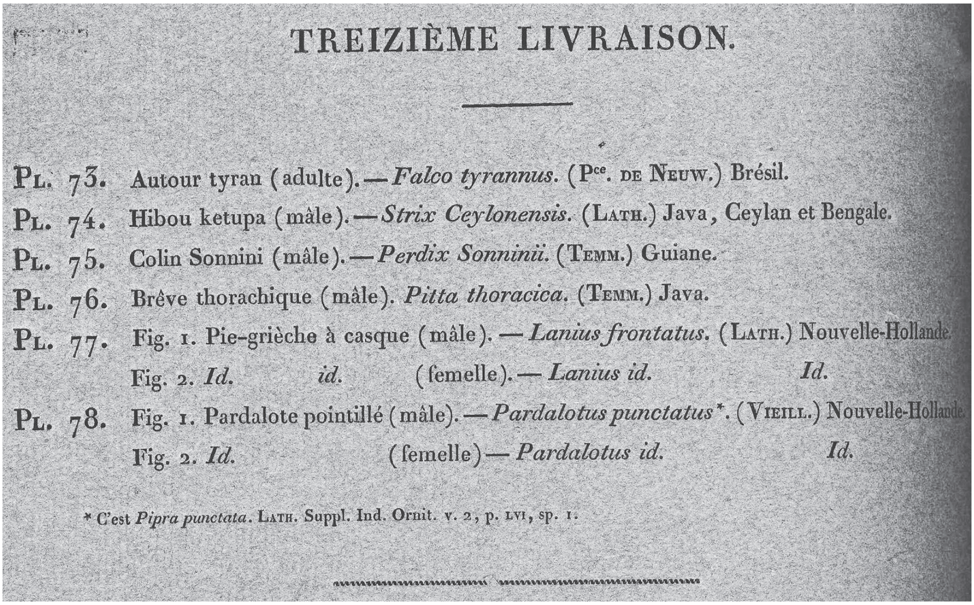
Coordination, explained in Art. 36 of the Code). Much time was indeed committed, but it is not clear that this was disproportionate to what might have been expected or to the value of what he produced. Behind Bock's observation was an acute awareness of the size of the corpus of serious ornithological literature, and it is fairly certain that that body has grown faster in the last 20 years. While much of the older, rarer, content may now be easier to access, thanks to BHL, some later works will be harder to access due to copyright restrictions.

In parallel, the ornithological community needs to set up and sustain a collaborative process to see that all new names do get added to ZooBank. However much encouraged, publishers and authors will not all register what they should unless registration becomes mandatory for a name to be validly published.

Discussions aimed at stimulating collaborative work on avian generic synonymies began over five years ago and numerous promises of help on specific families have been received and will be taken up! This suggests that this need is well understood and becoming more so due to the increasing quality of molecular studies and of the interpretations of their results. In addition, continued bibliographic research should be encouraged, and this is no doubt true for other fields of zoology so that interdisciplinary collaboration and data-sharing will be highly desirable.

The wrappers of books published as part-works, so often discarded when the work was complete and sent to the binder, are now often either completely missing or unrecorded. There is an enormous need to pool information on extant wrappers, at least for those critical to the dating of new taxa or to the spelling of their original names. Ideally we need illustrations of all such wrappers to be scanned for the Biodiversity Heritage Library (BHL), especially when the displayed content on the BHL website will be misinterpreted without such illustration. My favourite example is, of course, the "*Planches Coloriées*" of Temminck and Laugier, mentioned earlier. This is a special case because the first 20 parts (120 plates) appeared before the texts issued for them; no scientific names were on the plates, they appeared only on the wrappers accompanying each set of six plates (see Fig 3). Happily all twenty wrappers have now been located, with the images of their lists of contents published (Lebossé and Bour 2011, Dickinson 2012) and are available for use by BHL.

The wording of Art. 12.2.7 of the Code implies that a combination of text on a wrapper and a plate linked to a vernacular name on both meets the requirements for valid introduction of a name before 1931, so that the new names in these wrappers date from the issue of the plates. In *Index Animalium* Sherborn was not consistent in dating names from the *Planches Coloriées*. Years earlier Sherborn (1898) provided dates for each part, building on what was clear from the weekly issues of the *Bibliographie de la France*, but in *Index Animalium* he did not consistently apply those dates to the names in the first 20 parts; quite often he used dates that appear to come from published findings relating to when the subsequent texts for when those parts appeared. The discovery of the full set of wrappers permits consistent recognition of the original spellings from the wrappers and encourages dating from their appearance and not from the date of issue of the later text. However, what Sherborn did here may



TREIZIÈME LIVRAISON.

Pl. 73. Autour tyran (adulte). — *Falco tyrannus.* (Pce. de Neuw.) Brésil.
Pl. 74. Hibou ketupa (mâle). — *Strix Ceylonensis.* (Lath.) Java, Ceylan et Bengale.
Pl. 75. Colin Sonnini (mâle). — *Perdix Sonninii.* (Temm.) Guiane.
Pl. 76. Brêve thorachique (mâle). *Pitta thoracica.* (Temm.) Java.
Pl. 77. Fig. 1. Pie-grièche à casque (mâle). — *Lanius frontatus.* (Lath.) Nouvelle-Hollande.
Fig. 2. *Id.* *id.* (femelle). — *Lanius id.* *Id.*
Pl. 78. Fig. 1. Pardalote pointillé (mâle). — *Pardalotus punctatus**. (Vieill.) Nouvelle-Hollande.
Fig. 2. *Id.* (femelle) — *Pardalotus id.* *Id.*

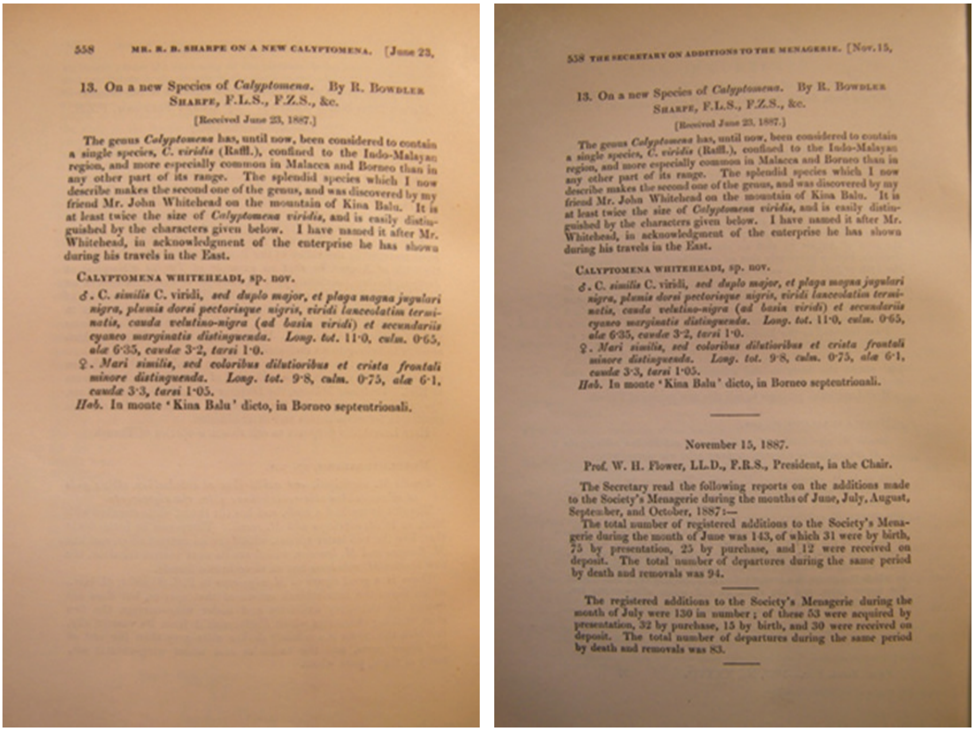
* C'est *Pipra punctata*. Lath. Suppl. Ind. Ornit. v. 2, p. lvi, sp. 1.

Figure 3. The wrapper for the 13th part of the *Planches Coloriées* showing vernacular and scientific names. All the wrappers for the first twenty parts, all that were published without supporting texts, have been depicted in *Zoological Bibliography*: 10 in vol. 1, No. 4, pp. 144–148 and 10 in Vol. 2, No. 1, 37–41 pp. These pages were published under the terms of the Creative Commons Attribution Licence and may be reproduced with proper attribution.

suggest that there are problems with the dating, in *Index Animalium*, of other older books that appeared in parts (Welter-Schultes et al. 2015; Dickinson et al. in press). Again it becomes clear that all existing datasets have some problems and that verification of name-records in ZooBank will require triage, with complex works referred to expert bibliographers.

It is also desirable that explanations of the findings drawn from all sets of wrappers be published as recently done by Lyal (2011) in the case of the *Biologia Centrali-Americana*. The same is true for research into dates for runs of journals; see, for example, Poggi (2010).

As a community, we must also encourage the managers of the Biodiversity Heritage Library (BHL) to help. For example, when journals – or books which were part-works – are scanned, every effort should be made to locate wrappers and to scan and display these alongside the content. The wrappers found by Lyal (2011) are available at the Natural History Museum and could be scanned. The photographs of the Temminck and Laugier wrappers can be made available to the Biodiversity Heritage Library (BHL, http://www.biodiversitylibrary.org/) without charge. Others may need to be specially photographed. Ultimately, notices of what is lacking in this regard should be circulated widely to prompt institutional libraries to locate and offer rare material of this kind from their holdings for scanning.



558 MR. R. B. SHARPE ON A NEW CALYPTOMENA. [June 23,

13. On a new Species of *Calyptomena*. By R. Bowdler Sharpe, F.L.S., F.Z.S., &c.

[Received June 23, 1887.]

The genus *Calyptomena* has, until now, been considered to contain a single species, *C. viridis* (Raffl.), confined to the Indo-Malayan region, and more especially common in Malacca and Borneo than in any other part of its range. The splendid species which I now describe makes the second one of the genus, and was discovered by my friend Mr. John Whitehead on the mountain of Kina Balu. It is at least twice the size of *Calyptomena viridis*, and is easily distinguished by the characters given below. I have named it after Mr. Whitehead, in acknowledgment of the enterprise he has shown during his travels in the East.

Calyptomena whiteheadi, sp. nov.

♂. C. *similis* C. viridi, *sed duplo major, et plaga magna jugulari nigra, plumis dorsi pectorisque nigris, viridi lanceolatim terminatis, cauda velutino-nigra (ad basin viridi) et secundariis cyaneo marginatis distinguenda. Long. tot.* 11·0, *culm.* 0·65, *alæ* 6·35, *caudæ* 3·2, *tarsi* 1·0.

♀. *Mari similis, sed coloribus dilutioribus et crista frontali minore distinguenda. Long. tot.* 9·8, *culm.* 0·75, *alæ* 6·1, *caudæ* 3·3, *tarsi* 1·05.

Hab. In monte 'Kina Balu' dicto, in Borneo septentrionali.

558 THE SECRETARY ON ADDITIONS TO THE MENAGERIE. [Nov. 15,

13. On a new Species of *Calyptomena*. By R. Bowdler Sharpe, F.L.S., F.Z.S., &c.

[Received June 23, 1887.]

The genus *Calyptomena* has, until now, been considered to contain a single species, *C. viridis* (Raffl.), confined to the Indo-Malayan region, and more especially common in Malacca and Borneo than in any other part of its range. The splendid species which I now describe makes the second one of the genus, and was discovered by my friend Mr. John Whitehead on the mountain of Kina Balu. It is at least twice the size of *Calyptomena viridis*, and is easily distinguished by the characters given below. I have named it after Mr. Whitehead, in acknowledgment of the enterprise he has shown during his travels in the East.

Calyptomena whiteheadi, sp. nov.

♂. C. *similis* C. viridi, *sed duplo major, et plaga magna jugulari nigra, plumis dorsi pectorisque nigris, viridi lanceolatim terminatis, cauda velutino-nigra (ad basin viridi) et secundariis cyaneo marginatis distinguenda. Long. tot.* 11·0, *culm.* 0·65, *alæ* 6·35, *caudæ* 3·2, *tarsi* 1·0.

♀. *Mari similis, sed coloribus dilutioribus et crista frontali minore distinguenda. Long. tot.* 9·8, *culm.* 0·75, *alæ* 6·1, *caudæ* 3·3, *tarsi* 1·05.

Hab. In monte 'Kina Balu' dicto, in Borneo septentrionali.

November 15, 1887.

Prof. W. H. Flower, LL.D., F.R.S., President, in the Chair.

The Secretary read the following reports on the additions made to the Society's Menagerie during the months of June, July, August, September, and October, 1887:—

The total number of registered additions to the Society's Menagerie during the month of June was 143, of which 31 were by birth, 75 by presentation, 25 by purchase, and 12 were received on deposit. The total number of departures during the same period by death and removals was 94.

The registered additions to the Society's Menagerie during the month of July were 130 in number; of these 53 were acquired by presentation, 32 by purchase, 15 by birth, and 30 were received on deposit. The total number of departures during the same period by death and removals was 83.

Figure 4. An example of two-state publication. Here the *Proceedings of the Zoological Society of London*, 1887, demonstrates a two-state situation. Left image: "first state". Right image: "second state". Images from Dickinson et al. (2011: 36) reproduced with permission from The Natural History Museum, London.

As regards journals, date research is easiest with sets in which the issue wrappers have been bound in (preferably in place rather than at the end of the volume). While the practice of binding these at the end is usually sufficient, it is unsafe to assume this is definitive. For many older journals issues did not actually end where the bound volume may suggest! Many journals that had a page or two of the final signature, of say eight pages, blank, later began the signature again at the start of the next issue so that pages would run on smoothly. In other words, some pages appeared in two "states", one with empty space on the page and one with that space filled. Examining such pages in their second state can falsely convince the reader that part of an article appeared in the previous issue, see, for example, Dickinson (2004) as illustrated above (Fig. 4). But even this can be deceptive because sometimes some text not accommodated in the final signature first appeared on the back wrapper before re-appearing on the first page or so of the first signature of the next issue!

To the extent that the Code (ICZN 1999) provides rules governing dates of publications, it takes priority seriously in recognition of its importance both for synonymies and for the validity of a name when precedence is shown to be an issue potentially requiring change. However, these rules should be rather more detailed and in particular should be less ambiguous. Suggestions for consideration when the Code is revised will be found in Dickinson et al. (2011: 15–23).

Editorial concern for the provision by journals of accurate dates of publication has declined although it is not clear that there was good reason for change. In 1990 the *Ibis*, the United Kingdom's senior periodical in ornithology, ceased the provision (annually in arrears) of day-dates of publication of the four issues per year. Since then, and prior to 2012 and the changed relevance of electronic publication, on more than one occasion release of a January issue, which included the introduction of the name of a new taxon, actually occurred in the previous December. One such case has been acknowledged editorially the other has not. Thus one can be backdated, but to backdate the other requires retained proof of a date of receipt. Other journals in ornithology have had similar lapses – perhaps when commercial publishers unfamiliar with the Code take over the production and publication. However, it is not apparent that commercial publishers are worse than associations or even institutions such as museums. What has changed is that most student zoologists get no teaching time on the subject of nomenclature and the Code, and nomenclature is seen as a tiresome inconvenience rather than a tool designed to promote international dialogue through provision of as much stability in nomenclature as taxonomic change will allow.

The opportunity to register in ZooBank would seem to provide for the accurate future determination of dates of publication of works from 1 January 2012 onwards. However, this will only be true if the precise rules published in 2011 (ICZN 2011) are fully respected; in addition it will depend on an involvement by the publisher. In the case of retrospective registration it will be essential that there is provision for a date to be corrected and this may not happen during basic validation so allowance must be made for research results to be considered later and, when convincing, for them to be taken into use in ZooBank.

Summary and conclusion

The tools that will facilitate needed future work are essentially available as a normal part of the array of digital programmes (e.g. web tools, spreadsheets and databases). However, they will certainly need some elaboration. Ornithology as a whole must take a look at itself and determine how collaboration can be mobilised and put behind such work. If this remains the task of a few individual enthusiasts then the challenges described here will not be met any time soon. Because nomenclature is the unsung foundation for taxonomy, it is taxonomic work, and thus accurate description of the living world, that will then ultimately suffer.

Acknowledgements

Grateful thanks for comments and corrections go to Steven Gregory, James Jobling, Ellinor Michel, Leslie Overstreet, Florence Pieters and Frank Steinheimer and two anonymous reviewers.

References

Alonso-Zarazaga MA, Fautin DG, Michel E (2016) The List of Available Names (LAN): A new generation for stable taxonomic names in zoology? In: Michel E (Ed.) Anchoring Biodiversity Information: From Sherborn to the 21st century and beyond. ZooKeys 550: 225–232. doi: 10.3897/zookeys.550.10043

Banks RC (2004) Ornithological nomenclature. In: del Hoyo J, Elliott A, Christie D (Eds) Handbook of the birds of the world, 9. Cotingas to pipits and wagtails. Lynx Edicions, Barcelona, 13–25.

Bock WJ (1994) History and nomenclature of avian family-group names. Bulletin of the American Museum of Natural History 222: 1–281.

Brisson (1760) Ornithologia sive synopsis methodica sistems Avium visionem in ordines Sectiones, Genera, Species ipsarumque Varietates. Cum accurata cujusque speciei descriptione, Citationibus Auctorum de iis tractantium, Nominibus eis ab ipsis E Nationibus impositis, Nominibusque vulgaribus. 3 vols, Leiden.

Browning MB, Monroe BL Jr (1991) Clarifications and corrections of the dates of issue of some publications containing descriptions of North American birds. Archives of Natural History 18(3): 381–405. doi: 10.3366/anh.1991.18.3.381

Cory CB, Hellmayr EC, Conover HB (1918–1950) Catalogue of birds of the Americas and the adjacent islands in the Field Museum of Natural History including all species and subspecies known to occur in North America, Mexico, Central America, South America, the West Indies, and islands of the Caribbean Sea, the Galapagos Archipelago, and other islands which may be included on account of their faunal affinities. Field Museum of Natural History. Zoology Series, 13. [15.vols]

Dickinson EC (2001) Systematic notes on Asian birds. 9. The "Nouveau recueil des planches coloriées" of Temminck & Laugier (1820–1839). Zoologische Verhandelingen 335: 7–54.

Dickinson EC (2003) The Howard & Moore complete checklist of the birds of the world. Christopher Helm, London.

Dickinson EC (2004) Confirmation of the date of description of a broadbill (*Calyptomena*, Aves) published in the Proceedings of the Zoological Society of London. Archives of Natural History 31(1): 168–169.

Dickinson EC (2011) Vieillot's Histoire naturelle des oiseaux de l'Amerique septentrionale, depuis Saint Domingue jusqu'à la baie de Hudson; contenant un grand nombre d'espèces décrites ou figurées pour la première fois. Zoological Bibliography 1(3): 136.

Dickinson EC (2012) The first twenty livraisons of '*Les Planches Coloriées d'Oiseaux*' of Temminck & Laugier (1820–1839): IV. Discovery of the remaining wrappers. Zoological Bibliography 2(1): 35–49.

Dickinson EC, Bruce MD, Gregory S, Peterson AP, Pittie A (2004) The dating of names proposed in the first Supplement to Thomas Jerdon's Catalogue of the birds of the peninsula of India. Bulletin of Zoological Nomenclature 61: 214–224.

Dickinson EC, Christidis L (2014) The Howard & Moore complete checklist of the birds of the world – Passerines. Aves Press Limited, Eastbourne.

Dickinson EC, David N, Bruce MD (in press) A review of the dates of publication of birds newly described from zoological collection on the "Voyage de la Coquille" by Lesson & Garnot. Zoological Bibliography 3.

Dickinson EC, David N, Overstreet LK, Steinheimer FD, Jansen JJFJ (2010) Histoire naturelle des pigeons or Les Pigeons; *Temminck versus* Knip. Archives of Natural History 37(2): 203–220. doi: 10.3366/anh.2010.0003

Dickinson EC, Overstreet LK, Dowsett RJ, Bruce MD (2011) Priority! The dating of scientific names in ornithology. A directory to the literature and its reviewers. Aves Press Limited, Northampton.

Dickinson EC, Remsen JV Jr (2013) The Howard & Moore complete checklist of the birds of the world – Non-passerines (4th. ed., vol. 1). Aves Press Limited, Eastbourne, 461 pp.

Evenhuis NL (2003a) Publication and dating of the journals forming the Annals and Magazine of Natural History and the Journal of Natural History. Zootaxa 385: 1–68.

Evenhuis NL (2003b) Dating and publication of the Encyclopédie Méthodique (1782-1832), with special reference to the parts of the Histoire Naturelle and details of the Histoire Naturelle des Insectes. Zootaxa 166: 1–48.

Evenhuis NL (2008) Preliminary catalog of dating sources for zoological works. Bishop Museum Technical Report 47: 1–172.

Evenhuis NL (2016) Charles Davies Sherborn and the "Indexer's Club". In: Michel E (Ed.) Anchoring Biodiversity Information: From Sherborn to the 21st century and beyond. ZooKeys 550: 13–32. doi: 10.3897/zookeys.550.9697

Hartert E (1909–23) Die Vögel der paläarktischen Fauna. R. Friedlander & Sohn, Berlin.

Hartert E, Steinbacher F (1932–34) Die Vögel der paläarktischen Fauna. Ergänzungsband. R. Friedlander & Sohn, Berlin.

ICZN (1905) International Commission for Zoological Nomenclature – Règles internationales del la nomenclature zoologique. International code of zoological nomenclature, Rudeval, Paris.

ICZN (1987) International Commission for Zoological Nomenclature – Official lists and indexes of names and works in zoology. The International Trust for Zoological Nomenclature, London.

ICZN (1999) International Commission for Zoological Nomenclature – International code of zoological nomenclature (4th ed). The International Trust for Zoological Nomenclature, London.

ICZN (2001) International Commission for Zoological Nomenclature – Official lists and indexes of names and works in zoology. Supplement 1986–2000. The International Trust for Zoological Nomenclature, London.

ICZN (2012) International Commission for Zoological Nomenclature – Amendment of Articles 8, 9, 10, 21 and 78 of the International Code of Zoological Nomenclature to expand and refine methods of publication. Bulletin of Zoological Nomenclature 69(3): 161–169.

Jerdon TC (1839 [= 1841]) Catalogue of the birds of the Peninsular of India, arranged according to the modern system of classification, with brief notes on their habits and geographical distribution, and descriptions of new, doubtful and imperfectly described species. JB Pharoah, Madras.

Knip P, Temminck CJ (1811) Les Pigeons. Livraisons 14 and 15. Paris.

Lebossé A, Bour R (2011) The first twenty livraisons of "Les Planches Coloriées d'Oiseaux" of Temminck & Laugier (1820–1839): I. The ten wrappers now known. Zoological Bibliography 1(4): 141–150.

Lepage D, Vaidya G, Guralnick R (2014) Avibase – a database system for managing and organizing taxonomic concepts. ZooKeys 420: 117–135. doi: 10.3897/zookeys.420.7089

Linnaeus C (1758) Systema naturae per regna tria naturae, secundum classes, ordines, genera, species, cum characteribus, differentiis, synonymis, locis (10th edition). Laurentii Salvi, Holmiae.

Linnaeus C (1766) Systema naturae per regna tria naturae, secundum classes, ordines, genera, species, cum characteribus, differentiis, synonymis, locis (12th edition). Holmiae.

Lyal CHC (2011) The dating of the Biologia Centrali-Americana. Zoological Bibliography 1(2): 67–100.

Lyal CHC (2016) Digitising legacy zoological taxonomic literature: Processes, products and using the output. In: Michel E (Ed.) Anchoring Biodiversity Information: From Sherborn to the 21st century and beyond. ZooKeys 550: 189–206. doi: 10.3897/zookeys.550.9702

Mathews GM (1925) The Birds of Australia (Supplement 5). Bibliography of the Birds of Australia. Books used in the preparation of this work with a few biographical details of authors and collectors. H.F. & G. Witherby, London.

Mathews GM (1927) Systema Avium Australasianarum. A systematic list of the birds of the Australasian region. Pt. I. British Ornithologists' Union, London.

Mathews GM (1930) Systema Avium Australasianarum. A systematic list of the birds of the Australasian region. Pt. II. British Ornithologists' Union, London.

Mayr E (1957) New species of birds described from 1941 to 1955. Journal für Ornithologie 98(1): 22–35. doi: 10.1007/BF01677166

Mayr E (1971) New species of birds described from 1956 to 1965. Journal für Ornithologie 112: 302–316. doi: 10.1007/BF01640689

Mayr E, Cottrell GW (1979) Introduction. In: Mayr E, Cottrell GW (Eds) Check-list of birds of the world. Revision of the work of James L. Peters, 1. Museum of Comparative Zoology, Cambridge, Massachusetts.

Mayr E, Greenway JC Jr (1960) Introduction. In: Mayr E, Greenway JC Jr (Eds) Check-list of birds of the world. A continuation of the work of James L. Peters, 9. Museum of Comparative Zoology, Cambridge, Massachusetts.

Mayr E, Paynter RA Jr (1964) Introduction. In: Mayr E, Paynter RA Jr (Eds) Check-list of birds of the world. A continuation of the work of James L. Peters, 10. Museum of Comparative Zoology, Cambridge, Massachusetts.

Mengel RM (1972) A catalogue of the Ellis collection of ornithological books in the University of Kansas libraries (Vol. 1. A–B). Lawrence, Kansas.

Mengel RM (1983) A catalogue of the Ellis collection of ornithological books in the University of Kansas libraries (Vol. 2. C–D). Lawrence, Kansas.

Moss S (2009) Birding past, present and future – a global view. In: Hoyo J. del., Elliott A, Christie D (Eds) Handbook of the birds of the world, 14. Bush-shrikes to Old World sparrows. Lynx Edicions, Barcelona.

Neave SA (1939–40) Nomenclator zoologicus. A list of the names of the genera and subgenera in zoology from the tenth edition of Linnaeus 1758 to the year 1935 (etc.). The Zoological Society of London 1 (A–C) [1939]; 2 (D–L) [1939]; 3 (M–P) [1940]; 4 (Q–Z and Supplement) [1940]; 5 (Supplement 1936–1945) [1950].

Nelson EC (2001) Index to Journal of the Society for the Bibliography of Natural History vols 1-9, continued as Archives of Natural History vols. 10-20 (1936-2001). Cumulative index to the Archives of Natural History formerly Journal of the Society for the Bibliography of Natural History, volumes 1–28, 1936–2001, and other publications. Science History Publications Limited, Cambridge, 87–119.

Olson SL, Browning MR (1992) Introduction. The Richmond index to genera and species of birds. GK Hall & Co., Boston, Massachusetts. doi: 10.1007/978-1-4612-2960-5_1 [Leaflet issued with the microfiches]

Padial JM, Miralles A, de la Riva I, Vences M (2010) The integrative future of taxonomy. Frontiers in Zoology 7: 1–16. doi: 10.1186/1742-9994-7-16

Pallas PS (1764) Adumbratiunculae Avium variarum praecedenti Elencho insertarum, sed quae in Systemate Naturae Illustr. Linnaei nondum extant". In: Vosmaer A (Ed.) Catalogue raisoné d'une collection supérieurement belle d'Oiseaux, tant exotique qu'Européens, de Quadrupedes et d'Insectes. Empaillés & arrangés avec beaucoup d'art en situations & attitudes extrèmement naturelles, & granatis de la corruption d'une façon particulière. Le tout rassemblé & arrangé, pendant une longue suite d'années, avec beaucoup de peines & à grand frais, par M. A. Vroeg. Pieter van Os, s'Gravenhage, 1–7.

Peters JL and successors (1931–1986) Check-list of birds of the world. 15 vols. Harvard University Press (later the Museum of Comparative Zoology), Cambridge, Massachusetts.

Peters JL (1931) Introduction. Check-list of birds of the world. 1. Harvard University Press, Cambridge, Massachusetts.

Peters JL (1934) Introduction. Check-list of birds of the world. 2. Harvard University Press, Cambridge, Massachusetts.

Pilsk SC, Kalfatovic MR, Richard JM (2016) Unlocking Index Animalium: From paper slips to bytes and bits. In: Michel E (Ed.) Anchoring Biodiversity Information: From Sherborn to the 21st century and beyond. ZooKeys 550: 153–171. doi: 10.3897/zookeys.550.9673

Poggi R (2010) Gli "Annali" pubblicati dal Museo Civico di Storia Naturale "Giacomo Doria" di Genova, storia del periodico ed indice generali dei primo cento volumi (1870–2009). Annali del Museo Civico di Storia Naturale "Giacomo Doria" 101: 1–629.

Pyle RL (2016) Towards a Global Names Architecture: The future of indexing scientific names. In: Michel E (Ed.) Anchoring Biodiversity Information: From Sherborn to the 21st century and beyond. ZooKeys 550: 261–281. doi: 10.3897/zookeys.550.10009

Richmond CW (1899) On the date of Lacépède's 'Tableaux'. Auk 14(4): 325–329. doi: 10.2307/4069359

Richmond CW (1902) List of generic terms proposed for birds during the years 1890 to 1900, inclusive, to which are added names omitted by Waterhouse in his Index Generum Avium. Proceedings of the United States National Museum 24: 663–730. doi: 10.5479/si.00963801.1267.663

Richmond CW (1908) Generic names applied to birds during the years 1901 to 1905, inclusive, with further additions to Waterhouse's Index Generum Avium. Proceedings of the United States National Museum 35: 583–655. doi: 10.5479/si.00963801.1656.583

Richmond CW (1917) Generic names applied to birds during the years 1906 to 1915, inclusive, with additions and corrections to Waterhouse's Index Generum Avium. Proceedings of the United States National Museum 53: 565–636. doi: 10.5479/si.00963801.2221.565

Richmond CW (1927) Generic names applied to birds during the years 1916 to 1922, inclusive, with additions to Waterhouse's Index Generum Avium. Proceedings of the United States National Museum 70: 1–44. doi: 10.5479/si.00963801.2664

Richmond CW (1992) In: O'Hara RJ (Ed.) The Richmond index to genera and species of birds. GK Hall & Co., Boston, Massachusetts. [Microfiches]

Sauer GC (1982) John Gould the bird man. A chronology and bibliography. Henry Sotheran Ltd., London.

Schulze FE (1926–54) Nomenclator Animalium generum et subgenerum. 5 vols., and Sonderdruck (supplement). Preußischen Akademie der Wissenschaften zu Berlin, later the Deutschen Akademie der Wissenschaften zu Berlin, Berlin.

Sclater WL (1924) Systema Avium Æthiopicarum. A systematic list of the birds of the Ethiopian region. Pt. I. British Ornithologists' Union/Taylor & Francis, London.

Sclater WL (1930) Systema Avium Æthiopicarum. A systematic list of the birds of the Ethiopian region. Pt. II. British Ornithologists' Union/Taylor & Francis, London.

Sharpe RB et al. (1874–1898) Catalogue of the birds in the British Museum. 27 volumes. Trustees of the British Museum (Natural History), London.

Sharpe RB (1875) Catalogue of the birds in the British Museum. II. Catalogue of the Striges, or nocturnal birds of prey in the British Museum. Trustees of the British Museum, London.

Sharpe RB (1899–1909) A handlist of the genera and species of birds (Nomenclator avium tum fossilium tum viventum). 5 vols. Trustees of the British Museum, London.

Sherborn CD (1898) On the dates of Temminck and Laugier's Planches coloriées. The Ibis (7) 4: 485–488.

Sherborn CD (1902) Index Animalium sive index nominum quae ad A.D.MDCCLVIII generibus et speciebus animalium imposita sunt. Sectio prima. MDCCLVIII-MDCCC. Cambridge University Press, Cantabrigiae.

Sherborn CD (1922–33) Index Animalium sive index nominum quae ad A.D.MDCCLVIII generibus et speciebus animalium imposita sunt. Sectio secunda. MDCCI-MDCCCL. 33 parts. Trustees of the British Museum, London.

Shindler K (2016) A magpie with a card-index mind – Charles Davies Sherborn 1861–1942. In: Michel E (Ed.) Anchoring Biodiversity Information: From Sherborn to the 21st century and beyond. ZooKeys 550: 33–56. doi: 10.3897/zookeys.550.9975

Sibley CG (1965) Molecular systematics: new techniques applied to old problems. L'Oiseau et la Revue Française d'Ornithologie 35(Special): 112–124.

Sibley CG, Monroe BL Jr (1990) Distribution and taxonomy of birds of the world. Yale University Press, New Haven, Connecticut.

Stone W (1933) In memoriam: Charles Wallace Richmond 1868–1932. The Auk 50: 1–22. doi: 10.2307/4076543

Stresemann E (1959) The status of avian systematics and its unsolved problems. The Auk 76: 269–280. doi: 10.2307/4081807

Temminck CJ (1808–1810) Histoire naturelle générale des pigeons. Livraisons, Paris, 1–13. [For later livraisons see Knip & Temminck (above)]

Temminck CJ (1820–1839) Nouveau recueil de planches coloriées d'oiseaux, pour servir de suite et de complément aux planches enluminées de Buffon. 102 livraisons. Levrault, Dufour et d'Ocagne, Paris.

Vieillot LP (1807) Histoire naturelle des oiseaux de l'Amerique septentrionale, depuis Saint Domingue jusqu'à la baie de Hudson; contenant un grand nombre d'espèces décrites ou figurées pour la première fois. Paris.

Waterhouse FH (1889) Index Generum Avium. A list of the genera and subgenera of birds. RH Porter, London. doi: 10.5962/bhl.title.55246

Wilson EO (2005) Systematics and the future of biology. Proceedings of the National Academy of Sciences 102(Suppl. 1): 6520–6521. doi: 10.1073/pnas.0501936102

Zimmer JT (1926) Catalogue of the Edward E. Ayer ornithological library. Field Museum of Natural History, Zoology Series 16(1/2): 1–706.

Zimmer JT, Mayr E (1943) New species of birds described from 1938 to 1941. The Auk 60: 249–262. doi: 10.2307/4079651

ZooKeys 550: 135–152 (2016)
doi: 10.3897/zookeys.550.9447
http://zookeys.pensoft.net

RESEARCH ARTICLE

A peer-reviewed open-access journal
ZooKeys
Launched to accelerate biodiversity research

Sherborn's influence on *Systema Dipterorum*

F. Christian Thompson[1], Thomas Pape[2]

1 *Department of Entomology, Smithsonian Institution, Washington, DC 20560, USA* **2** *Department of Biosystematics, Natural History Museum of Denmark, Universitetsparken 15, 2100 Copenhagen, Denmark*

Corresponding author: *Thomas Pape* (tpape@snm.ku.dk)

Academic editor: *E. Michel* | Received 24 February 2015 | Accepted 24 February 2015 | Published 7 January 2016

http://zoobank.org/79E2FE38-BD4F-4049-ABDB-1D6B43676D52

Citation: Thompson CF, Pape T (2016) Sherborn's influence on *Systema Dipterorum*. In: Michel E (Ed.) Anchoring Biodiversity Information: From Sherborn to the 21st century and beyond. ZooKeys 550: 135–152. doi: 10.3897/zookeys.550.9447

Abstract

Flies make up more than 10% of the planetary biota and our well-being depends on how we manage our coexistence with flies. Storing and accessing relevant knowledge about flies is intimately connected with using correct names, and *Systema Dipterorum* provides a single authoritative classification for flies developed by consensus among contributors. The 160,000 species of flies currently known are distributed among 160 recent families and some 12,000 genera, which with their synonyms encompass a total of more than a quarter of a million names. These names and their associated classification are shared with relevant global solutions. Sherborn appears to have done remarkably well indexing Diptera names with an overall error rate estimated to be close to 1%.

Keywords

Flies, nomenclator, taxonomic catalog, identification, biodiversity informatics infrastructure, quality assurance standard

Introduction

Flies are ubiquitous and dominant in most terrestrial ecosystems, by their numbers of species as well as by their immeasurable myriads of individuals. Flies come in a multitude of forms and with an exceedingly vast array of life habits and are often considered the ecologically most varied of the insect orders. Fly larvae in particular are ecologically versatile and have adapted to the harshest of habitats, from pools of crude oil and torrential mountain streams to the bacterial soups of pit-latrines and vertebrate carrion. Flies flourish in the highly disturbed environments created by human activities, often reaching nuisance levels, and flies not surprisingly interfere with man in numerous and varied ways. On the dark side, flies dominate among the blood-sucking pests, with some of the most potent of human diseases being transmitted by dipterous vectors, thereby causing suffering that goes beyond description. However, flies are also beneficial, for example through their processing and recycling of the large quantities of surplus organic material produced by many of our modern societies (Pape 2009). Flies are essential for the successful pollination of many flowering plants, including several crops: without flies we would have no chocolate (Young 1994). Flies are important as both specialist and generalist pollinators, and in cooler or more shaded habitats flies are more important pollinators than bees (National Research Council Committee on the Status of Pollinators 2007, Ssymank et al. 2008). Actually, certain flies appear to be such critical pollinators that some authorities apparently think they are bees (O'Toole and Raw 2004, cover photo)!

The significance of flies reaches deep into our culture. Disease-carrying flies have had tremendous impact on local demography and land-use far beyond any other group of insects, and in the case of West African sickle-cell anemia, flies—even if mediated through a parasite—have reached into our very genome by indirectly favoring a specific genetic mutation (Sabeti et al. 2006, Higgs and Wood 2008). Mimetic flower flies enrich our lives by their stunning similarity to various bees and wasps, and *Drosophila* fruit flies tell us about evolution and genetics from preliminary school to the most advanced research in human genetic disorders. Flies may be seen as part of an 'extended phenotype' of our civilization, with the archetypical 'fly' being embedded as a key element in many old tales and myths (Kohler 1994, Courtney et al. 2009).

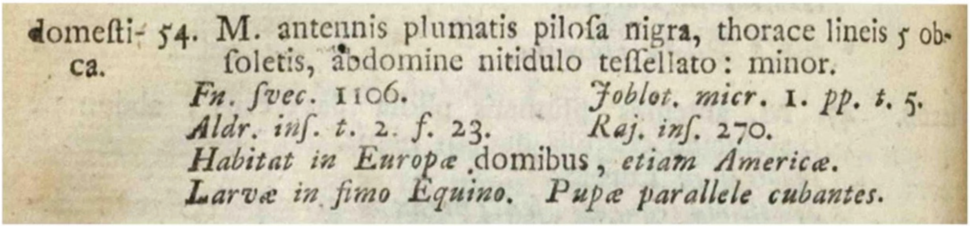
Flies are ancient. The earliest flies began diversifying in the Upper Triassic some 225 million years ago (Grimaldi and Engel 2005), and following three episodes of particularly rapid radiations, flies today constitute about 10% of the known (published) biodiversity, i.e., some 160,000 species (Wiegmann et al. 2011). This multitude of species is distributed among 160 recent families and some 12,000 genera (Pape et al. 2011; Pape and Thompson [online]). But estimates suggest that we may know as little as 10% of the species actually existing on our planet, and if that holds for Diptera, there may be one and a half million species of flies occurring worldwide, 90% of which have not yet been named.

Fly encyclopedia of past and present

What do people want to know about flies? People may be confronted with a fly that is strange to them, so they may want to know: "What is it?", "What does it do?" and "Where did it come from?" Or they may have a specific problem, for example with rotting oranges, and when learning that a fly maggot is the problem, they may want to know what fly it is. The resolution in each case starts with the identification of the fly, i.e., the first and crucial step leads to a name. With the right name, people get access to knowledge (Thompson 1996).

Knowledge about life (organisms over time), which was first essentially locked up in *Systema Naturae*, is now dispersed across hundreds of thousands of works, but maybe some day soon it will again be unified—or rather interconnected by means of an *Encyclopedia of Life* (http://www.eol.org). To build such an all-encompassing encyclopedia, we must first assemble the critical pieces of the biodiversity informatics infrastructure. Just as what we need for life in a modern society is transported on a system of airways, highways, seaways, and railways, biodiversity information must also be disseminated via a critical infrastructure. That infrastructure starts with nomenclators and catalogs, and the information is mediated by way of names. Information on identification and classification is disseminated through revisions and monographs. Most of the infrastructure of systematics remains in the traditional printed medium, but the migration to the online, digital medium of the Internet has begun. For flies, the first critical component of the biodiversity informatics infrastructure is the *Systema Dipterorum* (SD: see http://www.diptera.org; Evenhuis, Pape et al. 2010), a nexus of nomenclator and taxonomic catalog.

The past is our prologue as we build on the knowledge of our predecessors and learn from their mistakes. The official start of the modern understanding of flies and their classification has been deemed to be the 10th edition of Linnaeus' *Systema Naturae* in 1758 (ICZN 1999). Linnaeus produced the first comprehensive database of the natural world. He was first to apply a uniform set of names for all known life forms and to place those names into a "natural" hierarchical classification. Using a fixed suite of formal ranks, he divided the living world into three kingdoms, many classes, and even more orders, genera, and species. Flies were placed in the order Diptera (with some noteworthy exceptions) of the class Insecta. Linnaeus derived both his classification of flies and their name from Aristotle. For the order, the principal character Linnaeus used was the presence of only two wings, but he also noted the halteres, the modified second pair of wings, which is the most conspicuous unique autapomorphy of the dipteran clade. He divided the order into 10 genera and 191 species, and for each species he provided a single word, a specific name or epithet, then a diagnosis, distribution statement, summary of the biology, and references to where further information could be found. For flies, Linnaeus cited the works of 24 different authors. His entry for *Musca domestica* (Fig. 1) provides an example of his method, and this is largely what is still needed today by users of biodiversity information: a single, comprehensive, and authoritative reference work. Because Linnaeus was deemed the first, his nomenclature was by default perfect, but as knowledge of nature increased, the *Systema Naturae* quickly became difficult to maintain. Linnaeus' last edition of his *Systema* was the 12th (Linnaeus 1767). A German, Johann Friedrich Gmelin, took on the task of summarizing our knowledge of nature and produced a 13th edition of the *Systema* in a number of parts spanning the years 1783 to 1793 (with the flies appearing in Gmelin 1790). However, even before this, the students of Linnaeus had already decided to divide the work among themselves. Of those, Johann Christian Fabricius took responsibility for the insects and produced



domeſti- 54. M. antennis plumatis piloſa nigra, thorace lineis 5 ob-
ca. ſoletis, abdomine nitidulo teſſellato: minor.
Fn. ſvec. 1106. *Joblot. micr.* 1. *pp. t.* 5.
Aldr. inſ. t. 2. *f.* 23. *Raj. inſ.* 270.
Habitat in Europæ domibus, *etiam Americæ.*
Larvæ in fimo Equino. Pupæ parallele cubantes.

Figure 1. The entry for *Musca domestica* in *Systema Naturae*, 10th edition (Linnaeus 1758).

Table 1. Number of valid genera and valid species as well as total number of names for the three major Diptera nomenclators through time (fossils excluded).

	Linnaeus 1758	Fabricius 1805	*Systema Dipterorum* 2013
Genera	10	78	12,073
Species	191	1,151	160,042
Names	201	1,242	272,029
Authors	24	46	5,701

his *Systema Entomologiae* (Fabricius 1775). His system was identical in form to *Systema Naturae*, but was restricted to just the insects. Yet, as the number of insect species and the knowledge about them increased significantly over time, even this *Systema* on all insects could no longer be maintained. Fabricius realized this and, thus began a series of works devoted to one order each. Toward the end of his life, he provided the *Systema Antliatorum* (Fabricius 1805), devoted just to flies, fleas and a few other sucking insects, and in that work Fabricius changed the name of the order from Diptera to Antliata. Again, the format was the same as in Linnaeus' *Systema*, but with much more knowledge summarized. In 1805 Fabricius knew 1,151 species of flies, distributed among 78 genera as based on his own work and that of 46 other authors (Table 1). Unfortunately, Fabricius was not comprehensive as he did not include everything that was known at the time. Some 1,767 names were missing, represented in the works of 49 missing authors. After Fabricius, a growing cadre of naturalists took on the study of flies, and the number of species and names increased exponentially.

Comprehensive works like the *Systemae* of Linnaeus and Fabricius fell victim to the rapid increase in biodiversity information. We do understand what we have lost, and using modern technologies (computers and the Internet), we have begun to build their modern equivalents (e.g., Encyclopedia of Life, Species2000, ZooBank; see also Pyle 2016).

What information sources are currently available to help us build the new online *Systema Dipterorum*? As knowledge expanded and became larger than what one person could assimilate, as happened to Linnaeus and Fabricius, the universe of knowledge was subdivided into smaller, more manageable shares. This division was based on different approaches, some divided up the universe by the taxon, others by time or geography or people.

Division by taxon is simple: once Linnaeus (and then Gmelin) did all organisms, then Fabricius did all arthropods and finally all flies, but after that the division of labor was by geography, with Meigen doing all the flies of Europe and Wiedemann doing all the "exotic" (i.e., non-European) flies. Later, the amount of information became too large for comprehensive works, so a new format (catalogs) were invented and used as a summary of our knowledge, having merely a citation to the basic information. The first catalog for Diptera was produced by Osten Sacken (1858), who summarized knowledge about North American flies. Later Kertész (1902–1910) attempted to do this for the world but never finished. In recent times, dipterists have come together to produce regional catalogs of flies: America north of Mexico (Stone et al. 1965), Americas south of USA (Papavero 1966–1984; Amorim and Papavero 2008), Oriental (Delfinado and Hardy 1973–1977), Afrotropical (Crosskey 1980), Palaearctic (Soós and Papp 1984–1993), and Australasian and Oceanian regions (Evenhuis 1989). Add to that the fossils (Evenhuis 1994). With the continuing increase in papers describing new taxa and refining classifications, some of these regional treatments are greatly out-of-date.

Another division of knowledge was by the publications—works that included knowledge about organisms. One of the first to attempt making a bibliography of all works related to zoology (and geology) was Agassiz (1848–1854). Then Hagen (1862–1863) merely attempted to do all works related to arthropods up to and including the year 1862 (updated and revised by Horn and Schenkling 1928–1929), and he was followed by Derksen and Göllner-Scheiding (1965–1968) (with index by Derksen et al. 1975). These works covered knowledge-sources up to 1900. Evenhuis (1997) provided a selected index to some individual works on Diptera (not articles, but separately published "books") to 1930.

Just as bibliographies greatly facilitate our access to published works, a main portal to scientific names is embodied by the indexes built for those names. The monographs by Linnaeus and Fabricius were such indexes in addition to being taxonomic tools. After them, however, few comprehensive indexes were developed, and although high-profile initiatives are now underway to dynamically interconnect existing indexes in a way that streamlines the taxonomic enterprise (Pyle 2016), we are still vastly behind in aggregating content in interconnected systems.

Sherborn's contribution to dipterology

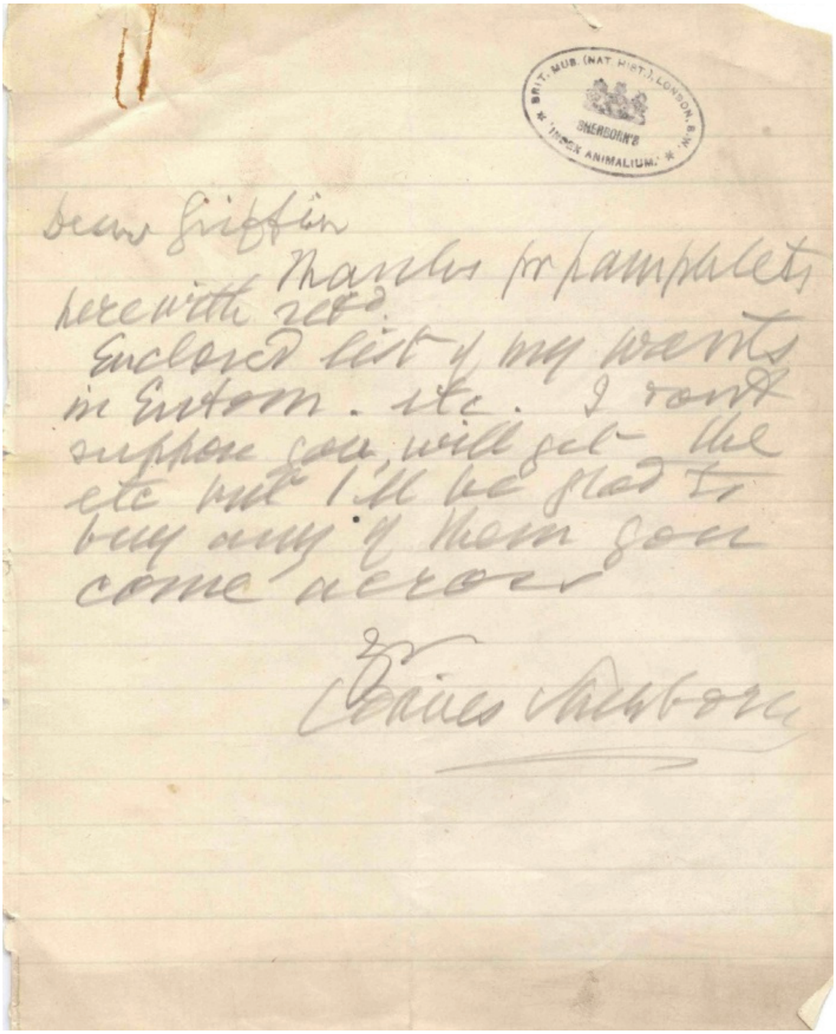
Sherborn (1902; 1922–1932) was the first to attempt a complete index of all scientific names relating to animal species (Evenhuis 2016). For practical reasons, he broke this task down by time. So, Sherborn first indexed all names published before 1800 and then those before 1850. Earlier zoologists, mainly curators at the British Museum (Natural History), had already agreed to develop and maintain an index of all new names applied to animals—the *Zoological Record*—which since 1864 has been an annual index of new names and taxonomic changes. All scientific names for species include two parts. One for the group that the author(s) feel the species belongs to (the genus-group name), and then the specific name itself (the epithet). Consequently, those genus-group names become a critical core to all scientific names. Hence, zoologists realizing that indexing all animal names would be a colossal undertaking decided to first concentrate on indexing the genus-group names. So, Louis Agassiz, who built the first bibliography of zoological works, also built a nomenclator of all genus-group names in Zoology (Agassiz 1846a,b). He was followed by Marshall (1873), Scudder (1882, 1884), Neave (1920–1996) and Schulze et al. (1920–1954). Sherborn, however, indexed both genus-group names as well as species-group names.

Today, how do we assess Sherborn's accomplishments? First, we need to appreciate the platform of nomenclatural legislation that Sherborn worked from. Today we have the 4th edition of the *International Code of Zoological Nomenclature*, but Sherborn worked under the 'Strickland Code' (Strickland et al. 1842) and he therefore evaluated names somewhat differently from today's standards. Second, while today we have computers, database software, etc., Sherborn worked only with slips of paper. Fortunately, Sherborn worked at the British Museum (Natural History) and, therefore, had access to the greatest library of scientific works on animals. Unfortunately, there were works that the British Museum did not have, so Sherborn was partially dependent on colleagues for helping him out for certain names. For example, Francis Griffin helped provide many rare entomological works (Fig. 2). Third, what did Sherborn produce? For each name, Sherborn provided six data elements (name, higher group [genus if species; order if genus], author, source [an abbreviated title and volume], year and page). He checked for two elements that in his view were crucial for validity—publication and diagnosis—and gave four data elements for bibliography (author, year, reference [title, volume, etc.] and place of publication). At the time Sherborn was working, his good friend Bernard Barham Woodward was preparing the catalog of the library of the British Museum (Natural History), so Sherborn (who was helping his friend in the cataloging of the library) decided that certain bibliographic details were unnecessary and unfortunately abbreviated most information.

How good was Sherborn? For a study of the species-group names proposed in the genus *Musca*, Sherborn indexed 1,807 names. Thompson and Pont (1994) found only 3 errors and 1 name that he missed. For all genus-group names proposed for flies (Order Diptera) Sherborn indexed 1,959. We have found another 43 which he missed and another 17 which he missed due to changes in the ICZN (plus a large number of orthographic variants, which by an unfortunate addition to the 4th version of the ICZN are to be deemed unjustified emendations; Evenhuis, O'Hara et al. 2010; O'Hara et al. 2011). Thus, for Diptera Sherborn appears to have done even better than his overall average (Welter-Schultes et al. 2016).

From *Systema Naturae* to the Web

The impediments that made the continued updating of Linneaus' original *Systema* impossible were the inflexibility of printing and the increased cost of disseminating knowledge by printing with fixed types and reproducing text in ink on paper. Today,


BRIT. MUS. (NAT. HIST.), LONDON, S.W.
SHERBORN'S
INDEX ANIMALIUM

Dear Griffin
Thanks for pamphlets
herewith recd.
Enclosed list of my wants
in Entom. etc. I dont
suppose you will get all
etc but I'll be glad to
buy any of them you
come across
Yr
C Davies Sherborn

Figure 2. Letter from Sherborn to Griffin accompanying one of his "Want" lists. The text of the message says: "Dear Griffin Thanks for pamphlets herewith recd. Enclosed list of my wants in Entom. etc. I don't suppose you will get all etc but I'll be glad to buy any of them you come across. Yours
C. Davies Sherborn".

computers are taking over the physical aspects of printing and provide an easy means for integrating the past with current knowledge, and they also allow for alternative dissemination media beyond paper. The Internet with the World Wide Web is a relatively new and ever more dominating medium, allowing anyone anywhere with a computer and online access to receive information in real time from anywhere in the world. The modern workflow in monographic taxonomy is at least potentially greatly enhanced (Penev et al. 2010, Smith et al. 2013), and the challenge is now neither the production nor the media, but what society wants of our science: systematics. [The accepted term for taxonomy and nomenclature and other aspects of our science today is (bio) systematics, a term that is directly derived from *Systema Naturae*. That is, the science of inferring the system inherent in the natural world.]

A Chinese proverb ascribed to Confucius states that wisdom begins with applying the correct names ("*If names be not correct, language is not in accordance with the truth of things*"; cf. Legge 1971). Linnaeus (1737) similarly stated that if you do not know the names, the knowledge about things has no value. Clearly, names and knowledge are intimately connected and work together to create meaningful communication. So, to deliver and decipher biodiversity information about flies, we are first developing the SD as an authority for information about the names of all flies. The names are then organized into a classification (or a taxonomy) just as Linnaeus and Fabricius did in their *Systemae*. For flies, the SD provides a single authoritative classification developed by consensus among contributors and derived from a more comprehensive taxonomy, which includes information on the characters used to generate the scientific hypotheses underlying the classification. The names and their classification are shared with global solutions such as the *Integrated Taxonomic Information System* (http://www.itis.gov/) and *Species 2000 Annual Checklist of Life* (Roskov et al. 2013; http://www.sp2000.org/). A substantial amount of the taxonomic information still remains in the traditional print medium, but the situation is rapidly changing. This is partly because an increasing quantity is now being produced online, with prime examples being serials like *Zootaxa* (http://www.mapress.com/zootaxa), and *ZooKeys* (http://pensoftonline.net/zookeys), but few contemporary journals with taxonomic content lack an online edition, and a growing number are online only, like *Biodiversity Data Journal* (http://www.biodiversitydatajournal.com/) and *European Journal of Taxonomy* (http://www.europeanjournaloftaxonomy.eu/index.php/ejt). Add to that the massive amount of legacy data being digitized through the *Biodiversity Heritage* Library (http://www.biodiversitylibrary.org), *Google® Books* (http://books.google.com), *AnimalBase* (http://www.animalbase.org), and *Gallica* [Bibliotheque National de France] (http://gallica.bnf.fr).

Our survival and well-being ultimately depends on accumulated knowledge about the 'nuts and bolts' of the natural world, and perhaps nowhere else in the natural sciences do we find a greater variety of different 'units' than in the biological discipline of taxonomy. Ever since Linnaeus, the basic unit of biological classifications is the species, and with our living world containing an estimated 5–12 million species, the need for names obviously is paramount.

Classifications are merely hierarchical groupings, and in evolutionary biology the basic unit is the species. Panzer (1792–1813), at least for entomology, was the first to recognize that if information dissemination is focused on this unit so that each unit is separate and independent from others, then new information can easily be integrated and new classifications easily generated. Panzer did this in the format of small booklets, species-by-species, with text and image on facing pages (Fig. 3). Today this is recognized as the concept of the online species page, some of the first examples of which

Figure 3. An example of the 'species pages' found in Panzer (1792–1813); here his entry for *Musca arcuata* Fabricius, 1781.

were placed at the USDA Diptera Website in April 1996 and are now incorporated into the *Encyclopedia of Life* (http://www.eol.org).

The last but most important set of components of the biodiversity infrastructure for users is identification tools. These range from the early paper-based diagnoses provided by Linnaeus and Fabricius to modern interactive, image-rich expert systems that run on hand-held tablets and smartphones, which from being available only for more conspicuous species like birds and whales, now are rapidly expanding to include applets for categories like forest pest insects, tree fungi, mushrooms and broadleaf weeds. For the flies, there are still only a few examples of identification tools that have left paper as the medium for storing and conveying relevant information. Primary examples of CD-ROM based identification systems for flies are the *Fruit Fly Expert Identification System* (Thompson 1999) for the identification of pestiferous fruit fly species, *Agromyzidae of the World* (Dempewolf 2004) for the identification of agromyzids of economic importance, and *On-The-Fly* (Hamilton et al. 2006) for the identification of the families of Australian flies, but keys to local faunas, like the dacine Tephritidae of Malesiana (White and Hancock 2004) and the Asilidae of Germany (Geller-Grimm 2003), are also available. Works accessible through the Internet, and usually by way of the World Wide Web, are of increasing importance because of obvious advantages in easy access and versatility in continuous updating (e.g., Walter and Winterton 2007). Noteworthy examples are the *Canadian Journal of Arthropod Identification* (http://www.biology.ualberta.ca/bsc/ejournal/ejournal.html) with nine of 25 works being devoted to Diptera, the online pages for the pestiferous fruit flies (Carroll et al. 2006a,b), the *Keys to the Medically Important Mosquito Species* (Walter Reed Biosystematics Unit [online]) and *MOSCHweb* (Cerretti et al. 2012) with a matrix-based interactive key to Palaearctic tachinid genera. Additional examples worthy of mention are Dempewolf (2004 [online version], Agromyzidae), Fetzner (2005) and Young (2005) (Tipulidae), and Meiklejohn et al. (2012, Sarcophagidae); see also Winterton (2009).

Systema Dipterorum: today and tomorrow

The SD is designed as a comprehensive online information source for all the critical information about scientific names of flies and the basic information about species of flies. This system grew out of a vision of a group of dipterists who wanted to capitalize on the knowledge that had been generated in preparing a series of regional catalogs of Diptera, which began in the 1960s with a catalog of Nearctic (or rather North American) Diptera that involved Canadian and U.S. fly specialists (Stone et al. 1965). A catalog for the Diptera of the Americas south of the United States was also initiated at this time, but it remained incomplete (Papavero 1966–1984) until revived by Amorim and Papavero (2008) and is now largely completed. The Oriental Diptera were cataloged soon after (Delfinado and Hardy 1973–1977), closely followed by the Afrotropical Diptera (Crosskey 1980). In 1984, at the International Congress of Entomology in Hamburg, dipterists gathered to celebrate the start of the effort to catalog the Palaearctic Region (Soós and Papp 1984–1993), the largest and historically most complex region. Subsequent years focused on completing the cycle with an Australasian/Oceanian catalog (Evenhuis 1989) and starting a series of world catalogs. Funding was successfully obtained from the U.S. National Science Foundation to do that last regional catalog (Evenhuis 1989), and private funding contributed to the production of the world fossil fly catalog (Evenhuis 1994), but attempts to secure funding for further world catalogs have been unsuccessful. USDA provided pilot-test project funds to develop new technologies for an expert identification and information system for fruit flies (Thompson et al. 1993), which provided the basis for the current SD. The original database software used was based on a Wang proprietary COBOL data-management system, and later migrated to FileMaker Pro, the current software system.

The SD is today a fully online system containing all the critical information about the system itself and its contents. What follows here is merely a snapshot of what was available online as of October 2013. Irregularly, the SD (initially as the *BioSystematic Database of World Diptera*) is archived to CD-ROM via the *Diptera Data Dissemination Disk* series (Thompson and Evenhuis 1999, Norrbom and Thompson 2004). As segments of information are completed and peer-reviewed, they have been published in traditional print format in the series *Myia* (e.g., Woodley 2001, Brake and Thompson 2011, Mathis 2011, Mathis and Barraclough 2011, Mathis and McAlpine 2011, Mathis and Sueyoshi 2011). Today, the components online are the nomenclator and reference files and all the appropriate supporting documentation for the system. The only major components not yet online are the species interface and online editing facilities for specialists.

The nomenclator and reference files contain all the essential nomenclatural details as well as minimal species information. For each name, information is provided about the original source and format of the name, correct spelling and type information if the name is available, the nomenclatural and taxonomic status of the name, the distribution, and a link to the original reference. The predecessor of *Systema Dipterorum*, the *BioSystematic Database of World Diptera* (BDWD), was built incrementally in recognition of the long path to perfection and the need to serve the user community. The

Table 2. *Systema Dipterorum* statistics as of October 2013 indicating number of records and the proportion reaching the quality assurance level at which they are ready for pre-publication peer review.

Number of records:
198,258 species-group names (160,042 valid)
23,437 genus-group names (12,073 valid)
32,900 references
Records compared to original literature:
29,493 species (~15%)
6,138 genera (~27%)
Records nomenclaturally and taxonomically reviewed:
11,509 species-group names (~6%)
2,462 genus-group names (~11%)

BDWD was initially based to a large extent on the published regional Diptera catalogs (and several other major sources as explicitly outlined in our online documentation). While those sources were all of a high quality, they were still secondary (some possibly even tertiary), and as our aim is to ultimately present records checked against their original source by a named authority, most records are still flagged as working records.

The SD continues to be built incrementally so as to provide useful information more quickly than having to wait until it is complete at optimal standards (the sources from which the SD was built are documented online). Each record includes a quality assurance standard indicator (these are also documented online) telling users how complete the record is especially in respect to our ultimate status of taxonomic and nomenclatural peer review by assigned specialists. Records meeting the ultimate level are identified by the name(s) of the specialist(s) and date of review. Currently only about 6% of records meet this highest level, but in reality most records are as good as those already published (that is, the source from which they derived) or better (Table 2).

With *ZooBank* (Polaszek et al. 2008, Pyle and Michel 2009) growing in capacity and having taken the first steps to become part of the nomenclatural legislation concerning names published in digital works (International Commission on Zoological Nomenclature 2012), a liaison between *ZooBank* and *Systema Dipterorum* within the *Global Names Architecture* (Pyle and Michel 2009) is an obvious next step. Names for the planetary biota are best made available to the user community by a global informatics infrastructure, and Diptera names proposed by practicing taxonomists are already to an increasing extent migrated into *ZooBank* by semi-automated routines of front-end taxonomic journals (Smith et al. 2013). The major task of the team behind *Systema Dipterorum* may increasingly be to provide nomenclatural 'vetting' plus taxonomic authority, of which the latter is a quality that can be sustained only by the dipterist community itself.

The planned species interface will differ from the nomenclator only in the way the user can query the information. At present, a user enters a name and the nomenclator returns nomenclatural and taxonomic information about that name. The species interface will allow queries about the species and some of its other attributes, such as distribution and biology. So, one can ask, for example, for a list of all the fruit flies known from Costa Rica or for a list of all the species that are known to attack a certain fruit. The challenge of the species interface will be to determine which attributes users want to query (e.g., Conservation status? Distribution? Economic importance? Hosts? Morphology?) and then encode that information. Today, the nomenclator includes only minimal distributional data for species.

The most important aspect of the whole SD enterprise is our team—the people who have contributed their expertise and labor to build the SD (and before that the BDWD).

The final aspect of the SD is its legal status, which is documented online under 'How to cite & copyrights'. The critical fact is that SD is a community enterprise built by dipterists for themselves and for all people. So the information is without copyright and is freely available to all. While at various times the master database may have resided physically in some institution, that master was always a product of the SD team and belonged to those people. When the SD first went online, it was hosted by the Smithsonian Institution; later it was transferred to the USDA, and most recently it is served by the Natural History Museum of Denmark. In the future it will keep migrating to the best place that is willing to properly maintain and improve it.

Acknowledgments

We express our thanks to the entire *Systema Dipterorum* team of contributors and to all others who have supplied information over the years, and to the Schlinger Foundation, the Global Biodiversity Information Facility, the United States Department of Agriculture, and our home institutions, all of which provided financial support at various times. Neal Evenhuis, Dan Bickel and Steve Gaimari skillfully commented on the manuscript.

References

Agassiz L (1846a) 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, nec non variis adnotationibus et emendationibus. [= Fasc. XII, 1847], Jent & Gassman, Soloduri [= Solothurn, Switzerland], 393 pp.

Agassiz L (1846b) 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, nec non variis adnotationibus et emendationibus. [= Fasc. XII, 1847], Jent & Gassman, Soloduri [= Solothurn, Switzerland], 393 pp.

Agassiz L (1848–1854) Bibliographia zoologiae et geologiae: a general catalogue of all books, tracts, and memoirs on zoology and geology (4 vols). Ray Society, London. [Vol. 4 edited by Strickland HE and Jardine W]

Amorim DS, Papavero N (2008) A journal for the systematics and biogeography of Neotropical Diptera, 250 years after the publication of the tenth edition of the Systema Naturae. Neotropical Diptera 1: 1–5.

Brake I, Thompson FC (2011) Contributions to the Systema Dipterorum (Insecta: Diptera). Myia 12: 1–572.

Carroll LE, Norrbom AL, Dallwitz MJ, Thompson FC (2006a) Pest fruit flies of the world – larvae. Version: 8th December 2006. http://delta-intkey.com/ffl/ [visited 20 March 2014]

Carroll LE, White IM, Freidberg A, Norrbom AL, Dallwitz MJ, Thompson FC (2006b) Pest fruit flies of the world. http://delta-intkey.com/ffa/ [visited 20 March 2014]

Cerretti P, Tschorsnig H-P, Lopresti M, Di Giovanni F (2012) MOSCHweb – a matrix-based interactive key to the genera of the Palaearctic Tachinidae (Insecta, Diptera). ZooKeys 205: 5–18. doi: 10.3897/zookeys.205.3409

Courtney GW, Pape T, Skevington JH, Sinclair BJ (2009) Biodiversity of Diptera. In: Adler PH, Foottit RG (Eds) Insect Biodiversity: Science and Society. Blackwell Publishing, Oxford, 185–222. doi: 10.1002/9781444308211.ch9

Crosskey RW (1980) Catalogue of the Diptera of the Afrotropical Region. British Museum (Natural History), London, 1437 pp.

Delfinado MD, Hardy DE (1973–1977) A catalog of the Diptera of the Oriental Region (vols 1–3). University Press of Hawaii, Honolulu.

Dempewolf M (2004) Arthropods of Economic Importance – Agromyzidae of the World (ETI). University of Amsterdam, Amsterdam. http://wbd.etibioinformatics.nl/bis/agromyzidae.php [visited 20 March 2014] [Also as CD-ROM]

Derksen W, Göllner-Scheiding U (1963) Index Litteraturae Entomologicae Serie II. Die Welt-Literatur über die gesamte Entomologie von 1864 bis 1900 (4 vols). Deutsche Akademie der Landwirtschaftswissenschaften zu Berlin, Berlin, 2623 pp.

Derksen W, Göllner-Scheiding U, Gaedike R (1975) Index Litteraturae Entomologicae Serie II – Die Welt-Literatur über die gesamte Entomologie von 1864 bis 1900 (vol. 5). Deutsche Akademie der Landwirtschaftswissenschaften zu Berlin, Berlin, 238 pp.

Evenhuis NL (1989) Catalog of the Diptera of the Australasian and Oceanian regions. Bishop Museum Press and E.J. Brill, Honolulu, 1155 pp.

Evenhuis NL (1994) Catalogue of the fossil flies of the world (Insecta: Diptera). Backhuys Publishers, Leiden, 600 pp.

Evenhuis NL (1997) Litteratura Taxonomica Dipterorum (1758–1930) being a selected list of the books and prints of Diptera taxonomy from the beginning of Linnaean nomenclature to the end of the year 1930; containing information on the biographies, bibliographies, types, collections, and patronymic genera of the authors listed in this work; including detailed information on publication dates, original and subsequent editions, and other ancillary data concerning the publications listed herein (2 vols). Backhuys Publishers, Leiden, 871 pp.

Evenhuis NL (2016) Charles Davies Sherborn and the "Indexer's Club". In: Michel E (Ed.) Anchoring Biodiversity Information: From Sherborn to the 21st century and beyond. ZooKeys 550: 13–32. doi: 10.3897/zookeys.550.9697

Evenhuis NL, O'Hara JE, Pape T, Pont AC (2010) Nomenclatural studies toward a world list of Diptera genus-group names – Part I: André-Jean-Baptiste Robineau-Desvoidy. Zootaxa 2373: 1–265. doi: 10.1201/EBK1420095012-c7

Evenhuis NL, Pape T, Pont AC, Thompson FC (2010) Flying after Linnaeus: Diptera names since *Systema Naturae* (1758). In: Polaszek A (Ed.) Systema Naturae 250 – The Linnaean Ark. CRC Press, 75–82.

Fabricius JC (1775) Systema entomologiae, sistens insectorum classes, ordines, genera, species, adiectis synonymis, locis, descriptionibus, observationibus. Kortii, Flensbvrgi et Lipsiae [= Flensburg and Leipzig], 832 pp.

Fabricius JC (1805) Systema Antliatorum secundum ordines, genera, species adiectis synonymis, locis, observationibus, descriptionibus. C. Reichard, Brunsvigae [=Brunswick], xiv + 15-372 + [1] + 30 pp.

Fetzner JW (2005) The crane flies (Diptera: Tipulidae) of Pennsylvania. http://iz.carnegiemnh.org/cranefly/idkeys.htm [visited 20 March 2014]

Geller-Grimm F (2003) Photographic atlas and identification key to the robber flies of Germany (Diptera: Asilidae). Ampyx Verlag, Halle (Saale). [CD-ROM]

Gmelin JF (1790) Caroli a Linné, systema naturae per regna tria naturae, secundum classes, ordines, genera, species; cum caracteribus, differentiis, synonymis, locis. Editio decima tertia, aucta, reformata [= Ed. 13]. Vol. 1: Regnum Animale, Pt 5. G.E. Beer, Lipsiae [= Leipzig], 2225–3020.

Grimaldi D, Engel MS (2005) Evolution of the insects. Cambridge University Press, New York, 755 pp.

Hagen HA (1862–1866) Bibliotheca entomologica. Die Litteratur über das ganze Gebiet der Entomologie bis zum Jahre 1862 (2 vols). W. Engelmann, Leipzig.

Hamilton JR, Yeates DK, Hastings A, Colless DH, McAlpine DK, Bickel D, Daniels G, Schneider M, Cranston P, Marshall S (2006) On The Fly—The Interactive Atlas and Key to Australian Fly Families. Australian Biological Resources Study (ABRS) and Centre for Biological Information Technology (CBIT). [CD-ROM]

Higgs DR, Wood WG (2008) Genetic complexity in sickle cell disease. PNAS 105(33): 11595–11596. doi: 10.1073/pnas.0806633105

Horn W, Schenkling S (1928–1929) Index litteraturae entomologicae. Serie: Die Welt-Literatur über die gesamte Entomologie bis inklusive 1863. 4 vols, W. Horn, Berlin-Dahlem.

ICZN (1999) International Code of Zoological Nomenclature. Fourth edition. International Trust for Zoological Nomenclature, London, 306 pp.

ICZN (2012) Amendment of Articles 8, 9, 10, 21 and 78 of the International Code of Zoological Nomenclature to expand and refine methods of publication. Zootaxa 3450: 1–7.

Kertész K (1902–1910) Catalogus dipterorum hucusque descriptorum (7 vols). G. Engelmann, Budapest.

Kohler RE (1994) Lords of the fly. *Drosophila* genetics and experimental life. Chicago University Press, Chicago, 344 pp.

Linnaeus C (1737) Critica botanica in qua nomina plantarum generica, specifica & variantia examini subjiciuntur, selectoria confirmantur, indigna rejiciuntur; simulque doctrina circa denominationem plantarum traditur. Seu Fundamentorum botanicorum pars IV Accedit Johannis Browallii De necessitate historiae naturalis discursus. Conrad Wishoff; Lugduni Batavorum [Leiden], 270 pp.

Linnaeus C (1758) Systema naturae per regna tria naturae, secundum classes, ordines, genera, species, cum characteribus, differentiis, synonymis, locis (10th ed., Vol. 1). Laurentii Salvii, Holmiae [= Stockholm], 824 pp.

Linnaeus C (1767) Systema naturae per regna tria naturae, secundum classes, ordines, genera, species, cum characteribus differentiis, synonymis, locis (12th ed.). Laurentii Salvii, Holmiae [= Stockholm] 1(2): 533–1327.

de Marschall AF (1873) Nomenclator zoologicus continens nomina systematica generum animalium tam viventium quam fossilium, secundum ordinem alphabeticum disposita. C. Ueberreuter, Vindobonae [= Vienna], 482 pp.

Mathis WN (2011) World catalog and conspectus on the family Heterocheilidae (Diptera: Schizophora). Myia 12: 281–289.

Mathis WN, Barraclough DA (2011) World catalog and conspectus on the family Diastatidae (Diptera: Schizophora). Myia 12: 235–266.

Mathis WN, McAlpine DK (2011) A catalog and conspectus on the family Coelopidae (Diptera: Schizophora). Myia 12: 171–205.

Mathis WN, Sueyoshi M (2011) World catalog and conspectus on the family Dryomyzidae (Diptera: Schizophora). Myia 12: 207–233.

National Research Council Committee on the Status of Pollinators (2007) Status of pollinators in North America. National Academies Press, Washington, DC, 322 pp.

Meiklejohn KA, Kavazos CRJ, Dowton M, Pape T, Wallman JF (2012) Australian *Sarcophaga* (sensu lato) (Diptera: Sarcophagidae). Institute for Conservation Biology and Environmental Management, University of Wollongong, Australia. http://keys.lucidcentral.org/keys/v3/Sarcophaga/ [visited March 2014]

Neave SA (1939–1996) Nomenclator Zoologicus; a List of the Names of Genera and Subgenera in Zoology from the Tenth Edition of Linnaeus, 1758, to the End of 1935 (with supplements) (10 vols). Zoological Society of London, London.

Norrbom AL, Thompson FC (Eds) (2004) Diptera Data Dissemination Disk – Volume 2. [CD-ROM]

O'Hara JE, Cerretti P, Pape T, Evenhuis NL (2011) Nomenclatural Studies Toward a World List of Diptera Genus-Group Names. Part II: Camillo Rondani. Zootaxa 3141: 1–268.

O'Toole C, Raw A (2004) Bees of the world. Facts on File, New York, 192 pp.

Osten Sacken CRR (1858) Catalogue of the described Diptera of North America. Smithsonian Miscellaneous Collections 3(1): 1–92.

Panzer GWF (1792–1813) Favnae insectorum germanicae initia oder Devtschlands Insecten. Felseckerschen, Nürnberg.

Papavero N (Ed.) (1966–1984) A catalogue of the Diptera of the Americas South of the United States. Departamento de Zoología, Secretaria de Agricultura, São Paulo. [102 separate fascicles]

Pape T (2009) Economic importance of Diptera. In: Brown BV, Borkent A, Cumming JM, Wood DM, Woodley NE, Zumbado M (Eds) A manual of Central American Diptera. Vol. 1. NRC Press, Ottawa, 65–77.

Pape T, Blagoderov V, Mostovski MB (2011) Order DIPTERA Linnaeus, (1758) In: Zhang Z-Q (Ed.) Animal biodiversity: An outline of higher-level classification and survey of taxonomic richness. Zootaxa 3148: 222–229.

Pape T, Thompson FC (Eds) (2014) [online]. Systema Dipterorum. Version 2.0. http://www.diptera.org [last visited 20 March 2014]

Penev L, Roberts D, Smith V, Agosti D, Erwin T (2010) Taxonomy shifts up a gear: New publishing tools to accelerate biodiversity research. ZooKeys 50: 1–4. doi: 10.3897/zookeys.50.543

Polaszek A, Pyle RL, Yanega D (2008) Animal Names for All: ICZN, ZooBank and the New Taxonomy. In: Wheeler QD (Ed.) The New Taxonomy. The Systematics Association Special Volume Series 76, 129–142.

Pyle RL (2016) Towards a Global Names Architecture: The future of indexing scientific names. In: Michel E (Ed.) Anchoring Biodiversity Information: From Sherborn to the 21st century and beyond. ZooKeys 550: 261–281. doi: 10.3897/zookeys.550.10009

Pyle RL, Michel E (2009) Unifying nomenclature: ZooBank and Global Names Usage Bank. Bulletin of Zoological Nomenclature 66: 298.

Robineau-Desvoidy J-B (1863) Histoire naturelle des diptères des environs de Paris. Oeuvre posthume du Dr Robineau-Desvoidy publiée par les soins de sa famille, sous la direction de M. H. Monceaux. Masson et fils, Paris, Wagner, Leipzig, and Williams & Norgate, London. Tome premier, xvi + 1,143 pp. Tome second, 920 pp.

Roskov Y, Kunze T, Paglinawan L, Orrell T, Nicolson D, Culham A, Bailly N, Kirk P, Bourgoin R, Baillargeon G, Hernandez F, De Wever A (Eds) (2013) Species2000 & ITIS Catalogue of Life, 2013 Annual Checklist. DVD, Species2000, Reading.

Sabeti PC, Schaffner SF, Fry B, Lohmueller J, Varilly P, Shamovsky O, Palma A, Mikkelsen TS, Altshuler D, Lander ES (2006) Positive Natural Selection in the Human Lineage. Science 312: 1614–1620. doi: 10.1126/science.1124309

Schulze FE, Kükenthal W, Heider DK (1926–1954) Nomenclator animalium generum et subgenerum. 5 vols, Preussischen Akademie der Wissenschaften, Berlin, ccxliv + 3,716 pp.

Scudder SH (1882) Nomenclator zoologicus. An alphabetical list of all generic names that have been employed by naturalists for recent and fossil names from the earliest times to the close of the year (1879) I. Supplemental list of genera in zoology. List of generic names employed in zoology and paleontology to the close of the year 1879, chiefly supplemental to those catalogued by Agassiz and Marschall, or indexed in the Zoological Record. Bulletin of the United States National Museum, 19[i], i–xxi, 1–376.

Scudder SH (1884) Nomenclator zoologicus. An alphabetical list of all generic names that have been employed by naturalists for recent and fossil names from the earliest times to the close of the year (1879) II. Universal index to genera in zoology. Complete list of generic names employed in zoology and paleontology to the close of the year 1879, as contained in the nomenclators of Agassiz, Marschall and Scudder, and in the Zoological Record. Bulletin of the United States National Museum, 19[ii], 340 pp.

Sherborn CD (1902) Index animalium sive index nominum quae ab A.D. MDCCLVIII generibus et speciebus animalium imposita sunt. Section prima. a Kalendis Ianuariis, MDCCLVIII usque ad finem Decembris, MDCCCL. Cantabrigae [= Cambridge], 1195 pp. [in 33 parts]

Sherborn CD (1922–1932) Index animalium sive index nominum quae ab A.D. MDCCLVIII generibus et speciebus animalium imposita sunt. Section secunda. a Kalendis Ianuariis, MDCCCI usque ad finem Decembris, MDCCC. London, 7056 pp. [in 33 parts]

Smith V, Georgiev T, Stoev P, Biserkov J, Miller J, Livermore L, Baker E, Mietchen D, Couvreur T, Mueller G, Dikow T, Helgen K, Frank J, Agosti D, Roberts D, Penev L (2013) Beyond dead trees: integrating the scientific process in the Biodiversity Data Journal. Biodiversity Data Journal 1: e995. doi: 10.3897/BDJ.1.e995

Soós Á, Papp L (Eds) (1984–1993) Catalogue of Palaearctic Diptera (14 vols). Akademiai Kiado and Hungarian Natural history Museum, Budapest and Elsevier Science, Amsterdam.

Ssymank A, Kearns CA, Pape T, Thompson FC (2008) Pollinating flies (Diptera): a major contribution to plant diversity and agricultural production. Biodiversity 9: 86–89. doi: 10.1080/14888386.2008.9712892 [special issue on "The value of Biodiversity to food and agriculture"]

Stone A, Sabrosky CW, Wirth WW, Foote RH, Coulson JR (Eds) (1965) A catalog of the Diptera of America north of Mexico. United States Department of Agriculture, Agricultural handbook 276, 1969 pp.

Strickland HE, Henslow JS, Phillips J, Shuckard WE, Richardson JB, Waterhouse GR, Owen R, Yarrell W, Jenyns L, Darwin C, Broderip WJ, Westwood JO (1842) Report of a Committee appointed "to consider the rules by which the nomenclature of Zoology may be established on a uniform and permanent basis". John Murray, London, for the British Association for the Advancement of Science.

Thompson FC (1996) Names: The keys to Biodiversity. Chapter 13. In: Reaka-Kudla ML, Wilson, D, Wilson EO (Eds) Biodiversity II. Understanding and protecting our biological resources. Joseph Henry Press, Washington, DC, 199–216.

Thompson FC (1999) Introduction [to Fruit Fly Expert Identification System and Systematic Information Database]. Myia 9: 1–6.

Thompson FC, Evenhuis NL (Eds) (1999) BioSystematic Database of World Diptera. Diptera Names Working Data Files. Diptera data dissemination disk, Volume 1. [CD-ROM]

Thompson FC, Norrbom AL, Carroll LE, White IM (1993) The fruit fly biosystematic information database. In: Aluja M, Liedo P (Eds) Proceedings of the International Symposium on Fruit Flies of Economic Importance 1990, Antigua, Guatemala (October 14–20). Springer-Verlag, New York, 3–7.

Thompson FC, Pont AC (1994) Systematic database of *Musca* names (Diptera). A catalog of names associated with the genus-group name *Musca* Linnaeus, with information on their classification, distribution, and documentation. Theses Zoologicae 20: 1–219.

Walter DA, Winterton S (2007) Keys and the crisis in taxonomy: extinction or reinvention? Annual Review of Entomology 52: 193–208. doi: 10.1146/annurev.ento.51.110104.151054

Walter Reed Biosystematics Unit (2014) Keys to the medically important mosquito species. http://www.wrbu.org/command_aors_MQkeys.html [last visited 20 March 2014]

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. In: Michel E (Ed.) Anchoring Biodiversity Information: From Sherborn to the 21st century and beyond. ZooKeys 550: 173–183. doi: 10.3897/zookeys.550.10041

White IM, Hancock DL (2004) Fauna Malesiana – Interactive key for dacine fruit flies (Diptera: Tephritidae: Dacini). ETI. University of Amsterdam, Amsterdam.

Wiegmann BM, Trautwein MD, Winkler I, Barr NW, Kim J-W, Lambkin C, Bertone MA, Cassel BK, Bayless KM, Heimberg AM, Wheeler BM, Skevington JH, Pape T, Sinclair BJ, Borkent A, Blagoderov V, Caravas J, Kutty SN, Peterson KJ, Schmidt-Ott U, Kampmeier GE, Thompson FC, Grimaldi DA, Beckenbach AT, Courtney GW, Friedrich M, Meier R, Yeates DK (2011) Episodic radiations in the fly tree of life. PNAS 108: 5690–5695. doi: 10.1073/pnas.1012675108

Winterton SL (2009) Diptera Biodiversity Informatics. In: Pape T, Bickel D, Meier R (Eds) Diptera Diversity: Status, Challenges and Tools. Koninklijke Brill NV, Leiden, 381–406. doi: 10.1163/ej.9789004148970.I-459.62

Woodley NE (2001) A world catalog of the Stratiomyidae (Insecta: Diptera). Myia 11: 1–473.

Young AM (1994) The chocolate tree, a natural history of cacao. Smithsonian Institution Press, Washington, DC & London, 200 pp.

Young CW (2005) The Crane Flies (Diptera: Tipulidae) of Pennsylvania. Carnegie Museum of Natural History, Pittsburgh. http://iz.carnegiemnh.org/cranefly [visited 20 March 2014]

ZooKeys 550: 153–171 (2016)
doi: 10.3897/zookeys.550.9673
http://zookeys.pensoft.net

RESEARCH ARTICLE

A peer-reviewed open-access journal
ZooKeys
Launched to accelerate biodiversity research

Unlocking *Index Animalium*: From paper slips to bytes and bits

Suzanne C. Pilsk[1], Martin R. Kalfatovic[1], Joel M. Richard[1]

1 *Smithsonian Libraries, Washington DC, USA*

Corresponding author: *Suzanne C. Pilsk* (PilskS@si.edu)

Academic editor: *E.Michel* | Received 23 March 2015 | Accepted 25 March 2015 | Published 7 January 2016

http://zoobank.org/F8D9D471-E11A-4EE9-AEE8-BDEB26F0DAB2

Citation: Pilsk SC, Kalfatovic MR, Richard JM (2016) Unlocking *Index Animalium*: From paper slips to bytes and bits. In: Michel E (Ed.) Anchoring Biodiversity Information: From Sherborn to the 21st century and beyond. ZooKeys 550: 153–171. doi: 10.3897/zookeys.550.9673

Abstract

In 1996 Smithsonian Libraries (SIL) embarked on the digitization of its collections. By 1999, a full-scale digitization center was in place and rare volumes from the natural history collections, often of high illustrative value, were the focus for the first years of the program. The resulting beautiful books made available for online display were successful to a certain extent, but it soon became clear that the data locked within the texts needed to be converted to more usable and re-purposable form via digitization methods that went beyond simple page imaging and included text conversion elements. Library staff met with researchers from the taxonomic community to understand their path to the literature and identified tools (indexes and bibliographies) used to connect to the library holdings. The traditional library metadata describing the titles, which made them easily retrievable from the shelves of libraries, was not meeting the needs of the researcher looking for more detailed and granular data within the texts. The result was to identify proper print tools that could potential assist researchers in digital form. This paper outlines the project undertaken to convert Charles Davies Sherborn's *Index Animalium* into a tool to connect researchers to the library holdings: from a print index to a database to eventually a dataset.

Sherborn's microcitation of a species name and his bibliographies help bridge the gap between taxonomist and literature holdings of libraries. In 2004, SIL received funding from the Smithsonian's Atherton Seidell Endowment to create an online version of Sherborn's *Index Animalium*. The initial project was to digitize the page images and re-key the data into a simple data structure. As the project evolved, a more complex database was developed which enabled quality field searching to retrieve species names and to search the bibliography. Problems with inconsistent abbreviations and styling of his bibliographies made the parsing of the data difficult. Coinciding with the development of the Biodiversity Heritage Library

(BHL) in 2005, it became obvious there was a need to integrate the database converted *Index Animalium*, BHL's scanned taxonomic literature, and taxonomic intelligence (the algorithmic identification of binomial, Latinate name-strings). The challenges of working with legacy taxonomic citation, computer matching algorithms, and making connections have brought us to today's goal of making Sherborn available and linked to other datasets. Partnering with others to allow machine-to-machine communications the data is being examined for possible transformation into RDF markup and meeting the standards of Linked Open Data. SIL staff have partnered with Thomson Reuters and the Global Names Initiative to further enhance the *Index Animalium* data set. Thomson Reuters' staff is now working on integrating the species microcitation and species name in the ION: Index to Organism Names project; Richard Pyle (The Bishop Museum) is also working on further parsing of the text. The *Index Animalium* collaborative project's ultimate goal is to successful have researchers go seamlessly from the species name in either ION or the scanned pages of *Index Animalium* to the digitized original description in BHL - connecting taxonomic researchers to original authored species descriptions with just a click.

Keywords
Metadata, Digitization, Linked Open Data

Background

The Smithsonian Libraries' collections support the varied museums and research centers that support the mandate for the "increase and diffusion of knowledge" established by the benefactor James Smithson. The diversity of the subject matter in the Libraries collection reflects the range of topics, disciplines and activities undertaken by Smithsonian researchers. The Libraries has developed along the lines of the Institution to support the vast array of topics that has become the largest complex of museums and research centers in the world with 20 libraries supporting 19 museums and 9 researcher centers. The Institution's natural history collections date back to the 18th century and have been collected to assist in the study and stewardship of the extensive specimen collections. The United States National Museum was established within the Institution in 1858 and moved into a separate, individual museum in 1910. The Natural History Library collections of Smithsonian Libraries have grown in conjunction with the National Museum to help researchers identify and document specimen collections. With a substantial amount of focus since its founding on classic collections-based research of systematics and taxonomy, the collection in the library supports the discovery of species and naming. The reliance on historical literature to perform the work has made for a strong library collection. The Smithsonian Libraries has grown to take on a role of providing authoritative information and creates innovate services for the curators, scientists, and researchers. (Smithsonian Institution Libraries. *Rare Books and Special Collections in the Smithsonian Institution Libraries*. 1995) This includes the move towards providing the necessary information in digital form alongside the traditional print collections.

To build and preserve along with the supporting of present day research needs, the Libraries looks for ways to have the information within the collection reach the needed patron whenever and wherever they may be. Acquiring digital data, electronic journals and resources and current research database subscriptions is one aspect of the Libraries reach. Scanning the holdings of collections to provide better access was another step towards the delivery of critical information to the researchers. The Libraries' first started digitization projects in 1996. By 1999, a full-scale digitization center was in place and rare volumes from the natural history collections, often of high aesthetic value, were the focus for the first years of the program. The resulting beautiful books made available for online display were successful to a certain extent; but it soon became clear that the data locked within the texts needed to be converted to more usable and re-purposable form via digitization methods that went beyond simple page imaging. There was a critical need to include text conversion elements. This "freeing the data locked on the page" began to be the goal of the Smithsonian Libraries' digitization program. The online version of reference sources began to be scanned with the text created into datasets and was the natural progression from the initial tomes with pleasing plates. Sherborn's *Index Animalium* was one of the Libraries first attempts at digitization for database conversion.

I.

Smithsonian Libraries is a traditional library with books on shelves with librarians and staff ready to assist the patron with their information requests. Traditional library description of monographs and serials is based on standards within the library and information sciences. The inventories of the holdings of large academic libraries require standardized practices and efficiencies of scale to accomplish sophisticated catalogues. The Libraries cover a wide range of topics from art to zoology and is geographically located across the United States and Panama. The data that is captured for each title assists in the physical allocation of material. Yet, the granularity of the descriptive data is effectively only at the title level and does not delve into the contents, chapter, article or page level with in each title, and does not index the specific details within the texts. Most libraries' tools dating back to the card catalogues to the current online integrated library catalogues, have found that the metadata describing titles has worked with limited success. The discovery aspect of this overarching or "high" level of metadata limits the results of inquiries but sufficed for physical discovery of the titles.

Taxonomic research requires the specific citation of descriptions of species. The International regulatory codes for identifying and naming of species require in-depth research on species and genus. The rules are quite clear that when naming species, the name is considered fully formed once the description is published and available (International Commission on Zoological Nomenclature's International Code, http://www.iczn.org/iczn/index.jsp; See specifically Article 8 and Article 11). Major natural history libraries and the Smithsonian's National Museum of Natural History Library, specifically, have served the function of ensuring that publications of species names are stored and the publications are available. Yet the librarian standards of description of

splendens Podontia, Guérin-M. in Duperry, Voy. "Coquille," Zool. II (2) 1 (1838), 144.
splendens Pseudoophonus (Gebl.), V. v. Mochulski, Kaefer Russl., *post* May 1850, 31.— [Ophonus.
splendens Pupa, Menke in L. Pfeiffer, Symb. Helic. I. 1841, 45.
splendens [Pycnodus], H. v. Meyer, N. Jahrb. f. Min. 1847, 574 [*n. n.*].
splendens Pycnodus, H. v. Meyer, Palaeontographica, I (5) Dec. 1849, 239.
splendens Rhizomys (Ruepp.), J. A. Wagner in Schreber, Säugth., Suppl. III. 1842, 368.— Bathyergus.
splendens Rissoa, O. G. Costa, Atti R. Accad. Sci. [Napoli] IV (Zool.) 1839, 192 [*n. n.*].
splendens Rissoa, O. G. Costa, Corrisp. Zool. I. 1839, 71 [*n. n.*].
splendens Saprinus (Payk.), W. F. Erichson, Jahrb. Insectenk. I. 1834, 178.—Hister.
splendens Sargus, J. W. Meigen, Klassif. Zweifl. Insekt. I (1) 1804, 144.
splendens Saxicola, Quoy & Gaimard in d'Urville, Voy. "Astrolabe," Zool. I. 1830, 197.
splendens Scarabaeus, Palisot de B., Ins. Afr. Amér. (—) 1809, 89.
splendens Sertularia, J. V. F. Lamouroux, Hist. Polyp. 1816, 191.
splendens Solen, J. C. Chenu, Illustr. Conch. I (18 & 19) 1843, pl. 8 [*n. et f.*].
splendens Sphenoptera, Laporte & Gory, H. N. Coléopt. II. 1839, 29.
splendens Sponsor, F. E. Guérin-M., Rev. Zool. (Soc. Cuv.) III. Dec. 1840, 357.
splendens Staphylinus, T. Marsham, Coleopt. Brit., Sept. 1802, 524.
splendens Sterculia, E. Blanchard in d'Orbigny, Amér. Mérid. (Ins.) 1842, 83.
splendens Tellina, J. De C. Sowerby, Tr. Geol. Soc. London [2] V. 1837, 136.
splendens Temnoscheila, G. R. Gray in Griffith's Cuvier, Anim. Kingd., Ins. II. 1832, 93.
splendens Tetragonus, Gory & Percheron, Mon. Cétoines (2) 1834, 67.
splendens Tityra (Wied), G. R. Gray, Gen. Birds, I. June 1846, 254.—Muscicapa.
splendens Toxotus, Laich.; Dejean, Catal. Coléopt. 1821, 112 [*n. n.*].
splendens Trochus, C. G. Giebel, De Geogn. Hercyn. 1848, 28 [*n. n.*].
splendens Turbo, O. G. Costa, Cat. Test. Sicilie, 1829, pp. ci & civ.
splendens Turdus, W. E. Leach, Zool. Miscell. II. 1815, 30.
splendens Turdus (Daud.), L. P. Vieillot, N. Dict. H. N., ed. 2, XX. 1818, 260.—Sturnus, [1800.
splendens Udorpes, V. v. Motschulsky, Bull. Soc. Imp. Nat. Moscou, XVIII. 1845 (1) 1845,
splendens Unio, G. A. Goldfuss, Petref. German. II (6) 1837, 183. [108.
splendens Venus, D. Solander, Catal. Portland Mus. 1786, 102 & 149 [*n. n.*].
splendens Volatinia (Vieill.), C. L. Bonaparte, Consp. gen. Avium, I. 1850, 474.—Fringilla.
splendescens Pagurus, R. Owen, Zool. "Blossom," 1839, 81.
splendicans Buprestis, Norw.; J. Sturm, Catal. Ins. Samml. 1826, 105.
splendidalis Epipagis (Cr.), J. Huebner, Verz. bekannt. Schmett. 1826, 358.—Phalaena,
splendidella Coleophora, F. Lienig, Isis (Oken), 1846, 296. [1781.
splendidissima Lampyris, F. Green, Doughty's Cab. N. H. II (3) 1832, 55 [*teste* C. W. R.].
splendidissima Laphria, C. R. W. Wiedemann, Aussereurop. Zweifl. Insekt. II. 1830, 645.
splendidiventris Buprestis, Laporte & Gory, H. N. Coléopt. I. 1838, 69.
splendidulana Coccyx, A. Guénée, Ann. Soc. Ent. France [2] III (2) Oct. 1845, 179.
splendidula Baridius, Cristofori & Jan, Cat. Mus. (III a, Coleopt.) 1832, 64 [*n. n.*].

Figure 1. Page from *Index Animalium* showing examples of microcitations.

these materials has fallen short of the needs of the taxonomic researcher in identifying exactly where descriptions are located within these publications – the page level metadata and the data within the page is lacking.

Smithsonian Libraries staff met with researchers from the taxonomic community to understand their path to the literature and identified resources used to connect to the library holdings. The traditional library metadata describing the titles on the shelves of libraries was not meeting the needs of the researcher looking for more detailed and granular data within the texts. Their own bibliographies and indexes were required to pinpoint the data needed. From those sources, data points had to be mapped to the library search interface (the online catalogue) with different terminology and assumptions. Each individual researcher had to interpret and translate access points to locate the desired material.

As seen in the other essays within this compilation, Charles Davies Sherborn stepped in to fill the data needs that the traditional library catalogues were not and could not meet. (Neal Evenhuis. "Sherborn: Work history and impact of bibliography, dating and zoological informatics.") The beauty of his *Index Animalium* are the microcitations for a species giving the genus, species, author, abbreviated title of publication, and the critical date and page specifics. This level of access within the texts of monographs and serials is the data that the libraries were failing to deliver. Smithsonian Libraries saw that Sherborn's *Index* was actually a data set that needed to be liberated off the printed page and made available digitally. The microcitations were needed in the electronic world to interact with taxonomists working in the digital world – writing, citing, and interacting with their research. The first task at hand was to scan and make available a fully searchable *Index Animalium*.

Funded by the Atherton Seidell Endowment Fund, SIL contracted to have the entire set of 30+ volumes scanned: cover to cover, over 9,000 page images. Subsequent to the imaging, the entire *Index* was re-keyed into a database. Spot checked and refined with the vendor, the final database has an accuracy rate of 99.995% and consists of over 430,000 lines of useful data. The *Index Animalium* electronic version is available at http://www.sil.si.edu/digitalcollections/indexanimalium/.

II.

The first goal of the digital e-version of Sherborn's Index, was to provide to the world a searchable version of the full tex of the index and the accompanying bibliographies. As the project continued, it became a mission to identify every volume that Sherborn examined in creating the Index. Once identified, the volumes could then be physically located with first preference being our own Smithsonian Libraries' collection. If the title was not in Smithsonian's holdings, a location would be sought within the realm of natural history libraries. This layer of access was to assist anyone using the online *Index Animalium*'s microcitations to be able to locate the book that Sherborn references.

Sherborn states in the Epilogue of *Index Animalium*, March 1922: "In any well-appointed Natural History Library there should be found every book and every edition of every book dealing in the remotest way with the subjects concerned." With over 7,700 titles listed, Sherborn gives the most comprehensive list of all important works in the study of zoology. The four bibliographies scattered throughout the multi-volume Index records every title that Sherborn examined. He included indications if the work had no systematic zoological name, no Linnaean names, inconsistent binomial names, no specific names mentioned, or if no new species were found in the texts.

Researching all the potentially related species is required for the study of species naming. The Smithsonian Libraries' online version of Sherborn's *Index* is aimed to facilitate the researcher locating all the texts that are referenced. Sherborn's bibliog-

ROUGH LIST OF BOOKS REFERRED TO IN THE COMPILATION OF THIS INDEX.

No sp. nn. = no specific names
No n. spp. = no new species
n. b. = not consistently binominal
n. L. = no Linnean names
n. z. = no systematic zoology

"J. M. A." Beschr. Naturalienkab. Meerspurg. 8vo. *Bregenz*, 1786. [No sp. nn.]
Aaskow, U. B. Tent. Tetrapod. Danicae. 8vo. *Havniae* I. 1764; II. 1766. [Names not used in a generic sense.]
Abbot, J. N. H. Lepid. Ins. Georgia. 2 vols. fo. *Lond*. 1797. [Names by J. E. Smith.]
Abel, J. C. A. M. Conch. Nat.-Kab. B. von Konstanz. 8vo. *Bregenz*, 1787. [n. b.]
Abhandl. v. d. Wickel-Raupe. 8vo. *Berl*. 1779. [No sp. nn.]
Abildgaard, S. Beschr. Stevens Klint. 8vo. *Copen*. 1764.
Acharius, E. *See* Rosenblad, E.
Acosta, J. de. Hist. Nat. Indias. 2 vols. 4to. *Madrid*, 1792. [No sp. nn.]
Acta (N.) eruditorum. 4to. *Leipz*. → 1782. [All reviews.]
Acta Helvetica. *See* Basel.
Acta Med. Suecica. 1 vol. 8vo. *Upsala*, 1783.
Adams, G. Micrgr. illust. Ed. 4. 8vo. *Lond*. 1771. [No sp. nn.]
Adams, G. (fil.). Essays Microscope. 4to. *Lond*. 1787.
—— —— —— Ed. 2, by Kanmacher. 4to. *Lond*. 1798.
Adanson, M. Voy. to Senegal. 8vo. *Lond*. 1759. [No sp. nn.]
Admiral, J. L'. Naauwk. Waarn. Insekten. fo. *Amsterdam*, n. d. [? 1762]; fo. *Amsterdam*, 1774. [No sp. nn.]
Afzelius, A. Account Nat. Prod. Sierra Leone. 8vo. *Lond*. 1794. [No sp. nn.]
Aikin, A. Journ. Tour N. Wales. 8vo. *Lond*. 1797. [No sp. nn.]
Albin, E. *See* Martyn, T., Aranei.
—— Nat. Hist. English Song-birds. New ed. 8vo. *Lond*. 1799. [No sp. nn.]
Albinus, B. S. Acad. Annot. 2 vols. 4to. *Leidae*, 1754–68. [No sp. nn.]
Alessandri, J., & P. Scattaglia. Descr. degli Anim. 4 vols. fo. *Venez*. 1771–75. [No sp. nn.]
Alexander, W. Tent. med. Cantharid. 8vo. *Edin*. 1769. [No sp. nn.]
Algemene Konst- en Letter-bode. 4to. *Haarlem*, 1788–93.
—— (Nieuwe) —— 1794–1800. [Reviews only.]
Algemeene Vaderland Letteroefningin. 8vo. *Amst*. 1761 → [Reviews.]
Allamand, J. N. S. See Buffon.

Figure 2. Example of the first page of the first bibliography (Sherborn. *Index Animalium*, Vol. 1, p. xi).

raphy, though a very comprehensive list of important titles, is not complete in the descriptions of these titles. His use of inconsistent abbreviation, "ibid" indications, use of shortened titles, and other idiosyncrasies has made identifying the exact titles challenging.

III.

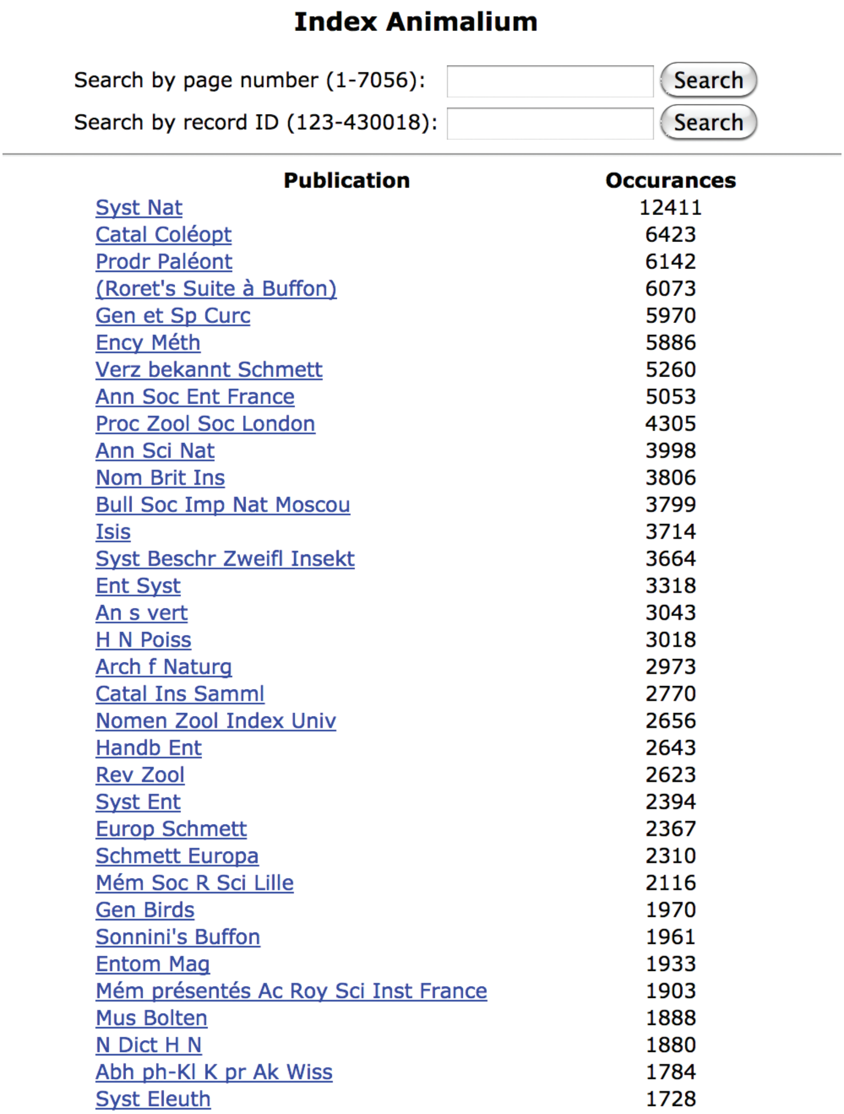
Smithsonian Libraries first foray into moving beyond pretty books to creating datasets faced many challenges including re-assessing the actual needs and deliverables of the project. The online project morphed from the initial basic scanning of the text – to a searchable database – to a goal of connecting each microcitation to the proper line in the bibliography – to the goal of having the microcitation connected to bibliography connected to physical location of the text. Difficulties emerge when computer-to-

—— Ad orat. aud. De primis anim. vert. 4to. *Jenae*, 1836.
—— Hist.-topogr. Taschenb. Jena. 8vo. *Jena*, 1836.
Zetterstedt, J. W. Orthopt. Suec. 8vo. *Lund.* 1821.
—— Resa Sveriges och Norriges Lappmarkes. 2 vols. 8vo. *Lund*, 1822.
—— Fauna Ins. Lappon. 8vo. *Hann.* 1828.
—— Resa genom Umea Lappmarker. 8vo. *Oreb.* 1833.
—— Ins. Lappon. descr. 6 pts. 4to. *Lips.* 1838–40. [Pp. 1–867, 1838; –1013, 1839; –1140, 1840.]
—— Dipt. Scand. 9 vols. 8vo. *Lund.* 1842–50. [The remainder of this work does not come within the scope of this Index.]
Zhivonishuii Sbornik. 1 vol. 4to. *St Petersb.* 1850. [B. M., popular.]
"Ziegler." [In molluscan literature his n. spp. were recorded by Menke and by Rossmaessler.]
Zieten, C. H. de. Les Pétrif. de Wurtemberg. 12 pts. fo. *Stoutg.* 1830(–33). [(1 & 2) pp. 1–16, 1830; (3 & 4) 17–32, 1831; (5–8) 33–64, 1832; (9–12) 65–96, 1833.]
Zigno, A. de. Sopra corpi organ. infusioni. Ed. 2. 8vo. *Padova*, 1842.
—— Sopra due foss. calc. Monte Padovani. 4to. *Pad.* 1845. [G. S.]
—— Sul cret. Ital. sept. 4to. *Pad.* 1846. [N. Saggi R. Acc. Sci. Padova, VI; no n. spp.]
—— Nota nov. prom. foss. il Bianconi. 8vo. *Ven.* 1846. [Atti Adun. I. R. Ist. Sci. Venezia; G. S.]
Zimmermann, C. Monogr. Carabiden. 8vo. *Berl.* 1831.
—— Das Harzgebirge. 8vo. *Darmst.* 1834. [G. S., lists.]
Zimmermann, E. A. W. Taschenb. d. Reisen. 8vo. *Leipz.* 1819. [*Teste* M. J. Rathbun.]
Zodiac. 12 nos. 4to. *Albany*, 1835–6. [B. M., no n. spp.]
Zoologe (Der). 8 pts. 8vo. *Eisenach*, 1795–97. [This is the XXI Part of "Compendiose Bibliothek d. gemeinnuetz. Kenntn. f. alle Staende," lent by the Univ. of Kiel to the B. M. (N. H.), 1907; I take the "B–n" to be Beckstein, but Gerritt Miller and Knud Andersen think it Borkhausen.]
Zoological Journ. (The). 20 pts. in 5 vols. 8vo. *Lond.* 1824–35; 5 pts. of supplementary plates.
Zoological Magazine or Journ. of Nat. Hist. 6 pts. 8vo. *Lond.* 1833–34.
Zoological Miscellany (Gray's). 6 pts. 8vo. *Lond.* 1821–44. [Pt. 1, pp. 1–40, Feb. 1831; the other parts are dated at sign. F, G, H, I, and K.]
Zoological Miscellany (Leach's). 3 vols. 8vo. *Lond.* 1814–17. [Dates uncertain; I adopt on my evidence I, 1814; II, 1815; III, 1817; it is almost certain that vol. I began Jan. 1814.]
Zoologisches Magazin. 8vo. *Kiel*, 1817–23.
Zoologist (The). I–VIII. 8vo. *Lond.* 1843–50.
Zuerich. Nat. Gesellsch. An die zuerisch. Jugend. 4to. *Zuer.* 1799–1850; Berichte, 1826–38; Mittheil. I, II, 1847–50.

Figure 3. One example page from *Index Animalium* bibliography.

computer resolving of microcitations and bibliography entries were attempted. Most problematic was the use of computer scripts against Sherborn's inconsistent notation made it impossible for clear connections of species citations to title citation to be made in a systematic way.

Simple Regular Expressions were used to break apart the re-keyed text of Sherborn based on the lessons learned by MBL WHOI Library's project for Neave's *Nomenclator Zoologicus*. (*Nomenclator Zoologicus* online version from uBio, Marine Biological Laboratory, Woods Hole Oceanographic Institution http://uio.mbl.edu/NomenclatorZoologicus/) Regular Expressions are a simple syntax particularly suited for identifying and dividing up textual data by looking at patterns, punctuation, and even character strings or sequences. David Remsen and Patrick Leary (formerly at MBL/WHOI) used these parsing techniques to isolate titles within the *Index's* species citations. Using those strings they used comparisons of strings to match against the bibliography citations.



Index Animalium

Search by page number (1-7056): Search

Search by record ID (123-430018): Search

Publication	Occurances
Syst Nat	12411
Catal Coléopt	6423
Prodr Paléont	6142
(Roret's Suite à Buffon)	6073
Gen et Sp Curc	5970
Ency Méth	5886
Verz bekannt Schmett	5260
Ann Soc Ent France	5053
Proc Zool Soc London	4305
Ann Sci Nat	3998
Nom Brit Ins	3806
Bull Soc Imp Nat Moscou	3799
Isis	3714
Syst Beschr Zweifl Insekt	3664
Ent Syst	3318
An s vert	3043
H N Poiss	3018
Arch f Naturg	2973
Catal Ins Samml	2770
Nomen Zool Index Univ	2656
Handb Ent	2643
Rev Zool	2623
Syst Ent	2394
Europ Schmett	2367
Schmett Europa	2310
Mém Soc R Sci Lille	2116
Gen Birds	1970
Sonnini's Buffon	1961
Entom Mag	1933
Mém présentés Ac Roy Sci Inst France	1903
Mus Bolten	1888
N Dict H N	1880
Abh ph-Kl K pr Ak Wiss	1784
Syst Eleuth	1728

Figure 4. uBio parsing showing count of abbreviations found in publication area of microcitation.

Documentation regarding the parsing of *Index Animalium* data is found at http://uio.mbl.edu/Sherborne/index.html.

As seen in some of the examples below, there were some high accuracy results at times and mixed results in others. Problems with titles that are extremely common in the field of taxonomy, that Sherborn abbreviated in a way that the researcher could recognize when reading the citations, fall short when attempting to use computerized matching. Sherborn was not consistent in his abbreviation within the microcitations. Within the bibliography, he was not consistent with title, author, editor, edition, vol-

Parsing Index Animalium

Index Animalium

Publication Occurances
Zool Miscell 1407

Text Examples

uID	publication
61791	Zool. Miscell. III. 1817, 120
62178	Zool. Miscell. III. 1817, 123
62211	Zool. Miscell. (Gray) (1) Feb. 1831, 23
62392	Zool. Miscell. III. 1817, 113
62541	Zool. Miscell. (Gray) (6) June 1844, 83
62553	Zool. Miscell. III. 1817, 76
62580	Zool. Miscell. (Gray) (6) June 1844, 82
62998	Zool. Miscell. (Gray), (1) Feb. 1831, 8
63018	Zool. Miscell. (4) May 1842, 71
63303	Zool. Miscell. (Gray) (6) June 1844, 85
68599	Zool. Miscell. (Gray) (6) June 1844, 82
63724	Zool. Miscell. II. 1815, 79
63840	Zool. Miscell. (Gray) (6) June 1844, 84
64089	Zool. Miscell. II. 1815, 25
64794	Zool. Miscell. (Gray) (6) June 1844, [84
65110	Zool. Miscell. I. 1814, 41
66101	Zool. Miscell. (Gray) (1) Feb. 1831, 29
66432	Zool. Miscell. (Gray) (6) June

Bibliography Examples

bID	bibliography
1373	— Miscell. Zool. 4to. *Hagae Com.* 1766.,Yes,
5384	Zoological Miscellany (Leach's). 3 vols. 8vo. *Lond.* 1814-17. [Dates uncertain ; I adopt on my evidence I, 1814 ; II, 1815 ; III, 1817 ; it is almost certain that vol. I began Jan. 1814.],,Yes
7262	Zoological Miscellany (Gray's). For 1821-44 *read* 1831-44.,,
7766	Zoological Miscellany (Gray's). For 1821-44 *read* 1831-44.,,

Figure 5. uBio parsing. Microcitation "Zool Miscell" found 1407 times in the *Index Animalium* and potentially matches four entries in Sherborn's bibliographies.

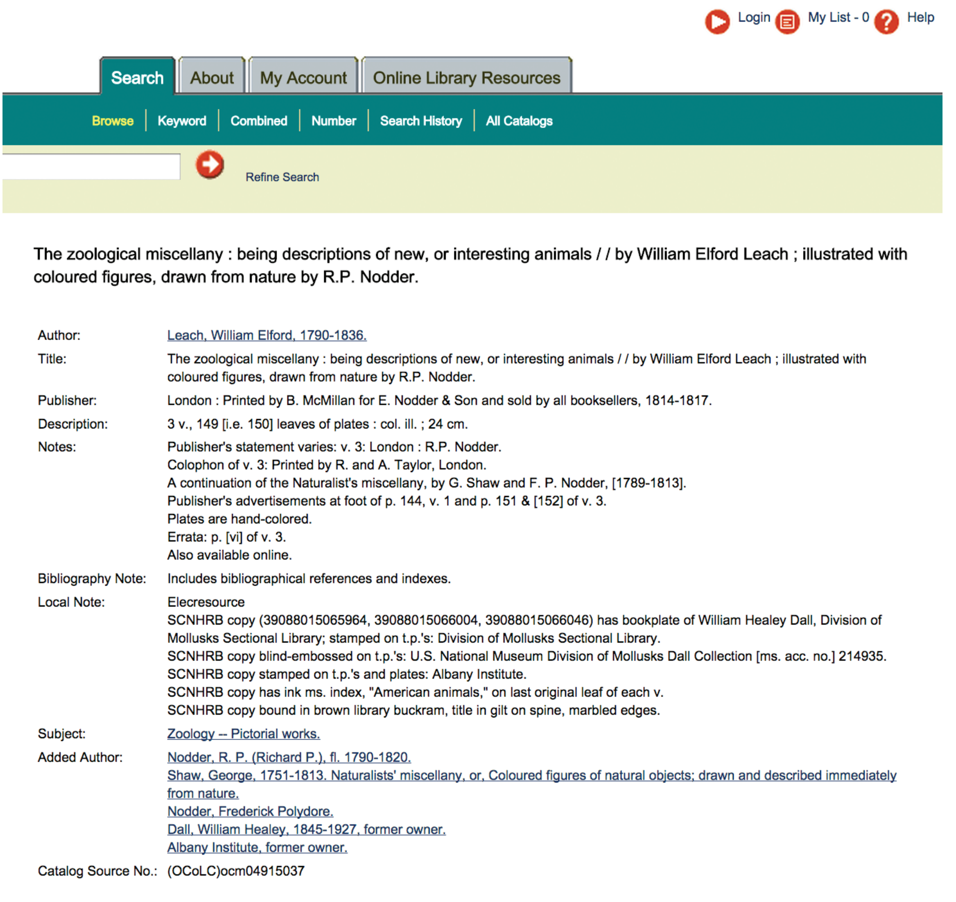
ume, publisher, or places of publication abbreviations. He was not always consistent on what he decided to abbreviate or how he formed the abbreviations. Having a systematic string matching between the citations to the bibliography did not prove to provide the clean matching that was needed for unambiguous one to one matching. The metadata fields do not line up for easy comparison and matching; for example: author to author versus author to editor. Century-old systems of notation translated into library standard database structure have been a road-block in speedily unlocking and connecting the data. (Pilsk, S.C., et al. "The Biodiversity Heritage Library: Advancing Metadata Practices in a Collaborative Digital Library" Journal of Library Metadata 10:136-155, 2010 doi: 10.1080/19386389.2010.506400) Connecting the index to the bibliography and the bibliography to library holdings has required many more hands and eyes than lines of script. Staff, interns and volunteers began to attempt to locate standard library records for each title in the bibliography.

IV.

A Microsoft Access database was constructed that contained only the bibliography from *Index Animalium*. Sherborn's bibliography entries were sorted by the greatest number of associated microcitations. These were searched and the full title, author, date and related identifiers were added to a database. Each line of data from the *Index's* bibliographies was matched against standard library data using Smithsonian's online

ID	21258
Bid	5384
Orig	Zoological Miscellany (Leach's). 3 vols. 8vo. Lond. 1814-17. [Dates uncertain ; I adopt on my evidence I, 1814 ; II, 1815; III, 1817 ; it is almost certain that vol. I began Jan. 1814.],,Yes
uBio Counts	1407
Has Species	Yes
Title	The zoological miscellany : being descriptions of new, or interesting animals
Author	Leach, William Elford
WorldCat ID	4915037
SIL SIRIS ID	120341
BHL Link	41372

Figure 6. Data from the Smithsonian Libraries database for one line in the *Index Animalium* Bibliography.


Login My List - 0 Help

Search | About | My Account | Online Library Resources

Browse | Keyword | Combined | Number | Search History | All Catalogs

Refine Search

The zoological miscellany : being descriptions of new, or interesting animals / / by William Elford Leach ; illustrated with coloured figures, drawn from nature by R.P. Nodder.

Author:	Leach, William Elford, 1790-1836.
Title:	The zoological miscellany : being descriptions of new, or interesting animals / / by William Elford Leach ; illustrated with coloured figures, drawn from nature by R.P. Nodder.
Publisher:	London : Printed by B. McMillan for E. Nodder & Son and sold by all booksellers, 1814-1817.
Description:	3 v., 149 [i.e. 150] leaves of plates : col. ill. ; 24 cm.
Notes:	Publisher's statement varies: v. 3: London : R.P. Nodder. Colophon of v. 3: Printed by R. and A. Taylor, London. A continuation of the Naturalist's miscellany, by G. Shaw and F. P. Nodder, [1789-1813]. Publisher's advertisements at foot of p. 144, v. 1 and p. 151 & [152] of v. 3. Plates are hand-colored. Errata: p. [vi] of v. 3. Also available online.
Bibliography Note:	Includes bibliographical references and indexes.
Local Note:	Elecresource SCNHRB copy (39088015065964, 39088015066004, 39088015066046) has bookplate of William Healey Dall, Division of Mollusks Sectional Library; stamped on t.p.'s: Division of Mollusks Sectional Library. SCNHRB copy blind-embossed on t.p.'s: U.S. National Museum Division of Mollusks Dall Collection [ms. acc. no.] 214935. SCNHRB copy stamped on t.p.'s and plates: Albany Institute. SCNHRB copy has ink ms. index, "American animals," on last original leaf of each v. SCNHRB copy bound in brown library buckram, title in gilt on spine, marbled edges.
Subject:	Zoology -- Pictorial works.
Added Author:	Nodder, R. P. (Richard P.), fl. 1790-1820. Shaw, George, 1751-1813. Naturalists' miscellany, or, Coloured figures of natural objects; drawn and described immediately from nature. Nodder, Frederick Polydore. Dall, William Healey, 1845-1927, former owner. Albany Institute, former owner.
Catalog Source No.:	(OCoLC)ocm04915037

Figure 7. Screen capture of the Smithsonian Institution Online Catalog SIRIS for William Elford Leach's Zoological Miscellany.

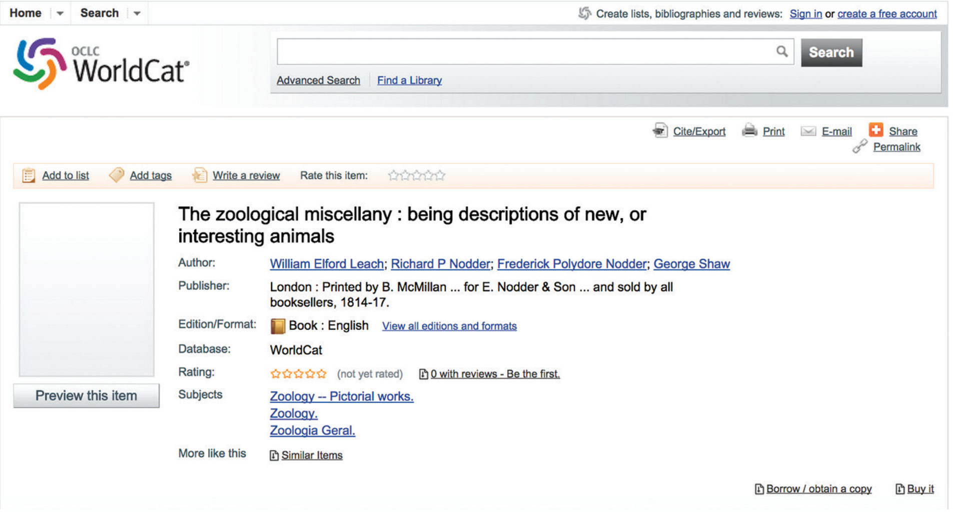


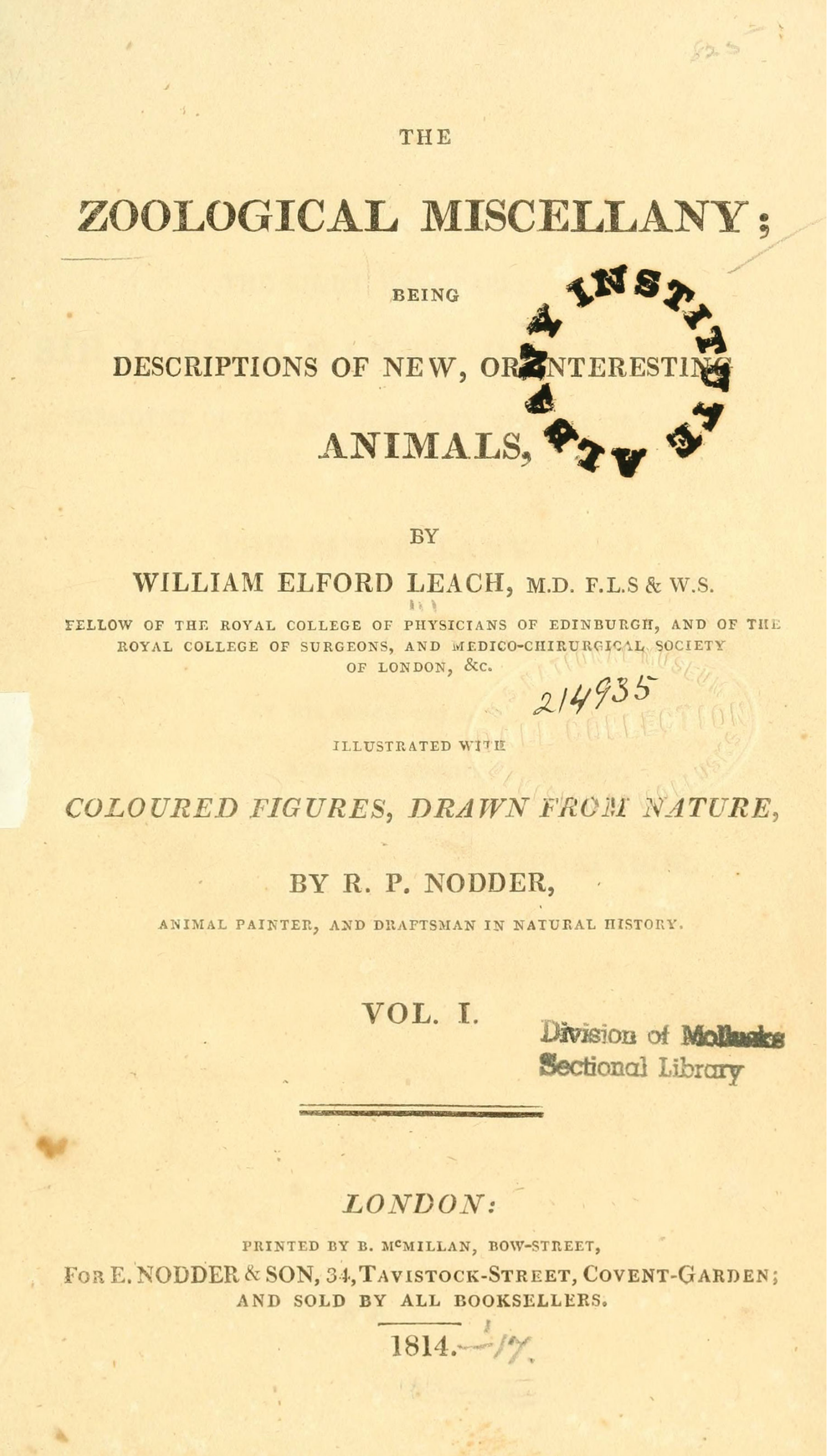
Figure 8. Screen capture from OCLC's WorldCat for William Elford Leach's Zoological Miscellany.

catalogue (SIRIS) and against the OCLC WorldCat catalogue. (Smithsonian Libraries' online catalog SIRIS is searchable via http://siris-libraries.si.edu.WorldCat is considered to be the largest network of library data. Listing library holdings from around the world, it contains metadata describing these titles following international standards. OCLC's WorldCat is available for searching http://www.worldcat.org/).

V.

The Smithsonian Libraries' *Index Animalium* project took a new direction as the Biodiversity Heritage Library (BHL) project began production. A large scale scanning project, BHL's mission is to digitize legacy natural history literature that is significant in the study and research of biodiversity. BHL is made up of a consortium of international natural history and botanical libraries. Specific funding of the BHL supports the scanning of the literature published before 1923 – titles that Sherborn referenced in his bibliographies. BHL ramped up fairly quickly and began to have full text scans online in 2005. Libraries participating in identifying and scanning the literature stretch across the globe and continue to produce millions of pages of online text ever year. Biodiversity Heritage Library information can be found at http://biodivlib.wikispaces.com/ and the collection is searchable http://www.biodiversitylibrary.org/.

As more and more of the literature becomes available online via the BHL, the *Smithsonian Libraries Index Animalium* project has, once again, shifted in the goal of service to the taxonomic researcher. Instead of getting the researcher to the library shelf for the text, it is becoming more desirable to deliver the fully scanned text to the researcher. Currently the matching of scanned titles is underway with identified titles in


THE

ZOOLOGICAL MISCELLANY;

BEING

DESCRIPTIONS OF NEW, OR INTERESTING

ANIMALS,

BY

WILLIAM ELFORD LEACH, M.D. F.L.S & W.S.

FELLOW OF THE ROYAL COLLEGE OF PHYSICIANS OF EDINBURGH, AND OF THE ROYAL COLLEGE OF SURGEONS, AND MEDICO-CHIRURGICAL SOCIETY OF LONDON, &c.

ILLUSTRATED WITH

COLOURED FIGURES, DRAWN FROM NATURE,

BY R. P. NODDER,

ANIMAL PAINTER, AND DRAFTSMAN IN NATURAL HISTORY.

VOL. I.

LONDON:

PRINTED BY B. McMILLAN, BOW-STREET,

FOR E. NODDER & SON, 34, TAVISTOCK-STREET, COVENT-GARDEN; AND SOLD BY ALL BOOKSELLERS.

1814.—17.

Figure 9. Title page from Biodiversity Heritage Library for William Elford Leach's Zoological Miscellany.

the *Index's* bibliographies. The anticipated result will have the researcher "click and go" from Sherborn's *Index* online to the BHL scanned text online. Apparently seamless, the connections being made behind the scenes match the online *Index's* microcitation to the full title record and resolve to the proper title in BHL.

VI.

Partnerships forming over the use of Sherborn's *Index Animalium* distribute the work into more functional pieces to achieve the seamless online research tool. Richard Pyle of the Bishop Museum is applying some reviews to the re-keyed text and providing complete citations for items that have partial data in the Smithsonian database. Another partnership is with Thomson Reuter's staff working on ION: Index to Organism Names. Thomson Reuters Index to Organism Names (ION) http://www.organism-names.com/ is a free online service to search the names included in *Zoological Record*, a continuously updated database of biological taxonomic research. As Nigel Robinson's presentation "Sherborn's *Index Animalium* Integration into ION: Access to All" demonstrated, the parsing of the microcitations and identifying the full text is underway increasing the data in ION and providing the connection needed for the taxonomic

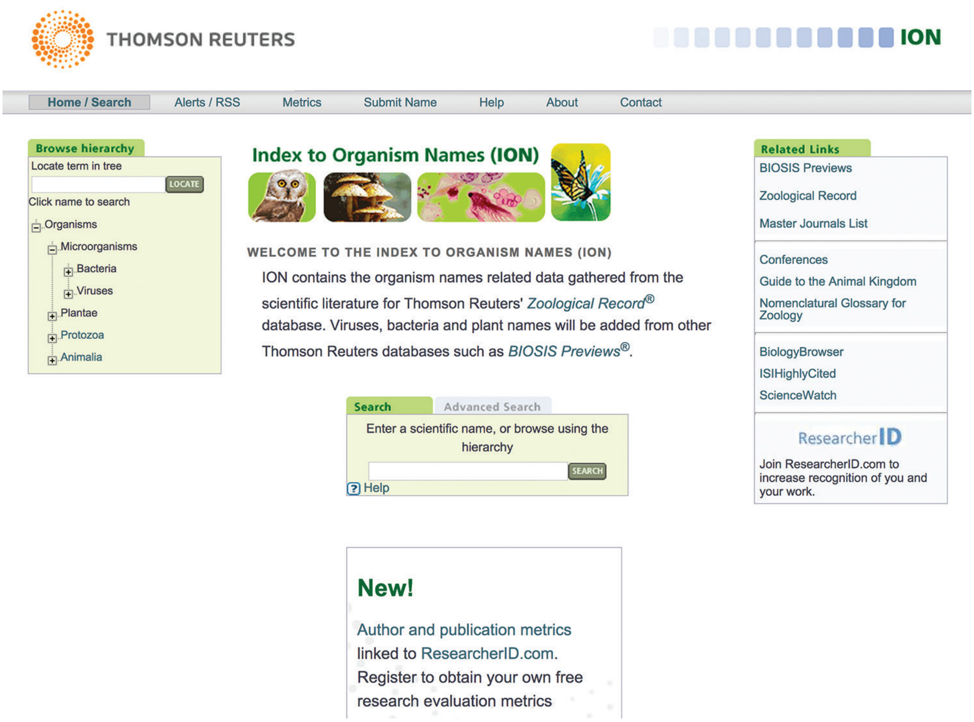


Figure 10. ION home page screen capture.

splendens Tellina, J. De C. Sowerby, Tr. Geol. Soc. London [2] V. 1837, 136.
splendens Temnoscheila, G. R. Gray in Griffith's Cuvier, Anim. Kingd., Ins. II. 1832, 93.
splendens Tetragonus, Gory & Percheron, Mon. Cétoines (2) 1834, 67.
splendens Tityra (Wied), G. R. Gray, Gen. Birds, I. June 1846, 254.—Muscicapa.
splendens Toxotus, Laich.; Dejean, Catal. Coléopt. 1821, 112 [*n. n.*].
splendens Trochus, C. G. Giebel, De Geogn. Hercyn. 1848, 28 [*n. n.*].
splendens Turbo, O. G. Costa, Cat. Test. Sicilie, 1829, pp. ci & civ.
splendens Turdus, W. E. Leach, Zool. Miscell. II. 1815, 30. [1800.
splendens Turdus (Daud.), L. P. Vieillot, N. Dict. H. N., ed. 2, XX. 1818, 260.—Sturnus,
splendens Udorpes, V. v. Motschulsky, Bull. Soc. Imp. Nat. Moscou, XVIII. 1845 (1) 1845,
splendens Unio, G. A. Goldfuss, Petref. German. II (6) 1837, 183. [108.
splendens Venus, D. Solander, Catal. Portland Mus. 1786, 102 & 149 [*n. n.*].
splendens Volatinia (Vieill.), C. L. Bonaparte, Consp. gen. Avium, I. 1850, 474.—Fringilla.

Figure 11. Sherborn citation from page 6101 of the Index Animalium: *Turdus splendens*, W.E. Leach, Zool. Miscell. II. 1815, 30.

Smithsonian ID	362382
Smithsonian IA Page Image ID	SIL34_02_24_0193
IA Page	6101
Species	splendens
Genus	Turdus
Author Forename	W.E.
Author Last name	Leach
Abbreviated Publication Name	Zool. Miscell.
Volume	II.
Year	1815
Page	30

Figure 12. Parsed Data from Smithsonian Libraries Sherborn Database.

researcher. Slides from Nigel Robinson's presentation are available at http://www.slideshare.net/iczn/4-sherborns-index-animalium-integration-into-ion.

The ION team working on *Index Animalium* at Thomson Reuters is looking at supplementing the *Zoological Record* dataset. Sherborn's data back fills ION with taxonomic names for 1758 to 1864. To achieve the data extraction from *Index Animalium*, Robinson reports that there are challenges in parsing and properly identifying the data elements. The review of the data is needed since Sherborn's use of commas, brackets and notations all have meanings that need to be carefully interpreted so as to not lose the intention. As the project has progressed, inconsistencies are coming to light that can now be documented. With this detailed look, the ION team is finding re-keying errors, as well as errors made by Sherborn, and the typesetting done based on Sherborn's initial transcriptions.

ION's management classification protocol is also being added to the microcitation so that the data can be processed and incorporated into the systems already in place at Thomson Reuter. The species and genus identified in Sherborn are being folded into the overall

Genus	Turdus
Species	splendens
Name string	Turdus splendens
Author last name	Leach
Author forename	W.E.
Abbreviated publication name	Zool. Miscell.
Full Publication name	The Zoological Miscellany
Volume	II
Year	1815
Page	30

Figure 13. Focusing on the species and genus names, ION is examining the publication abbreviation against the internal Zoological Record data and resolving the microcitations.

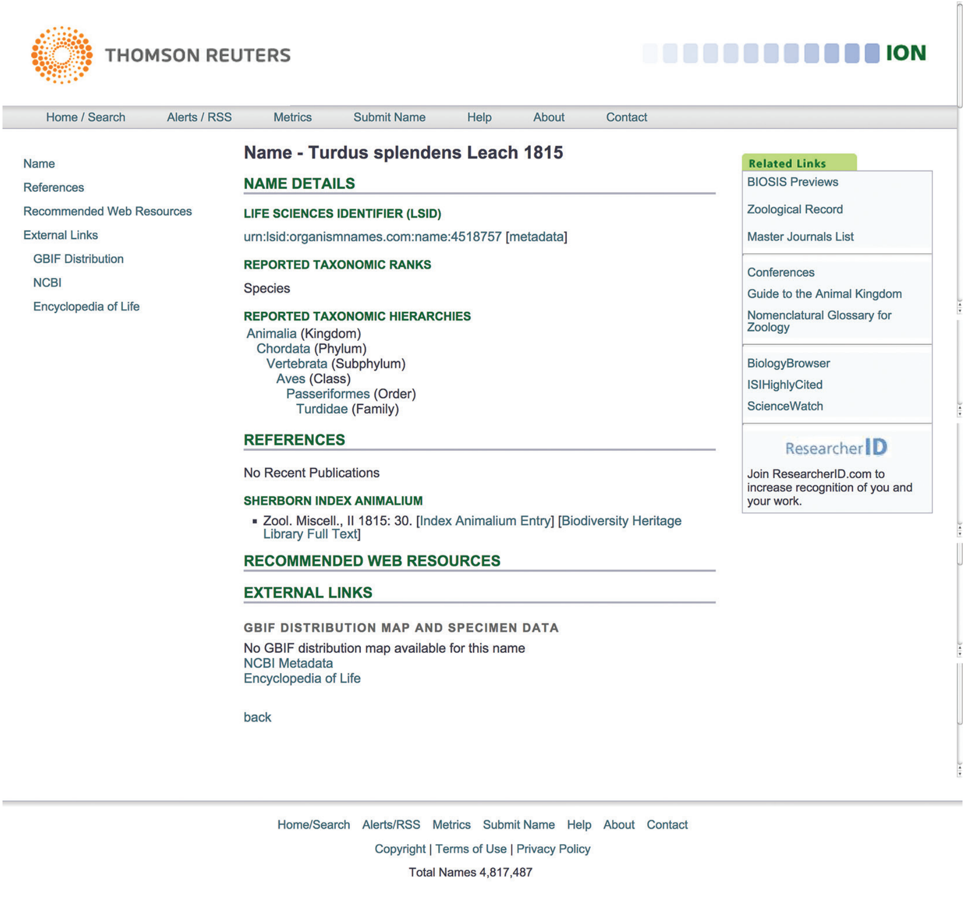


Figure 14. Results of searching in ION for *Turdus splendens*.

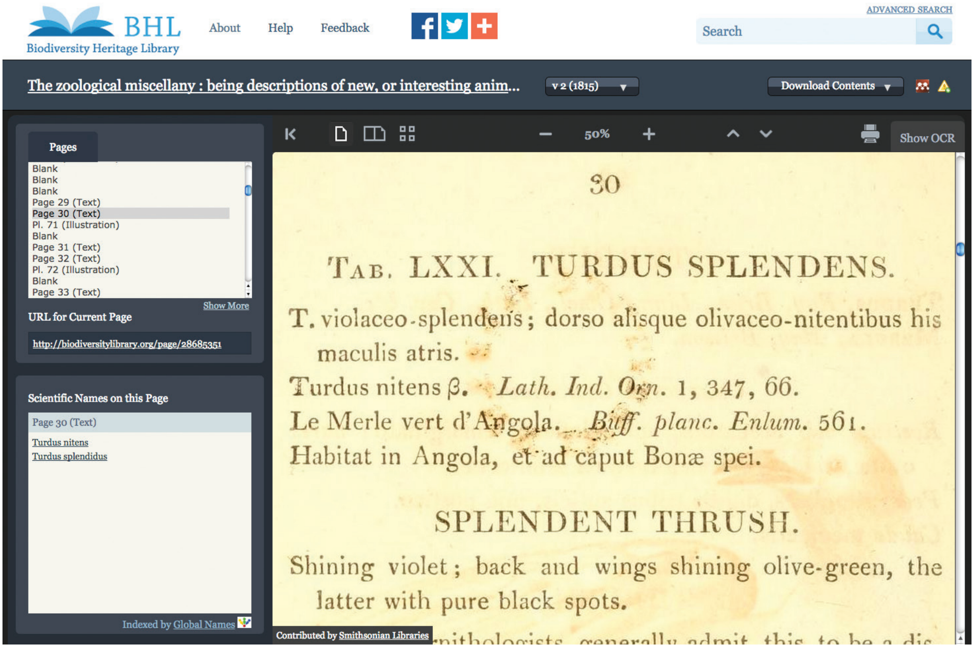


Figure 15. Results of a direct link from ION *Turdus splendens* to the page in the Biodiversity Heritage Library.

delivery of data via the ION search. Robinson's presentation illustrated the parsing with an example of one line of data from Smithsonian Libraries Sherborn Database that teases out the identification of the citation in the Smithsonian database and the various elements.

(View the page of *Index Animalium* for this reference http://www.sil.si.edu/digitalcollections/indexanimalium/volumes/pagedisplaypage.cfm?filename=SIL34_02_24_0193)

From this breakdown and reconfigurations, ION is able to map data into ION and integrate with the existing ION content to form a nomenclator of names for the literature published from 1758 onwards.

BHL provides stable consistent page identifiers for all titles scanned. In this example *Turdus splendens* page identifier http://biodiversitylibrary.org/page/28685351 is a persistent identifier allowing ION to create a direct link into the Biodiversity Heritage Library. The results are the "click and go" for the user to reach the page of the text Sherborn cites.

VII.

The challenges of working with legacy taxonomic citations, computer matching algorithms, and making connections have not stopped the attempts to continually improve the reach of Sherborn's unique and critical data to the researcher. New developments and constant revisiting of the goals has brought us to yet another shift to today's goal of making Sherborn's *Index* available and linked to other datasets. The Smithsonian Libraries is exploring a different data structure than a relational database currently in use. Partnering with others in the world of metadata development and information sharing has led to an attempt to allow machine-to-machine communications. The *Index* is being looked at as the data set of the elements it contains. These data points are being examined for possible transformation into RDF mark up and meeting the standards of Linked Open Data. This will allow for broader discovery and access than a stand-alone database. Linked Open Data is primarily aimed at consumption by computer software, but the availability of such data allows the offering of an online research tool geared towards the general population of natural history researchers.

Linked Data is based on the concept of triples or a sentence made up of three parts: subject, predicate, and object. The subject is an identifiable "thing" that can be assigned a unique identifier. The predicate can be considered the "verb" with a controlled vocabulary that has a term defined and assigned a unique identifier. The object is the last "thing" in the triple that subject is connected. A possible triple that would be created from Sherborn's *Index Animalium* is diagrammed below. In this scheme, each species is presented as an identifier with related microcitation data pointing to the

Subject	<genus>Turdus <species>splendens
Predicate	Authored by
Object	<lastname>Leach<forename>William Elford
Subject	<genus>Turdus <species>splendens
Predicate	Published in
Object	The zoological miscellany : being descriptions of new, or interesting animals

Figure 16. Example of potential triples from Index Animalium's citation for "splendens Turdus".

scanned title and page at BHL. The goal of providing the Index as an open data set in the RDF would allow others to reuse, repurpose, and mine the data.

The details of creating a complete open linked data set out of *Index Animalium* are still being discussed and explored. Smithsonian Libraries, dedicated to providing data in an open platform, is already beginning to work on providing some Open Linked Data in a new project based off the *Taxonomic Literature: A selective guide to botanical publications and collections with dates, commentaries and types*, 2nd edition. Known by most as TL2, the entire 15 volume set has been scanned and OCRed. The data is currently available for searching and the break down into triples has begun. Smithsonian Libraries TL2 online (http://www.sil.si.edu/digitalcollections/tl-2) allows for reading or searching the entire text of the literature of systematic botany published between 1753 and 1940. Incorporating *Index Animalium*, Smithsonian's goal is to create a TL3: an online resource containing both botanical and zoological linked open data resource for taxonomic research.

VIII.

A project that began to simply provide a URL for anyone in the world to read Sherborn's *Index Animalium* has grown and changed as the fast paced world of knowledge sharing has adapted to the technology available. The *Index* has matured from the pieces of paper of Charles Davies Sherborn's carefully indicated notes of species citations to a linked data structure. The overarching goal of providing access has been achieved but there is room for it to improve by making the information usable, repurpose-able, and integrated into the researcher's workflow.

References

Alonso-Zarazaga MA, Fautin DG, Michel E (2016) The List of Available Names (LAN): A new generation for stable taxonomic names in zoology? In: Michel E (Ed.) Anchoring Biodiversity Information: From Sherborn to the 21st century and beyond. ZooKeys 550: 225–232. doi: 10.3897/zookeys.550.10043

Alonso-Zarazaga MA, Bouchet P, Pyle RL, Kluge N, Fautin DG (2016) Manual for proposing a Part of the List of Available Names (LAN) in Zoology. In: Michel E (Ed.) Anchoring Biodiversity Information: From Sherborn to the 21st century and beyond. ZooKeys 550: 283–298. doi: 10.3897/zookeys.550.10042

Dickinson EC (2016) Reinforcing the foundations of ornithological nomenclature: Filling the gaps in Sherborn's and Richmond's historical legacy of bibliographic exploration. In: Michel E (Ed.) Anchoring Biodiversity Information: From Sherborn to the 21st century and beyond. ZooKeys 550: 107–134. doi: 10.3897/zookeys.550.10170

Evenhuis NL (2016) Charles Davies Sherborn and the "Indexer's Club". In: Michel E (Ed.) Anchoring Biodiversity Information: From Sherborn to the 21st century and beyond. ZooKeys 550: 13–32. doi: 10.3897/zookeys.550.9697

Kalfatovic MR (1998) Re-Defining the Library Meme: Memory and Imagination. In: Wolf MT, Ensor P, Thomas MA (Eds) Information Imagineering: Meeting at the Interface. American Library Association, Chicago, 155–65.

Licklider JCR (1965) Libraries of the Future. MIT Press, Cambridge, 219 pp.

Lyal CHC (2016) Digitising legacy zoological taxonomic literature: Processes, products and using the output. In: Michel E (Ed.) Anchoring Biodiversity Information: From Sherborn to the 21st century and beyond. ZooKeys 550: 189–206. doi: 10.3897/zookeys.550.9702

Marcum D (2013) The Biodiversity Heritage Library: Ithaka S+R Case Study. http://sr.ithaka.org/sites/default/files/reports/SR_BHL_20140129.pdf

Page RDM (2006) Taxonomic names, metadata, and the Semantic Web. Biodiversity Informatics 3: 1–15. doi: 10.17161/bi.v3i0.25

Page RDM (2008) Biodiversity informatics: the challenge of linking data and the role of shared identifiers. Briefings in bioinformatics 9(5): 345–54. doi: 10.1093/bib/bbn022.

Page RDM (2016) Surfacing the deep data of taxonomy. In: Michel E (Ed.) Anchoring Biodiversity Information: From Sherborn to the 21st century and beyond. ZooKeys 550: 247–260. doi: 10.3897/zookeys.550.9293

Pilsk SC, Person MA, deVeer JM, Furfey JF, Kalfatovic MR (2010) The Biodiversity Heritage Library: Advancing Metadata Practices in a Collaborative Digital Library. Journal of Library Metadata 10: 136–155. doi: 10.1080/19386389.2010.506400

Remsen D (2016) The use and limits of scientific names in biological informatics. In: Michel E (Ed.) Anchoring Biodiversity Information: From Sherborn to the 21st century and beyond. ZooKeys 550: 207–223. doi: 10.3897/zookeys.550.9546

Remsen DP, Norton C, Patterson DJ (2006) Taxonomic informatics tools for the electronic Nomenclator Zoologicus. The Biological Bulletin 210(1): 18–24. doi: 10.2307/4134533.

Robinson N (2011) Sherborn's *Index Animalium* Integration into ION: Access to All. Presentation at "Anchoring Biodiversity Information: From Sherborn to the 21st Century and Beyond." Symposium, 28 October 2011. http://www.slideshare.net/iczn/4-sherborns-index-animalium-integration-into-ion

Seitchek C (2003) Century-old biology text gets an e-life. The Torch 3(8): 6–7.

Sherborn CD (1903–33) Index animalium : sive, Index nominum quae ab A.D. MDCCLVIII generibus et speciebus animalium imposita sunt, societatibus eruditorum adiuvantibus. British Museum, London. http://www.sil.si.edu/digitalcollections/indexanimalium/ and http://biodiversitylibrary.org/bibliography/14658

Shindler K (2016) A magpie with a card-index mind – Charles Davies Sherborn 1861–1942. In: Michel E (Ed.) Anchoring Biodiversity Information: From Sherborn to the 21st century and beyond. ZooKeys 550: 33–56. doi: 10.3897/zookeys.550.9975

Smithsonian Libraries (1995) Rare Books and Special Collections in the Smithsonian Institution Libraries. Smithsonian, Washington, DC, 106 pp.

Stafleu A, Cowan RS (1976-1988) Taxonomic Literature: a Selective Guide to Botanical Publications and Collections with Dates, Commentaries and Types (7 volumes). Bohn, Scheltema and Holkema, Utrecht. Online at http://www.biodiversitylibrary.org/item/103414 and http://www.sil.si.edu/digitalcollections/tl-2 [database format]

Thompson CF, Pape T (2016) Sherborn's influence on Systema Dipterorum. In: Michel E (Ed.) Anchoring Biodiversity Information: From Sherborn to the 21st century and beyond. ZooKeys 550: 135–152. doi: 10.3897/zookeys.550.9447

Weitzman AL, Lyal CHC, Garnett T (2004) The Biologia Centrali-Americana Centennial: A vision for electronic access to taxonomic resources, the information interface between libraries and systematic biology. A talk presented at The National Museum of Natural History, Smithsonian Institution, June 2004.

Weitzman AL, Lyal CHC (2004) An XML schema for taxonomic literature – taXMLit. http://www.sil.si.edu/digitalcollections/bca/documentation/taXMLitv1-3Intro.pdf

Weitzman AL, Lyal CHC (2006) INOTAXA – INtegrated Open TAXonomic Access and the "Biologia Centrali-Americana". Proceedings of the Contributed Papers Sessions Biomedical and Life Sciences Division, Special Libraries Association, 8 pp.

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. In: Michel E (Ed.) Anchoring Biodiversity Information: From Sherborn to the 21st century and beyond. ZooKeys 550: 173–183. doi: 10.3897/zookeys.550.10041

ZooKeys 550: 173–187 (2016)
doi: 10.3897/zookeys.550.10041
http://zookeys.pensoft.net

RESEARCH ARTICLE

A peer-reviewed open-access journal
ZooKeys
Launched to accelerate biodiversity research

Sherborn's *Index Animalium*: New names, systematic errors and availability of names in the light of modern nomenclature

Francisco Welter-Schultes[1], Angela Görlich[1], Alexandra Lutze[1]

1 *Zoologisches Institut der Universität, Berliner Strasse 28, D-37073 Göttingen, Germany*

Corresponding author: *Francisco Welter-Schultes* (fwelter@gwdg.de)

Academic editor: *E. Michel* | Received 25 May 2015 | Accepted 25 May 2015 | Published 7 January 2016

http://zoobank.org/C111F4AF-4C13-4C88-B784-BB1F9DE87AEC

Citation: 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. In: Michel E (Ed.) Anchoring Biodiversity Information: From Sherborn to the 21st century and beyond. ZooKeys 550: 173–187. doi: 10.3897/zookeys.550.10041

Abstract

This study is aimed to shed light on the reliability of Sherborn's *Index Animalium* in terms of modern usage. The AnimalBase project spent several years' worth of teamwork dedicated to extracting new names from original sources in the period ranging from 1757 to the mid-1790s. This allowed us to closely analyse Sherborn's work and verify the completeness and correctness of his record. We found the reliability of Sherborn's resource generally very high, but in some special situations the reliability was reduced due to systematic errors or incompleteness in source material. *Index Animalium* is commonly used by taxonomists today who rely strongly on Sherborn's record; our study is directed most pointedly at those users. We recommend paying special attention to the situations where we found that Sherborn's data should be read with caution.

In addition to some categories of systematic errors and mistakes that were Sherborn's own responsibility, readers should also take into account that nomenclatural rules have been changed or refined in the past 100 years, and that Sherborn's resource could eventually present outdated information. One of our main conclusions is that error rates in nomenclatoral compilations tend to be lower if one single and highly experienced person such as Sherborn carries out the work, than if a team is trying to do the task. Based on our experience with extracting names from original sources we came to the conclusion that error rates in such a manual work on names in a list are difficult to reduce below 2–4%. We suggest this is a natural limit and a point of diminishing returns for projects of this nature.

Keywords

Sherborn

Introduction

Appreciation of Sherborn's tremendous work grows when we understand the extent to which Sherborn's index data can be used for nomenclatural and taxonomic purposes today. However, some of the names listed in the *Index Animalium* should be used with caution. *Index Animalium* remains a core source of nomenclatural bibliographic information, but is not without errors.

In the AnimalBase project we manually checked original sources for approximately 40,000 names that were new between 1757 and 1795 and compared our results with those in Sherborn's *Index Animalium*. For each examined work we extracted all new names under the present-day nomenclatural rules (4th edition of the ICZN Code), and compared our results with Sherborn's list extracted from the same work.

Total numbers of names

It is crucial to know how to read the *Index Animalium*, as not all names were marked as new. Sherborn listed 420,000 names referenced between 1757 and 1850. 70% of these (300,000) were listed as new names from their original publications, however 30% of these names were not new.

Taxonomic productivity, as indicated by number of new taxa described (Fig. 1), decreased temporarily around 1810 and increased enormously after 1835. After 1850 the rate did not continue to increase, but stayed at the same high level until around 1910 when it fell to levels of the 1820s. In the AnimalBase project we checked new names published between 1757 (the starting point for Code-regulated zoological nomenclature, ICZN Articles 3.1 and 3.2) and the mid-1790s, which was the historical limit for AnimalBase work for funding reasons. The numbers of listed names in this period listed 10-30% more names than found by AnimalBase, mainly because Sherborn included more *nomina nuda* (unavailable new names mentioned without description) than AnimalBase, and in addition many subsequent uses of previously established names.

Understanding Sherborn's style

Sherborn changed his standards various times in the course of his work. This sometimes gives the impression that he had a team and difficulties in reaching a common standard, however we know that he worked completely on his own (Evenhuis 2016, Shindler 2016). Initially he seemed to have recorded all subsequent uses of names, but this was given up (this was because he started with Linné (1766) and did not remove the entries when he decided to add those of the 1758 work). For a slightly longer period he recorded when specific names were placed in different genera. We have not been

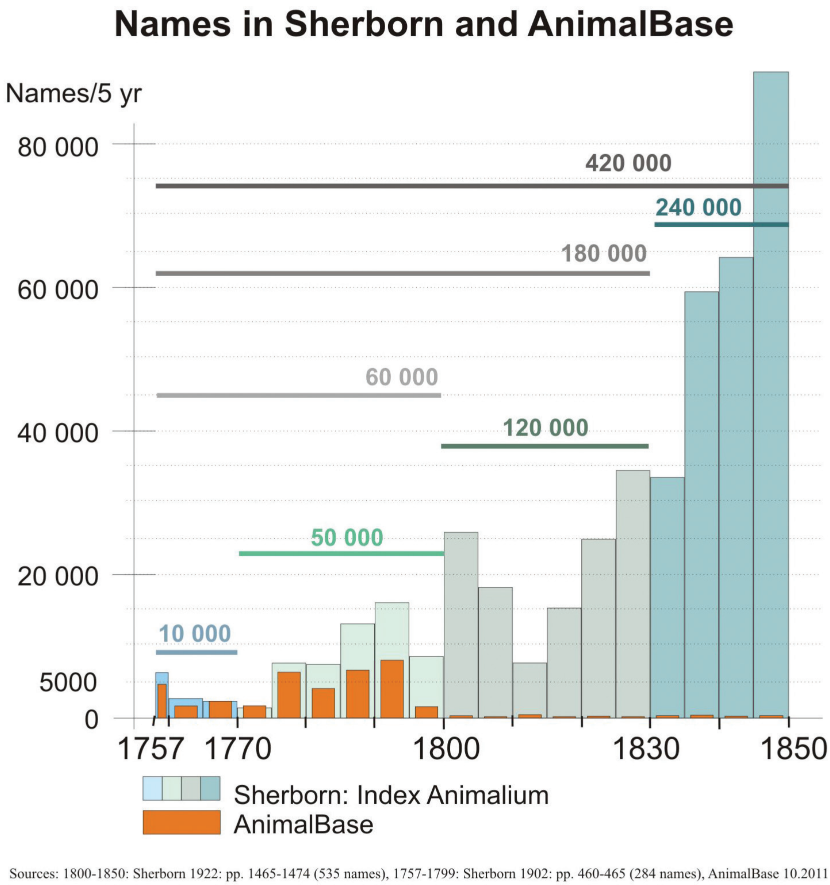


Figure 1. Number of names. Total number of taxonomic names listed in Sherborn's *Index Animalium* and in AnimalBase (2011) plotted in 5-year intervals. It must be taken into account that Sherborn's numbers included a proportion of 30% of names that were not new, while in AnimalBase this proportion was much lower (less than 5 %).

able to explain why *nomina nuda* or varieties were occasionally included. At later stages he only recorded new names and nothing more. When Sherborn was extracting names established in the 1790s he finally seemed to have reached a more stable standard.

We aim to shed light on Sherborn's errors and their sources, as this will help taxonomists who use Sherborn's data today to use it with greater reliability. With an understanding of systematic error sources, users can know when Sherborn's record is at its most reliable, and when to proceed with some caution. We found that Sherborn's data were consistent with our own finds at an average rate of 80–90%. The degree of reliability of Sherborn's data differed by work and by animal group and depended on factors that we discuss below.

Bibliographic methods and materials

How did Sherborn mark a name that was classified as an available new name, what was a subsequent use, a misspelling or a *nomen nudum*? Not all names in the *Index Animalium* were listed as new. We determined the rate of new names at roughly 70% of the listed names. 28–30% of the listed names were marked as subsequent uses and 2.5% as *nomina nuda*. Here are some examples of how to read Sherborn's *Index Animalium* (our own comments below, after a hyphen):

aenea Nitidula, J. C. Fabricius, Syst. Ent. 1775, 78.
- New name, available.

carocolla Helix, Linnaeus, Syst Nat., ed. 10, 1758, 769 ; ed. 12, 1767, 1243.
- New name, available, with one subsequent use recorded for 1767.

haliaëtos Falco, J. F. Gmelin in Linn. Syst. Nat., ed. 13, I. 1788, 263 ; varr. arundinaceus, carolinensis, cayenensis.
- No new name *F. haliaetos*, but three new names of varieties (Sherborn regarded these names as unavailable; today they are regarded as available because of ICZN Art. 45.6.4)

aegyptius Lygaeus (L.), J. C. Fabricius, Ent. Syst. IV. 1794, 155.—Cimex.
- No new name, only a subsequent use of *Cimex aegyptius* Linnæus, 1758.

abbreviator Cryptus, J. C. Fabricius, Syst. Piezat. 1804, 84.—Ichneumon, 1798.
- No new name, only a subsequent use of *Ichneumon abbreviator* Fabricius, 1798.

aterrima Megilla (Panz.), J. C. Fabricius, Syst. Piezat. 1804, 331.—Apis, 1798.
- No new name, only a subsequent use of *Apis aterrima* Panzer, 1798.

abnormis Terebratula, M. Tuomey, Rep. Geol. S. Carolina, 1848, 209 [n. n.].
- No new name, not made available (nomen nudum).

carolina Manduca, J. Huebner, Exot. Schmett., Tab. Manduca [n. et f.].
- Name without description but with figure (nomen et figura) (Sherborn regarded this name as unavailable; today the name is regarded as available because of ICZN Art. 12.2.7).

Differences in nomenclature between Sherborn's times and today

Nomenclatural rules applied 100 years ago were different than those in force today. Sherborn began his work in the 1890s. The rules of zoological nomenclature were internationally fixed in 1905 (Blanchard et al. 1905), and not always in Sherborn's sense. This generated errors for which Sherborn was not responsible as the criteria for availability of names, corrections of incorrect Latin, authorships for names, unavailability of non-binominal works had changed underneath him. In the following century the rules were continuously modified and refined. With every new edition of the ICZN Code the number of available names has changed again, despite all efforts to keep the status of those names stable which were made available under previous rules. Thousands of animal names still suffer from unclear regulations and different interpretations of ambiguous rules in the Code.

We looked closely at each name in an original copy of a published work and did not rely on Sherborn's list. For example, Sherborn regarded names as available that had no description but listed only a host plant, which would be unavailable today. In contrast, he regarded names as unavailable that had no description but had a bibliographic reference to a description, but these would be regarded as available today. Sherborn usually did not list names of fossils or names of varieties, however today these are regarded as available names, in the technical sense of the Code.

Names for varieties were not usually listed as new by Sherborn. Names for varieties were mentioned from very early dates, for example those mentioned in Linné's publications. Later (sometime between 1780 and 1800) Sherborn discontinued including variety names. This shows that over time Sherborn modified his own standards. As today these variety names are regarded as available, the consequences of Sherborn's decision to exclude them created a systematic hole in the *Index Animalium*, with those names missing or listed from subsequent sources with incorrect dates.

Names with host plants

Names that were only mentioned with a host plant, but without description, were generally listed as new by Sherborn who regarded such names as available. Today a new name that was published only with a host plant and not with a description, figure or indication, is regarded as unavailable (ICZN Code Art. 12.3).

Names with figures

Names that were mentioned without description and that were only equipped with a figure were usually not listed as new by Sherborn as he did not regard them as available. Today a new name which was only published with a figure in the original source, is regarded as available (ICZN Art. 12.2.7).

Names with references

New names that were mentioned without descriptions, but with bibliographical references to descriptions, were usually not listed as new by Sherborn as he regarded such names as unavailable. Today a new name published with a bibliographical reference in the original source is regarded as available (ICZN Art. 12.2.1).

Non-binominal works

Names that were mentioned in important non-binominal or not consistently binominal works were inconsistently listed as new and available by Sherborn. Today a new name established in a non-binominal work is regarded as unavailable (ICZN Art. 11.4). The ICZN Code however does still not contain a precise definition of what exactly is a binominal name, so still today some works are under debate from the point of view of binominality. By attributing the new names to those works, Sherborn missed recording the first sources in binominal works where those names were later actually made available. Some examples for such non-binominal works are Geoffroy (1762), Hasselquist (1762), Gronovius (1763, 1781), Tunstall (1771), Zimmermann (1777), Meuschen (1778), and Martyn (1784).

Correct spellings

Incorrect Latin was often corrected by Sherborn, who did not always copy the orthography exactly as used in the original source. This applied to some cases where Sherborn regarded the Latin names as incorrectly spelled. Today incorrect Latin is not corrected (ICZN Art. 32.5). Sherborn corrected incorrect Latin in only relatively few cases, but it makes *Index Animalium* unreliable to some degree. In questionable cases (if a name sounds unusual or has an unusual spelling) it is always necessary to consult the original sources. For example, *compunctus* was corrected to *compunctor*. The rate of such debatable names among all names is probably less than 1%, but these few names accumulate significantly among the names a taxonomist will consult when having been confronted with a spelling problem in the literature.

Authorship

At the beginning of Sherborn's work there were no universally agreed rules for nomenclature, and conventions for authorship of taxonomic names deviated substantially from today's rules. Sherborn often attributed names to people to which names were attributed in original sources, but who could have been people who had not contributed to the descriptions. These people are usually called authors of manuscript names. Many manuscript names of molluscs were created by shell dealers (for example, Ziegler or Parress) who labelled specimens with self-invented names without having published anything. After the adoption of the first global rules of zoological nomenclature in 1905 it was determined that the authorship must be attributed to the scientist who published the first description (Blanchard et al. 1905). This basic rule has been maintained in all subsequent Codes (ICZN Art. 50). From 1905 onwards it has been globally accepted that the names should be attributed to the first scientists who have published a description, which in the mollusc examples were often E. A. Rossmässler or L. Pfeiffer.

Error sources: understanding circumstances under which new names were systematically missed

Sherborn's individual (i.e., non-systematic) error rate was remarkably low. His usual rate of overlooked names was 1–2%, and in carefully compiled works, it was sometimes even less. This was a very low rate for any human endeavour, let alone one of such monumental scale requiring detailed work over many decades. It is all the more impressive when we realise that today we have comparable failure rates, despite having many computer tools and in some cases teams of people to help with these issues.

The proportion of overlooked names from each source work depended on its style. We observed up to 40% overlooked names in chaotically arranged works such as Forskål (1775), Hartmann (1821) or Swainson (1840). We use the expression "chaotic work" for publications in which it is very difficult to detect a systematic arrangement of species accounts, different names for the same species were used in the same work, species were classified in different genera in the same work without visible preference, new names were not indicated as such and it was hardly visible if a name was new or not, and no systematic or alphabetic index was present.

However, even in some well-organised works Sherborn overlooked 2–5% of the new names for no apparent systematic reason. We were surprised to observe this in the extracted names of some of Fabricius's insect works which all had a clear style. Our own error rates ranged at the same levels or higher. Sometimes names in *Index Animalium* were listed twice, with both correct and incorrect information.

There were however various systematic error sources that produced more damage to the reliability of the record.

Languages

Publications in languages that Sherborn had not mastered were a visible and significant problem. Sherborn was certainly well-versed in Latin, but had obvious difficulties in understanding all other languages except English. The problem was that only a fraction of the relevant works he had to consult was written in English or Latin (Fig. 2).

In the late 1700s German and Latin were the most commonly used languages for scientific publications. Even at that time, Latin was in continuous decline. Between 1760 and 1790 English was less frequently used than Swedish. Danish and Dutch were also important in this epoch. The importance of French in binominal zoological literature increased dramatically after Buffon's death in 1788. Italian gained importance after 1820. This means that, at a minimum, Sherborn understood only one third of the published texts in the works from which he had to extract new names. Recognising names of new taxa was usually no problem, but understanding if a name was new or only a subsequent use of a previously established name (often transferred to a different genus) usually required understanding the context of descriptions.

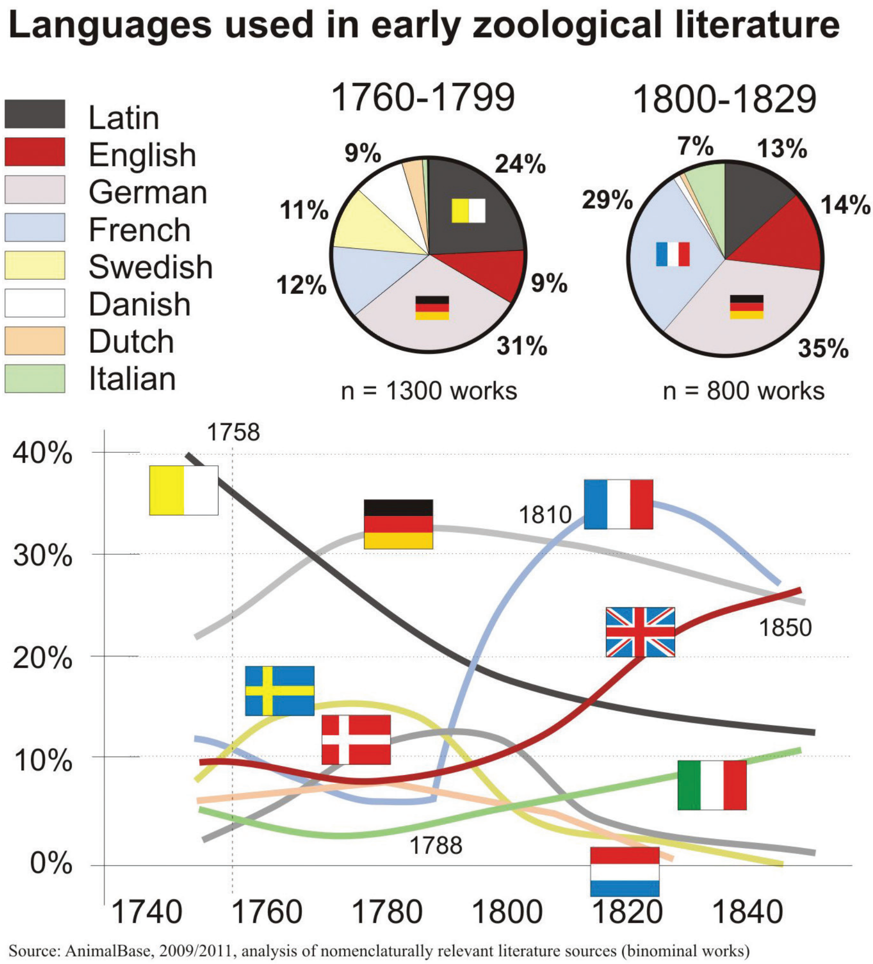


Figure 2. Languages used in early zoological literature. Language analysis of 2100 arbitrarily selected binominal zoological works published between 1758 and 1850.

In our survey of Sherborn's error rates we were able to see that if images or Latin diagnoses were absent and texts were only in French, German, Dutch, Danish or Swedish, Sherborn often failed to draw the appropriate conclusions and consequently erroneously classified many names as new.

Inaccessible works

Sherborn had unparalleled access to literature in London, but despite all efforts to get all published zoological literature, Sherborn did not have all relevant works at his disposal. Names established in missing works were not included in the *Index Animalium* and this often had tremendous impact: It distorted Sherborn's work to a visible extent.

A striking example in the molluscan record is Férussac and Deshayes (1819–1851): *Histoire naturelle des mollusques*. This was an important basic work for molluscs in which 1% of the currently used names for species in Europe were established (and 5% of the names published before 1840). Sherborn did not have access to this work, so he missed many important names or listed them incorrectly as new from subsequent works. In combination with varieties having been ignored in most works, and careless research in difficult works like Hartmann (1821), this had the effect that today the *Index Animalium* is not widely considered as a reliable source for molluscs.

Difficult works: Careless research as a strategy to save time

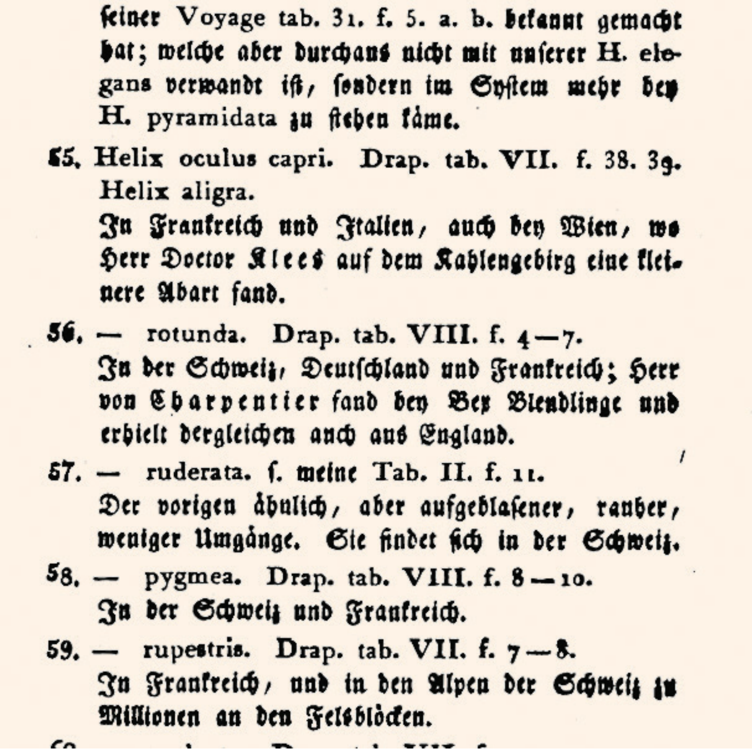
Sherborn had to analyse difficult and chaotically arranged works, works for which extracting one new name took much longer than in carefully compiled works. Difficult works became increasingly common after 1790. One example is Hartmann (1821), written in German in Fraktur script with 64 new names of molluscs, of which 6 are used today (Fig. 3). New names in this work were not indicated to be new, so it was very difficult to see in the work which names were new and which ones had previously been established. In such a case it is necessary to check every single mentioned name, to see if it had previously been established or not. Without a computer this is a time-consuming process.

Of those that were new, it was difficult to see if they had a description, because no Latin diagnoses were given and many different topics were discussed in the text. Some names were subsequently used and placed into a different genus, so that it requires either additional research to find such previously established names in the database, or special malacological knowledge, or understanding Hartmann's German Fraktur text in which he explained that the name was subsequently used from a previous work where it had been classified in a different genus.

Sherborn listed 24 names as new, only 17 of which were correct (two are used today), five were attributed to incorrect works (three are used today).

Hartmann (1821): new names and names extracted by Sherborn

- 17 names were listed correctly (among these were *Acicula* Hartmann, 1821 and *Hydrobia* Hartmann, 1821)
- 5 species were missed: 2 new species were overlooked, 1 name with figure was marked as a nomen nudum, 2 names were incorrectly listed as available for previous works where they had been nomina nuda. Two of the five names of this fraction are used today: *Ciliella ciliata* (Hartmann, 1821), *Discus ruderatus* (Hartmann, 1821) (Fig. 3).
- 41 varieties were missed. Sherborn systematically ignored all names established for varieties. Two of these 41 names are used today for species: *Radix ampla* (Hartmann, 1821), *Trochulus clandestinus* (Hartmann, 1821).


seiner Voyage tab. 31. f. 5. a. b. bekannt gemacht
hat; welche aber durchaus nicht mit unserer H. ele-
gans verwandt ist, sondern im System mehr bey
H. pyramidata zu stehen käme.

55. Helix oculus capri. Drap. tab. VII. f. 38. 39.
Helix aligra.
In Frankreich und Italien, auch bey Wien, wo
Herr Doctor Klees auf dem Kahlengebirg eine klei-
nere Abart fand.

56. — rotunda. Drap. tab. VIII. f. 4—7.
In der Schweiz, Deutschland und Frankreich; Herr
von Charpentier fand bey Bex Blendlinge und
erhielt dergleichen auch aus England.

57. — ruderata. s. meine Tab. II. f. 11.
Der vorigen ähnlich, aber aufgeblasener, rauher,
weniger Umgänge. Sie findet sich in der Schweiz.

58. — pygmea. Drap. tab. VIII. f. 8—10.
In der Schweiz und Frankreich.

59. — rupestris. Drap. tab. VII. f. 7—8.
In Frankreich, und in den Alpen der Schweiz zu
Millionen an den Felsblöcken.

Figure 3. Example of a new name established in a chaotically arranged early zoological work. Detail of Hartmann (1821: p. 231), with the original description of *Helix ruderata* Hartmann, 1821 (currently *Discus ruderatus*). For the non-insider it is very difficult to see that a new name was established here. It is necessary to understand precisely the content of the German text: "s. meine Tab. II. f. 11. Der vorigen ähnlich, aber aufgeblasener, rauher, weniger Umgänge" (= see my plate II, figure 11. Similar to the previous one, but more inflated, rougher, fewer whorls).

- 1 species had an incorrectly cited genus (*Clausilia parvula* was incorrectly listed in *Auricella*). A specific name listed with an incorrect genus cannot be recognised in the list.
- 1 nomen nudum was listed as new (*Planorbis dubius*). This name had no description.
- 5 names were established in previous sources and subsequently used by Hartmann, but were listed as new for Hartmann by Sherborn. In three cases Hartmann gave bibliographical references which were not verified by Sherborn, in one case a German explanation was not understood. Three names are used today: *Acicula lineata* (Draparnaud, 1801), *Oligolimax annularis* (Studer, 1820), *Pomatias* Studer, 1789.

Incomplete research (not checking all cited sources, not verifying the presence of a true description in a foreign language) was probably a strategy to avoid spending too much time on difficult works with low numbers of new names. We know from Sherborn's correspondence he felt a sense of continuous pressure and responsibility for his huge task (Evenhuis 2016, Shindler 2016) and he may have chosen to adopt a strategy of avoiding situations of greatly diminishing returns.

In the AnimalBase team we did not skip such difficult works, as our objective was to test the existing historical data. This had the effect that some highly-skilled team members had to spend a significant amount of time with extremely low output in terms of numbers of new taxonomic names extracted per time unit.

We were often surprised to observe that Sherborn did not always verify bibliographical references given in the original sources, to see if a name was established in a previous work and thus, not new in the reference in front of him. Similar issues are faced by authors of applications to the ICZN Commission, and are often caught by careful work by the editorial work on the *Bulletin of Zoological Nomenclature* (E. Michel 2013). Checking sources was perhaps easier for the AnimalBase team than it was for Sherborn, when a digitised book was often only one mouse-click away, and more time consuming for Sherborn who had to go to the shelf, or request loans from far away libraries, to get the printed book.

The danger zones for high error rates when working with *Index Animalium*

How dangerous is it to rely on Sherborn's work, and where do we have to pay special attention? This question cannot be answered easily or generally. Taxonomists work with Sherborn's *Index Animalium* under various different objectives. The main objective of our team was to extract all available names from a published work. Doing this we had to be aware that Sherborn's list was very slightly incomplete. We also knew that if no Latin diagnoses were given and works were in German, Swedish or French, we had to pay more attention. But this was probably an unusual form of working with the *Index*.

More commonly, taxonomists will look in *Index Animalium* for the original source of a name that was established before 1851. Here they will be confronted with the danger of not finding the name, or finding the name recorded from an incorrect source.

In 2–3% of the cases the name will not be found, and in some taxonomic groups more. Sherborn missed various important works with different impact in various groups. In those animal groups where many varieties were named before 1851 the danger of missing names is generally higher. As a general rule, Sherborn's record provides higher reliability for insects than for molluscs, but also within insect taxa there seem to be differences. It is always necessary to ask colleagues working in the same area and with some experience in bibliographic work, for their judgement on the reliability on Sherborn's *Index*.

Taxonomists working on fossil species will also know the extent to which they can rely on Sherborn, as this can vary significantly.

Finding the name recorded with an incorrect source will happen more often. This also differs by taxonomic group, for the same reason just explained above. If important standard works were missed, the names were attributed to incorrect subsequent sources. If non-binominal works were cited as original sources, such as Gronovius (1763, 1781), Meuschen (1778) or Martyn (1784), it is highly probable that the source will be incorrect. This is easy to see, however in the next step without Sherborn's help it is very difficult to find the correct source in such a case.

Others might consult Sherborn to determine the orthography of a name's original spelling. These researchers may run the danger of finding the name incorrectly spelled in the *Index*. This danger is generally very low, however in disputed names, for example in unusual Latin words, incorrect information may have accumulated in *Index Animalium*. We came to the general conclusion that if the spelling of a name has subsequently been the subject of debate, it is always necessary to consult the original source and not rely on the *Index* alone.

Again others might consult the *Index* to see if a name was actually made available in a certain work or not. They will run the risk of finding an unavailable name being incorrectly marked as available in the *Index*, or an available name recorded as unavailable. Insects such as aphids or moths with specific host plants are particularly problematic in this regard, as they were often presented without description in the original sources. In Sherborn's *Index*, no indication is visible that these names were not made available in those sources. Generations of scientists have relied on Sherborn's information and it comes often as a great surprise to see that a name of a well-known and frequent species established more than 200 years ago cannot be attributed to a certain work because no description was given. The moths in [Denis and Schiffermüller] (1775) are the best example. Hundreds of names of this extremely important work were incorrectly regarded as available for many decades. For many names their unavailability from that work and their correct original sources are yet to be discovered. Interestingly, also some Linnean names presented in his 1758 work with host plants have this problem.

Taxonomists should probably not consult Sherborn to verify the correct authorship for a name. For this task it is necessary to consult the original source and to apply the modern rules, especially if various different authorships were cited in various literature sources. As a general rule it could perhaps be said, if two or more authorships were given for a name in various different recent literature sources, the authorship cited by Sherborn would quite probably be the incorrect one.

Final conclusions

The main conclusion we can derive for future projects aimed to extract complete lists of names from original sources is that if the size of the project passes a certain limit, then doing this without errors is effectively impossible. We would not recommend attempting production of a 'perfect list'. This applies particularly to the idea of establishing Lists of Available Names (LANs, Code Art. 79). In lists of 2000 or 3000 names it might be possible with reasonable costs in terms of time and energy to have a list free of error in the spellings and the original sources of the names. If the number of names passes this barrier, the time needed to get the record complete will increase rapidly towards infinity.

Manual compilation of large nomenclators will always face an error rate of at least 2-4%, no matter how thoroughly they are researched and how many people contribute. If more people contribute, like the AnimalBase team, the increase in coverage is counterbalanced by the difficulty in maintaining a common standard, which potentially leads to higher error rates. If only one person contributes, like in Sherborn's case, nobody is there for a control, but the experience shows that the record can be more reliable. If more data are to be recorded besides name and original source, such as in the ZooBank project, then the error rates per entry would increase accordingly. If any such projects are aimed to provide official data resources they will need regulations what to do in those cases when the data record contains errors.

Sherborn was confronted with many problems that we also had in our own work. This included the difficulty in maintaining a common standard over time. We came to the conclusion that anyone who intends to repeat Sherborn's job will inevitably be fascinated and awed by what he achieved, and particularly by his low non-systematic error rates.

Acknowledgements

We would like to thank three anonymous reviewers who helped improving our manuscript. The AnimalBase project was funded by the German Science Foundation (DFG) (2003–2005, 2008–2011). The contribution to the Sherborn Meeting 28 Oct 2011 in London was supported by the Biodiversity Heritage Library for Europe.

References

Blanchard R, von Maehrenthal F, Stiles CW (1905) Règles internationales de la nomenclature zoologique adoptées par les Congrès International de Zoologie; International Rules of Zoological Nomenclature; Internationale Regeln der Zoologischen Nomenklatur. FR de Rudeval, Paris, 57 pp.

Denis JNCM, Schiffermüller I (1775) Ankündung eines systematischen Werkes von den Schmetterlingen der Wienergegend, herausgegeben von einigen Lehrern am k. k. Theresianum. Wien, 322 pp. http://resolver.sub.uni-goettingen.de/purl?PPN574458115

Evenhuis NL (2016) Charles Davies Sherborn and the "Indexer's Club". In: Michel E (Ed.) Anchoring Biodiversity Information: From Sherborn to the 21st century and beyond. ZooKeys 550: 13–32. doi: 10.3897/zookeys.550.9697

Forskål P (1775) Descriptiones animalium avium, amphibiorum, piscium, insectorum, vermium; quæ in itinere orientali observavit Petrus Forskål. Post mortem auctoris edidit Carsten Niebuhr. Adjuncta est materia medica Kahirina atque tabula maris rubri geographica. Möller, Hauniæ, 164 pp. http://www.biodiversitylibrary.org/item/18564

Forster JR (1781) Indische Zoologie oder systematische Beschreibungen seltener und unbekannter Thiere aus Indien, mit 15 illuminirten Kupfertafeln erläutert. Nebst einer kurzen vorläufigen Abhandlung über den Umfang von Indien und die Beschaffenheit des Klima, des Bodens und des Meeres daselbst, und einem Anhange, darin ein kurzes Verzeichniß der Thiere in Indien mitgetheilt wird. Zoologia Indica selecta tabulis XV aeneis illustrata. Praemittitur de finibus et indole aeris, soli, marisque Indici brevis lucubratio. Seqvitur ad calcem brevis enumeratio animalium Indiae. Gebauer, Halle, 42 pp. http://resolver.sub.uni-goettingen.de/purl?PPN600894029

Geoffroy EL (1762) Histoire abrégée des insectes qui se trouvent aux environs de Paris; dans laquelle ces animaux sont rangés suivant un ordre méthodique. Durand, Paris, 523 pp. [Tome premier], 690 pp. [Tome second]. http://resolver.sub.uni-goettingen.de/purl?PPN367362899

Gronovius LT (1763) Zoophylacii Gronoviani fasciculus primus exhibens animalia quadrupeda, amphibia atque pisces, quae in museo suo adservat, rite examinavit, systematice disposuit, descripsit, atque iconibus illustravit. Lugduni Batavorum, 136 pp. http://resolver.sub.uni-goettingen.de/purl?PPN50020554X

Gronovius LT (1781) Zoophylacii Gronoviani fasciculus tertius exhibens vermes, mollusca, testacea, et zoophyta, quae in museo suo adservavit, examini subjecit, systematice disposuit atque descripsit. Lugduni Batavorum, Haak, 241–380. http://resolver.sub.uni-goettingen.de/purl?PPN50020554X

Hartmann W (1821) System der Erd- und Flußschnecken der Schweiz. Mit vergleichender Aufzählung aller auch in den benachbarten Ländern, Deutschland, Frankreich und Italien sich vorfindenden Arten. Neue Alpina 1: 194–268.

Hasselquist F (1762) Reise nach Palästina in den Jahren von 1749 bis 1752. Auf Befehl Ihro Majestät der Königinn von Schweden herausgegeben von Carl Linnäus. Aus dem Schwedischen. Koppe, Rostock, 606 pp. http://resolver.sub.uni-goettingen.de/purl?PPN483918202

ICZN (1999) International Commission on Zoological Nomenclature. Fourth Edition. The International Trust for Zoological Nomenclature, London.

Linnæus C (1758) Systema naturæ per regna tria naturæ, secundum classes, ordines, genera, species, cum characteribus, differentiis, synonymis, locis (Tomus I) – Editio decima, reformata. Salvius, Holmiæ, 824 pp. http://resolver.sub.uni-goettingen.de/purl?PPN362053006

Martyn T (1784) The universal conchologist, exhibiting the figure of every known shell, accurately drawn, and painted after nature: with a new systematic arrangement by the author. - Le conchyliologiste universel, montrant la figure de chaque coquille, aujourd'hui connu. Soigneusement dessinée et peinté d'apres nature. Le tout arrangé selon le systeme de l'auteur. London, 39 pp. http://resolver.sub.uni-goettingen.de/purl?PPN583629644

Meuschen FC (1778) Musevm Gronovianvm. Sive index rervm natvralivm tam mammalivm amphibiorvm piscivm insectorvm conchyliorvm zoophytorvm plantarvm et mineralivm exqvisitissimorvm qvam arte factarvm nonnvllarvm. Inter quae eminet herbarivs siccvs plantarvm a Tovrnefortio Claitonio Linnaeo aliisqve botanicis collectarvm. Qvae omnia mvlta cvra et magnis svmptibvs sibi comparavit vir amplissimvs & celeberrimvs Lavr. Theod. Gronovivs. Haak, Meerburg; Lvgdvni Batavorvm, 251 pp. http://www.mdz-nbn-resolving.de/urn/resolver.pl?urn=urn:nbn:de:bvb:12-bsb10076155-2

Shindler K (2016) A magpie with a card-index mind – Charles Davies Sherborn 1861–1942. In: Michel E (Ed.) Anchoring Biodiversity Information: From Sherborn to the 21st century and beyond. ZooKeys 550: 33–56. doi: 10.3897/zookeys.550.9975

Swainson W (1840) A treatise on malacology, or shells and shell-fish. Longman, London, 419 pp. http://www.biodiversitylibrary.org/item/67905

Tunstall M (1771) Ornithologia Britannica seu avium omnium Britannicarum tam terrestrium, quam aquaticarum catalogus, sermone Latino, Anglico & Gallico redditus: cum subjicitur appendix, aves alienigenas, in Angliam raro advenientes, complectens. London, 4 pp.

Zimmermann EAW (1777) Specimen zoologiae geographicae, quadrupedum domicilia et migrationes sistens. Dedit, tabulamque mundi zoographicam adjunxit. Haak, Lugduni Batavorum, 685 pp. http://resolver.sub.uni-goettingen.de/purl?PPN572532466

ZooKeys 550: 189–206 (2016)
doi: 10.3897/zookeys.550.9702
http://zookeys.pensoft.net

RESEARCH ARTICLE

A peer-reviewed open-access journal
ZooKeys
Launched to accelerate biodiversity research

Digitising legacy zoological taxonomic literature: Processes, products and using the output

Christopher H. C. Lyal[1]

1 *Life Sciences Department, The Natural History Museum, Cromwell Road, London SW7 5BD, UK*

Corresponding author: *Christopher H. C. Lyal* (c.lyal@nhm.ac.uk)

Academic editor: *E. Michel* | Received 26 March 2015 | Accepted 26 March 2015 | Published 7 January 2016

http://zoobank.org/3A5F049A-E528-43D1-B419-534F4A76DC8E

Citation: Lyal CHC (2016) Digitising legacy zoological taxonomic literature: Processes, products and using the output. In: Michel E (Ed.) Anchoring Biodiversity Information: From Sherborn to the 21st century and beyond. ZooKeys 550: 189–206. doi: 10.3897/zookeys.550.9702

Abstract

By digitising legacy taxonomic literature using XML mark-up the contents become accessible to other taxonomic and nomenclatural information systems. Appropriate schemas need to be interoperable with other sectorial schemas, atomise to appropriate content elements and carry appropriate metadata to, for example, enable algorithmic assessment of availability of a name under the Code. Legacy (and new) literature delivered in this fashion will become part of a global taxonomic resource from which users can extract tailored content to meet their particular needs, be they nomenclatural, taxonomic, faunistic or other.

To date, most digitisation of taxonomic literature has led to a more or less simple digital copy of a paper original – the output of the many efforts has effectively been an electronic copy of a traditional library. While this has increased accessibility of publications through internet access, the means by which many scientific papers are indexed and located is much the same as with traditional libraries. OCR and born-digital papers allow use of web search engines to locate instances of taxon names and other terms, but OCR efficiency in recognising taxonomic names is still relatively poor, people's ability to use search engines effectively is mixed, and many papers cannot be searched directly. Instead of building digital analogues of traditional publications, we should consider what properties we require of future taxonomic information access. Ideally the content of each new digital publication should be accessible in the context of all previous published data, and the user able to retrieve nomenclatural, taxonomic and other data / information in the form required without having to scan all of the original papers and extract target content manually. This opens the door to dynamic linking of new content with extant systems: automatic population and updating of taxonomic catalogues, ZooBank and faunal lists, all descriptions of a taxon and its children instantly accessible with a single search, comparison of classifications used in different publications, and so on. A means to do this is through marking up content into XML, and the more atomised the mark-up the greater the possibilities for data retrieval and integration. Mark-up requires XML that accommodates the required content elements and is interoperable with other XML schemas, and there are now several written to do this, particularly TaxPub, taxonX and taXMLit, the last of these being the most atomised. We now need to automate this process as far as possible. Manual and automatic data and information retrieval is demonstrated by projects such as INOTAXA and Plazi. As we move to creating and using taxonomic products through the power of the internet, we need to ensure the output, while satisfying in its production the requirements of the Code, is fit for purpose in the future.

Keywords

XML, taxonomy, digitisation, nomenclature, legacy literature, zoology, botany

Introduction

The primary source of taxonomic and nomenclatural information is taxonomic literature. Despite the growing number of databases the original published source is authoritative and provides datable interpretable content linked axiomatically to known taxon concepts. A challenge for today is how to extract the content of the last few hundred years' publications and make it accessible in the same way that at least some novel taxonomic and nomenclatural statements are accessible (see also Pilsk et al. 2015). One means of addressing this is establishing how the taxonomic and nomenclatural acts proposed in legacy literature might be most effectively extracted and linked to (or, even more effectively, incorporated into) modern databases. The accelerating digitization of legacy literature provides an opportunity to optimise access to taxonomic and nomenclatural information. This paper focusses on zoological taxonomy, but the same arguments (with some different actors, such as MycoBank, IPNI, Index Fungorum and the appropriate Codes of Nomenclature) apply to other taxonomic domains.

The availability of both legacy literature and new publications as digital items has many advantages over solely paper format, including simple portability – it is quite possible to carry the entire corpus of publications on a taxonomic group on a single hard drive. It also provides easy availability on the internet (Pyle 2016); digitisation by large-scale initiatives such as Biodiversity Heritage Library (BHL) (http://www.biodiversitylibrary.org/), AnimalBase (http://www.animalbase.org/), the Bibliography of New Zealand Terrestrial Invertebrates-Online (BUGZ) (http://www.bugz.org.nz/WebForms/about.aspx) and Google has led to a massive increase in accessibility of otherwise difficult-to-obtain literature around the world. Ideally the process will accelerate and become increasingly efficient, with a long-term goal of creating an online digital version of all relevant literature. However, to date most digitisation of taxonomic literature has led to a more or less simple digital copy of a paper original – the output of the many efforts has effectively been an electronic copy of a traditional library. While this has increased accessibility of publications through internet access, the means by which many scientific papers are indexed and located is much the same as with traditional libraries.

While digitisation of literature is on the face of it a simple concept, it is driven by multiple objectives. These include improving security of paper copies by providing digital surrogates, improving and increasing access to publications including rare and old items, improving searchability of content by providing machine-readable versions, pursuing institutional or individual open-access policies, facilitating sharing or selling of reprints, commercial benefit, and providing a flexible publication medium. The objective of the individual or initiative responsible for the digitisation may dictate the methods used, any mark-up system employed, and the metadata provided with the digital object. This in turn may determine the uses to which the digital object may be put; for example, King and Morse (2013) show that XML mark-up employed for one use may hinder use in another area of work, looking particularly at a contrast between taxonomists who use mark-up to exploit the documents' contents (that are relevant to them) and computer scientists who wish to explore the documents for multiple uses. In one example they noted that the scope of a marked-up text (and the editorial decisions taken by the digitisers) led to only species from Central America being tagged by the schema elements for taxonomic names; species names in the resource from other geographic regions were not tagged, and therefore undiscoverable by text mining tools created for discovery of names.

The potential of using XML-markup is being exploited increasingly in new publications, throughout the workflow of manuscript preparation to publication (Blagoderov et al. 2010). XML-markup enables content to be both displayed effectively and repurposed through delivery to relevant datasets (Penev et al. 2010). The potential for exploring that potential for legacy literature, only available as printed or OCRed text, displaying it in a way that allows users to search on content types rather than simply key terms, and access and download relevant data rather than have to manually extract it from blocks of text, has also been demonstrated by the Plazi (http://plazi.org/) and INOTAXA (http://www.inotaxa.org/jsp/index.jsp) projects (Agosti et al. 2007; Weitzman and Lyal 2006, 2015; Lyal and Weitzman 2008).

In addition to the current approaches to digitisation, and to enable us to build on the huge amount of work already done, we should consider what properties are required of future taxonomic information access systems and then develop them. Ideally the content of each new digital publication should be accessible in the context of all previous published data, and the user able to retrieve nomenclatural, taxonomic and other data or information in the form required without having to review all of the original papers and extract target content manually. Rather than building forward from the past we should be building back from the future, and bring together the many efforts across digital taxonomic information to support this. Focussing as far as possible on taxonomic and nomenclatural content, this paper will consider what such a system should deliver, and what its properties might be.

Building the System

Requirements

There are two general workflows and sets of components to be considered, although these are strongly interlinked. Firstly, there is the workflow necessary to discover relevant content within the digitised literature, and populate a taxonomic information retrieval system with extraction from or links to the legacy literature and other relevant resources. Secondly, there is the workflow that will operate for a user seeking information among the available databases and resources on the internet. To a large extent the requirements of the latter system will dictate the operation of the first. A third workflow should be mentioned, the production of new literature with an appropriate XML mark-up so that the contents are directly accessible together with the contents of legacy literature.

Broad requirements of the first system include:

- Generation of accurate re-purposable information from the legacy literature source;
- Generation of this information only once, not multiple times;
- Automatic population and updating of taxonomic catalogues, ZooBank, species lists, Encyclopedia of Life (EoL), etc. from digitized literature;

And of the second include:

- Accessibility of all relevant information from any (connected) source;
- Accessibility of new digital content in the context of all previous published data.

Questions that users of the system should be able to find answers to include:

- What is the correct name for an entity sought?
- What is the nomenclatural history of that name? Is it available under the relevant Code?
- What descriptions of the taxon are there?

While these are by no means all the questions that one might ask, the scope of this paper is largely limited to taxonomic and nomenclatural content. The arguments presented would, however, apply equally to components of non-taxonomic publications (physiology, immunology, ecology, biology, distribution, etc.).

In order to deliver a system that can respond to these questions, the components of that system and how they might be linked together must first be understood. Having developed the concept of the system it must be built and populated, and the content must be updated continuously and kept available to users so that current needs are met as far as possible.

The requirements of a system for this area in terms of detailed content retrievability were determined in part by a user-needs survey (Parr and Lyal 2007).

Auto-population of sectorial systems

One of the requirements listed above is the facility to populate databases and other systems with appropriate content from the digitised literature. This needs a little further discussion, since it dictates how digitisation should be performed and elements within the digital content should be tagged. Examples of this already exist for digitised legacy literature and new publications. Both INOTAXA and Plazi expose their content to EoL, and Plazi further supplies data to ZooBank. The journal *ZooKeys* supplies data to GBIF, ZooBank, EoL and Species-ID, and the sister publication for open access biodiversity data, *Biodiversity Data Journal*, implements for the first time a pre-submission mark-up of various types of highly atomised biodiversity data. This allows for automated export of Darwin Core Archives of treatments, occurrences etc. and their direct indexing by GBIF and EOL (http://biodiversitydatajournal.com/about#Globallyuniqueinnovations). However, rather than simply deliver names and the associated citations to ZooBank, for example, much more is possible with appropriate data and metadata capture from the publication. Requirements for the availability of new names under the ICZN are set out in the Code itself (an abbreviated list is presented in Table 1). If these requirements could be turned into a set of standard queries, and the schema built to enable the appropriate data and metadata from the publication to be delivered, it

Table 1. Some of the criteria for availability of names under the ICZN, and potential to extract information to determine availability. Note that some of these are required or appropriate only for the current amended Code (http://www.nhm.ac.uk/hosted-sites/iczn/code/), and the date of publication of the original text will dictate what criteria are applicable.

Criterion for availability	Requirement
Publication is obtainable in numerous identical copies	metadata
Publication: If non-paper, produced by a method that assures widely accessible electronic copies with fixed content and layout, and registered in ZooBank.	metadata
Publication not excluded by Article 9	metadata
The name is published using the Latin Alphabet	metadata
For species-group names, name agrees in gender with the genus name	markup + algorithm
For family-group names, name has a permitted ending	markup + list
For family-group names, name has suffix appropriate for rank given	markup + list + algorithm
For family-group names, name is based on the genus name stated	markup + algorithm
Name not already registered	markup + ZooBank search
Name contains more than one letter	markup + algorithm
Genus in which new species-group name is placed (if applicable)	markup
Name not published as a synonym but as a valid name	markup
Valid genus name on which new family-group name is based	markup
Type species of new genus-group name (including original combination, author and date)	markup
Description of taxon, or bibliographic reference to a description, is part of publication	markup + algorithm

would open the door to automatic assessment of nomenclatural availability through algorithms built into ZooBank, and populate ZooBank with not just the name but also the availability status with reasons for that status.

This clarifies a requirement that when legacy literature is taken from a non-digital to a digital state, not only should the original be viewable and searchable, but there should be the potential for subsets to be viewed, extracted on request, and be able to be analysed separately. Extending this beyond the elements related to taxonomy and nomenclature, the ability to extract components in order to be able to work with these independently opens the door to those components being repurposed in different applied contexts (e.g. taxa grouped according to where they have been recorded geographically or with what other taxa they were collected, by classification etc.). This also, of course, opens the door to relevant content being extracted (ideally automatically) to populate and enhance sector-critical systems (e.g. ZooBank, GBIF). This can be done through appropriate mark-up in XML.

A first step to this increased functionality is the user-needs assessment mentioned above. While general requirements are fairly straightforward to assess, there are differences in perception of priorities and significant needs, and consequently how the texts need to be marked-up to enable these to be met. One distinction that is already clear is whether mark-up should focus on taxon treatments (the policy of Plazi) or the complete content (the policy of INOTAXA). Given that relevant content can be found outside taxon treatments in the corpus of taxonomic literature, for a full system satisfying the requirements above it appears that full-text mark-up is required. This is discussed further in section 2.2.4 below. Analysis of different XML mark-up systems by Penev et al. (2011) revealed that the technology was only part of the issue, and that editorial policies of the groups managing the systems were also significant. These include the sections of the texts to be marked up, and also the level of interpretation of ambiguous content and how such interpretation is exposed to the user. Such editorial policies should also be harmonised, at least to a set of agreed practices or practice alternatives.

The functional components

The basic workflow of a system to acquire, put into a suitable format, retrieve and utilize legacy literature is given in Fig. 1. The system is self-reinforcing; the more populated the vocabularies and glossaries generated through extraction of the digitised elements become, the stronger they will be as a tool to assist the automated or semi-automated OCR and mark-up processes.

Many components of the system already exist. These are the initiatives, organisations and individuals digitising and hosting taxonomic literature, databases of names and other taxonomic and nomenclatural information, museum databases and the aggregators such as GBIF that make specimen data widely available and are developing interoperable systems, and people working on standards, particularly with Biodiversity

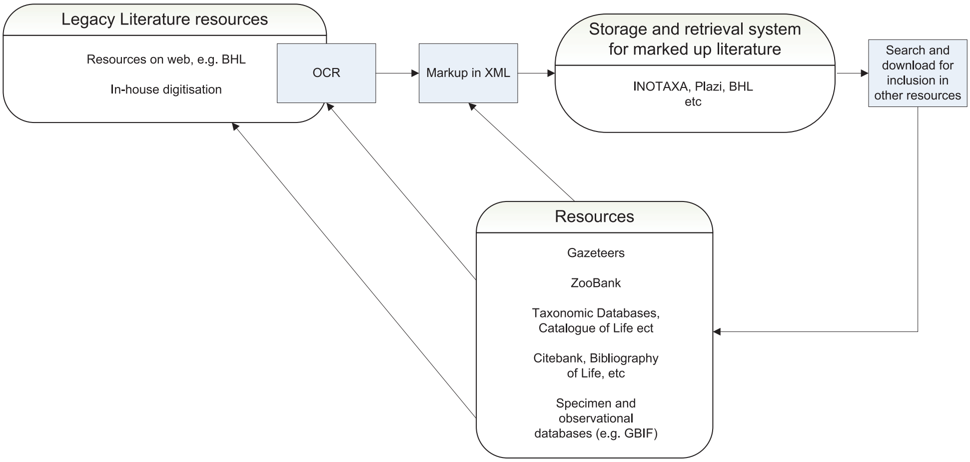


Figure 1. Outline workflow to acquire, put into a suitable format, retrieve and utilize legacy literature.

Standards (TDWG). Some of these are already communicating more or less effectively; others less so or not at all.

Patterson et al. (2010) set out a vision of access to the varied and massive amount of biological information on the internet through using taxonomic names to index content, and in particular the Global Names Architecture (http://www.globalnames.org/) (GNA) (see also Pyle 2016). A core of the GNA is the Global Names Usage Bank (GNUB) which is intended to index all published statements about life on Earth. They identify a first iteration of this as ZooBank (http://zoobank.org), the registry for names of animals developed by the International Commission on Zoological Nomenclature (Pyle and Michel 2008, 2009). They also refer to the Global Names Index (http://gni.globalnames.org/) (GNI), which is a simple index of all unique name strings (whether correctly or incorrectly spelled), with or without author attributions. Other actors (at a generic level) identified include Biodiversity Library's CiteBank (http://citebank.org/), containing citations for biodiversity publications, and taxonomic catalogues of species. Page (2013) uses taxon names to link a range of different taxonomic products and resources to texts at the article level, using a system 'BioNames' (http://bionames.org) that combines classifications from GBIF and GenBank, images from EoL, animal names from the Index of Organism Names (ION) and bibliographic data from sources including BHL and CrossRef. Many of these actors are built up by collaboration among many contributors; GBIF names are derived from nearly 15,000 datasets of observations and specimen records, plus the names from the Catalogue of Life (CoL) (http://www.catalogueoflife.org/); the latter is compiled from 143 contributing databases.

Records of taxonomic names can be considered in three classes: *name lists*, containing name strings, with no taxonomic or nomenclatural information (such as stored by the GNI); *nomenclators* such as ZooBank, which are lists of names with nomenclatural information but no taxonomic interpretation (i.e., will include all names as 'equal', not distinguishing between junior and senior synonyms, for example); and *catalogues and* *checklists*, which are lists of names that include taxonomic information such as synonymies, validity and current combinations. What is perhaps surprising is that there is no clear agreement of what data are collected where or, indeed, how they should be stored in a database. Taxonomic databases and catalogues are the least consistent in content. They will almost always contain the valid name as understood by the compiler, although whether this has been checked either against the most recent expert taxonomic treatment or consensus, or for compliance with the ICZN, is rarely stated. Such databases may contain valid names and combinations, obsolete combinations, original combination and orthography, obsolete ranks and systematic positions, junior synonyms, subgenera, misspellings, type taxa and full citations for each – or they may not. ZooBank includes only nomenclatural acts as covered by the Code (mostly 'original descriptions' of new scientific names for animals, but other acts may include emendations, lectotypifications, and other acts as governed by the ICZN Code), publications that contain nomenclatural acts, authors of those publications, and type specimens for scientific names of animals (the last is considered provisional and is not yet fully implemented in ZooBank). The data quality is generally good, reaching excellent where the entries have been checked. The names accessible through the GNI are just that – names, with no qualitative information. GBIF is similar to the GNI in part of its name provision; it serves names from taxonomic databases contributing to the Catalogue of Life but also names attached to specimens and observations that may or may not be reconciled to those for CoL.

Digitised legacy taxonomic literature

There is a large quantity of digitised content on the web. At the time of writing, the BHL made more than 45.4 million pages (from 156,114 volumes – 91,167 titles) freely available. Another resource, BUGZ, contains a literature database of more than 16,000 articles and full-text indexing of more than 200,000 pages. These are not the only sources. Therein, oddly, lies a problem. There is a very large number of both large and small-scale libraries, ranging from these major initiatives to individual researchers' web-sites. Consequently, it can be problematic for a human user to find a required article, and the user may need specialised understanding of the content and priorities of different libraries, including where to find particular collections. This issue is at least partially addressed by the ReFindIt tool (http://refindit.org/index.html), an integral part of Bibliography of Life (http://biblife.org/). ReFindIt searches across multiple bibliographies such as CrossRef, Mendeley, GNUB, Refbank, DataCite and DOIs. Some articles can be found directly through web browser searches, but this is not the case for the contents of some of the larger initiatives, which are only findable using the search systems of the initiatives themselves. These search systems can be restrictive, and may only allow searches using pre-determined criteria requiring information not necessarily available to all users (such as the full name string of the journal in only one form – a problem given the uncertainty often associated with journal names and their abbreviations – see Pilsk et al. 2015). Another issue with some initiatives is that digitisation has been carried out, for very good reason, at volume level, and consequently the metadata are at this and not the article level. This problem is compounded by the difficulty of automatically identifying the beginning and end of articles during digital capture or subsequent parsing. There is thus a mismatch between the metadata served and the search unit often employed by the user, who is often looking for articles or even taxon treatments.

There are various possible solutions to the problems outlined above. BHL, for example, has a browse function using a number of elements (titles, authors, subjects, origin of publication (using a map), publication year, language, contributors and collections). Another effective means would be to expose all digital content to direct search by internet browsers; Google books uses this method, as do many smaller collections. However, while this would improve current accessibility, it does not necessarily provide easy access within a workflow since, unless the search criteria are very rigorous, sorting through multiple irrelevant search results may mean the user abandons the process before finding a required resource. In addition, links between related papers (separate parts of the same paper published at different times, supplementary works, etc.) are generally not emplaced in the digital resources (Page 2016).

A second step is to facilitate machine-location of digitised texts. This might require the adoption of standard search routines that could operate as a web service across different sites. It entails adding metadata to digitised objects that would enable searches for units or entities that are of particular value to the expected users. Such units, for taxonomists, will include article-level publications (articles, chapters) and text elements within articles (for example taxon treatments). It will also require increased use of unique identifiers such as DOIs (Page 2016).

For inclusion in a workflow from digitisation to an interoperable system, large literature resources such as BHL have major advantages in addition to the very large content (and thus the easier location by users setting their own priorities). These include the standardisation of OCR that has taken place, the facility to download in one of a number of formats easily, and the persistence of the source.

A relevant although non-technical issue is copyright, which prevents open access to many publications. This has been discussed recently by Agosti and Egloff (2009) and Patterson et al. (2014). The former authors, considering taxonomic descriptions, opined that these are in the public domain and can be used for scientific research without restriction, whether or not contained in copyrighted publications. Patterson et al. (2014) focussed on scientific names in catalogues and checklists, and concluded, similarly, that these are not creative under at least US and EU copyright law and could be freely shared, although they noted the importance of providing credit for the original compilers.

Optical Character Recognition and accuracy of the digitised content

The majority of digitised legacy literature on the web is provided as an image of the page plus an underlying text generated by Optical Character Recognition (OCR), so that the software can find words or phrases within the text in response to a search. In some cases there is no OCR, while other sources provide only the OCR. While a scanned document may appear accurate as an image, an underlying (or alternative) OCRed text may not fully correspond. For this reason a search of a PDF may not retrieve all instances of a term, even when the term can be seen plainly on the screen, an issue compounded by non-Roman alphabets not properly OCRed in that language.

OCR techniques are not as effective for taxonomic literature as one might wish. There are particular problems associated with using OCR on digitised legacy literature, where character recognition may be compromised by factors including old fonts, foxing of the pages, transfer of print or image elements between pages, and translucency of paper so the image displays text from the obverse and converse of the page. However, while OCR can, subject to resolution of these problems, recognise most terms, the effectiveness of many programmes in recognising technical terms, taxonomic names and geographic place names is relatively poor. Consequently searches on such terms, even within specialist sites like BHL, do not reveal the full content. These problems have been addressed by, among others, Morse et al. (2009) and Akella et al. (2012); the systems developed need to be embedded effectively in functioning workflows, and currently for the majority of digitised and OCRed legacy literature the issues still remain.

To assist OCR software to recognise non-standard terms, specialised glossaries and vocabularies would be of value, although use of contextual information and machine learning are also both important. For example, a list of Parties for the names of authors, editors, collectors etc. (including the variations in the possible strings involved resulting from use of initials or full names, order of initials etc.) would facilitate recognition of more complex names. Other glossaries that would assist in text recognition are geographical place names, technical terms and of course taxonomic names; these, in addition to facilitating accurate OCR, would, through their distribution in the text, facilitate recognition of different text elements: a paragraph with many geographical place names is likely to be either a paragraph outlining distribution, or a list of specimen localities. A list of journal names and abbreviations would assist in good OCR of bibliographic elements (since references are often in a variety of languages, which can also pose a problem for OCR). Such a list would have to be open, since abbreviations of journal names are far from standard in zoology; it might form a part of a master bibliography as discussed below. As indicated above, such glossaries and vocabularies can be strengthened by the products of OCR and mark-up.

Even with the assistance of tools such as glossaries and technical vocabularies, text may contain terms of use for searches but which OCR software cannot simply recognise, such as abbreviations of author names and generic names, and these may need manual interpretation. This is also an issue for parsing text into XML, as discussed below.

Marking up the content

In order to transform OCRed content into a resource that can be searched effectively for classes of content (e.g. names, places, taxonomic acts etc.), it needs to be marked up into XML or other system that permits computer searching. This paper is not the place to address whether XML or RDF is the most appropriate system for development (Penev et al. 2011). Instead, it deals with the currently-developing systems, which are largely based on XML. The simple requirements are that mark-up accommodates the required content elements and is interoperable with other XML schemas relevant to the topic so that relevant content can be accessed wherever needed. Weitzman and Lyal (2007) attempted to map the various potentially interacting standards in play for taxonomic content, indicating where interoperability is required. Such interoperability might be achieved either by appropriate mapping between elements used in different schemas, or by re-use of schema components. Both positions have positive and negative aspects, and both approaches are in use. There are several schemas and DTDs written to manage taxonomic literature, particularly TaxPub, taxonX and taXMLit, the last of these being the most atomised. These have been compared by Penev et al. (2011).

Different schemas provide different levels of atomisation of the content. The greater the atomisation of the mark-up, the greater are the possibilities for data retrieval and integration, although this can carry a cost burden, since the greater the atomisation the greater the effort needed to parse content into the schema. However, this must be driven by the long-term goals and requirements. Some of the elements required have been identified in Table 1 – those providing sufficient data and metadata to automatically assess nomenclatural availability of a name. These elements would have to be separately tagged. Other elements are those which users would typically wish to search for, as outlined in the Introduction, and include both publications and subsets of publications. Within a publication the ‘top level’ is article-level (including chapters), these necessarily including the full publication metadata such as date of publication (both cited date and ‘true’ date, the second of which may require annotation) and other metadata that will enable a user (either human or automated) to assess its publication status and some of the availability criteria under the ICZN.

An important subset of publications at the article level, and that to which most attention has been paid to date, is the taxon treatment, since this is a text element of particular interest to taxonomist target users (Weitzman and Lyal 2004; Kirkup et al. 2005; Agosti et al. 2007; Sautter et al. 2007; Lyal and Weitzman 2008; Penev et al. 2011). For a system to have the most effective functionality, a search should retrieve all treatments of a taxon and, in some circumstances, its hierarchical children (subspecies, infrasubspecies, synonyms); this of course requires all of these treatments to have been digitised and marked up, and there is an issue of cost-effectiveness that will dictate to what extent treatments are digitised and to what extent they are captured in databases. For taxonomic and nomenclatural purposes the digitisation of a treatment should include in its identified sub-elements all components that will support identification of nomenclatural availability and validity under the ICZN.

Within articles the automated recognition of taxon treatments for mark-up is reliant on text and formatting recognition. Curry and Connor (2008) identified several standard text elements within treatments: name, synonyms, diagnosis, description, distribution, material examined and discussion. Unfortunately such text elements are neither uniform in formatting nor universal in their presence, so different publications may need different 'rules' to allow their recognition. The largest resource offering publications broken down into taxon treatments, Plazi, relies on manual intervention to recognise beginning and ends of treatments. Progress is required on use of OCR systems that retain text formatting, coupled with a natural language programming approach and perhaps use of catalogue resources to improve recognition of both article and treatment boundaries. Less work has been done on recognition of other large text elements such as bibliographies or checklists, but the priorities for these would be usefully developed and mechanisms to recognise them explored.

Other subsets of taxonomic publications that might be required might include taxon hierarchy (which may be an implicit construct from a publication rather than and explicit section within that publication), bibliography, diagnosis and description, biological associations, specimen data and character statements, taxon citation ([name] [author] [date] [nomenclatural/taxonomic act]), the original description citation ([name][author][date][reference]), subsequent taxonomic or nomenclatural changes citations, nomenclatural, taxonomic and other data or information.

Some of these subsets are more strongly structured data than cursive text, such as citations, bibliographic records and specimen data. In such cases the content, once marked up and made available through a suitable interface including both human searchability and web services, should be downloadable and repurposable, and consequently delivered or extractable in a common standardised format (e.g. Darwin Core for the specimen data). BHL already does this with the full content of publications, making the text available not only as HTML but also as downloadable PDF, OCRed text etc. INOTAXA makes specimen data extracted from text available as a spreadsheet and exposes taxon treatments and subsets to harvest by EoL, while Plazi extracts specimen data in Darwin Core and supplies it to GBIF. A concomitant to this is that content that is not required by the user (e.g. the text around specimen data) should be capable of exclusion from the retrieved content, and ideally the user should be able to retrieve such data without manually reviewing the whole paper and retrieving the required data by copying text items. This serves to define elements that need to be recognised in the schema.

A further requirement for successful mark-up is the population of implicit terms. For example, specimen data listed in papers is often incomplete for many of the specimens; the country may only be provided for the first of multiple localities within that country, or specimens with similar data to others may simply be listed "as previous, but dates …". If the specimen data are to be downloadable the country and all other data must be available for all the localities not just the first. Similarly, terms such as 'loc cit' and 'ibid' must be replaced as automatically as possible with the full citation.

Despite the work done in automating mark-up to date (Sautter et al. 2007a; Penev et al. 2011), the process is unlikely ever to be fully automated. Experience has shown that texts can hold many ambiguities that require interpretation, and for which automation is unlikely to be either cost-effective or successful, although semi-automation is certainly possible and has been used both in the Plazi and INOTAXA projects. Future work should focus on increasing the proportion of automation and reducing human input. This requires the process to accommodate the workflows of those most likely to be able to resolve ambiguities, who are likely to be taxonomic experts. That said, progress has been made with using crowd-sourcing to mark-up legacy documents (Thomer et al. 2012), and a system ultimately may make use of automation, crowd-sourcing and expert review and annotation.

A required resource for a mark-up workflow is a location where marked-up texts can be stored. This may have to incorporate several versions of a text, given that mark-up may not be completed for a text, and may even be done by more than one person. This repository may be the same initiative where the original document was sourced, or could be another place. However, ideally it should be fully accessible for users, including whatever search system is put in place for accessing content according to user needs. This last point implies a single gateway for searches, which could either search locally or, more effectively, across sites (where marked up text in appropriate formats is also exposed to searching). The search system itself will need to be in line with the requirements outlined above and capable of refinement as other user demands develop. Two systems have been developed in recent years that satisfy at least some of the needs, the Plazi and INOTAXA systems. Both work with a single repository of documents marked up within their projects to particular standards.

The global bibliography

Locating texts, whether marked up or not, ideally will involve indexing in some manner, and this may be a function of a global bibliography. This has already been mentioned above in the context of assisting mark-up through identifying beginning and end of articles, but also has a function here. The bibliography will have to include standards for citation of both library and taxonomic sectors, which do seem to differ, the latter including abbreviations and contractions understood (and used) by the sector but not appearing in library catalogues. Many journals, for example, have 'standard' abbreviated formats required by some publishers for use in bibliographies, and some taxonomist names are similarly abbreviated in a standard fashion [e.g. 'L.' for Carolus Linnaeus (= Linnaeus, C. = Linnaeus, = C. Linnaeus, = Carl Linnaeus, = Carl Nilsson Linnaeus = Carl von Linné = Carolus a Linné, etc.) (but not for his son)]. A vocabulary of abbreviations and alternate name strings for the same entity, accompanied by suitable unique identifiers, will enable some mark-up issues to be resolved automatically. There will still be a requirement for de-duplication and interpretation, although the global bibliography itself will assist in resolving them, especially if combined with the Global Names Architecture. For example, the microcitation 'Smith 1995' is undoubtedly ambiguous, but if combined with a taxonomic name ('Aus bus' Smith 1995') which can be found in a taxonomic database (using, perhaps, the Global Names Architecture) and then linked through that to a bibliographic reference, the microcitation (and author) can be resolved to known entities and the taxon name can be added to the metadata associated with the reference in the global bibliography.

As discussed above, automated recognition of articles within volumes is not straightforward; textual or formatting cues differ widely between publications, and many OCR conversion programmes (which allow the text to be searched) do not record original text formatting, thus stripping potential cues from the machine-readable content. Article recognition might be facilitated through access to bibliographic databases as discussed above (using web services); the title namestring might be used in conjunction with the page number to find the beginning of the article, and the final page number from the reference to find the end, for example. This requires the population of (open-access) bibliographic databases. There have been some attempts at compiling such databases, but this is a major task and needs automation where possible. Botanical literature between 1753 and 1940 has been captured by Stafleu and Cowan (1976–1988), Stafleu and Mennaga (1993–2008) and Donn and Nicholson (2008, 2008a), and is now online in database format (http://www.sil.si.edu/digitalcollections/tl-2) (Pilsk et al. 2015). This gives information not only on publications but also provides standard abbreviations of author names, which is of great value in disambiguating these. The BHL developed CiteBank (http://citebank.org/), and currently collects citation details when users choose to download a part of a digitised volume – users are permitted to download any set of pages they required from within a digitised volume; however, to do this they must enter the citation of the article being downloaded, which is then retained as metadata. Under the recent ViBRANT project the 'Bibliography of Life' (http://biblife.org/) was created in order to store, de-duplicate, parse, curate and share references, linked to a set of free services; this currently holds more than 215,000 references.

Interoperability and the information network

Implicit in all of the above is interoperability. How this is achieved is largely beyond the scope of this paper, but clearly all of the different databases and repositories in any system must be able to communicate and share data. There must also be the ability for users at any point to annotate records. This is discussed at some length in the recent Global Biodiversity Informatics Outlook (http://www.biodiversityinformatics.org/). One key component that has already been alluded to is that all of the participating actors must agree to some standard elements for their schemas and databases. Taxonomic name strings must be atomised in the same way, taxonomic acts, authors, citations, specimen data and so on must all be recorded in such a way that interoperability is possible. Without this the system cannot deliver benefits.

Sociological factors

Irrespective of technological advances, the major barriers to progress are likely to be sociological. Across the communities of scientists who might be expected to contribute to building a system there are very widely differing levels of understanding of what is needed, different skill levels in biodiversity informatics and other relevant technologies, and different levels of understandings of the current possibilities for data sharing and how these data may be used. At one end are individuals, usually highly experienced taxonomists, who still compile their information on stand-alone databases, spreadsheets or word-processed documents; at the other are computer-focussed builders of innovative bibliographic research tools, but who may have little engagement with taxonomic research. No tool is easily incorporated into a workflow unless it delivers what the users need more simply and effectively than it is delivered by familiar methods. This means that tools must be simple to use and not change too much or too rapidly. What is needed is not a set of beta-version products as generated by a succession of independent research projects but relatively stable well-documented production tools. This is not a straightforward requirement; we are all familiar with the truism that it is simpler to get a grant to develop a novel system than to obtain money to populate an existing one, and developments are happening so rapidly that it is difficult to harmonise them all. That said, BHL is an example of a component that has placed itself in multiple workflows effectively, and is growing by adding functionalities. ZooBank similarly has embedded itself in workflows, to the extent that for some journals and authors inclusion of ZooBank registration numbers is best practice, and initiatives such as NCBI (GenBank), BHL and GBIF are building it into their functionalities. It is also establishing cross-links to legacy literature – an important development which will support the building of the system discussed in this paper.

A further sociological factor is the independence of database compilers. Each database being compiled is created for a particular purpose and owned by one or more individuals. This has led to some taxa being covered by more than one taxonomic database (sometimes at global and regional levels), while others are not covered at all. In some cases the same information has been collected multiple times from primary and secondary literature. As noted above, each of these databases may be compiled using non-interoperable systems and without clarity or even consistency on the nomenclatural and taxonomic elements it contains. Assistance to use standardised systems that can connect to others to download and share information would gain more uniformity among database owners and facilitate their work. An analogous system is that of genealogical research, a very popular hobby in some parts of the world. There are many individual researchers but the tools they use are fairly standardised, many use a common (GEDCOM) format to exchange content, and many have easy links to online resources to find and download relevant data, and to upload content. Such programmes also exist for taxonomists, such as SpeciesFile (http://software.speciesfile.org/HomePage/Software/SoftwareHomePage.aspx) or Mantis (http://140.247.119.225/Mantis/index.htm) but as yet none can be considered as standard tools.

Summary

Taxonomic research and its nomenclatorial supporting structure are embracing the digital environment. However, to an extent we are still treating each resource – databases, checklists, taxonomic publications, faunas and floras etc. – as separate stand-alone items. Instead we should be mainstreaming the idea of bringing all of these together in a digital environment. One means to do that is through use of XML. Both legacy and new literature can be marked-up into dedicated schemas, and the more atomised the mark-up the greater the possibilities for data retrieval and integration. Mark-up requires XML that accommodates the required content elements and is interoperable with other XML schemas, and there are now several written to do this, particularly TaxPub, taxonX and taXMLit, the last of these being the most atomised. A need now is to automate this process as far as possible. With such mark-up and display in an appropriate platform, the door is opened to dynamic linking of new content with extant systems: automatic population and updating of taxonomic catalogues, ZooBank and faunal lists, all descriptions of a taxon and its children instantly accessible with a single search, comparison of classifications used in different publications, and so on. To move to such a model will require an agreement on vision and wider acceptance of both standards and desirable properties of digitised output.

Acknowledgments

The ideas expressed in this paper have been germinated in discussion with many colleagues and friends, and reflect those discussions. I am particularly grateful to Anna Weitzman, Dave Roberts, David Morse, Dauvit King, Adrian Hine, Andy Polaszek and Vince Smith for their insights and eagerness to discuss the movement of our discipline into the information age. I also thank Ellinor Michel, Neal Evenhuis, Lyubomir Penev and Thomas Pape for very helpful comments on the manuscript. Errors and misunderstandings remain my own.

References

Agosti D, Egloff W (2009) Taxonomic information exchange and copyright: the Plazi Approach. BMC Research Notes 2: 53. doi: 10.1186/1756-0500-2-53

Agosti D, Klingenberg C, Sautter G, Johnson N, Stephenson C, Catapano T (2007) Why not let the computer save you time by reading the taxonomic papers for you? Biológico (São Paulo) 69(suplemento 2): 545–548. http://hdl.handle.net/10199/15441

Akella L, Norton CN, Miller H (2012) NetiNeti: discovery of scientific names from text using machine learning methods. BMC Bioinformatics 13(1): 211 doi: 10.1186/1471-2105-13-211

Blagoderov V, Brake I, Georgiev T, Penev L, Roberts D, Rycroft S, Scott B, Agosti D, Catapano T, Smith VS (2010) Streamlining taxonomic publication: a working example with Scratchpads and ZooKeys. ZooKeys 50: 17–28. doi: 10.3897/zookeys.50.539

Curry GB, Connor RCH (2008) Automated extraction of data from text using an XML parser: An earth science example using fossil descriptions. Geosphere 4(1): 159–169. doi: 10.1130/GES00140.1

Dorr LJ, Nicolson DH (2008) A selective guide to botanical publications and collections with dates, commentaries and types. Regnum Vegetabile 149 (Supplement 7 [F – FRER]): 1–469.

Dorr LJ, Nicolson DH (2008a) A selective Guide to botanical publications and collections with dates, commentaries and types. Regnum Vegetabile 150 (Supplement 8 [FRES – G]): 1–558.

King D, Morse D (2013) Document mark-up for different users and purposes. 7th Metadata and Semantic Research Conference, 19–22 November 2013, Thessaloniki, Greece. Springer International Publishing, 355–360. doi: 10.1007/978-3-319-03437-9_34

Kirkup D, Malcolm P, Christian G, Paton A (2005) Towards a digital African Flora. Taxon 5(2): 457–466. doi: 10.2307/25065373

Lyal CHC, Weitzman L (2008) Releasing the content of taxonomic papers: solutions to access and data mining. Proceedings of the BNCOD Workshop "Biodiversity Informatics: challenges in modelling and managing biodiversity knowledge". http://biodiversity.cs.cf.ac.uk/bncod/LyalAndWeitzman.pdf

Morse D, Dil A, King D, Willis A, Roberts D, Lyal C (2009) Improving search in scanned documents: Looking for OCR mismatches. In: Bernardi R, Chambers S, Gottfried B (Eds) Proceedings of the workshop on advanced technologies for digital libraries (AT4DL 2009), 58–61. http://purl.org/bzup/publications/9788860460301

Page RD (2013) BioNames: linking taxonomy, texts, and trees. PeerJ 1: e190. doi: 10.7717/peerj.190

Page RDM (2016) Surfacing the deep data of taxonomy. In: Michel E (Ed.) Anchoring Biodiversity Information: From Sherborn to the 21st century and beyond. ZooKeys 550: 247–260. doi: 10.3897/zookeys.550.9293

Parr CS, Lyal CHC (2007) Use cases for online taxonomic literature from taxonomists, conservationists, and others. TDWG Annual Conference, Slovakia.

Patterson DJ, Cooper J, Kirk PM, Pyle RL, Remsen DP (2010) Names are the key to the big new biology. Trends in Ecology and Evolution 25(12): 686–691. doi: 10.1016/j.tree.2010.09.004

Patterson DJ, Egloff W, Agosti D, Eades D, Franz N, Hagedorn G, Rees JA, Remsen DP (2014) Scientific names of organisms: attribution, rights, and licensing. BMC Research Notes 7: 79. doi: 10.1186/1756-0500-7-79

Penev L, Agosti D, Georgiev T, Catapano T, Miller J, Blagoderov V, Roberts D, Smith VS, Brake I, Rycroft S, Scott B, Johnson NF, Morris RA, Sautter G, Chavan V, Robertson T, Remsen D, Stoev P, Parr C, Knapp S, Kress WJ, Thompson FC, Erwin T (2010) Semantic tagging of and semantic enhancements to systematics papers: ZooKeys working examples. ZooKeys 50: 1–16. doi: 10.3897/zookeys.50.538

Penev L, Lyal C, Weitzman A, Morse D, King D, Sautter G, Georgiev T, Morris R, Catapano T, Agosti D (2011) XML schemas and mark-up practices of taxonomic literature. ZooKeys 150: 89–116. doi: 10.3897/zookeys.150.2213

Pilsk SC, Kalfatovic MR, Richard JM (2016) Unlocking *Index Animalium*: From paper slips to bytes and bits. In: Michel E (Ed.) Anchoring Biodiversity Information: From Sherborn to the 21st century and beyond. ZooKeys 550: 153–171. doi: 10.3897/zookeys.550.9673s

Pyle RL (2016) Towards a Global Names Architecture: The future of indexing scientific names. In: Michel E (Ed.) Anchoring Biodiversity Information: From Sherborn to the 21st century and beyond. ZooKeys 550: 261–281. doi: 10.3897/zookeys.550.10009

Pyle R, Michel E (2008) Zoobank: Developing a nomenclatural tool for unifying 250 years of biological information. Zootaxa 1950: 39–50 http://www.mapress.com/zootaxa/2008/f/zt01950p050.pdf

Pyle R, Michel E (2009) Unifying nomenclature: ZooBank and Global Names Usage Bank. Bulletin of Zoological Nomenclature 66: 298.

Sautter G, Böhm K, Agosti D (2007) A Quantitative Comparison of XML Schemas for axonomic Publications. Biodiversity Informatics 4: 1–13. https://journals.ku.edu/index.php/jbi/article/view/36

Sautter G, Böhm K, Agosti D (2007a) Semi-automated XML markup of biosystematic legacy literature with the GoldenGATE editor. Pacific Symposium on Biocomputing 12: 391–402. http://psb.stanford.edu/psb-online/proceedings/psb07/sautter.pdf

Stafleu A, Cowan RS (1976–1988) Taxonomic Literature: a Selective Guide to Botanical Publications and Collections with Dates, Commentaries and Types (ref QK 97 S72 1976) (7 volumes). Bohn, Scheltema and Holkema, Utrecht. http://www.biodiversitylibrary.org/item/103414, http://www.sil.si.edu/digitalcollections/tl-2/history.cfm [in database format at]

Stafleu F, Mennega EA (1993) A selective guide to botanical publications and collections with dates, commentaries and types. Regnum Vegetabile 130 (Supplement 2 [BE – BO]): 1–464.

Stafleu F, Mennega EA (1995) A selective guide to botanical publications and collections with dates, commentaries and types. Regnum Vegetabile 133 (Supplement 3 [BR – CA]): 1–550.

Stafleu F, Mennega EA (1997) A selective guide to botanical publications and collections with dates, commentaries and types. Regnum Vegetabile 134 (Supplement 4 [CE – CZ]): 1–614.

Stafleu F, Mennega EA (1998) A selective guide to botanical publications and collections with dates, commentaries and types. Regnum Vegetabile 135 (Supplement 5 [DA – DI]): 1–432.

Stafleu F, Mennega EA (2000) A selective guide to botanical publications and collections with dates, commentaries and types. Regnum Vegetabile 137 (Supplement 6 [DO – E]): 1–518.

Thomer A, Vaidya G, Guralnick R, Bloom D, Russell L (2012) From documents to datasets: A MediaWikibased method of annotating and extracting species observations in century-old field notebooks. In: Blagoderov V, Smith VS (Eds) No specimen left behind: mass digitization of natural history collections. ZooKeys 209: 235–253. doi: 10.3897/zookeys.209.3247

Weitzman AL, Lyal CHC (2004) An XML schema for taxonomic literature – taXMLit. http://www.sil.si.edu/digitalcollections/bca/documentation/taXMLitv1-3Intro.pdf

Weitzman AL, Lyal CHC (2007) Models for Integrating TDWG Taxonomic Literature: Summary and Next Steps. TDWG Annual Conference, Slovakia.

Weitzman AL, Lyal CHC (2006) INOTAXA – INtegrated Open TAXonomic Access and the "Biologia Centrali-Americana". Proceedings of The Contributed Papers Sessions Biomedical And Life Sciences Division, SLA, 8 pp. http://dbiosla.org/events/past_sla_conference/Baltimore/inotaxa.pdf

Weitzman AL, Lyal CHC (2015) Extending Floras and Faunas to include Users' views. In: Watson MF, Lyal C, Pendry C (Eds) Descriptive Taxonomy: The Foundation of Biodiversity Research. Systematics Association Special Volume, CUP, Cambridge, 247–259.

ZooKeys 550: 207–223 (2016)
doi: 10.3897/zookeys.550.9546
http://zookeys.pensoft.net

RESEARCH ARTICLE

A peer-reviewed open-access journal
ZooKeys
Launched to accelerate biodiversity research

The use and limits of scientific names in biological informatics

David Remsen[1]

1 *Department of Marine Resources, Marine Biological Laboratory, 7 MBL Street, Woods Hole, MA 02543*

Corresponding author: *David Remsen* (dremsen@mbl.edu)

Academic editor: *E. Michel* | Received 6 March 2015 | Accepted 9 March 2015 | Published 7 January 2016

http://zoobank.org/A812E05B-5DC3-4BE0-9551-79A80A3B99C8

Citation: Remsen D (2016) The use and limits of scientific names in biological informatics. In: Michel E (Ed.) Anchoring Biodiversity Information: From Sherborn to the 21st century and beyond. ZooKeys 550: 207–223. doi: 10.3897/zookeys.550.9546

Abstract

Scientific names serve to label biodiversity information: information related to species. Names, and their underlying taxonomic definitions, however, are unstable and ambiguous. This negatively impacts the utility of names as identifiers and as effective indexing tools in biological informatics where names are commonly utilized for searching, retrieving and integrating information about species. Semiotics provides a general model for describing the relationship between taxon names and taxon concepts. It distinguishes syntactics, which governs relationships among names, from semantics, which represents the relations between those labels and the taxa to which they refer. In the semiotic context, changes in semantics (i.e., taxonomic circumscription) do not consistently result in a corresponding and reflective change in syntax. Further, when syntactic changes do occur, they may be in response to semantic changes or in response to syntactic rules. This lack of consistency in the cardinal relationship between names and taxa places limits on how scientific names may be used in biological informatics in initially anchoring, and in the subsequent retrieval and integration, of relevant biodiversity information. Precision and recall are two measures of relevance. In biological taxonomy, recall is negatively impacted by changes or ambiguity in syntax while precision is negatively impacted when there are changes or ambiguity in semantics. Because changes in syntax are not correlated with changes in semantics, scientific names may be used, singly or conflated into synonymous sets, to improve recall in pattern recognition or search and retrieval. Names cannot be used, however, to improve precision. This is because changes in syntax do not uniquely identify changes in circumscription.

These observations place limits on the utility of scientific names within biological informatics applications that rely on names as identifiers for taxa. Taxonomic systems and services used to organize and integrate information about taxa must accommodate the inherent semantic ambiguity of scientific names.

The capture and articulation of circumscription differences (i.e., multiple taxon concepts) within such systems must be accompanied with distinct concept identifiers that can be employed in association with, or in replacement of, traditional scientific names.

Keywords
Taxonomic name services, taxon concepts, identifiers, relevance, search and retrieval

Introduction

Scientific names are labels for taxa that are governed by formalized rules of nomenclature. These rules were introduced to establish clarity, stability, economy and uniqueness to the fragmented landscape of pre-Linnaean nomenclature (Mayr 1991). Sherborn's *Index Animalium* (IA) represents a monumental attempt to capture key data elements regarding the source and orthography of (nearly) all zoological names for species from the beginning of formalized Linnaean zoological nomenclature in 1758 through 1850. An index, in the sense of *Index Animalium*, is a list of terms linked to, or pointing to, a greater volume of values, data, information or knowledge that pertain to the term. *Index Animalium* links zoological names to their originating bibliographic citation. (Pilsk et al. 2016) It also links two separate records when a species described with one name was subsequently moved to a new genus and established a new binomial name. The primary function, of the more than 9,000 pages, however, is as an authoritative reference that provides the correct spelling of a name and pointer to its original description (CWR 1903, Alonso-Zarazaga et al. 2016). Much of the value and respect that IA has received is derived from the enormous amount of work required to compile and verify the names and associated publications. Biologists rely on this reference when they need to consult the original work (Evenhuis 2016).

The use and value of IA, however, extends beyond its referential value and, in an age of increasingly vast amounts of digitized biodiversity information being accessible online, serves as an immensely valuable resource in establishing order within a substantially larger index of biodiversity information: the entire corpus of recorded biodiversity knowledge.

Throughout the past 250 years, nearly all information about taxonomic groups such as species has been linked through a name, nearly always a scientific name. (Grimaldi and Engel 2005). Scientific names are not the sole means to label species information. Informal and provisional names also play supporting roles (Murray and Stackebrandt 1995) but only when their use can be unambiguously linked to a species through a scientific name. Practically speaking, scientific names form the basis for referring to species and they label biodiversity information across the entire spectrum of biodiversity knowledge (Thompson and Pape 2016).

Names label voucher specimens in natural history museums, for instance, and are used to identify biological observations at all scales, from molecules to ecosystems, providing the key biological context to associated metacontent such as the observation locality and date Figure 1. Scientific names are used in all manner of publications and

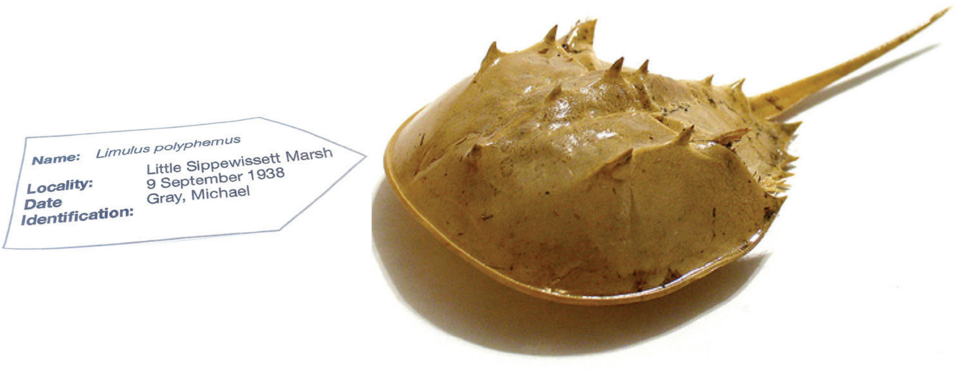


Figure 1. Scientific names label information about species.

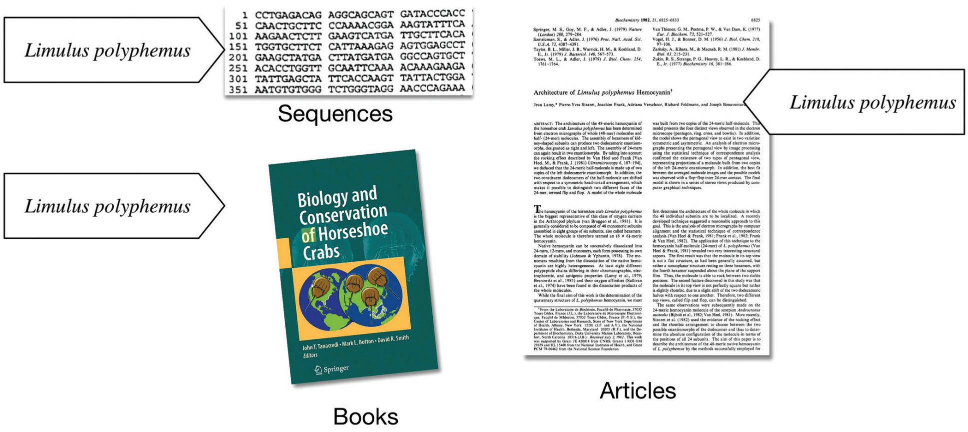


Figure 2. All information of a species is linked by a name.

communications be they scientific, agricultural, commercial, medical, legislative or social. Increasingly, these communications are taking place online and in digital environments and legacy information is being retrospectively digitized and also placed in online data stores (Patterson et al. 2010). Without a name associated with an information or data object, the taxonomic link is effectively lost.

Discussion

Given the ubiquitous linkage between biodiversity information and scientific names, there must exist an enormous and virtual super-index of names tied to the world's species information. Such an index, assembled and presented within a Sherborn-like data store, would, in principle, link to all, or nearly all, information related to all described species. This implies a far more important and central role for names as mediators to biodiversity information. As more and more retrospective and prospective information is placed in online data stores, such an index is becoming increasingly realistic (Pyle 2016).

Indexing and search engines like Google and Yahoo generate billions of dollars in revenue by processing countless electronic data stores and producing searchable indexes (Varian 2007). An index of names, such as provided through *Index Animalium*, combined with similar technologies could, in principal, provide access to whatever online information is associated with the names in the index. Broaden the list through the consolidation of similar indexes within the zoological, microbial and botanical domains, and it's not inconceivable that a comprehensive list of names could be assembled. Such a collection would provide the means to discover and access the complete wealth of recorded biodiversity knowledge (Patterson et al. 2010). The legacy of Linneaus and Sherborn appear to have provided the framework for the systematic organization and delivery of biodiversity knowledge in the digital age (Patterson et al. 2006).

There are limits to this utility however, and these limits are inherent within biological nomenclature and its relationship to the taxa they label. *Index Animalium* and other similar compilations represent a list of names, not lists of species. This distinction has ramifications that place limits on helping us utilize their ubiquity in labeling biodiversity information. Semiotics refers to the study of how we use signs or symbols, such as names, to confer meaning to objects in the real world and provides a broader framework to this understanding. Semiotics is divided into several sub-domains that include:

- *Semantics*, which refers to the relationship between signs and the things to which they refer; their meaning.
- *Syntactics*, which refers to the relationships among signs or symbols within formal structures.

These two terms have analogs in taxonomy. Nomenclature, particularly formalized scientific nomenclature governs much of the syntax domain while semantics is the realm of taxonomy, which links names with taxon definitions or *circumscriptions* (Dubois 2005, Franz et al. 2008, Witteveen 2014).

The triangle of reference, or semiotic triangle (Ogden and Richards 1923), is a model of how syntax and semantics are related to the objects they represent (van Rijsbergen 1979).

In the model (Figure 3), there is no direct relationship (dotted line) between symbols (i.e., *names*) and the real-world objects (the *referent*) they represent (A). Meaning, or the relationship between the name and the object, is conveyed only through a concept that exists in the mind of the user of the name. In taxonomy, a biologist (B) determines a specimen is sufficiently distinct to constitute a new species and documents the concept or idea of this novelty to a publication and assigns a name to it. Another person (C) subsequently reading the name, perhaps as a label on a specimen, evokes the concept originally described by the biologist, to refer to the specimen. Accurate communication occurs when there is congruence between both concepts among the writer and the reader.

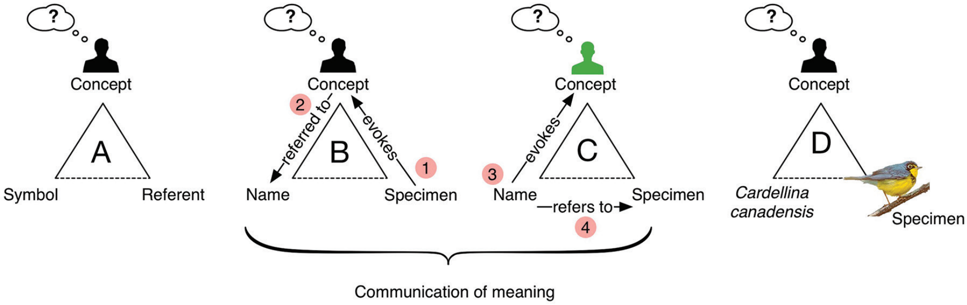


Figure 3. The semiotic triangle describes how names communicate meaning.

In biological taxonomy, a species name refers to a concept anchored by a specimen but created in the mind of a biologist. The function of the name is to facilitate communication. Communication is facilitated, however, only when the concepts (not the objects) are approximately congruent. Success is not black and white, but can be partial – whether partial is good enough is contingent on context-specific inference needs that the reciprocal concept alignment must fulfill. Thus, two persons look at the same avocado and one declares it a fruit, because it is derived from floral ovaries, while another declares it is not a fruit because it is not sweet. This conflict occurs when there is no congruency in the concepts invoked through the use of the name. Similar issues occur within taxonomy. In the simple case above, the term 'fruit' is associated with two definitions, or, more formally, the *cardinality* between syntax and semantics is one-to-two, or more generally, one-to-many (1:N). The same object evokes the same name but refers to two concepts according to two individuals. It is the relationship between the name and the concept that is important. Cardinality between syntax and semantics has a direct impact on the use and limits of scientific names as identifiers in biological informatics (Franz 2014).

Identifiers such as names have utility in information discovery and retrieval that is directly proportional to the degree of correlation between the term and the associated meaning or, in the semiotic context, in the correlation between syntax and semantics. Laypersons may think of scientific names as stable and unique, where a single Latin binomial name refers to one species and remains that way for all time. In other words, that there is a stable one-to-one relationship between a name (syntax) and the taxon (semantics) that it labels (Stringer 2002). This is an important informatics pre-condition if we are to rely on names as a means to search for and retrieve relevant information related to taxa (Thau et al. 2010).

Relevance in information retrieval is measured as a combination of two factors: precision and recall (van Rijsbergen 1979). Precision refers to the exactness, or quality in an information retrieval instance. Recall is a measure of quantity or completeness. For example, in Figure 4, below, a search for articles on ants misses some relevant articles but also accidentally returns articles on plants.

A search result, therefore, can produce two kinds of relevance errors.

1. Recall Is the proportion of relevant documents that are retrieved.
Recall =A/(A+D) or 7/10 =70%
2. Precision is the proportion of retrieved documents that are relevant.
Precision=A/(A+B) or 7/11 = 63%

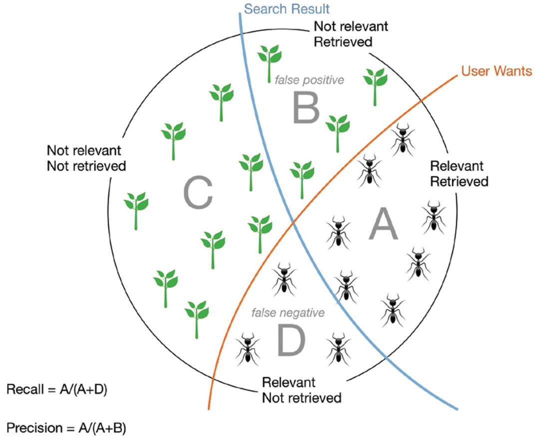


Figure 4. Precision vs. recall in search results.

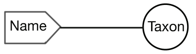
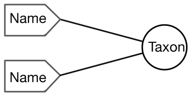
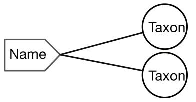
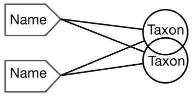
- A false positive error occurs when the system returns a result that is non-relevant. This is an error of precision.
- A false negative error occurs when the system fails to return a relevant result. This is an error of recall.

Perfect identifiers

Based on the above, we can define a perfect identifier as one that returns 100% relevant results; that is, zero false positive, and zero false negative, results. This is easy to understand in a relational database system that uses internal unique identifiers to ensure that all relevant records are returned in queries. Relational integrity within a database management system relies on a 1:1 relationship between a primary key and the object it represents. Integrity would be lost if two identifiers referred to the same object or if the same identifier referred to two objects. For an identifier to be a perfect identifier both the cardinality and correlation between syntax and semantics is exactly 1:1. From a taxonomic standpoint, this would require a single, unique name to refer to a single, distinct taxon. Any change or difference in semantics should be linked to a corresponding change in syntax (Lepage et al. 2014). This is not, however, the reality of biological taxonomy (Berendsohn and Geoffroy 2007, Franz et al. 2008).

Laypersons are often surprised to learn that scientific names are neither stable nor unique identifiers for taxa. The underlying causes for this instability have their roots in both syntax and semantics (Witteveen 2014) but the common consequence is a departure from the tight 1:1 correlation that is required to maintain the relational integrity between a name and the taxon to which it refers. This cardinal relationship dictates the utility and application of scientific names within biological information retrieval. It

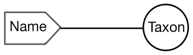
Table 1. Summary of cardinal relationships between names and taxa.

				Impact	
Cardinality	**Abbrev.**	**Diagram**	**Example**	**Recall**	**Precision**
One to One	1:1		Stable taxon	No	No
Many-to-One	N:1		Synonyms	Yes	No
One-to-Many	1:N		Homonyms/ Polysemes	No	Yes
Many-to-Many	N:N		Taxon Concept	Yes	Yes

is this relationship between syntax and semantics that dictates whether the impact on relevance will fall on precision or on recall, or on both.

There are four cardinal relationships possible between syntax (names) and semantics (taxa) in this regard and they are summarized in Table 1 below.

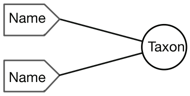
The relationship between a scientific name and the taxon to which it refers always falls into one of the four conditions in this table. Each of these conditions is represented within biological taxonomy and imposes informatics challenges that, in many cases, may be mitigated.

One-to-One (1:1)	**Cardinality**		**Impact**	**Result**
	Syntax One Name	Semantics One Meaning	No impact on precision/recall	Maximum relevance

The perfect identifier, as defined above, returns no false positive or false negative results when applied in a search. Thus, a search by name returns all and only the relevant related objects. In biology, there are many taxa that are so under-studied that they are only known from their original description and none or very few subsequent references (Thessen et al. 2012). The name alone, so long as it is a unique name, is sufficient to locate all related material. As noted, under these conditions:

- Recall – A single name will ensure no false negatives will be missed
- Precision – A single taxon labeled with the name will ensure no false positives are included in the results.

In reality Latinized scientific names are complex and easily misspelled such that this pure one-to-one condition is not as easily met. When this occurs, multiple synonyms refer to the same taxon.

Many-to-One (N:1)	Cardinality		Impacts	Result
	Syntax	Semantics	Recall	False negatives
	Multiple names	One meaning		

Synonyms are multiple names associated with a single taxon. Rules of nomenclature dictate that only one name is the correct label for a taxon. Any others must be "sunk" in synonymy (Del Hoyo and Pedrola-Monfort 2010). When this occurs it is clear that a single name may no longer be used to retrieve all information related to the taxon. This has the following impact on relevance.

- Recall – Synonyms impact recall because the use of a single name will result in false negative results.
- Precision – Synonyms in a N:1 condition do not impact precision because, by definition, only a single concept is involved. Thus, false positive results are not possible through matching any of the names.

Synonyms fall into several classes based on their origin:

Orthographic or lexical synonyms

Variations in spelling represent one class of synonyms although they are often not formally referred as such. The names "*Loligo pealeii*," "*Loligo pealei*" and "*Loligo pealii*" for example, have all been used to refer to a particular species of squid (*Loligo pealeii* LaSeur, 1821) although only the first spelling is correct. "*Pomatomus saltator*" and "*Pomatomus saltatrix*" represent variants based on differences in Latin gender applied to the species name. While only one is syntactically correct, they are both regularly used (David and Gosselin 2002, Welter-Schultes et al. 2015). This conflation of syntax impacts recall when data stores containing variant orthographies are searched.

Nomenclatural synonyms

Nomenclatural synonyms represent a syntactic change without an associated change in semantics. This may occur when two names are discovered to refer to the same original publication or to the specimens that form the basis for the description. For example, the name *Taraxacum officianale* F.H. Wigg shares the same type specimen as the name *Leontodon taraxacum* L. The two names therefore, refer to the same taxon (Kirschner and Štěpánek 2011).

The binomial name of scientific names result in a change in syntax when a taxon is moved to a different genus or if a name is not published according to formal nomenclatural rules. (Blackwelder 1967, International Commission on Zoological Nomenclature 1999, Dubois 2000, McNeill et al. 2005) When *Drosophila melanogaster* was proposed to belong within the genus *Sophophora*, a new binomial, *Sophophora melanogaster* was created. These syntactic changes are not reflective of a change in taxonomic circumscription (van der Linde and Yassin 2010).

Taxonomic synonyms

Taxonomic synonyms are the result of a change in circumscription that occurs when two, formerly distinct taxa, are merged. This may occur due to broad variation within a species giving rise to multiple, correctly published species descriptions that are ultimately deemed to belong to the same taxon. For example, *Antilocapra anteflexa* Gray, 1855, is an antelope who's description was based on a pair of horns. It has since been determined, and is now generally accepted to be, a variant of a previously-described species, *Antilocapra americana* Ord, 1815 (O'Gara 1978). Syntactic and pragmatic rules result in one name being applied to the newly merged taxon while other names, which may include orthographic variants and nomenclatural synonyms linked to the grouped taxa, fall into synonymy (Polaszek 2008).

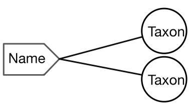
In all of these cases, information tied to a single taxon may be labeled with multiple different labels. This will result in false negative results in search and retrieval across data stores containing multiple names for the taxon.

Mitigation of synonyms

Different approaches have been applied to overcome the impact on recall inherent to synonymy.

- "Fuzzy" name-matching services are used to group orthographic variants and misspellings (Rees 2014).
- Taxonomic names servers, such as provided by uBio, iPlant and ITIS offer thesaurus-like services that provide the list of related names that can be used to conflate a search and improve recall (Boyle et. al. 2013, Remsen 2014, Integrated Taxonomic Information System 2014)

Homonyms

One-to-Many (1:N)	Cardinality		Impacts	Result
	Syntax	Semantics		
	One name	Multiple meanings	Precision	False positives

Homonyms are two identically-spelled names that refer to two distinct taxa. For example, the genus *Aotus* refers to both a legume and a primate (Remsen and Patterson 2010). The species *Agathis montana* may be either a wasp or a conifer (Encyclopedia of Life 2014). Secondary homonyms are a consequence of transferring a species to a new genus that already contains the constructed binomial (Dickinson 2016).

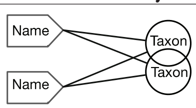
The word *homograph* is similarly used to refer to two identically-spelled words and broadens the definition outside of biological taxonomy. *Cancer*, for example, is both a genus of crab and a medical condition. The result, however is the same: a one-to-many (1:N) relationship between syntax (one name) and semantics (two taxa). This has the following impact on relevance:

- Recall – Homonyms do not impact recall in this condition, because, by definition, only a single name is relevant and false negative matches are not possible.
- Precision – Homonyms impact precision because the name is ambiguous and can produce false positive results when a match is made to a non-target taxon.

Mitigation of homonyms

There are two ways to improve precision when a name is too ambiguous; syntactic and pragmatic. The syntactic approach is to change the cardinality between the names and taxa from one-to-many to one-to-one. This is achieved by changing the syntax to two distinct forms. In the case of *Aotus* for example, the legume may be formally referred to as *Aotus* Smith, 1805 while the primate is *Aotus* Illiger, 1811. While the name components remain identical, the appended authorship information renders them distinct. The usage of this form of the name improves precision, effectively moving the burden of relevance to recall. The use of the more precise term supports the distinction of the two taxa but results in the minting of a synonym. The result would be fewer false positive but a potential increase in false negatives as the use of the more refined name would miss relevant results labeled only with the homonym. In general, improvements in precision result in a decrease in recall and vice-versa; Resolving this would require changing all the retrospective ambiguous use of the name *Aotus* with the more precise amended form (Rees 2012).

The pragmatic approach relies on analytic techniques that try to identify context to disambiguate the term. For example, the term "monkey" or "pea" in the vicinity of the use of the name *Aotus*, could help disambiguate the usage and improve precision (Boulis and Ostendorf 2005).

One-to-Many	**Cardinality**		**Impacts**	**Result**
	Syntax	Semantics	Precision	False positives
	One name	Multiple meanings		

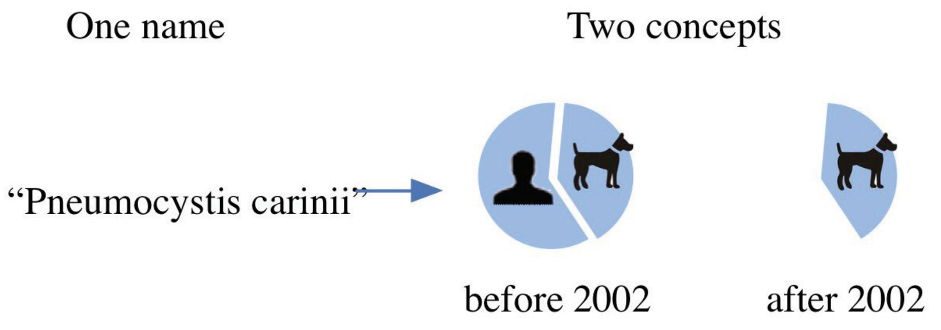


Figure 5. A polyseme is a single name referring to more than one overlapping or included concept.

Polysemy (literally "many meanings") is a condition similar to homonymy and refers to a single name that refers to two taxa. Instead of consisting of entirely distinct taxa, however, the circumscriptions overlap. This occurs when taxa are lumped and split and result in two or more taxon concepts (Remsen and Patterson 2010, Franz and Cardona-Duque 2013). A polyseme impacts relevance in a manner similar to a homonym but is much more common. Impacts on relevance are as follows:

- Recall – Recall is not impacted. Syntactic ambiguity is not a factor here as there is only a single name.
- Precision – Polysemes impact precision because a single taxon name refers to two or more different circumscriptions for a taxon.

Pneumocystis carinii Delanoe & Delanoe, 1912, is a fungal pathogen responsible for a deadly pneumonia (PCP) in HIV-infected patients. It was originally described in dogs and rats and later found to occur in humans. In 2002 genetic analysis determined that the human pathogen was distinct from the one that infects dogs. A new name, *Pneumocystis jiroveci*, was applied to the human pathogen (Stringer 2002). The result was a taxonomic split where the original name, *Pneumocystis carinii* subsequently referred to just part of the original circumscription and the new name *Pneumocystis jiroveci*, to the remaining part. Figure 5 illustrates this split.

The name, *Pneumocystis carinii*, now refers to two different circumscriptions; one that contains dog and human pathogens and one that contains only dog pathogens. A search using the name can return results that may refer to either use of the name, corresponding to false positives that impact precision.

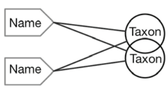
Mitigation of polysemes

Rules of nomenclature do not support reflective syntax changes due to changes in circumscription. When a taxon is split, the original name is carried on to refer to one of the resultant parts.

Berendsohn (1995) has suggested that the name be concatenated with the annotation "sensu" followed by the author of the split to denote the circumscription reference with a unique label. In this case, the taxon would be known by two names:

- "*Pneumocystis carinii* Delanöe & Delanöe, 1912" (original) or plain "*Pneumocystis carinii*".
- "*Pneumocystis carinii* sensu Stringer, 2002" (new).

This syntax would provide the means to distinguish the two circumscriptions in any future application but it leaves all previous applications ambiguous since the earlier application of the name can, in the context of the subsequent split, refer to either of the two new concepts. Any previous applications of the name would have to be re-assessed and re-labeled for any retrospective precision improvements. In some cases, this can be inferred through re-inspection and reasoning, using both manual and automated methods (Lepage et al. 2014). For example, pre-2002 medical literature referring to *P. carinii* in human patients can be reasonably inferred to refer to the more precise taxon, *P. jiroveci*.

Many-to-Many	**Cardinality**		**Impact**	**Result**
	Syntax	Semantics	Precision & Recall	False positives False negatives
	Many	Many		

Polysemes were introduced as referring to a single name referring to multiple, related taxa; a condition that results from splitting a taxon concept into two or more new circumscriptions. Polysemy, however, is not the only result of semantic changes. When *Pneumocystis carinii* was divided into the human pathogen, *Pneumocystis jiroveci* and a reduced, non-human-infecting *Pneumocystis carinii,* the net result was three distinct circumscriptions labeled with two names.

Table 2. The result of a taxonomic split on syntax and semantics.

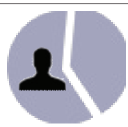
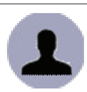
Syntax		
Pneumocystis carinii		Original taxon that infects both dogs and humans
		New taxon that only infects dogs
Pneumocystis jiroveci		New taxon that only infects humans

Table 3. Lumped and split taxon and use of names to impact relevance where P=Precision and R=Relevance.

Taxon infects	Names	Semantics	P	R
Dogs and humans	***Pneumocystis carinii*** *- Synonym: Pneumocystis jiroveci*		Y	Y
Dogs only	***Pneumocystis carinii*** *- Syn.: Pneumocystis carinii (part)*		N	Y
Humans only	***Pneumocystis jiroveci*** *- Syn.: Pneumocystis carinii (part)*		N	Y

The relationship between the names and the circumscriptions corresponds to a many-to-many (N:N). The two names and three concepts are all inter-related (Franz 2014).

The net result of the split, and the resultant impact on relevance in search, is summarized in Table 3 above. Synonymy can be used to positively improve precision and recall for the lumped taxon when applied to search and retrieval. When taxa are split, synonymy may improve recall by retrieving otherwise false negative results tied to the use of synonyms. As one of the names is a polyseme, however, synonymy cannot improve precision.

Summary

Scientific names link nearly all information related to a species but the relationship between nomenclatural syntax and taxonomic semantics is inherently ambiguous. Informatics processes that rely on data-gathering methods linked to taxon names are susceptible to this ambiguity and run the risk of providing imprecise or incomplete sets of data to subsequent downstream processes.

Sets of related scientific names may be used, as in today's array of taxonomic name servers, to improve recall in search and retrieval for information tied to a taxon. The ambiguity of scientific names that occurs when the same name refers to two distinct, or overlapping taxa, however means that, in many cases, a single name returns an imprecise result and this is something that cannot be rectified through the use of name services.

Comprehensive taxonomic thesauri are required to model the relationships between names and taxa. Nomenclatural databases that currently capture the objective syntactic properties of names could improve their relevance by cataloging nomenclatural synonyms, as attempted in Index Animalium. Effectively modeling semantics requires a clean division between these syntactic aspects of taxonomy and the subsequent subjective processes that result in changes in circumscription (Lepage et al. 2014). An ideal system enables the identification, modeling and exposure of a complete array of circumscription changes for any and all taxa and are coupled with services that allow these to be embedded within informatics processes (Tuominen et al. 2011, Chem et al. 2014).

Acknowledgements

The author would like to extend his thanks and graditude to reviewers, Nico M Franz, Ph.D., Associate Professor and Curator of Insects, Arizona State University, and John Todd, Ph.D., Curator of Molluscs, Natural History Museum, London. An additional thanks to Dr. Ellinor Michel, ICZN Executive Secretary, for serving as general editor and expediting the whole Sherborn publication process.

References

Alonso-Zarazaga MA, Fautin DG, Michel E (2016) The List of Available Names (LAN): A new generation for stable taxonomic names in zoology? In: Michel E (Ed.) Anchoring Biodiversity Information: From Sherborn to the 21st century and beyond. ZooKeys 550: 225–232. doi: 10.3897/zookeys.550.10043

Berendsohn WG (1995) The concept of "potential taxa" in databases. Taxon 44: 207–212. doi: 10.2307/1222443

Berendsohn WG, Geoffroy M (2007) Networking taxonomic concepts – uniting without 'unitary-ism. In: Curry GB, Humphries CJ (Eds) Biodiversity Databases: Techniques, Politics, and Applications. Systematics Association Special Volume 73. CRC Press, Boca Raton, 13–22. doi: 10.1201/9781439832547.ch3

Blackwelder RA (1967) Taxonomy: A text and reference book. Wiley, New York, 714 pp.

Boulis C, Ostendorf M (2005) Text classification by augmenting the bag-of-words representation with redundancy-compensated bigrams. Feature Selection in Data Mining. SIAM conference on Data Mining.

Boyle B, Hopkins N, Lu Z, Garay JA, Mozzherin D, Rees T, Matasci N, Narro ML, Piel WH, Mckay SJ, Lowry S (2013) The taxonomic name resolution service: an online tool for automated standardization of plant names. BMC Bioinformatics 14(16): 14–16. doi: 10.1186/1471-2105-14-16

Chem M, Yu S, Franz N, Bowers S, Ludäscher B (2014) Euler/X: A Toolkit for Logic-based Taxonomy Integration. Retrieved 10 21, 204, from Cornell University LIbrary at arXiv.org. http://arxiv.org/abs/1402.1992

CWR (1903) Review of Sherborns *Index Animalium*. The Auk 20(4): 448–450.

David N, Gosselin M (2002) Gender agreement of avian species-group names under Article 31.2.2 of the ICZN Code. Bulletin of the British Ornithologists Club 122(1): 14–49.

Del Hoyo A, Pedrola-Monfort J (2010) Taxonomic clarification in W Mediterranean Androcymbium (Colchicaceae): A. wyssianum sunk in the synonymy of A. gramineum and A. europaeum restored. Willdenowia 40(1): 47–53. doi: 10.3372/wi.40.40103

Dickinson EC (2016) Reinforcing the foundations of ornithological nomenclature: Filling the gaps in Sherborn's and Richmond's historical legacy of bibliographic exploration. In: Michel E (Ed.) Anchoring Biodiversity Information: From Sherborn to the 21st century and beyond. ZooKeys 550: 107–134. doi: 10.3897/zookeys.550.10170

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

Encyclopedia of Life (2014) Homonyms on EoL. http://eol.org/collections/6375 [retrieved 10 20 2014]

Evenhuis NL (2016) Charles Davies Sherborn and the "Indexer's Club". In: Michel E (Ed.) Anchoring Biodiversity Information: From Sherborn to the 21st century and beyond. ZooKeys 550: 13–32. doi: 10.3897/zookeys.550.9697

Franz NM (2014) Explaining taxonomy's legacy to computers – how and why? Retrieved 10 20, 2014, from Slide Share web presentation: http://www.slideshare.net/taxonbytes/franz-2014-explaining-taxonomys-legacy-to-computers-how-and-why

Franz NM (2014) Names Are Not Good Enough: Reasoning over Taxonomic Change in the Andropogon Complex. Semantic Web Journal, 25.

Franz NM, Cardona-Duque J (2013) Description of two new species and phylogenetic reassessment of Perelleschus Wibmer & O'Brien, 1986 (Coleoptera: Curculionidae), with a complete taxonomic concept history of Perelleschus sec. Franz & Cardona-Duque, 2013. Systematics and Biodiversity 11: 209–236. doi: 10.1080/14772000.2013.806371

Franz NM, Chen M, Yu S, Bowers S, Ludäscher B (2014) Names Are Not Good Enough: Reasoning over Taxonomic Change in the Andopogon Complex. http://www.semantic-web-journal.net/content/names-are-not-good-enough-reasoning-over-taxonomic-change-andropogon-complex [retrieved October 15, 2014]

Franz NM, Peet RK, Weakley AS (2008) On the use of taxomomic concepts in support of biodiversity research and taxonomy. In: Wheeler QD (Ed.) The New Taxonomy. Systematics Association Special Volume Series 74, Taylor and Francis, Boca Raton, 63–86. doi: 10.1201/9781420008562.ch5

Grimaldi D, Engel MS (2005) Evolution of the insects (Cambridge Evolution Series). Cambridge University Press, Cambridge, 772 pp.

Integrated Taxonomic Information System (2014) Integrated Taxonomic Information System. http://www.itis.gov [retrieved Sept 29, 2014]

ICZN (1999) International Code of Zoological Nomenclature (4th ed.). The International Trust for Zoological Nomenclature, London.

Kirschner J, Štěpánek J (2011) Typification of Leontodon taraxacum L. (≡ Taraxacum officinale F.H. Wigg.) and the generic name Taraxacum: A review and a new typification proposal. Taxon 60(1): 216–220.

Lepage D, Vaidya G, Guralnick R (2014) Avibase – a database system for managing and organizing taxonomic concepts. ZooKeys 420: 117–135. doi: 10.3897/zookeys.420.7089

Linnaeus C (1758) Systema naturae per regna tria naturae: secundum classes, ordines, genera, species, cum characteribus, differentiis, synonymis, locis. In: Salvius L (Ed.) http://www.biodiversitylibrary.org/bibliography/542 [retrieved Febr 18, 2013]

Mayr E (1991) Principles of Systematic Zoology. McGraw-Hill, New York, 475 pp.

McNeill J, Barrie FR, Burdet HM, Demoulin V (2005) International code of botanical nomenclature (Vienna Code) adopted by the seventeenth International Botanical Congress, Vienna, Austria, July 2005. International Association for Plant Taxonomy, Vienna, AT, 568 pp.

Murray RG, Stackebrandt E (1995) Taxonomic Note: Implementation of the Provisional Status Candidatus for Incompletely Described Procaryotes. International Journal of Systematic and Evolutionary Microbiology 45(1): 186–187. doi: 10.1099/00207713-45-1-186

O'Gara BW (1978) Mammalian Species. The American Society of Mammalogists 90: 1–7.

Ogden CK, Richards IA (1923) The Meaning of Meaning: A Study of the Influence of Language upon Thought and of the Science of Symbolism. Harcourt, San Diego, 363 pp.

Page RD (2008) Biodiversity informatics: the challenge of linking data and the role of shared identifiers. Briefings in Bioinformatics 9(5): 345–354. doi: 10.1093/bib/bbn022

Patterson DJ, Cooper J, Kirk PM, Pyle RL, Remsen DP (2010) Names are key to the big new biology. Trends in Ecology and Evolution 25(12): 686–691. doi: 10.1016/j.tree.2010.09.004

Patterson DJ, Remsen DP, Marino WA, Norton C (2006) Taxonomic Indexing - Extending the role of taxonomy. Systematic Biology 55(3): 367–373. doi 10.1080/10635150500541680

Pilsk SC, Kalfatovic MR, Richard JM (2016) Unlocking *Index Animalium*: From paper slips to bytes and bits. In: Michel E (Ed.) Anchoring Biodiversity Information: From Sherborn to the 21st century and beyond. ZooKeys 550: 153–171. doi: 10.3897/zookeys.550.9673

Polaszek A (2008) Comments on the proposed conservation of the usage of the generic name of Drosophila Fallén, 1823 (Insecta, Diptera) 1 (Case 3407). Bulletin of Zoological Nomenclature 65(1): 55.

Pyle RL (2016) Towards a Global Names Architecture: The future of indexing scientific names. In: Michel E (Ed.) Anchoring Biodiversity Information: From Sherborn to the 21st century and beyond. ZooKeys 550: 261–281. doi: 10.3897/zookeys.550.10009

Rees T (2012) IRMNG Homonyms. CSIRO, Producer. http://www.cmar.csiro.au/datacentre/irmng/homonyms.htm [retrieved Febr 21, 2013]

Rees T (2014) Taxamatch, an Algorithm for Near ('Fuzzy') Matching of Scientific Names in Taxonomic Databases. PLoS ONE 9(9): e107510. doi: 10.1371/journal.pone.0107510

Remsen DP (2014) uBIo Project. uBio Project: http://www.ubio.org [retrieved Sept 29, 2014]

Remsen DP, Patterson D (2010) The All-Genera Index: Strategies for Managing the BIG Index of All Scientific Names. In: Polaszek A (Ed.) Systema Naturae 250 – The Linnaean Ark. CRC Press, Boca Raton, 149–162. doi: 10.1201/EBK1420095012-c14

Remsen DP, Norton C, Patterson DJ (2006) Taxonomic Informatics Tools for the Electronic Nomenclator Zoologicus. Biological Bulletin 210: 18–24. doi: 10.2307/4134533

Stringer JR (2002) A New Name for Pneumocystis from Humans and New Perspectives on the Host-Pathogen Relationship. Emerging Infectious Diseases 8(9): 891–896. doi: 10.3201/eid0809.020096

Thau D, Bowers S, Ludäscher B (2010) Towards Best-Effort Merge of Taxonomically Organized Data – ICDE Workshops. IEEE, Long Beach, 151–154.

Thessen AE, Patterson DJ, Murray SA (2012) The Taxonomic Significance of Species That Have Only Been Observed Once: The Genus Gymnodinium (Dinoflagellata) as an Example. PLoS ONE 7(8): e44015. doi: 10.1371/journal.pone.0044015

Thompson CF, Pape T (2016) Sherborn's influence on *Systema Dipterorum*. In: Michel E (Ed.) Anchoring Biodiversity Information: From Sherborn to the 21st century and beyond. ZooKeys 550: 135–152. doi: 10.3897/zookeys.550.9447

Tuominen J, Laurenne N, Hyvönen E (2011) Biological Names and Taxonomies on the Semantic Web – Managing the Change in Scientific Conception. The Semanic Web: Research and Applications Lecture Notes in Computer Science 6644: 255–269.

van der Linde K, Yassin A (2010) The fruit fly formerly know as Drosophila. New Scientist 206(2765): 24–25. doi: 10.1016/S0262-4079(10)61485-6

van Rijsbergen CJ (1979) Information Retrieval (2nd Edition). University of Glasgow, Glasgow, 224 pp.

Varian HR (2007) The Economics of Internet Search. 'Angelo Costa' Lectures Serie 7: 17.

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. In: Michel E (Ed.) Anchoring Biodiversity Information: From Sherborn to the 21st century and beyond. ZooKeys 550: 173–183. doi: 10.3897/zookeys.550.10041

Witteveen J (2014) Naming and contingency: the type method of biological taxonomy. Retrieved 10 19, 2014, from Archive for Preprints in Philosophy of Science: http://philsci-archive.pitt.edu/id/eprint/10890

ZooKeys 550: 225–232 (2016)
doi: 10.3897/zookeys.550.10043
http://zookeys.pensoft.net

RESEARCH ARTICLE

A peer-reviewed open-access journal
ZooKeys
Launched to accelerate biodiversity research

The List of Available Names (LAN): A new generation for stable taxonomic names in zoology?

Miguel A. Alonso-Zarazaga[1], Daphne Gail Fautin[2], Ellinor Michel[3]

1 *Departamento de Biodiversidad y Biología Evolutiva, Museo Nacional de Ciencias, Naturales (MNCN-CSIC), Madrid, Spain* **2** *Department of Ecology and Evolutionary Biology, and Natural History Museum (Biodiversity Institute), University of Kansas, Lawrence, Kansas USA & International Commission on Zoological Nomenclature* **3** *Department of Life Sciences, The Natural History Museum, Cromwell Road London SW7 5BD, UK*

Corresponding author: *Miguel A. Alonso-Zarazaga* (zarazaga@mncn.csic.es)

Academic editor: *E. Michel* | Received 25 May 2015 | Accepted 25 May 2015 | Published 7 January 2016

http://zoobank.org/7BD917A6-4797-4852-AB98-CCEFA6C7A70E

Citation: Alonso-Zarazaga MA, Fautin DG, Michel E (2016) The List of Available Names (LAN): A new generation for stable taxonomic names in zoology? In: Michel E (Ed.) Anchoring Biodiversity Information: From Sherborn to the 21st century and beyond. ZooKeys 550: 225–232. doi: 10.3897/zookeys.550.10043

Abstract

The *List of Available Names in Zoology* (LAN) is an inventory of names with specific scope in time and content, presented and approved in parts, and constituted as a cumulative index of names available for use in zoological nomenclature. It was defined in Article 79 in the fourth edition of the *International Code of Zoological Nomenclature*. The LAN is likely to gain importance with the development of the online Official Registry for Zoological Nomenclature (ZooBank) as it is potentially a source of many nomenclaturally certified names. Article 79 describes the deliberative process for adding large numbers of names to the LAN simultaneously, detailing steps and chronology for submission of a candidate Part to the LAN and consideration of a candidate Part by the public and Commission, but it is largely mute about the contents of a candidate Part. It does make clear that a name within the scope of a Part but not on the LAN has no nomenclatural standing, even if it had previously been considered available, thereby preventing long-forgotten names from displacing accepted ones and the accumulation of *nomina dubia*. Thus, for taxa on the LAN, nomenclatural archaeology – the resurrecting of old unused names to replace by priority names in current usage – will not be worthwhile. Beyond that, it has been unclear if Article 79 is intended to document every available name known within the scope of the Part, or if its intention is to pare the inventory of available names within the scope of the Part. Consideration by the Commission and two committees to deal with the LAN have defined steps to implement Article 79 with the latter intent. Procedures for consideration of a candidate Part are defined in a manual, published as an appendix in this volume.

Keywords
Scientific names, stable nomenclature, taxonomy, bibliography, Sherborn

Introduction

The fourth edition of the *International Code of Zoological Nomenclature* (International Commission on Zoological Nomenclature 1999; hereafter "the Code") introduced the concept of a *List of Available Names in Zoology* (LAN) as a way to deal with the plethora of available names that has accumulated over the more than two and a half centuries of zoological nomenclature since the founding datum of the field by Linnaeus' 10th edition of *Systema Naturae* (1758). The Code defined the LAN as an inventory of names with specific scope in time and content, presented and approved in parts, to constitute a cumulative index of names available for use in zoological nomenclature. The LAN, which was envisioned as a major step in stabilizing nomenclature (cf. Scoble 2004), has taken on additional significance with the development of the online Official Registry for Zoological Nomenclature, with its online presence called ZooBank (http://zoobank.org), because the LAN can potentially serve as a source of many nomenclaturally certified names. It is an idea of long standing, discussed and advanced in fora and articles (e.g., ICZN 1990, Savage 1990), and was welcomed as "the second major change in the Code" in a review of the 4th Edition (Ferraris and Eschmeyer 2000: 908). Polaszek and Michel (2010), for example, proposed that the *Official Lists* and the LAN would play a key role in populating ZooBank.

Article 79 of the fourth edition of the Code describes a procedure for simultaneous addition of large numbers of names to the LAN (the prescription for creating a LAN was novel in the fourth edition). Article 79 deals in considerable detail with who may submit a candidate Part of the LAN for consideration, and with how the candidate Part is to be considered (including timing), but it does not define the content of a candidate Part of LAN. Creating a procedure to implement Article 79 has been the job of a Standing Committee of the *International Commission on Zoological Nomenclature* (hereafter "the Commission") headed by Commissioner Alonso-Zarazaga, and of an *ad hoc* Committee headed by Commissioner Fautin that was appointed to deal with an application made under Article 79 (as required by Article 79.2.1).

A procedure for adding a Part to the LAN has taken time to establish. A major motive for crafting this procedure clearly, precisely and comprehensively is that it sets precedent. This contrasts with many actions of the Commission. Article 80.5 of the Code, for example, states "An Opinion applies only to the particular case before the Commission and is to be rigidly construed; no conclusions other than those expressly specified are to be drawn from it." Thus, an Opinion rendered by the Commission applies only to the Case in question, but implementation of Article 79 is a Commission action that stipulates a procedure, and therefore Parts of the LAN adopted by that procedure set precedent.

Aspects addressed in the Code

Facets of the procedure stipulated in Article 79 for simultaneously adding large numbers of names to the LAN include 1) who may submit a candidate Part for consideration, 2) the scope of a candidate Part, 3) what those who have proposed a Part are to do, and 4) what is to be done with the candidate Part by the Commission.

Submission of a candidate Part of the LAN must be by "an international body of zoologists." The scope of the candidate Part must be specified in terms of taxon, rank(s), and time period covered. As for Sherborn's list in *Index Animalium*, the bibliographic source of each name must be provided, but so must details of any relevant actions by the Commission, and the "status" and details of its type, which for species involves citing how type specimens were designated and their repositories. The language of Article 79, therefore, would preclude a latter-day Sherborn: he acted single-handedly, dealt with all animal taxa, and was not concerned with typification.

Article 79 contains considerable detail on timing. Once a candidate Part is open for "comments by zoologists" for 12 months, the community is notified by means of a notice published in the *Bulletin of Zoological Nomenclature*. At the end of the year of review, the *ad hoc* Committee dealing with the particular candidate Part, considering the public input, recommends a vote to the entire Commission. This vote, which must take place no less than two years from the date of publication of the notice (that is, in most cases at least a year after the period of public comment has closed), must be either to abandon further consideration of the candidate Part of the LAN or to consider a candidate Part of the LAN revised in light of comments received; there is no option for the Commission to accept the candidate Part of the LAN at that time. If consideration of the candidate Part continues, another 12-month period of public input on the revised candidate Part follows (subsequent to notice), after which the *ad hoc* Committee again recommends a vote to the entire Commission. This vote, too, must take place no less than two years from the date of publication of the notice, but this vote is either to abandon further consideration of the candidate Part or to accept it. Thus, the entire process of considering a candidate Part that is eventually approved for inclusion in the LAN takes a minimum of four years. When the Commission votes to add a candidate Part to the LAN, notice to that effect must promptly be published in the *Bulletin of Zoological Nomenclature*.

The Code also comments on what the LAN is not. Article 80.8 distinguishes between the LAN and the *Official Lists*, of which there are four, as defined in the Glossary, the relevant ones being the *Official List of Family-Group Names in Zoology,* the *Official List of Generic Names in Zoology*, and the *Official List of Specific Names in Zoology*. These lists are compiled by the addition of names singly or at most in small numbers, by contrast with the LAN, which is assembled by simultaneously adding large numbers of names. In case of conflict between the status of a name as given in the LAN and on one of the *Lists*, the former takes priority, as it does also in case of conflict between the status of a name as given in the LAN and an Opinion of the Commission (Article

80.8). However, although the LAN supersedes other actions of the Commission, according to Article 79.5, "If there are exceptional circumstances and only when an entry in the *List of Available Names in Zoology* is a cause of confusion, the Commission may amend the entry by use of its plenary power and publish its ruling in an Opinion." Nonetheless, some confusion still exists about these fundamentally different documents, the LAN and the *Official Lists* and *Indexes* (e.g. Gregory 2010).

Aspects not addressed in the Code

Despite details of who may propose a candidate Part and stipulating actions along a time line leading to rejection or adoption of a candidate Part, the Code provides few details about the desired contents of a candidate Part of the LAN. Article 79.4.3 does state "No unlisted name within the scope (taxonomic field, ranks, and time period covered) of an adopted Part of the *List of Available Names in Zoology* has any status in zoological nomenclature despite any previous availability." Thus, any name discovered subsequent to the adoption of a Part of the LAN does not compete for priority, etc., so Article 79 has the effect of preventing nomenclatural archeology as an end rather than a means on one side and getting rid of *nomina dubia* on the other. Of course, any omission or error can become the subject of an appeal to the Commission under Article 79.5 or 79.6.

Central to defining the content of a candidate Part of the LAN is understanding of what the LAN is meant to be. The main choices are whether the LAN is intended to contain all available names known or only some (carefully selected) available names; both positions have been advocated to us by current and past Commissioners. The universe of names of animals includes vernacular names, manuscript names, names decreed by the Commission to be unavailable for a diversity of reasons, etc. These are represented by the largest circle in Figure 1. Available names are a subset of all names, represented by the intermediate circle in Figure 1. If the LAN is to be an inventory of all available names, assembling it is a purely bibliographic exercise that seeks to uncover every name ever made available. The innermost circle in Figure 1 represents the subset of available names that are potentially valid. Nor does this inventory include names of doubtful application – those termed *nomina dubia* in the glossary of the Code. By this reasoning, junior objective synonyms and primary homonyms are to be excluded from the LAN, but it must include names considered to be subjective synonyms (they are still available and are retained in case they are needed in the future, e.g., for cryptic species).

What names belong on the LAN? What did the framers of Article 79 have in mind by establishing the elaborate procedure to assemble the LAN described in the Code? The first principle stated in the Introduction to the fourth edition of the Code is "The Code refrains from infringing upon taxonomic judgment, which must not be made subject to regulation or restraint," so assembling the LAN cannot involve excluding any available names based on taxonomy.

Retaining all available names – that is, holding on to names the significance of which may never be known – risks a Type I error: we must continue to inventory and

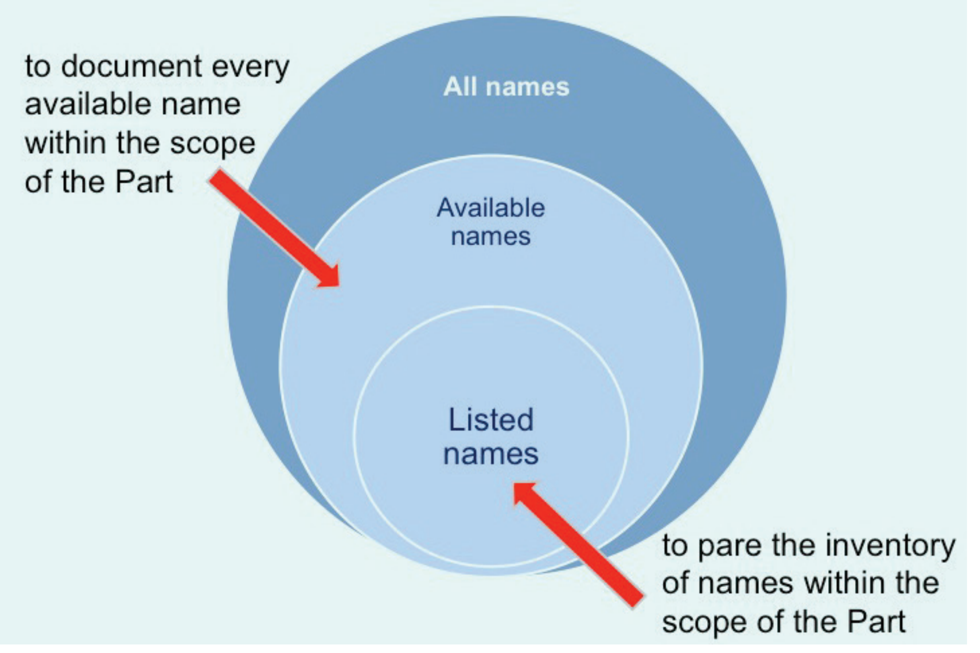


Figure 1. A schematic representation of the universe of animal names, the two possible choices in compiling a LAN and the contents of the two resulting lists.

deal with names of uncertain application in the hope we might someday associate at least some with a taxon. Paring names to those that we can currently associate with taxa risks a Type II error: we might discard a name that could eventually be associated with a taxon. Which type of error is the more costly involves considering the effort involved both in rectifying and in not rectifying the error. Retaining all available names requires not only continuing to inventory many names of uncertain application, but also, if the taxon to which a hitherto doubtful name applies is determined, redefining and retypifying the name, effectively redescribing the taxon. Restricting the LAN to a subset of all available names means effort need not be expended in carrying along names of uncertain application, but if a taxon that had had a name is rediscovered, that taxon must be described, with a new name assigned to it. This seems to us the more parsimonious procedure.

We have come down on the side of the LAN being a subset of the known available names. We consider that the taxon with a name that would not merit placement on the LAN would have to be redescribed extensively to fix the name, associating it unambiguously with the concept (and thereby meriting placement on the LAN); this would require as much effort as describing the taxon anew. Moreover, discarding the old name would not be disruptive because that name would not have been unambiguously used for the taxon, at least in a very long time. Thus paring the list will ultimately save effort, including the increasingly precious time of taxonomists. This interpretation is consistent with Article 79.4.3 ("No unlisted name within the scope (taxonomic field, ranks, and time period covered) of an adopted Part of the *List of Available Names in Zo-*

ology has any status in zoological nomenclature despite any previous availability"), with Recommendation 79A ("If for taxonomic and historical purposes an author desires to cite a name that is no longer available because it is not included in the relevant Part of the *List of Available Names in Zoology* adopted by the Commission, it should be made clear that it no longer has a status in zoological nomenclature"), and with item 15 of the Introduction ("Names within the scope of such an adopted list but not listed in it will be treated as unavailable"). Some names that were previously available will not end up on the LAN. In this way, each Part of the LAN will become a new datum for zoological nomenclature. This is similar to what the microbiologists did in making a new start for bacterial nomenclature on 1 January 1980, although, of course, Parts of the LAN will take effect at different times (Sneath 2003). Once accepted according to Article 79, the names on the Parts of the LAN can be entered into ZooBank with the assurance that they have been certified through a lengthy process of public vetting. If mandatory registration becomes part of making a name or act available (e.g., Krell 2009), ZooBank would achieve the same economy of effort that the Bacteriological Code has effected.

Article 79 lays out the requirements and timing of consideration of a candidate Part of the LAN. Implementation of Article 79, including stipulation of the contents of the candidate Part, are the subject of a Manual developed by the Standing Committee headed by Commissioner Alonso-Zarazaga, approved in the ICZN Session of November 20th 2013 held in Singapore, that accompanies this article as an Appendix and that will be posted on the Commission website (http://iczn.org).

The lengthy vetting process helps minimize the risk of a name in wide use inadvertently being omitted from the LAN. However, should the process fail, Article 79.6 provides that "If the Commission determines that there is a previously available name within the scope of an adopted Part of the *List of Available Names in Zoology* that has been omitted from the *List*, in exceptional circumstances the Commission may by use of the plenary power add an appropriate entry to that Part of the *List* and record this in an Opinion. The availability of the name thereby becomes restored."

An inventory that constitutes the candidate Part of the LAN may contain all available names within the scope of the Part. A name not on the candidate Part may have been intentionally excluded, or it may simply have been overlooked. So that members of the public reviewing such an inventory can distinguish between these two possibilities, the implementation document for Article 79 includes the requirement that the compilers of a candidate Part of the LAN include also an inventory of available names they do not want placed on the LAN. The form of this inventory is not stipulated, but the names in the two categories must clearly be differentiated: there may be two separate lists or the names may be on a single list but be distinguished by typeface, an indication such as an asterisk, etc. This inventory of names will become part of the public nomenclatural record – but these names, under Article 79.4.3 have no "status in zoological nomenclature despite any previous availability." Some concern has been raised about the minimum requirements that names placed in the list to be deleted should meet (e.g. Eschmeyer 2003, Dayrat 2005 and references therein) and also about how the placement of senior homonyms on it will affect junior homonyms in other taxonomic groups, for which there is no rule at present.

The LAN was designed to help stabilize nomenclature and to relieve the burden on taxonomists of dealing with names of uncertain application that have accumulated during more than two centuries of modern taxonomic science. A side benefit is that the LAN would deter the practice of what has been termed "taxonomic vandalism" (Polaszek 2010). It allows dealing with large numbers of homonymy and priority problems at the stroke of the pen: one by one, they would take a very long time and much more effort both for the proponents and for the Commission (cf. Bouchard et al. 2011). Stability achieved this way would be effective and everlasting. Creation of this kind of list is time-consuming, requiring that all the relevant literature be checked (Steiner and Kabat 2004), and so is not currently academically rewarding. Taxonomy would benefit greatly if institutional recognition were given for the considerable and lasting value of work on a LAN Part. No Part of the LAN has yet been approved as of writing this paper: the Part on the Species Group Names of Phylum Rotifera (Segers et al. 2012) is still being considered; two others on Family and Genus Group Names in Aphidoidea are at an impasse for submission.

Thus, while the LAN is ultimately more than a modernised version of Sherborn's *Index Animalium*, it also meets the need for creating order among an array of names with very different levels of usage and taxonomic effectiveness (Evenhuis 2016). This is a necessary step in streamlining taxonomic work (Lyal 2016, Page 2016, Penev et al. 2016) and allowing organismal names to efficiently function as handles for all biological information (Pyle 2016).

Acknowledgements

We thank members of the Standing and Ad hoc Committees for input, and the Commission for long discussions of this topic. The Sherborn meeting provided an opportunity to present this overview of the LAN to additional colleagues and explore their perspectives on its implementation; we thank them all for their input.

References

Bouchard P, Bousquet Y, Davies AE, Alonso-Zarazaga MA, Lawrence JF, Lyal CHC, Newton AF, Reid CAM, Schmitt M, Ślipiński SA, Smith ABT (2011) Family-group names in Coleoptera (Insecta). ZooKeys 88: 1–972. doi: 10.3897/zookeys.88.807

Dayrat B (2005) Towards integrative taxonomy. Biological Journal of the Linnean Society 85: 407–415. doi: 10.1111/j.1095-8312.2005.00503.x

Eschmeyer WN (2003) Introduction to the Series Annotated Checklist of Fishes. Annotated Checklist of Fishes 1: 1–5.

Evenhuis NL (2016) Charles Davies Sherborn and the "Indexer's Club". In: Michel E (Ed.) Anchoring Biodiversity Information: From Sherborn to the 21st century and beyond. ZooKeys 550: 13–32. doi: 10.3897/zookeys.550.9697

Ferraris CJ Jr, Eschmeyer WN (2000) Review. International Code of Zoological Nomenclature: Fourth Edition. Copeia 2000(3): 907–908.

Gregory SMS (2010) The two 'editions' of Duméril's Zoologie analytique, and the potential confusion caused by Froriep's translation Analytische Zoologie. Zoological Bibliography 1(1): 6–8.

ICZN (1990) [Reports on general sessions and open meeting of the Commission, University of Maryland, 4–5 July 1990]. Bulletin of Zoological Nomenclature 47(4): 246–253.

ICZN (1999) International Code of Zoological Nomenclature, 4th edition. The International Trust for Zoological Nomenclature, London, 306 pp.

Lyal CHC (2016) Digitising legacy zoological taxonomic literature: Processes, products and using the output. In: Michel E (Ed.) Anchoring Biodiversity Information: From Sherborn to the 21st century and beyond. ZooKeys 550: 189–206. doi: 10.3897/zookeys.550.9702

Krell F-T (2009) ZooBank and the next edition of the Code – challenges and new developments in the 250th year of zoological nomenclature. Aquatic Insects 31(Supplement 1): 269–282. doi: 10.1080/01650420903083393

Page RDM (2016) Surfacing the deep data of taxonomy. In: Michel E (Ed.) Anchoring Biodiversity Information: From Sherborn to the 21st century and beyond. ZooKeys 550: 247–260. doi: 10.3897/zookeys.550.9293

Penev L, Paton A, Nicolson N, Kirk P, Pyle RL, Whitton R, Georgiev T, Barker C, Hopkins C, Robert V, Biserkov J, Stoev P (2016) A common registration-to-publication automated pipeline for nomenclatural acts for higher plants (International Plant Names Index, IPNI), fungi (Index Fungorum, MycoBank) and animals (ZooBank). In: Michel E (Ed.) Anchoring Biodiversity Information: From Sherborn to the 21st century and beyond. ZooKeys 550: 233–246. doi: 10.3897/zookeys.550.9551

Polaszek A (2010) Preface. In: Polaszek A (Ed.) Systema Naturae 250 – The Linnaean Ark. CRC Press, London, 7–12. doi: 10.1201/EBK1420095012

Polaszek A, Michel E (2010) Linnaeus-Sherborn-ZooBank. In: Polaszek A (Ed.) Systema Naturae 250 – The Linnaean Ark. CRC Press, London, 163–172. doi: 10.1201/EBK1420095012-c15

Pyle RL (2016) Towards a Global Names Architecture: The future of indexing scientific names. In: Michel E (Ed.) Anchoring Biodiversity Information: From Sherborn to the 21st century and beyond. ZooKeys 550: 261–281. doi: 10.3897/zookeys.550.10009

Savage JM (1990) Meetings of the International Commission on Zoological Nomenclature. Systematic Zoology 39: 424–425. doi: 10.1093/syszoo/39.4.424

Scoble MJ (2004) Unitary or unified taxonomy? Philosophical Transactions of the Royal Society of London (B) 359: 699–710. doi: 10.1098/rstb.2003.1456

Segers H, de Smet WH, Fischer C, Fontaneto D, Michaloudis E, Wallace RL, Jersabek CD (2012) Towards a List of Available Names in Zoology, partim Phylum Rotifera. Zootaxa 3179: 61–68.

Sneath PHA (2003) A short history of the Bacteriological Code. http://www.the-icsp.org/misc/Code_history.htm [viewed 30 August 2014]

Steiner G, Kabat AR (2004) Catalog of species-group names of Recent and fossil Scaphopoda (Mollusca). Zoosystema 26(4): 549–726.

ZooKeys 550: 233–246 (2016)
doi: 10.3897/zookeys.550.9551
http://zookeys.pensoft.net

RESEARCH ARTICLE

A peer-reviewed open-access journal
ZooKeys
Launched to accelerate biodiversity research

A common registration-to-publication automated pipeline for nomenclatural acts for higher plants (International Plant Names Index, IPNI), fungi (Index Fungorum, MycoBank) and animals (ZooBank)

Lyubomir Penev[1,2], Alan Paton[4], Nicola Nicolson[4], Paul Kirk[4], Richard L. Pyle[6], Robert Whitton[6], Teodor Georgiev[2], Christine Barker[4], Christopher Hopkins[4], Vincent Robert[5], Jordan Biserkov[2], Pavel Stoev[2,3]

1 *Institute for Biodiversity and Ecosystem Research, Bulgarian Academy of Sciences, Sofia, Bulgaria* **2** *Pensoft Publishers, Sofia, Bulgaria* **3** *National Museum of Natural History, Sofia, Bulgaria* **4** *International Plant Name Index (IPNI) and Index Fungorum (IF), Royal Botanic Gardens Kew, UK* **5** *Mycobank, CBS Fungal Biodiversity Centre, Utrecht, The Netherlands* **6** *ZooBank, Bishop Museum, Honolulu, USA*

Corresponding author: *Lyubomir Penev* (penev@pensoft.net)

Academic editor: *E. Michel* | Received 10 March 2015 | Accepted 10 March 2015 | Published 7 January 2016

http://zoobank.org/3679C31D-07CF-4D76-999A-E3DDA7E9C932

Citation: Penev L, Paton A, Nicolson N, Kirk P, Pyle RL, Whitton R, Georgiev T, Barker C, Hopkins C, Robert V, Biserkov J, Stoev P (2016) A common registration-to-publication automated pipeline for nomenclatural acts for higher plants (International Plant Names Index, IPNI), fungi (Index Fungorum, MycoBank) and animals (ZooBank). In: Michel E (Ed.) Anchoring Biodiversity Information: From Sherborn to the 21st century and beyond. ZooKeys 550: 233–246. doi: 10.3897/zookeys.550.9551

Abstract

Collaborative effort among four lead indexes of taxon names and nomenclatural acts (International Plant Name Index (IPNI), Index Fungorum, MycoBank and ZooBank) and the journals PhytoKeys, MycoKeys and ZooKeys to create an automated, pre-publication, registration workflow, based on a server-to-server, XML request/response model. The registration model for ZooBank uses the TaxPub schema, which is an extension to the Journal Tag Publishing Suite (JATS) of the National Library of Medicine (NLM). The indexing or registration model of IPNI and Index Fungorum will use the Taxonomic Concept Transfer Schema (TCS) as a basic standard for the workflow. Other journals and publishers who intend to implement automated, pre-publication, registration of taxon names and nomenclatural acts can also use the open sample XML formats and links to schemas and relevant information published in the paper.

Keywords
Taxon names, nomenclatural acts, pre-publication registration, International Plant Name Index (IPNI), Index Fungorum, Mycobank, ZooBank

Introduction

The process of indexing nomenclatural acts from published literature has a long tradition, in some cases dating as far back as the middle of 19th century for different taxonomic groups. As a result there are several nomenclatural indexes that aim to be comprehensive for their focal taxa, for example, Index Kewensis in botany, Index Fungorum or MycoBank in mycology, and Zoological Record and *Index Animalium* in zoology. Sherborn's effort in *Index Animalium* surely stands as the giant among these efforts due to the sheer scale of described animal diversity (Evenhuis 2016; Miller 2016; Taylor 2016; Dickinson 2016; Pilsk et al. 2016; Welter-Schultes et al. 2016). Taxonomists use these indexes to trace nomenclatural acts through the literature and to help ensure they have considered relevant published works. They can also, along with a far broader audience, use them as an authoritative source of information including the correct spelling and authorship of the name. This improves information retrieval and fidelity (Lyal 2016; Page 2016; Pyle 2016). Increasingly these indexes are being used as the basis for a Global Names Architecture (www.globalnames.org, Patterson et al. 2010). Names from these indexes are either used directly in, or could be linked to, almost all information resources which contain information about organisms. Names also form the foundation for taxonomic concepts, which could be considered as a set of related names. Thus, these indexes occupy a vital role in connecting the occurrence of names in literature and diverse information systems with differing taxonomic concepts (Patterson et al. 2010).

Historically, these indexes have been compiled by a team of editors scanning the relevant literature. This is an inefficient process. The lists become outdated even as they are being produced, because newly described taxa are continually being added to the list. However, the increase in electronic publication of nomenclatural acts made possible by recent changes to the nomenclatural codes in zoology (ICZN 2012) and botany (Knapp et al. 2011, ICNafp 2012) provides an opportunity to reduce the time spent by index editors on keying information already processed by the author and publisher. By reducing the scanning times and manual compilation, index editors would be freed to spend more time ensuring the data quality of the index, facilitating greater linkages with other information resources. Therefore, the indexes should switch focus from *post-publication indexing* to *pre-publication registration*.

Electronic registration of nomenclatural acts in trusted online registries would have the advantage of ensuring nomenclatural novelties published according to the relevant code would be broadly disseminated and available for linkage into other systems. Registration needs to be developed in accordance with the revisions of the biological codes of nomenclature to make the most efficient use of developing web technologies. Mandatory registration would ensure that all new nomenclatural acts governed under the code were captured and treated consistently (Polaszek et al. 2005). However, if such a system were to be open, and work with broadly agreed standards, it could also open up the indexing and registration process to a broader range of actors, thus improving the scope and speed of data capture within the indexes, the linkage between indexes, and facilitating the creation of new indexes or registries which could complement or cover gaps between existing resources.

This paper deals with a specific and important part of the registration process, namely a common model for an automated, prior to publication, machine-to-machine, XML-based registration and associated workflow between publishers and indexes who could act as registries in further streamlining the process of registration and making it cost efficient.

Current status

There are several ways as to how registration (or indexing if registration is not yet mandated by the relevant code) can be best implemented. Different options and the relationship to the publication process have been extensively reviewed by Pyle and Michel (2008) and Morris et al. (2011). The concept of an automated registration model was first presented by several of the authors of this article at the Sherborn meeting in London in October 2011 and at the Biosystematics 2013 Conference in Vienna, in February 2013, with an expansion at the Digital Nomenclature Workshop in London in January 2015.

Despite the visible progress in recent years, four major questions remain to be answered:

1. When exactly should the registration of a nomenclatural act take place – before or after publication?
2. Who should be responsible for the registration of the act – authors, registry curators or publishers?
3. How is registration actually effected?
4. Who validates the accuracy of the bibliographic metadata for any registered act?

The International Botanical Congress in Melbourne in July 2011 had a major impact on streamlining the process by amending the International Code of Nomenclature for algae, fungi, and plants (ICNafp) such that, from 1 January 2013 to be validly published all new names of fungi must be registered before publication and identifiers for each name included in the publication (Miller et al. 2011, Knapp et al. 2011, McNeill et al. 2012, see also ICNafp 2012).

Shortly thereafter, The International Commission on Zoological Nomenclature voted in favour of a revised version of the amendment to the International Code of Zoological Nomenclature that was first proposed in 2008. The purpose of the amendment is to expand and refine the methods of publication allowed by the Code, particularly in relation to electronic publication. The amendment establishes an Official Register of Zoological Nomenclature (with ZooBank as its online version), allows electronic publication after 2011 under certain conditions, and disallows publication on optical discs after 2012. The requirements for electronic publications are that the work be registered in ZooBank before it is published, that the work itself states the date of publication and contains evidence that registration has occurred, and that the ZooBank registration states both the name of an electronic archive intended to preserve the work and the ISSN or ISBN associated with the work. Registration of new scientific names and nomenclatural acts is not required. The Commission confirmed that ZooBank was ready to handle the requirements of the amendment [International Commission on Zoological Nomenclature (ICZN) 2012].

The current situation with indexing and registration in the three domains of eukaryotic organisms can be summarized as follows:

FUNGI

- Post-publication Indexing in Index Fungorum (IF) and MycoBank (MB)
- Pre-publication registration mandatory for fungi since 1st of January 2013
- Record identifiers must be published in the protologue
- Three official registries are approved: MycoBank, Index Fungorum, Fungal Names

PLANTS

- Post-publication indexing is a well-established practice of the International Plant Names Index (IPNI) which covers seed plants, ferns and lycophytes but not bryophytes or algae
- Pre-publication indexing and inclusion of IPNI record identifiers in protologues piloted with Phytokeys, PLoS ONE and Kew Bulletin

ANIMALS

- Post-publication indexing is a well-established practice of Zoological Record (now published by Thomson Reuters)
- Pre-publication registration in ZooBank mandatory since 1st of January 2012 for e-only publications
- Record identifiers should be published in the original description

Registration of many new nomenclatural acts might be a tedious and extremely time-consuming process if done “by hand”, especially in the recently introduced but increasingly submitted "turbo-taxonomic" papers, combining molecular data, concise morphological descriptions and digital imaging (Butcher et al. 2012, Riedel et al. 2013). The numbers of new taxa described in such papers may count in hundreds, for example 178 new species of parasitic wasps (Butcher et al. 2012) and 101 new species of *Trigonopterus* weevils (Riedel et al. 2013). The ultimate record is held by the paper of Marsh et al. (2013) describing 277 new braconid wasps from Costa Rica. This paper is remarkable also because it became the first "turbo-taxonomic" paper where all 277 new species were registered in Zoobank automatically in just a few seconds, saving a great deal of time to the authors, publisher and the registry.

Which nomenclatural acts are subject of registration?

There are significant differences in the scope and number of nomenclatural acts that are tracked by the current indexes and registries (Table 1). In several cases, acts are treated differently by the biological codes. For example, new suprafamilial names and new combinations are governed by the ICNafp, but not by ICZN.

Table 1. Nomenclatural acts that are recorded by the indexing services and could potentially be a subject of pre-publication registration in botany, mycology and zoology.

Taxonomic / nomenclatural act	IPNI (botany: vascular plants)	Index Fungorum (mycology)	MycoBank (mycology)	ZooBank (zoology)
New taxon:				
- suprafamilial	-	+	+	
- familial	+	+	+	+
- infrafamilial	+	+	+	+
- generic	+	+	+	+
- infrageneric	+	+	+	+
- specific	+	+	+	+
- infraspecific	+	+	+	+
- hybrids[1]	+	+	+	n/a
New replacement name	+	+	+	
New combination	+	+	+	
Tautonym[2]	+	+	+	n/a
Typifications[3]				
- holotype	+	+	+	
- lectotype	-	+	+	
- neotype	-	+	+	
- epitype	-	+	+	n/a

1 Hybrids need not be treated as a category of new taxon, but there needs to be a mechanism of flagging the ranks above as hybrids where necessary

2 Tautonyms are not validly published in ICNafp. IPNI (and IF) record tautonyms if published, but such cases should be picked up at the indexing stage.

3 IPNI normally does not record new lectotypifications, but does include typifications of new taxa at generic rank and below.

The registration workflow

In our view, the registration (or indexing in groups where registration is not yet mandated by the code) of nomenclatural acts and the quality control of the bibliographic metadata in these registries should be a primary responsibility of publishers and registry curators and, to a lesser extent, of authors. Registration of a nomenclatural act could be initiated by an author, at the pre-submission or pre-acceptance for publication stage. However, we prefer the publisher-initiated model as it avoids registry curators curating data that may never be published according to the rules of the relevant code. Such a practice may lead to "over-saturation" of the registries with names that are not validly published, causing confusion. Focusing on names accepted for publication also allows these curators more time to focus on the published act and this may allow these specialist staff to assist publication by identifying inconsistencies with the relevant code. Moreover, the publishers' role is essential in checking and correcting the pre-publication registration details against the ultimately published information. The model presented below could easily be adapted for author initiation, though we envisage that there would be a greater curatorial overhead and a greater likelihood of errors being created. However, we accept that the model needs to be flexible and allow alternatives if it is to receive community support.

In the "journal-centric" model, the registration of taxonomic and nomenclatural acts involves two main classes of actors: (1) publishers, or editors, and (2) registry curators. The publisher takes the responsibility for initiating the registration of nomenclatural acts so that the workflow can be performed following a common stepwise model (see also Fig. 1):

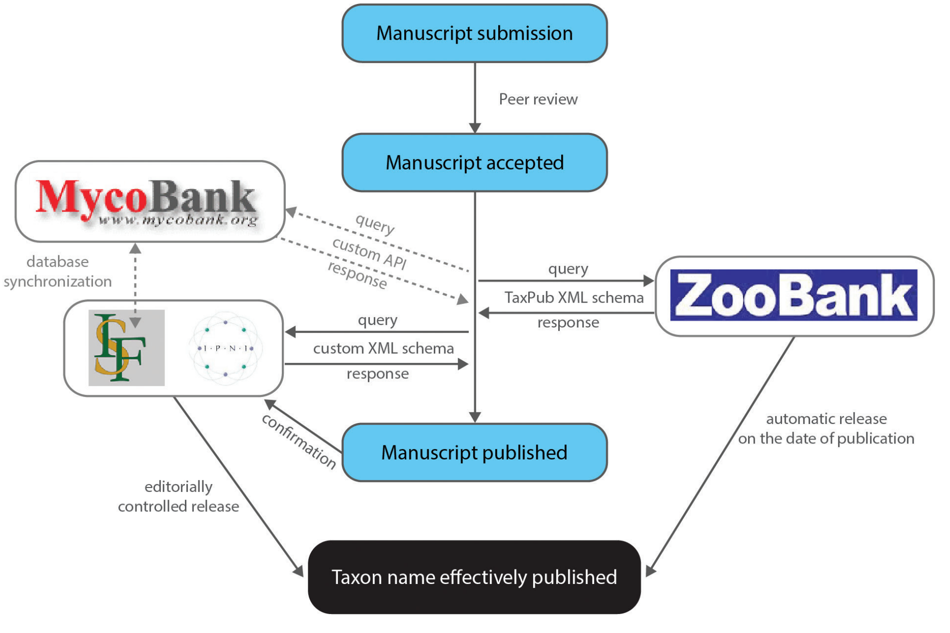


Figure 1. Automated registration process and validation of finally published data and metadata between publisher and registry. Abbreviation on logos: IPNI - International Plant Name Index, IF - Index Fungorum.

Step 1. XML message from the publisher to the registry on acceptance of the manuscript containing the type of act, taxon names, and preliminary bibliographic metadata; the registry will store the data *but not make these publicly available before the final publication date*.

Step 2a. Response XML report containing the unique identifier of the act as supplied by the registry and/or any relevant error messages.

Step 2b. Error correction and de-duplication performed manually: human intervention, at either registry's or publisher's side (or at both).

Step 3. Inclusion of registry supplied identifiers in the published treatments (protologues, nomenclatural acts).

Step 4. Making the information in the registry publicly accessible upon publication, providing a link from the registry record to the article.

The registration process should be as automated as possible. There are several reasons to maximize automation of registration, the most significant being:

- Increasing cases of bulk, "turbo-taxonomic", descriptions of new taxa within a single paper, sometimes counted in hundreds, which creates significant overhead on the authoring and editorial process.
- Decreased risk of errors caused by human intervention (e.g. re-typing).
- Disambiguation of the dates of acceptance and publication of a manuscript.
- Efficient and accurate validation of final published data and metadata through automated export from the publisher to the registry on the day of publication.

The automated registration process

Within the framework of the EU FP7 project pro-iBiosphere, and in close collaboration with Zoological Record, ZooBank, IPNI, MycoBank and Index Fungorum, as well as with the Global Names project (www.globalnames.org) we are developing a workflow and associated XML formats to streamline the registration of nomenclatural acts within the pre-publication process. The workflow was piloted by IPNI for higher plants and ZooBank for animals and the journals PhytoKeys and ZooKeys, respectively. The formats differ between the two main biological codes, ICNafp and ICZN, hence we describe these separately below.

Automated indexing with the International Plant Name Index (IPNI)

The pre-publication indexing of new plant taxa and nomenclatural acts in IPNI and inclusion of the IPNI identifiers in the protologues was first trialled in the journal PhytoKeys since the publication of its first issue in 2010 (Penev et al. 2010). With the pro-iBiosphere project the workflow has been piloted to include an automated registration module. The pilot project uses a custom XML format illustrated by a new genus *Lettowia* description and new combination *Lettowia nyassae* (Oliv.) H. Rob., comb. nov. in the paper of Robinson and Skvarla (2013) (Appendices 1 and 2). The emphasis of the pilot was to understand the workflow; as this is scaled up to production use with a broader range of partners, IPNI will move to use the Taxon Concept Schema standard to encode the data exchanged. This will enable broader adoption.

The XML query is submitted to IPNI's Application Programming Interface (API) through a POST request and replied back with automatically inserted IPNI identifiers.

Automated registration with Index Fungorum

The registration workflow of Index Fungorum (IF) will adopt that of IPNI after the IF system has moved to Royal Botanic Gardens Kew to run alongside IPNI.

Automated registration with MycoBank

The following methods of the MycoBank API are enough for a straightforward implementation:

1. SearchMycoBankWithFilters
2. InsertUserProfile
3. UpdateUserProfile
4. InsertMycobankRecord
5. UpdateMycobankRecord

Using the combinations (1, 2, 3) and (1, 4, 5) one can implement the Upsert (Update if exists, Insert otherwise) semantics required for the the Common query/response registration model.

As there are multiple fungi registries (MycoBank, Index Fungorum, Fungal Names), another approach would be to perform the registration with only one of them and rely on the synchronization mechanisms (currently being built) to propagate the information to the other databases.

Automated registration with ZooBank

Similarly to the case of PhytoKeys, ZooKeys was the first journal that implemented a mandatory registration of new taxon names in zoology, since the publication of its first issue in 2008 (Penev et al. 2008). The automated registration with ZooBank is based on a slightly different approach than that with IPNI and uses the TaxPub XML schema (Catapano 2010) as a basic standard. Upon acceptance and producing the XML version of the manuscript, we upload it on the ZooBank server through the ZooBank's interface (see Suppl. material 3 for the submitted TaxPub XML format). Then a software tool at ZooBank harvests the TaxPub XML and registers the title, authors and new taxon names. The tool also checks if some or all authors have been previously registered and inserts their current (or newly registered) ZooBank UUIDs. In case in the ZooBank database there are authors with identical names (homonyms), the interface displays these so that the operator at the editorial office could disambiguate the overlapping authors' names by selecting the right one. The whole TaxPub XML is sent back with inserted UUIDs for the article, authors and new names (Suppl. material 4). In case the manuscript XML has been changed after the registration process, it can be uploaded again and the new data will replace the previous ones. At the day of publication, the names and the bibliographic metadata are made publicly available in ZooBank.

What other journal publishers should do to use the workflow?

The registration workflow and XML formats published in this article are free to use for anyone who would like to implement it. To ensure broader adoption of the registration model, the data exchanged through the workflow should be encoded in a standard. For zoology, journals should adopt the TaxPub XML schema (Catapano 2010; open source available at: https://github.com/tcatapano/TaxPub/releases/tag/v0.5-beta) which encodes publications as required by the zoological code. For botany, the registration workflow uses currently a custom XML format based on the Taxon Concept Transfer Schema (TCS) XML schema (http://www.tdwg.org/standards/117/) which encodes names; in the future, IPNI and Index Fungorum will implement TCS as a basic standard for registration/indexing of new names and other nomenclatural acts.

The Suppl. materials 1–4 show some data encoded for the pilot project using a custom XML format – whilst this shows the kind of data that will be exchanged, it should not be used as a template – the TCS and TaxPub standards should be used as reference.

Once the editorial workflow is defined, and structured data can be produced according to these standards, journal editors should contact registries for access to their Application Programming Interfaces (APIs).

The authors of this article, staff at the registries and at Pensoft are available to consult journals who intend to implement the automated registration process. Future changes to the automated registration workflow will be published on the Wiki page of the pro-iBiosphere project at http://wiki.pro-ibiosphere.eu/wiki/Pilot_2.

Acknowledgements

The pro-iBiosphere project (Coordination & Policy Development in Preparation for a European Open Biodiversity Knowledge Management System, Addressing Acquisition, Curation, Synthesis, Interoperability & Dissemination, Contract no. RI-312848, www.pro-ibiosphere.eu) supported Pensoft and Royal Botanical Gardens Kew in developing, testing and implementation of the automated registration workflow. Pensoft has received also financial support by the EU FP7 projects ViBRANT (Virtual Biodiversity Research and Access Network for Taxonomy, www.vbrant.eu, Contract no. RI-261532) for designing the basic concept of the workflow. We are thankful also to Nigel Robinson from Zoological Record for the discussions of the early stages of the process. The work of *ZooBank* team was supported by the Global Names NSF project (DBI-1062441).

References

Butcher B, Smith M, Sharkey M, Quicke D (2012) A turbo-taxonomic study of *Thaialeiodes* (*aleiodes*) and *aleiodes* (*arcaleiodes*) (Hymenoptera: Braconidae: Rogadinae) based largely on COI barcoded specimens, with rapid descriptions of 179 new species. Zootaxa 3457: 1–232.

Catapano T (2010) TaxPub: An extension of the NLM/NCBI Journal Publishing DTD for taxonomic descriptions. Proceedings of the Journal Article Tag Suite Conference 2010. http://www.ncbi.nlm.nih.gov/books/NBK47081/#ref2

Dickinson EC (2016) Reinforcing the foundations of ornithological nomenclature: Filling the gaps in Sherborn's and Richmond's historical legacy of bibliographic exploration. In: Michel E (Ed.) Anchoring Biodiversity Information: From Sherborn to the 21st century and beyond. ZooKeys 550: 107–134. doi: 10.3897/zookeys.550.10170

Evenhuis NL (2016) Charles Davies Sherborn and the "Indexer's Club". In: Michel E (Ed.) Anchoring Biodiversity Information: From Sherborn to the 21st century and beyond. ZooKeys 550: 13–32. doi: 10.3897/zookeys.550.9697

International Commission of Zoological nomenclature (ICZN) (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. doi: 10.3897/zookeys.219.3944

ICNafp (2012) International Code of Nomenclature for algae, fungi, and plants (Melbourne Code). http://www.iapt-taxon.org/nomen/main.php

Knapp S, McNeill J, Turland N (2011) Changes to publication requirements made at the XVIII International Botanical Congress in Melbourne – what does e-publication mean for you? PhytoKeys 6: 5–11. doi: 10.3897/phytokeys.6.1960

Lyal CHC (2016) Digitising legacy zoological taxonomic literature: Processes, products and using the output. In: Michel E (Ed.) Anchoring Biodiversity Information: From Sherborn to the 21st century and beyond. ZooKeys 550: 189–206. doi: 10.3897/zookeys.550.9702

Marsh P, Wild A, Whitfield J (2013) The Doryctinae (Braconidae) of Costa Rica: genera and species of the tribe Heterospilini. ZooKeys 347: 1–474. doi: 10.3897/zookeys.347.6002

Barrie FR, Buck WR, Demoulin V, Greuter W, Hawksworth DL, Herendeen PS, Knapp S, Marhold K, Prado J, Prud'homme Van Reine WF, Smith GF, Wiersema JH (2012) International Code of Nomenclature for algae, fungi, and plants (Melbourne Code), Adopted by the Eighteenth International Botanical Congress Melbourne, Australia, July 2011 (electronic edition). International Association for Plant Taxonomy. http://www.iapt-taxon.org/nomen/main.php

Miller CG (2016) Sherborn's foraminiferal studies and their influence on the collections at the Natural History Museum, London. In: Michel E (Ed.) Anchoring Biodiversity Information: From Sherborn to the 21st century and beyond. ZooKeys 550: 71–81. doi: 10.3897/zookeys.550.9863

Miller JS, Funk VA, Wagner WL, Barrie F, Hoch PC, Herendeen P (2011) Outcomes of the 2011 Botanical Nomenclature Section at the XVIII International Botanical Congress. PhytoKeys 5: 1–3. doi: 10.3897/phytokeys.5.1850

Morffe J, Rodríguez N (2013) *Batwanema* gen. n. and *Chokwenema* gen. n. (Oxyurida, Hystrignathidae), new nematode genera as parasites of Passalidae (Coleoptera) from the Democratic Republic of Congo. ZooKeys 361: 1–13. doi: 10.3897/zookeys.361.6351

Morris PJ, Macklin JA, Croft J, Nicolson N, Whitbread G (2011) Fungal nomenclature 4. Letter of concern regarding Props. (117119) to amend the ICBN to require pre-publication deposit of nomenclatural information. Mycotaxon 116(1): 513–517. doi: 10.5248/116.513

Page RDM (2016) Surfacing the deep data of taxonomy. In: Michel E (Ed.) Anchoring Biodiversity Information: From Sherborn to the 21st century and beyond. ZooKeys 550: 247–260. doi: 10.3897/zookeys.550.9293

Patterson DJ, Cooper J, Kirk PM, Pyle RL, Remsen DP (2010) Names are key to the big new biology. Trends in Ecology and Evolution 25(12): 686–691. doi: 10.1016/j.tree.2010.09.004

Penev L, Erwin T, Thompson FC, Sues H-D, Engel MS, Agosti D, Pyle R, Ivie M, Assmann T, Henry T, Miller J, Ananjeva NB, Casale A, Lourenzo W, Golovatch S, Fagerholm H-P, Taiti S, Alonso-Zarazaga M (2008) ZooKeys, unlocking Earth's incredible biodiversity and building a sustainable bridge into the public domain: From "print-based" to "web-based" taxonomy, systematics, and natural history. ZooKeys Editorial Opening Paper. ZooKeys 1: 1–7. doi: 10.3897/zookeys.1.11

Penev L, Kress WJ, Knapp S, Li DZ, Renner S (2010) Fast, linked, and open – the future of taxonomic publishing for plants: launching the journal PhytoKeys. PhytoKeys 1: 1–14. doi: 10.3897/phytokeys.1.642

Pilsk SC, Kalfatovic MR, Richard JM (2016) Unlocking *Index Animalium*: From paper slips to bytes and bits. In: Michel E (Ed.) Anchoring Biodiversity Information: From Sherborn to the 21st century and beyond. ZooKeys 550: 153–171. doi: 10.3897/zookeys.550.9673

Polaszek A, Alonso-Zarazaga M, Bouchet P, Brothers DJ, Evenhuis N, Krell F-T, Lyal CHC, Minelli A, Pyle RL, Robinson NJ, Thompson FC, van Tol J (2005) ZooBank: the open-access register for zoological taxonomy: Technical Discussion Paper. Bulletin of Zoological Nomenclature 62(4): 210–220.

Pyle RL (2016) Towards a Global Names Architecture: The future of indexing scientific names. In: Michel E (Ed.) Anchoring Biodiversity Information: From Sherborn to the 21st century and beyond. ZooKeys 550: 261–281. doi: 10.3897/zookeys.550.10009

Pyle RL, Michel E (2008) ZooBank: Developing a nomenclatural tool for unifying 250 years of biological information. In: Minelli A, Bonato L, Fusco G (Eds) Updating the Linnaean Heritage: Names as Tools for Thinking about Animals and Plants. Zootaxa 1950: 39–50.

Riedel A, Sagata K, Surbakti S, Tänzler R, Balke M (2013) One hundred and one new species of *Trigonopterus* weevils from New Guinea. ZooKeys 280: 1–150. doi: 10.3897/zookeys.280.3906

Robinson H, Skvarla J (2013) *Lettowia*, a new genus of Vernonieae from East Africa (Asteraceae). PhytoKeys 25: 47–53. doi: 10.3897/phytokeys.25.5556

Taylor MA (2016) '*Where is the damned collection?*' Charles Davies Sherborn's listing of named natural science collections and its successors. In: Michel E (Ed.) Anchoring Biodiversity Information: From Sherborn to the 21st century and beyond. ZooKeys 550: 83–106. doi: 10.3897/zookeys.550.10073

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. In: Michel E (Ed.) Anchoring Biodiversity Information: From Sherborn to the 21st century and beyond. ZooKeys 550: 173–183. doi: 10.3897/zookeys.550.10041

Supplementary material 1

XML response of IPNI

Authors: Lyubomir Penev, Alan Paton, Nicky Nicolson, Paul Kirk, Richard Pyle, Robert Whitton, Teodor Georgiev, Christine Barker, Christopher Hopkins, Vincent Robert, Jordan Biserkov, Pavel Stoev

Data type: (measurement/occurence/multimedia/etc.)

Explanation note: XML query sent from Pensoft to IPNI on the day of acceptance of the manuscript for publication [exemplified with the paper of Robinson and Skvarla (2013)].

Supplementary material 2

IPNI response XML

Authors: Lyubomir Penev, Alan Paton, Nicky Nicolson, Paul Kirk, Richard Pyle, Robert Whitton, Teodor Georgiev, Christine Barker, Christopher Hopkins, Vincent Robert, Jordan Biserkov, Pavel Stoev

Data type: (measurement/occurence/multimedia/etc.)

Explanation note: XML response of IPNI to the query in Suppl. material 1. The response is sent back to Pensoft and contains the registration numbers of the new genus name and the new combination [exemplified with the paper of Robinson and Skvarla (2013)].

Supplementary material 3

XML response of TaxPub

Authors: Lyubomir Penev, Alan Paton, Nicky Nicolson, Paul Kirk, Richard Pyle, Robert Whitton, Teodor Georgiev, Christine Barker, Christopher Hopkins, Vincent Robert, Jordan Biserkov, Pavel Stoev

Data type: (measurement/occurence/multimedia/etc.)

Explanation note: TaxPub XML of a ready-to-publish manuscript submitted from Pensoft to ZooBank [exemplified with the paper of Morffe and Rodríguez (2013)].

Supplementary material 4

TaxPub response XML

Authors: Lyubomir Penev, Alan Paton, Nicky Nicolson, Paul Kirk, Richard Pyle, Robert Whitton, Teodor Georgiev, Christine Barker, Christopher Hopkins, Vincent Robert, Jordan Biserkov, Pavel Stoev

Data type: (measurement/occurence/multimedia/etc.)

Explanation note: TaxPub XML returned from ZooBank to Pensoft containing UUIDs of the article, authors and new taxon names [exemplified with the paper of Morffe and Rodríguez (2013)].

ZooKeys 550: 247–260 (2016)
doi: 10.3897/zookeys.550.9293
http://zookeys.pensoft.net

RESEARCH ARTICLE

A peer-reviewed open-access journal
ZooKeys
Launched to accelerate biodiversity research

Surfacing the deep data of taxonomy

Roderic D. M. Page[1]

1 *Institute of Biodiversity, Animal Health, and Comparative Medicine, College of Medical, Veterinary, and Life Sciences, University of Glasgow, Glasgow G12 8QQ, UK*

Corresponding author: *Roderic D. M. Page* (Roderic.Page@glasgow.ac.uk)

Academic editor: *E. Michel* | Received 28 January 2015 | Accepted 24 April 2015 | Published 7 January 2016

http://zoobank.org/D425561E-1578-43ED-B611-F57AD3DAA2BB

Citation: Page RDM (2016) Surfacing the deep data of taxonomy. In: Michel E (Ed.) Anchoring Biodiversity Information: From Sherborn to the 21st century and beyond. ZooKeys 550: 247–260. doi: 10.3897/zookeys.550.9293

Abstract

Taxonomic databases are perpetuating approaches to citing literature that may have been appropriate before the Internet, often being little more than digitised 5 × 3 index cards. Typically the original taxonomic literature is either not cited, or is represented in the form of a (typically abbreviated) text string. Hence much of the "deep data" of taxonomy, such as the original descriptions, revisions, and nomenclatural actions are largely hidden from all but the most resourceful users. At the same time there are burgeoning efforts to digitise the scientific literature, and much of this newly available content has been assigned globally unique identifiers such as Digital Object Identifiers (DOIs), which are also the identifier of choice for most modern publications. This represents an opportunity for taxonomic databases to engage with digitisation efforts. Mapping the taxonomic literature on to globally unique identifiers can be time consuming, but need be done only once. Furthermore, if we reuse existing identifiers, rather than mint our own, we can start to build the links between the diverse data that are needed to support the kinds of inference which biodiversity informatics aspires to support. Until this practice becomes widespread, the taxonomic literature will remain balkanized, and much of the knowledge that it contains will linger in obscurity.

Keywords

Biodiversity informatics, identifiers, DOI, literature, taxonomy, dark taxa, data cleaning, data integration

Introduction

Bergman (2001) coined the term “deep web” to refer to the part of the web that is largely hidden from search engine crawlers, such as dynamically generated web pages. A major challenge facing web crawlers is how to “surface” that deep web so that it becomes accessible to search engines. By analogy, while much of the scientific literature is readily discoverable, taxonomic literature remains largely obscure.

As an example of the consequences of this obscurity, consider the fate of the name *Leviathan* as used for a recently discovered fossil whale described in *Nature* (Lambert et al. 2010a). Unbeknownst to the authors, the name *Leviathan* was previously used for an extinct mammoth (Koch 1841). Once the homonymy was uncovered, an erratum was published in the same journal (Lambert et al. 2010b). The erratum cites the original publication of *Leviathan* (Koch 1841), but if the reader visits the corresponding page on the journal *Nature*’s website there is no link to the digitised version of this publication, despite it being available in the Biodiversity Heritage Library (http://biodiversitylibrary.org/). The lack of a link is symptomatic of the poor state of digital access to taxonomic literature. Reading the list of literature cited in a modern taxonomic paper online, it is striking that while citations of papers on phylogenetics, ecology, or methodology typically include links directly to that article (for example, using Digital Object Identifiers, DOIs), the citations to taxonomic literature are mostly devoid of such links. In an age when most readers will expect any publications worth reading to be online, the absence of links to the primary taxonomic literature consigns it to a ghetto where only the most determined and well-resourced readers will dare to venture. For many readers the lack of a link means a dead-end in their search for information.

Unless we want the taxonomic literature to linger in obscurity we need to make it easily findable and accessible. An obvious starting point would be if taxonomic databases linked to the digitised taxonomic literature. However, most taxonomic databases are little more than online collections of 5 × 3 index cards, a technology Linnaeus himself pioneered (Müller-Wille and Charmantier 2011). Often databases simply present the user with lists of names, divorced from the associated taxonomic literature (such as the original publication of the name). When literature is cited, it is typically as a text string, lacking either an identifier or a link that the researcher can use to locate the publication. This is not to deny the value of the heroic efforts of indexers such as Charles Davies Sherborn (Cleevely 2009), but it is striking how persistent conventions from the print world remain, despite the Internet removing many of the physical constraints of that medium. For example, the recent publication of the Rotifer List of Available Names (LAN) (Segers et al. 2012) comprises a list of references in abbreviated form (“microcitation”) instead of the full publication details, and the list lacks any bibliographic identifiers. Part of the goal of publishing such a List of Available Names is to enable others to scrutinise it, yet the authors of the list provide virtually no assistance to the reader in locating the corresponding literature.

If we accept that the key documents of taxonomy are the publications that contain the names, descriptions, nomenclatural changes, and taxonomic revisions, then a major challenge is to "surface" these documents so that readers can discover them. This means changing practices that have served the community well in the pre-digital era, but which are now hindering its progress. One of the key changes will be the adoption of globally unique identifiers for the taxonomic literature.

Globally unique identifiers

The taxonomic community's experience with globally unique identifiers has been mixed. Several factors have contributed to this. The first is the saga of Life Science Identifiers (LSIDs) (Martin et al. 2005) which seemed a promising technology for identifying data in biology, but in the end the biodiversity community were the only major adopters. This was compounded by the lack of reusing existing identifiers. Every project employing LSIDs created their own identifiers for their data, and rarely, if ever, used LSIDs from other projects. For example, both the Index of Organism Names (ION, http://www.organismnames.com/) and ZooBank (http://zoobank.org/) have records for the genus name *Tyrannobdella*, each with their own LSID (urn:lsid:organismnames.com:name:4439403 and urn:lsid:zoobank.org:act:43D55B49-C888-4D6B-AF6F-61238EC1339B, respectively). Neither database acknowledges the existence of the other by using the other's identifier. Furthermore, neither ION nor ZooBank use the most obvious identifier for the PLoS One paper that published *Tyrannobdella* (Phillips et al. 2010), namely the DOI: 10.1371/journal.pone.0010057. ION represents the reference as a text string:

"Tyrannobdella rex n. gen. n. sp. and the evolutionary origins of mucosal leech infestations. PLoS ONE, 5(4) 2010: e1057, 1–8."

ZooBank mints its own identifier for the PLoS One paper: urn:lsid:zoobank.org:pub:8D431ED1-B837-4781-A591-D3886285283A (since this was written ZooBank has added the DOI for this article). Ironically, the only thing that links these two records together is the taxonomic name "*Tyrannobdella*".

A consequence of the failure to reuse existing identifiers is that the biodiversity informatics community has created a large amount of data identified by a technology few people understand (LSIDs, which by default wouldn't work in a web browser) and with very few cross-links. This lack of links means each database is effectively another silo, and hence many of the expected benefits of serving biodiversity data in RDF (Page 2006) have not materialised.

This experience may encourage a healthy scepticism about the utility of identifiers, but I would argue that this is because we've overlooked the importance of their reuse. If different databases insist on minting their own identifiers and not using (or linking to) existing identifiers, then our data will remain in silos. Reusing identifiers will help establish links between databases, and it is these links that will be the basis of many of the hoped-for inferences we can make in biodiversity informatics (Page 2008).

Bibliographic identifiers

Taxonomic databases often contain names devoid of references to the literature. Names by themselves are of little value; it is the literature, specimens, and data derived from those specimens that are the primary data of taxonomy. Yet much of this information remains hard to obtain (even discovering that it exists can be challenging). Many citations to the taxonomic literature are obscure unless you are familiar with the conventions. For example, if you are searching for the original publication of the name *Tachyglossus* Illiger, 1811 (a genus of spiny anteaters) then *Nomenclator Zoologicus* (Neave 1939; Remsen et al. 2006) gives this as "Prod., 114." I suspect that most readers will find this less than helpful. The citation refers to page 114 of "Caroli Illigeri D. Acad. Reg. Scient. Berolinens. et Bavaricae Sod. Museo Zoologico Berolin. praefecti professoris extraord. Prodromus systematis mammalium et avium : additis terminis zoographicis utriusque classis, eorumque versione germanica." Given the length of the title of Illiger's work, one can see the desirability of abbreviating it for a printed list such as *Nomenclator Zoologicus*. But there are many ways to abbreviate a citation, which can result in a plethora of ways the same publication is cited in different databases (sometimes even within the same database).

One approach to tackling the plethora of ambiguous, if not downright obscure, citations is to use globally unique identifiers to refer to the publications. In the case of the "Prodromus systematis mammalium et avium" (Illiger 1811), this publication has recently acquired a DOI (10.5962/bhl.title.42403) assigned by the Biodiversity Heritage Library. DOIs are widely used in the publishing industry to identify articles (such as this the one you are currently reading), and are increasingly being used as identifiers for other digital objects, such as data sets (e.g., the DataCite project http://datacite.org/). By providing unique, stable identifiers for articles, the publishing industry has simplified the task converting lists of literature cited into clickable links. DOIs have been in use to identify the scientific literature for over a decade, but taxonomic databases have been slow to adopt these identifiers.

The utility of identifiers

Using existing bibliographic identifiers has several immediate advantages. It all but eliminates ambiguity in citations. Given that the same citation can be represented multiple ways (consider the bewildering and completely unnecessary proliferation of citation styles for different journals), matching citations using their representation as strings of characters is fraught with problems. Citation strings can also "mutate" over time (Specht 2010) and these mutations can propagate by "copy and paste" citation (Simkin and Roychowdhury 2011). Consistent use of globally unique identifiers mitigates this problem.

Identifiers provide additional value if they come with supporting services. For example, DOIs can be resolved to both human- and machine-readable content, which enables tools to be built that can consume DOIs and automatically populate databases with bibliographic information (most bibliographic management software makes use of these services). There are also services that take a bibliographic citation and find the corresponding DOI; publishers utilise these to add links to the list of literature cited in an article.

But the real value from identifiers becomes apparent when they are shared, that is, when different databases use the same identifiers for the same entities, instead of minting their own. Reusing identifiers can enable unexpected connections between databases. For example, the PubMed biomedical literature database has a record (PMID:948206) for the paper "Monograph on "*Lithoglyphopsis*" *aperta*, the snail host of Mekong River Schistosomiasis" (Davis et al. 1976). The PubMed record contains the abstract for the paper, but no link to where the user can obtain a copy of the paper. Actually, this reference is in a volume scanned by BHL, and has been extracted by BioStor (http://biostor.org/reference/102054). If PubMed was linked to BHL, users of PubMed could go straight to the content of the article. But this is just the start. The Davis et al. (1976) paper also mentions museum specimens in the collection of the Academy of Natural Sciences of Drexel University, Philadelphia. Metadata for these specimens has been aggregated by GBIF, and the BioStor page for this article displays those links. In an ideal world we should be able to go from PubMed to BioStor to GBIF. But in many ways the real power will come from traversing these links in the other direction. At present, a user of GBIF simply sees metadata for these specimens and a locality map. They are unaware that these specimens have been cited in a paper (Davis et al. 1976) which shows that the snails host the Mekong River schistosome. This connection would be trivial to make if the reciprocal link was made from GBIF to BioStor. Furthermore, the link from BioStor to PubMed would give us access to Medical Subject Headings (MeSH http://www.nlm.nih.gov/mesh/) for the paper. Hence we could imagine ultimately searching GBIF using queries from a controlled vocabulary of biomedical terms.

Making these connections requires not only that we have digital identifiers, but also that wherever possible we reuse existing identifiers. If we restrict ourselves to project-specific identifiers then we stymie attempts to create a network of connected data on biodiversity.

It is worth exploring ways we can reuse identifiers. One approach is to include links to existing identifiers wherever possible. For example, if a database includes an article that has a DOI, then that database should store the DOI as one of its fields. This is the easiest form of reuse, and doesn't prevent the database minting its own identifiers. This approach makes sense if we are adding data that hasn't yet been linked to existing identifiers, or if identifiers may only become available later (e.g., after a database entry has been created, a publisher subsequently digitises the print archive of a journal and issues DOIs for each article). A more powerful example of reuse is when a database incorporates existing identifiers into its own identifiers. The BBC is an excellent example of this: their music and nature sites reuse "slugs" from external resources, such as MusicBrainz and Wikipedia, respectively (Raimond et al. 2010). The "slug" is the part of the URL after the domain name (and any site-specific details). Hence, given that the URL for the Wikipedia page for the Komodo dragon (*Varanus komodoensis*) is http://en.wikipedia.org/wiki/Komodo_dragon, the BBC reuse the slug "Komodo_dragon" to create the URL http://www.bbc.co.uk/nature/life/Komodo_dragon. Similarly, instead of minting a completely new identifier for a journal, we can make use of the journal's ISSN to create a URL (e.g., http://bionames.org/issn/1313-2989). Reusing identifiers in this manner makes it easier to find equivalent entries in different databases (Raimond et al. 2010).

Identifiers and community

> "This may not be much of a revelation to many, but is a notion that is sinking home more deeply for me of late. By "Community", I don't necessarily mean the online community … I mean the taxonomic community." David Shorthouse "The community is dead" http://ispiders.blogspot.co.uk/2009/06/community-is-dead.html

There are many reasons why communities may or may not form, but arguably a community that shares an interest in a given topic benefits from having a standard way to refer to the things they care about. The increasing adoption of standard bibliographic identifiers such as DOIs makes it easier to build social bookmarking tools around the scientific literature (such as CiteULike http://www.citeulike.org/ and Mendeley http://www.mendeley.com/) because it becomes easier to determine how many members of the network have bookmarked the same paper.

Taxonomic communities are likely to be small and taxon-focussed. But this does not mean that these are the only communities that taxonomists can engage with, or that people outside the taxonomic community will not share the interests of those working on a particular taxon. Using bibliographic identifiers we can discover networks of people interested in particular topics that may intersect with taxonomists (obvious examples are people interested in ecology, conservation and evolutionary biology). By making publications the unit of sharing, companies such as Mendeley have grasped perhaps better than most that the connection between researchers is often not a direct social link, but rather shared interest in the same publication (formalised by patterns of citation and co-citation). For this reason, I suspect that attempts to build communities around taxa (Harman et al. 2009) may be ultimately less successful than embedding the taxonomic literature in the growing social networks assembling around scientific publications.

Identifiers and impact

The taxonomic community has long felt disadvantaged by the role of citation-based "impact factor" in assessing the importance of taxonomic research (Garfield 2001;

Krell 2000; Werner 2006) especially as much of the taxonomic literature appears in relatively low impact journals. A common proposal is to include citations to the taxonomic authority for every name mentioned in a scientific paper (Wägele et al. 2011). Regardless of the merits of this idea, the difficulty of locating bibliographic details for much of the taxonomic literature, coupled with the lack of identifiers such as DOIs means such proposals will be hard to implement, and likely to merely populate the literature cited section of papers with even more bibliographic dead ends.

At the same time, the concern about impact may help motivate the use of identifiers such as DOIs. There is a growing "altmetrics" movement (http://altmetrics.org/manifesto/) that aims to provide metrics for the post-publication impact of a publication in terms of activity such as social bookmarking, and commentary on web sites (Yan and Gerstein 2011). Gathering these metrics is greatly facilitated by using standard bibliographic identifiers (otherwise, how do we know whether two commentators are discussing the same article or not?). If taxonomic literature is be part of this burgeoning conversation it needs to be able to be identified unambiguously.

Making the taxonomic literature findable

The first step towards improving the current generation of taxonomic databases would be to associate the taxonomic literature with existing digital identifiers, such as DOIs. Admittedly, this will not always be straightforward. Although DOIs are the bibliographic identifier of choice, and CrossRef provides tools for locating an existing DOI for a reference, it is not always straightforward to find a DOI for a publication. Part of the difficulty in citing the older literature is that many of the conventions we take for granted in modern scientific articles are lacking. Modern articles have titles, and are published in journals that usually have an unambiguous name, volume number, and pagination. This triplet is usually unique, and makes it relatively easy to locate an article in a bibliographic database (Page 2009). However, these conventions need not apply to older publications. For example, (Bennett and Jarvis 2004) cite the following paper:

Ogilby W (1838) On a collection of Mammalia procured by Captain Alexander during his journey into the country of the Damaras. Proceedings of the Zoological Society of London 1838:5–15.

This journal has been digitised by both Wiley and BHL. Wiley makes pages 5-15 available as an article with the doi: 10.1111/j.1096-3642.1838.tb01402.x and attributes the authorship to Richard Owen, not W. Ogilby. On inspection we see that pages 5–15 comprise two articles, one by Ogilby and one by Owen. The first paragraph of page 5 contains the text:

"A selection of the Mammalia procured by Captain Alexander during his recent journey into the country of the Damaras, on the South West Coast of Africa, was exhibited, and Mr. Ogilby directed the attention of the Society to the new and rare species which it contained."

Subsequent authors have transformed this sentence into the article title "On a collection of Mammalia procured by Captain Alexander during his journey into the country of the Damaras". Note also that in this case, there is a mismatch between the granularity at which taxonomists cite the literature and the granularity at which Wiley has assigned the identifier (the DOI corresponds to two articles). Perhaps the most obvious example of this mismatch is exemplified by the BHL, which typically recognises units at the scale of journal volume, or individual pages, but not at article level (Page 2011a).

Discovering existing identifiers for the taxonomic literature will sometimes be difficult, for a multitude of reasons. For example, taxonomic databases often store an abbreviated (or even corrupted) version of the citation, the citation may be translated from its original language, or the journal may have been renamed and the new name applied retrospectively to older issues (Page 2011c). All of this makes creating the mapping tedious, but this mapping need only be done once.

Kinds of identifiers

While DOIs are the best-known bibliographic identifier, there are several others that are relevant to the taxonomic literature (Page 2009). DOIs are themselves based on Handles (http://hdl.handle.net) an identifier widely used by digital repositories such as DSpace (http://www.dspace.org/). A number of journals, such as the *Bulletins* and *Novitates of the American Museum of Natural History* are available in DSpace repositories and consequently have Handles. Other major archives such as JSTOR (http://www.jstor.org/) and CiNii (http://ci.nii.ac.jp/) have their own unique identifiers (typically integer numbers that are part of a URL). Having a variety of identifiers complicates the task of finding existing identifiers for a particular publication. Whereas for some identifiers, such as DOIs and CiNii NAIDs, (National Institute of Informatics Article IDs) there are OpenURL resolvers for this task, for other identifiers there may be no obvious way to find the identifier other than by using a search engine.

Identifiers also exist for aggregations of publications, such as journals. The practice of abbreviating journal titles has led to a plethora of ways to refer to the same journal. For example, the BioStor database (Page 2011a) has the following entries for the *Bulletin of Zoological Nomenclature*:

Bulletin of Zoological Nomenclature
The Bulletin of Zoological Nomenclature
Bull. Zool. Nom.
Bull.zool. Nom.
Bull. Zool. Nom

Bull, Zool. Nom.
Bull Zool. Nom.
Bull. Zool.nom.
Bull. Zool Nom.
Bull., Zool. Nom.
Bull. Zool. . Nom.
Bulletin Zoological Nomenclature
Bull Zoological Nomenclature
Bull Zool Nomen
Bull. Zool. Nomencl
Bull Zool Nom.
Bulletin of Zoological Nomeclature
Bulletin Zool. Nom.
Bull. Zool. Nomencl.

This practice of abbreviating journal names (motivated by the desire to conserve space on the printed page) complicates efforts to match citations to identifiers. One approach to tackling this problem is to map abbreviations to journal-level globally unique identifiers, such as International Standard Serial Numbers (ISSNs) (for the *Bulletin of Zoological Nomenclature* the ISSN is 0007-5167). In addition to reducing ambiguity, there are web services that take ISSNs and return the history of name changes for a journal, which in turn can help clarify the (often complicated) history of long-lived journals.

How much taxonomic literature has been digitised?

To assess the extent of taxonomic digitisation I harvested the metadata associated with the LSID for each record in the ION database. This database records names published under the International Code of Zoological Nomenclature. Over 4 million records have been harvested and imported into BioNames (http://bionames.org) (Page 2013), over a million of which have an associated bibliographic citation. In order to locate identifiers for these citations I attempted to parse each one into its constituent components (e.g., title, journal, volume, pagination) and used OpenURL resolvers to find the corresponding record in databases such as CrossRef and BioStor. To complement this approach I have harvested metadata for some 300,000 journal articles and stored these in Mendeley, then used approximate string matching to compare these to records in ION. This work is on-going, current results can be seen at http://bionames.org/dashboard. To date BioNames has over 60,000 articles with DOIs that publish new names, and if we consider all potential bibliographic identifiers (DOIs, Handles, PubMed, URLs, PDFs) then approximately 20% of all ICZN names are linked to publications that have a digital presence.

Access to the literature

Of course, having the literature digitised is not the same as having ready access to it. Numerous parties are undertaking digitisation efforts, and the results are being made available under a wide range of conditions. Some output is available under explicitly open access licenses (MacCallum 2007), such as content from BHL and the journals published by Pensoft and the Public Library of Science. Some publishers, notably Taylor and Francis, and Wiley are digitising back catalogues of journals and making them available to subscribers. Archives such as JSTOR and CiNii have a mixture of free and subscription-based content. Many smaller journals, often published by scientific societies are providing their contently for free online, if not explicitly under an open license. Note that it is something of a misconception that the bulk of BHL's content is pre-1923. In fact, for several key taxonomic journals its coverage extends into the 21st century, in places overlapping with content made available by the original publishers.

Discussion

As a final motivation to surface deep taxonomic data, consider the rise of "dark taxa" in genomics databases (Page 2011b). A growing percentage of "taxa" in GenBank lack a formal scientific name; in 2010 dark taxa comprised over 80% of invertebrate taxa added that year (Parr et al. 2011). Many of the most recent dark taxa are a product of DNA barcoding projects, and at the time of writing these sequences have been "suppressed" by GenBank, that is, they are still in the database but do not feature in search results. But there is still a background trend towards increasing numbers of unidentified sequences in GenBank. A significant challenge will be determining whether these dark taxa represent newly discovered taxa, or come from known taxa but have not been identified as such (Hibbett and Glotzer 2011; Nagy et al. 2011).

It is clear that some dark taxa do, in fact, have names. For example, consider the frog "*Gephyromantis* aff. *blanci* MV-2005" (NCBI tax_id 321743), which has a single sequence AY848308 associated with it. This sequence was published as part of a DNA barcoding study (Vences et al. 2005). If we enter the accession number AY848308 into Google we find two documents, one the supplementary table for (Vences et al. 2005), the other the a subsequent paper by (Vences and Riva 2007) that describes the frog with this sequence as a new species, *Gephyromantis runewsweeki*. This is a relatively straightforward example, and the taxonomic description is freely available online. But it still required significant time to track down the species description for this one example.

A key question facing attempts to find names for dark taxa is whether the methods available can be scaled to handle the magnitude of the problem. One could argue that newer technologies such as DNA barcoding make classical taxonomy less relevant, and perhaps the effort in digitising older literature and exposing the taxonomic names it contains is misplaced. A counter argument would be that the taxonomic literature potentially contains a wealth of information on ecology, morphology and behaviour, often for taxa in areas that have been subsequently altered by human activity. Furthermore, as technologies such as barcoding uncover previously overlooked variation, older taxonomic names previously sunk in synonymy may yet become relevant. For example, several taxa have been synonymised with the silvery mole-rat *Heliophobius argenteocinereus* Peters, 1846 (Peters 1846) but DNA sequence data has revealed several clades within that species (Faulkes et al. 2011). Consequently, rather than coin new names for these clades we can rescue older names from synonymy. Hence DNA barcoding may give a new lease of life to old names.

Names may have a special place in the hearts of taxonomists (Patterson et al. 2010) but the pace of biodiversity discovery is outstripping our ability to put names on taxa, as evidenced by the rise of dark taxa in GenBank. There are increasing calls to adopt less formal taxonomic naming schemes (Schindel and Miller 2010), or to focus on describing biodiversity without necessarily naming it (Deans et al. 2012; Maddison et al. 2011). Underpinning much of this call to "ramp up" the rate of biodiversity description will be identifiers, assigned to the entities that taxonomy deals with, including specimens, genotypes, phenotypes, publications, and, yes, taxonomic names. As I have argued previously (Page 2008), in many ways taxonomists have been doing this already but without using web-friendly identifiers. Examples include lists of collection acronyms (Leviton et al. 1985) and author names. The issue now is how do we scale these activities to accommodate the deluge of data we are accumulating as we digitise life and our efforts to document it?

Acknowledgements

I thank Ellinor Michel for the invitation to speak at the Sherborn meeting, and for her patience as I eventually got around to writing the promised manuscript. I thank the reviewers, Donat Agosti, Ken Johnson, and Rich Pyle for their constructive comments on the manuscript.

References

Bennett NC, Jarvis JUM (2004) *Cryptomys damarensis*. Mammalian Species 756(1): 1–5. doi: 10.1644/756

Bergman MK (2001) WHITE PAPER: The Deep Web: Surfacing Hidden Value. Journal of Electronic Publishing 7(1). doi: 10.3998/3336451.0007.104

Cleevely RJ (2009) Charles Davies SHERBORN: geologist and scientific bibliographer. Archives of Natural History 36: 357–359. doi: 10.3366/e0260954109001089

Davis GM, Kitikoon V, Temcharoen P (1976) Monograph on "*Lithoglyphopsis*" *aperta*, the snail host of Mekong River schistosomiasis. Malacologia 15(2): 241–87.

Deans AR, Yoder MJ, Balhoff JP (2012) Time to change how we describe biodiversity. Trends in ecology & evolution 27(2): 78–84. doi: 10.1016/j.tree.2011.11.007

Faulkes CG, Bennett NC, Cotterill FPD, Stanley W, Mgode GF, Verheyen E (2011) Phylogeography and cryptic diversity of the solitary-dwelling silvery mole-rat, genus *Heliophobius* (family: Bathyergidae). In: Kitchener A (Ed.) Journal of Zoology 285(4): 324–338. doi: 10.1111/j.1469-7998.2011.00863.x

Garfield E (2001) Taxonomy is small, but it has its citation classics. Nature 413(6852): 107. doi: 10.1038/35093267

Harman K, Scott B, Rycroft S, Roberts D, Smith V (2009) Scratchpads: a data-publishing framework to build, share and manage information on the diversity of life. BMC Bioinformatics 10(Suppl 14): 6. doi: 10.1186/1471-2105-10-S14-S6

Hibbett D, Glotzer D (2011) Where are all the undocumented fungal species? A study of *Mortierella* demonstrates the need for sequence-based classification. The New phytologist 191(3): 592–596. doi: 10.1111/j.1469-8137.2011.03819.x

Illiger JKW (1811) Caroli Illigeri D. Acad. Reg. Scient. Berolinens. et Bavaricae Sod. Museo Zoologico Berolin. praefecti professoris extraord. Prodromus systematis mammalium et avium: additis terminis zoographicis utriusque classis, eorumque versione germanica. Sumptibus C. Salfeld, Berolini, 301 pp. doi: 10.5962/bhl.title.42403

Koch AC (1841) Description of Missourium, or Missouri leviathan : together with its supposed habits and Indian traditions concerning the location from whence it was exhumed; also, comparisons of the whale, crocodile and missourium with the leviathan, as described in by. Prentice and Weissinger, Louisville, Kentucky. doi: 10.5962/bhl.title.35985

Krell FT (2000) Impact factors aren't relevant to taxonomy. Nature 405(6786): 507–508. doi: 10.1038/35014664

Lambert O, Bianucci G, Post K, de Muizon C, Salas-Gismondi R, Urbina M, Reumer J (2010a) The giant bite of a new raptorial sperm whale from the Miocene epoch of Peru. Nature 466(7302): 105–108. doi: 10.1038/nature09067

Lambert O, Bianucci G, Post K, de Muizon C, Salas-Gismondi R, Urbina M, Reumer J (2010b) The giant bite of a new raptorial sperm whale from the Miocene epoch of Peru. Nature 466(7310): 1134–1134. doi: 10.1038/nature09381

Leviton AE, Gibbs RH, Heal E, Dawson CE (1985) Standards in Herpetology and Ichthyology: Part I. Standard Symbolic Codes for Institutional Resource Collections in Herpetology and Ichthyology. Copeia 1985(3): 802–832.

MacCallum CJ (2007) When Is Open Access Not Open Access? PLoS Biology 5(10): e285. doi: 10.1371/journal.pbio.0050285

Maddison DR, Guralnick R, Hill A, Reysenbach A-L, McDade LA (2011) Ramping up biodiversity discovery via online quantum contributions. Trends in Ecology & Evolution 27(2): 72–77. doi: 10.1016/j.tree.2011.10.010

Martin S, Hohman MM, Liefeld T (2005) The impact of Life Science Identifier on informatics data. Drug discovery today 10(22): 1566–1572. doi: 10.1016/S1359-6446(05)03651-2

Müller-Wille S, Charmantier I (2011) Natural history and information overload: The case of Linnaeus. Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences 43(1): 4–5. doi: 10.1016/j.shpsc.2011.10.021

Nagy LG, Petkovits T, Kovács GM, Voigt K, Vágvölgyi C, Papp T (2011) Where is the unseen fungal diversity hidden? A study of *Mortierella* reveals a large contribution of reference collections to the identification of fungal environmental sequences. The New phytologist 191(3): 789–794. doi: 10.1111/j.1469-8137.2011.03707.x

Neave SA (1939) Nomenclator Zoologicus: a list of the names of the genera and subgenera in zoology from the tenth edition of Linnaeus 1758 to the end of 1935. The Zoological Society of London, London.

Page RDM (2006) Taxonomic names, metadata, and the Semantic Web. Biodiversity Informatics 3: 1–15. doi: 10.17161/bi.v3i0.25

Page RDM (2008) Biodiversity informatics: the challenge of linking data and the role of shared identifiers. Briefings in bioinformatics 9(5): 345–354. doi: 10.1093/bib/bbn022

Page RDM (2009) bioGUID: resolving, discovering, and minting identifiers for biodiversity informatics. BMC Bioinformatics 10(Suppl 14): 5. doi: 10.1186/1471-2105-10-S14-S5

Page RDM (2011a) Extracting scientific articles from a large digital archive: BioStor and the Biodiversity Heritage Library. BMC bioinformatics 12: 187. doi: 10.1186/1471-2105-12-187

Page RDM (2011b) Dark taxa: GenBank in a post-taxonomic world. http://iphylo.blogspot.co.uk/2011/04/dark-taxa-genbank-in-post-taxonomic.html

Page RDM (2011c) Orwellian metadata: making journals disappear. http://iphylo.blogspot.co.uk/2011/09/orwellian-metadata-making-journals.html

Page RDM (2013) BioNames: linking taxonomy, texts, and trees. PeerJ 1: e190. doi: 10.7717/peerj.190

Parr CS, Guralnick R, Cellinese N, Page RDM (2011) Evolutionary informatics: unifying knowledge about the diversity of life. Trends in Ecology & Evolution 27(2): 94–103. doi: 10.1016/j.tree.2011.11.001

Patterson DJ, Cooper J, Kirk PM, Pyle RL, Remsen DP (2010) Names are key to the big new biology. Trends in Ecology & Evolution 25(12): 686–691. doi: 10.1016/j.tree.2010.09.004

Peters WCH (1846) Über neue Säugethiergattungen aus den Ordnungen der Insectenfresser und Nagethiere. Bericht über die zur Bekanntmachung geeigneten Verhandlungen der Konigl.Preuss.Akademie der Wissenschaften zu Berlin 1846: 257–259.

Phillips AJ, Arauco-Brown R, Oceguera-Figueroa A, Gomez GP, Beltrán M, Lai Y-T, Siddall ME (2010) *Tyrannobdella* rex n. gen. n. sp. and the evolutionary origins of mucosal leech infestations. PLoS ONE 5(4): e10057. doi: 10.1371/journal.pone.0010057

Raimond Y, Scott T, Sinclair P, Miller L, Betts S, McNamara F (2010) Case Study: Use of Semantic Web Technologies on the BBC Web Sites. http://www.w3.org/2001/sw/sweo/public/UseCases/BBC/

Remsen DP, Norton C, Patterson DJ (2006) Taxonomic informatics tools for the electronic Nomenclator Zoologicus. The Biological bulletin 210(1): 18–24. doi: 10.2307/4134533

Schindel DE, Miller SE (2010) Provisional nomenclature: the on-ramp to taxonomic names. In: Polaszek A (Ed.) Systema Naturae 250 – The Linnaean Ark. CRC Press, 109–115.

Segers H, Smet WHDE, Fischer C, Fontaneto D, Wallace RL, Jersabek CD (2012) Towards a List of Available Names in Zoology, partim Phylum Rotifera. Zootaxa 68: 61–68.

Simkin MV, Roychowdhury VP (2011) Theory of citing. Springer Optimization and Its Applications. Physics and Society; Disordered Systems and Neural Networks; Applications 57(4): 463–505. doi: 10.1007/978-1-4614-0754-6_16

Specht CG (2010) Mutations of citations. The Scientist. http://www.the-scientist.com/?articles.view/articleNo/29252/title/Opinion--Mutations-of-citations/

Vences M, Riva IDL (2007) A new species of *Gephyromantis* from Ranomafana National Park, south-eastern Madagascar (Amphibia, Anura, Mantellidae). Spixiana 30(1): 135–143.

Vences M, Thomas M, van der Meijden A, Chiari Y, Vieites DR (2005) Comparative performance of the 16S rRNA gene in DNA barcoding of amphibians. Frontiers in zoology 2(1): 5. doi: 10.1186/1742-9994-2-5

Werner YL (2006) The case of impact factor versus taxonomy: a proposal. Journal of Natural History 40(21-22): 1285–1286. doi: 10.1080/00222930600903660

Wägele H, Klussmann-Kolb A, Kuhlmann M, Haszprunar G, Lindberg D, Koch A, Wägele JW (2011) The taxonomist - an endangered race. A practical proposal for its survival. Frontiers in zoology 8(1): 25. doi: 10.1186/1742-9994-8-25

Yan K-K, Gerstein M (2011) The spread of scientific information: insights from the web usage statistics in PLoS article-level metrics. PLoS ONE 6(5): e19917. doi: 10.1371/journal.pone.0019917

ZooKeys 550: 261–281 (2016)
doi: 10.3897/zookeys.550.10009
http://zookeys.pensoft.net

RESEARCH ARTICLE

A peer-reviewed open-access journal
ZooKeys
Launched to accelerate biodiversity research

Towards a Global Names Architecture: The future of indexing scientific names

Richard L. Pyle[1]

1 *Bernice Pauahi Bishop Museum, 1525 Bernice Street, Honolulu, HI 96817, USA*

Corresponding author: *Richard L. Pyle* (email address)

Academic editor: *Ellinor Michel* | Received 19 May 2015 | Accepted 20 May 2015 | Published 7 January 2016

http://zoobank.org/AD5B8CE2-BCFC-4ABC-8AB0-C92DEC7D4D85

Citation: Pyle RL (2016) Towards a Global Names Architecture: The future of indexing scientific names. In: Michel E (Ed.) Anchoring Biodiversity Information: From Sherborn to the 21st century and beyond. ZooKeys 550: 261–281. doi: 10.3897/zookeys.550.10009

Abstract

For more than 250 years, the taxonomic enterprise has remained almost unchanged. Certainly, the tools of the trade have improved: months-long journeys aboard sailing ships have been reduced to hours aboard jet airplanes; advanced technology allows humans to access environments that were once utterly inaccessible; GPS has replaced crude maps; digital hi-resolution imagery provides far more accurate renderings of organisms that even the best commissioned artists of a century ago; and primitive candle-lit microscopes have been replaced by an array of technologies ranging from scanning electron microscopy to DNA sequencing. But the basic paradigm remains the same. Perhaps the most revolutionary change of all – which we are still in the midst of, and which has not yet been fully realized – is the means by which taxonomists manage and communicate the information of their trade. The rapid evolution in recent decades of computer database management software, and of information dissemination via the Internet, have both dramatically improved the potential for streamlining the entire taxonomic process. Unfortunately, the potential still largely exceeds the reality. The vast majority of taxonomic information is either not yet digitized, or digitized in a form that does not allow direct and easy access. Moreover, the information that is easily accessed in digital form is not yet seamlessly interconnected. In an effort to bring reality closer to potential, a loose affiliation of major taxonomic resources, including GBIF, the Encyclopedia of Life, NBII, Catalog of Life, ITIS, IPNI, ICZN, Index Fungorum, and many others have been crafting a “Global Names Architecture” (GNA). The intention of the GNA is not to replace any of the existing taxonomic data initiatives, but rather to serve as a dynamic index to interconnect them in a way that streamlines the entire taxonomic enterprise: from gathering specimens in the field, to publication of new taxa and related data.

Keywords
Taxonomy, Carl Linnaeus, Charles Davies Sherborn, Biodiversity Data, Global Names Usage Bank, Global Names Index, ZooBank Biodiversity Library

Introduction

Although biological taxonomy is sometimes referred to as the "oldest profession" (Hedgpeth 1961, Chmielewski and Krayesky 2013), its current incarnation began with the start of modern nomenclature in the middle part of the eighteenth century (Linnaeus 1753, 1758). Throughout this time, the fundamental unit of taxonomy has been the "species", the concept for which has eluded a clear consensus definition (e.g., Wheeler and Meier 2000). Linnaeus himself was a creationist, and therefore saw species as the work of God (Linnaeus 1736:18; translation from Wilkins 2009:41):

Species tot sunt diversae quot diversas formas ab initio creavit infinitum Ens. [There are as many species as the Infinite Being produced diverse forms in the beginning.]

This is not at all surprising, given that Darwin's concept of evolution was not proposed until a century after the start of modern nomenclature (Darwin 1859). But even then, Darwin opted not to attempt a precise definition of "species", writing (p. 40):

Hence, in determining whether a form should be ranked as a species or a variety, the opinion of naturalists having sound judgment and wide experience seems the only guide to follow. We must, however, in many cases, decide by a majority of naturalists, for few well-marked and well-known varieties can be named which have not been ranked as species by at least some competent judges.

This idea was reflected by the definition of species by Regan (1926: 75):

A species is a community, or a number of related communities, whose distinctive morphological characters are, in the opinion of a competent systematist, sufficiently definite to entitle it, or them, to a specific name. [often paraphrased as, "a species is what a competent taxonomist says it is"]

Many modern taxonomists have dismissed this definition as unscientific or too reliant on the notion of what "competent" means, and as a result, debates regarding a more precise and biologically meaningful definition of species have continued over the decades well into modern times (publications too numerous to cite, but see Wilkins 2009 for a review).

Regardless of its merit, acceptance, or adoption, a variant of this definition, effectively "a species is what a community of taxonomists says it is" is the *de-facto* species definition that has been applied since the time of Linnaeus. Taxonomists have asserted individual species circumscriptions over the course of centuries, and those circumscriptions that have met with approval by subsequent taxonomic communities have endured the test of time. In the modern context, while there are certainly species that are subject to ongoing debate, the vast majority of species have achieved some level of stability.

In stark contrast to the dynamic, ongoing, and seemingly endless debates about what a "species" is, the nomenclatural system used by taxonomists during the past two and a half centuries has been remarkably consistent, universal, and stable. The primary reason for this consistent and universal stability has to do with the Codes of scientific Nomenclature (e.g., ICZN 1999, Lapage et al. 1990, McNeill et al. 2012), which have enjoyed near-universal adoption for more than a century. A major reason for the contrast between "species" and scientific names is that the former are, and likely always will be subjective in their core nature; whereas the latter leverage the objectivity of the nomenclatural codes to reduce matters of opinion and dispute to a minimum. In effect, the Linnaean nomenclatural system represents a stable scaffolding against which which the ever-changing landscape of species can be reliably referenced.

It is the objective and largely stable nature of scientific names of organisms that makes them well-suited for large-scale indexing of the sort that Charles Davies Sherborn (1861–1942) dedicated his life to. Whereas the majority of the nearly 4,400 species circumscriptions described by Linnaeus in his 1758 *Systema Naturae* bear very little resemblance to the species boundaries asserted by modern biologists, most of the scientific names he established are not only available under the current Code, but are in current use (though often in combination with different generic names than what Linnaeus used). Even when historical scientific names have been synonymized by later workers, they remain available (when Code-compliant), and therefore potentially relevant centuries after their establishment. Although catalogs of species (e.g., Linnaeus 1758) may begin to lose their taxonomic relevance almost immediately after publication, the scientific names established within such catalogs retain their nomenclatural relevance indefinitely. Ultimately, this is why the career-long labors of Sherborn have retained their value well beyond his own life, up until today and continuing indefinitely into the future.

The more things change, the more they stay the same

The system of scientific nomenclature is not the only aspect of the taxonomic enterprise that has remained relatively constant over the centuries. Certainly there have been some improvements to the way taxonomists do their jobs. For example, it once required months to journey across the seas aboard sailing ships, whereas now almost any part of the world can be reached within a few hours aboard modern jet airplanes (Figure 1). Early naturalists had to rely on crude maps drawn by sailors to figure out where their specimens were collected, whereas the Global Positioning System (GPS) and digital mapping tools such as Geographic Information Systems (GIS) and Google Earth allow modern taxonomists to pinpoint the collection location for a specimen

Figure 1. In centuries past, months-long journeys aboard sailing ships were required for taxonomists to reach their destinations (left, Thomas Whitcombe). Today, almost any part of the world can be reached aboard modern aircraft (right, R. L. Pyle).

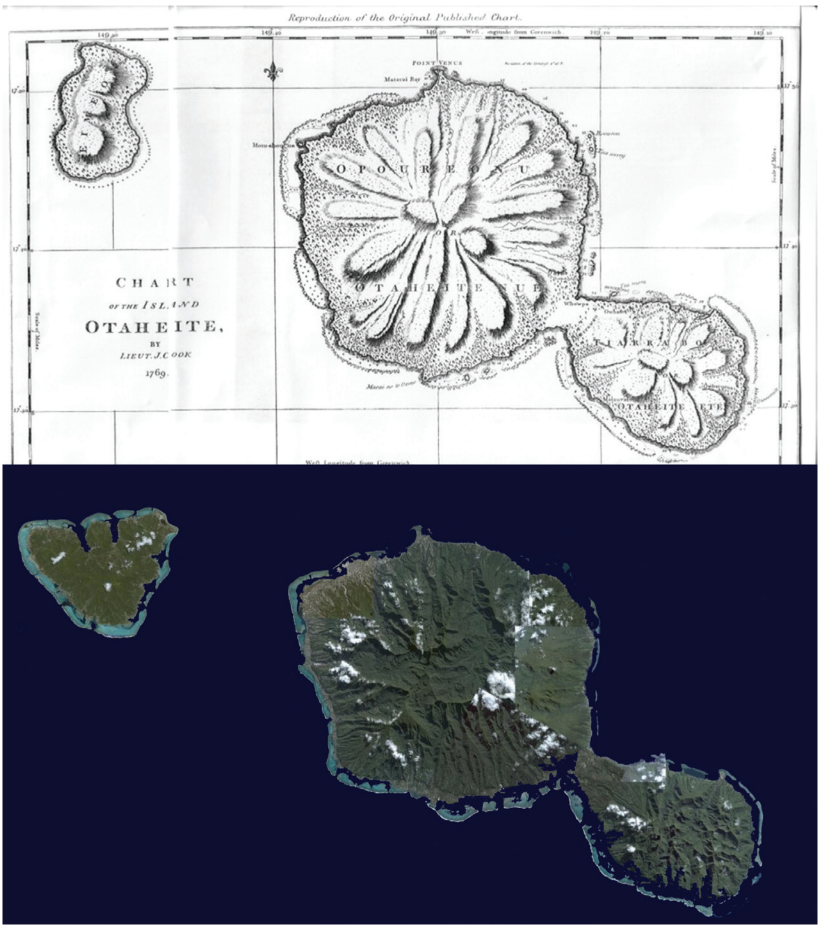


Figure 2. Early taxonomists had only crude maps to plot the locations of their specimens; in this case the French Polynesian islands of Tahiti and Moorea (top, from Prévost D’Exiles 1746–1789). Today, highly accurate maps and satellite imagery can pinpoint particular locations within a few meters (bottom, Landsat).

Figure 3. Highly trained artisans once labored to produce detailed hand-painted illustrations of specimens (top, from Jordan and Evermann 1903). Modern digital cameras can generate far more accurate and detailed images almost instantly and with minimal skill (bottom, R. L. Pyle). Both images depict *Bodianus sanguineus* Jordan and Evermann 1903.

within a meter or so (Figure 2). A century ago, highly skilled illustrators painstakingly created colorful works of art by hand, based on direct observations and descriptions of the color and form of living organisms, whereas modern digital imaging technology allows us to generate extremely high quality photographs of living and freshly prepared specimens in an instant (Figure 3). To examine his specimens, Linnaeus used primitive candle-lit microscopes with hand-ground optics, whereas today we can generate high-resolution three-dimensional images of the internal and external structures of organisms using 3D photogrammetry and CT scanning without even dissecting them, create crisp images of tiny structures using electron microscopy, and read the very code of life by sequencing DNA (Figure 4). Finally, the technology we use to access the environments in which organisms live has changed dramatically from centuries past (Figure 5).

However, despite these important technological advancements in the tools of the trade for taxonomy, the fundamental process remains the same today as it was centu-

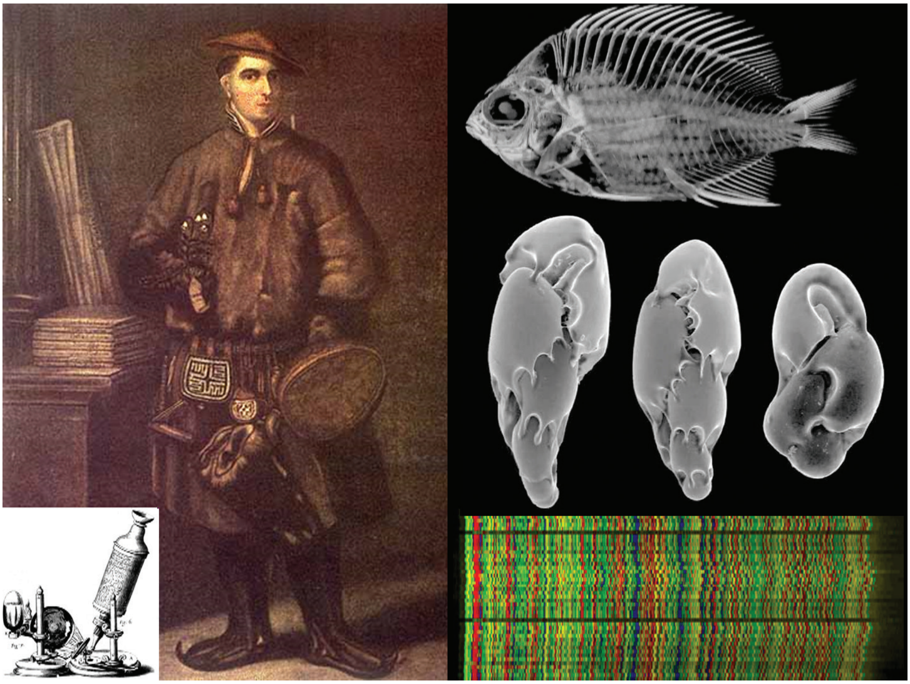


Figure 4. Carl Linnaeus used candle-lit microscopes with primitive optics to examine his specimens (left, H. Kingsbury). Modern technology allows us to generate high-resolution 3D CT scans of the internal structures of specimens without displacing a single scale (right top, Digimorph; *Chromis abyssus*), capture crisp images of tiny organisms through electron microscopy (right middle, NOAA; single-celled foraminifera), and read DNA sequences (right bottom, BOLD, unspecified taxon).

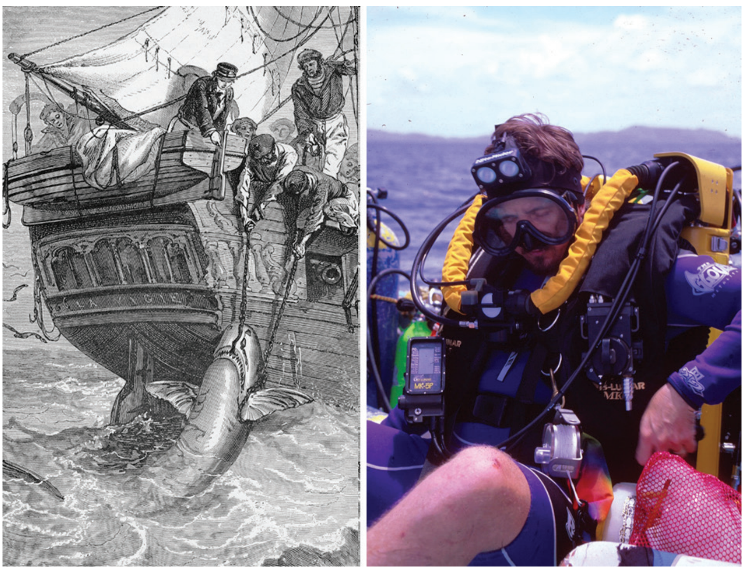


Figure 5. Methods of collecting specimens from the field have advanced from earlier eras (left, from C.Delon, 1889) to modern high-tech equipment of today (right, Ken Corben).

Figure 6. Despite many technological advancements in the tools of the taxonomic trade, the fundamental paradigm for the taxonomic enterprise remains almost unchanged from centuries ago (left, from Bates 1863; right Bishop Museum).

ries ago: impassioned naturalists seek financial support from governments and private entities to travel the globe to discover new species of organisms; they take detailed notes and acquire specimens, which they carefully transport back to Museums; they create color images, dissect, poke, prod, count, and measure their biological treasures; they write detailed descriptions that are printed on paper in books and periodicals. These fundamental steps in the taxonomic process, while aided by advanced technology, have remained fundamentally unchanged since the time of Linnaeus (Figure 6).

A revolution in information technology

There is one aspect of technological change that has been truly revolutionary, which is the means by which taxonomists manage and communicate the information of their trade. The rapid advancement in recent decades of computer database management software, and of information dissemination via the Internet, have both dramatically improved the potential for streamlining the entire taxonomic process. Less than two decades ago, graduate students in taxonomy spent untold hours in libraries, scouring through pages and pages of paper documents to find original descriptions and key taxonomic revisions. Today, with a few searches on Google and with the extremely useful Biodiversity Heritage Library, many original sources are only a few mouse clicks away. And the ease of access is not limited to digitized literature; specimens, images, and vast amounts of biological information are freely available through the Internet. One of the last remaining barriers to information availability – the pay-walls behind which many

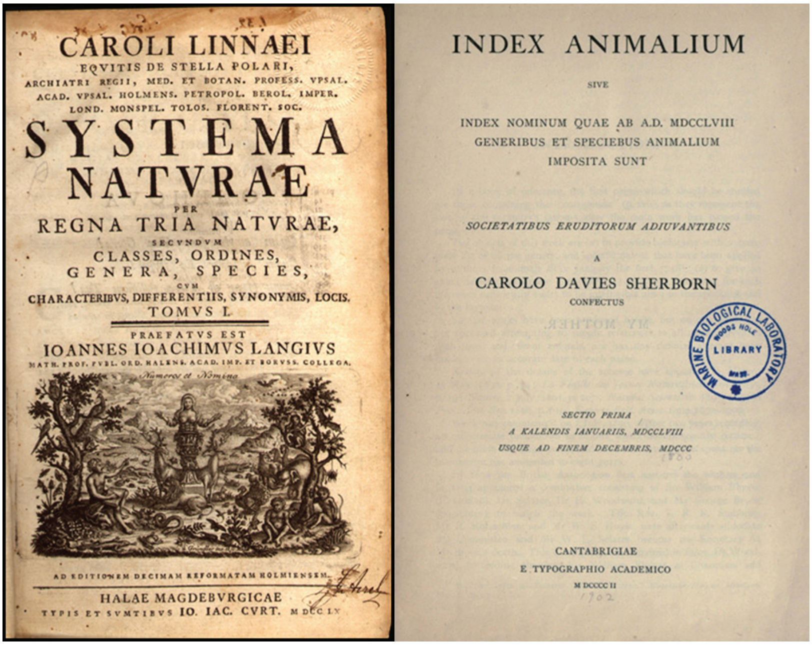


Figure 7. At the start of modern zoological nomenclature, Linnaeus' tenth edition of *Systema Naturae* contained almost 4,400 species-group names (left). By 1850, the number of species names for animals had reached nearly 430,000 – an increase of two orders of magnitude.

newly published research hides – is gradually eroding through an increasing demand for open-access models of publication.

This revolution in digital information technology is extremely fortunate, given the rate at which species have been (and continue to be) described. In 1758, the tenth edition of Linnaeus' *Systema Naturae* contained nearly 4,400 species-group names. At the time, this compilation represented the entire catalog of all known animals. In the century that followed, the number of scientific names for species had increased by two orders of magnitude, as represented in the volumes of Sherborn's *Index Animalium* (Figure 7). Today, the online edition of the Catalogue of Life includes more than 2.7 million names (representing more than 1.5 million species). Sherborn spent most of his professional career compiling what is effectively one sixth the number of names that likely exist in biology today. Without the electronic information revolution, the Catalogue of Life would be far less complete than it currently is.

As exciting as the electronic information revolution is, however, in the context of taxonomy there is still far more potential than there is reality in terms of harnessing the power of information technology. The vast majority of taxonomic information either remains non-digitized, or is digitized in a form that does not allow direct and easy access. Moreover, much (if not most) of the information that is easily accessed in digital form is not yet seamlessly interconnected. At present, the total biodiversity knowledge-base for all life forms is scattered across an estimated half-billion pages of printed literature, thousands of natural history collections housing billions of specimens, hundreds of thousands of digital databases and websites, and hundreds of millions of DNA sequences. Consumers of this knowledge-base, which includes tens of thousands of taxonomists, hundreds of thousands of biologists, a hundred million citizen scientists, governmental resource managers and policy makers, and ultimately much of the total human population, have not had easy access to this information (Heidorn 2008; IISE 2010; Thessen and Patterson 2011; Fontaine et al. 2012). There are many excellent websites containing valuable information, including nomenclatural, taxonomic, biogeographic, life-history and ecological information about species, not to mention genetic data, images and videos, and countless other data sources. To find all this information – even when it is readily available through the Internet – usually requires multiple web searches and visits to dozens of different online resources.

The next step in the information revolution for biodiversity information involves not just the digitization of content, but will involve the cross-linking and more seamless integration of existing digital resources.

The global names architecture

In an effort to bring reality closer to potential, a loose affiliation of major taxonomic resources, including the Global Biodiversity Information Facility (GBIF; http://www.gbif.org), the Encyclopedia of Life (EOL; http://eol.org), the former U.S. National Biological Information Infrastructure (NBII), Catalog of Life (CoL; http://www.catalogueoflife.org), the Integrated Taxonomic Information System (ITIS; http://www.itis.gov), the International Plant Names Index (IPNI; http://ipni.org), the International Commission on Zoological Nomenclature (ICZN; http://iczn.org), Index Fungorum (IF; http://www.indexfungorum.org), and many others have been crafting a "Global Names Architecture" (GNA). The intention of the GNA is not to replace any of the existing taxonomic or other biodiversity data initiatives, but rather to serve as a dynamic suite of web services and two primary indexes (GNI and GNUB, described below) that interconnect existing data systems in a way that streamlines the entire taxonomic enterprise: from gathering specimens in the field, to publication of new taxa and related data.

The basic premise behind the GNA is that scientific names of organisms represent the key to integrating disconnected biological data, to allow efficient and effective coordination between biological research and exploration activities, and broader understanding and management of biodiversity (Patterson et al. 2010). Throughout the vast global biological knowledge base – including natural history collections, historical and modern literature, observational databases, multimedia (image and video) resources, genetic data repositories, nomenclators, taxonomic catalogs, data aggregators, and major Internet search engines – the majority of data are given taxonomic context through simple text-string scientific names (Figure 8).

Unfortunately, sources of imprecision and ambiguity severely limit the use of these text-string names for cross-linking digital data content. For example, there are ap-

Figure 8. The icons around the periphery represent examples of where biological data tagged with scientific names currently exist. The cluster of names in the center represent examples of distinct text-strings that have been used to represent the same species within different data sources.

proximately two million scientifically described species (Chapman 2009; Trontelj and Fišer 2009; Mora et al. 2011); yet the Global Names Index (GNI) – the component of the GNA that indexes text-string scientific names – already contains nearly 20 million distinct name-strings (and this index is far from complete). The more than ten-fold discrepancy between species and text-string names results from several factors, including synonymy (multiple names for the same species), alternate nomenclatural combinations (the same species epithet combined with different generic and subgeneric names, and applied at different ranks), alternate spellings (orthographic variations, misspellings, abbreviations, etc.), and inconsistent formatting of names (e.g., with and without qualifiers, alternate formats for authorships and/or year, etc.). Additional confusion results from homonymy (the same name assigned to different species).While text-string names are generally easy to interpret and disambiguate by a human, they represent a substantial barrier to electronic cross-linking of data.

To overcome the limitations of text-string scientific names, the GNA includes a core component called the Global Names Usage Bank (GNUB). GNUB is a highly normalized database system, the primary purpose of which is to index and assign persistent globally unique identifiers (GUIDs) to Agents, References, and Taxon Name Usage (TNU) instances (among other relevant data objects). Agents are people and organizations, and in the context of GNUB mostly represent Authors of References. References include all published literature, as well as many forms of unpublished documentation (e.g., unpublished reports and manuscripts, specimen labels, herbarium sheets, field notes, etc.). Any static documentation source can be a Reference in the GNUB architecture. A TNU is any usage or treatment of a scientific name within a Reference. TNUs are the foundation for all Code-governed nomenclatural acts, taxon concept definitions, taxonomic treatments, synonymies and classifications, and any other forms of taxonomic assertions. The subset of TNUs that represent the establishment of new scientific names (i.e., original descriptions) are called "Protonyms" (Pyle 2004). Every scientific name (at every rank) has one Protonym TNU, and all subsequent TNUs refer back to the Protonym. For example, the fish genus *Gasterosteus* was described by Linnaeus (1758), so the TNU for that name in that Reference is the Protonym. Linnaeus (1766) also treated the genus *Gasterosteus*, and that TNU links back to the Protonym TNU in Linnaeus (1758). Protonyms apply to names at all ranks. For example, Linnaeus 1766 established the new species *Gasterosteus saltatrix*. Whereas the TNU for *Gasterosteus* within this publication is not the Protonym for that genus, the TNU for the species epithet *saltatrix* within Linnaeus 1766 is the Protonym for that species-group name (because this publication is the original description of the species, but not the genus).

The core elements of a TNU include the following items (see Table 1; not all elements are required for all TNUs):

1) A unique and persistent identifier for the TNU itself;
2) A link to the Reference (including page, if applicable) in which the TNU appears;
3) A recursive link to the Protonym-TNU for the name represented by the TNU;
4) An indication of the taxonomic rank at which a name was treated (e.g., "genus" for the TNU for *Gasterosteus* within Linnaeus 1766, and "species" for the TNU for *saltatrix* within Linnaeus 1766);
5) The exact spelling (as best as can be represented using UTF-8 encoding) of the name as used within the Reference (e.g., Regan 1909 spelled the genus *Gasterosteus* as "*Gastrosteus*", and Günther 1860 spelled the species *saltatrix* as "*saltator*");
6) A link to the TNU (within the same Reference) representing the immediate parent taxon (e.g., the Protonym TNU for the species *saltatrix* within Linnaeus 1766, would link to the non-Protonym TNU for the genus *Gasterosteus* as treated by Linnaeus 1766);

In cases where a name is treated as a junior synonym of another name, a link to the TNU (within the same Reference) representing the senior synonym as asserted by the indicated Reference for this junior synonym (e.g., Günther 1860 treated the name *Pomatomus skib* (Lacepède 1802) as a junior synonym of *Temnodon saltator* (Linnaeus 1766), so the TNU for the Günther 1860 treatment of *skib* links to the TNU representing the Günther 1860 treatment of *saltator* [=*saltatrix* Linnaeus 1766]).

By building an index of all TNUs across historical literature (starting with the Protonym TNU for each name), GNUB data services can efficiently perform powerful analyses

Table 1. Examples of Taxon Name Usage instances (TNUs). Records representing Protonyms are highlighted in yellow. Records representing treatments of a name as a synonym is highlighted in grey. Ellipses (…) in the Parent column represent links to higher-rank TNUs not included in the sample below. This table is highly simplified and does not represent the actual GNUB data model.

TNUID	Reference	Protonym	Rank	Spelling	Parent	Valid	Representative Usage
1	**Linnaeus 1758:295**	**1**	**Genus**	***Gasterosteus***	…	**1**	**Protonym of genus *Gasterosteus***
2	Linnaeus 1766:489	1	Genus	*Gasterosteus*	…	2	Subsequent usage of *Gasterosteus*
3	**Linnaeus 1766:491**	**3**	**Species**	***Saltatrix***	**2**	**3**	**Protonym of species *G. saltatrix***
4	Schöpf 1788:167	1	Genus	*Gafterofteus*	…	4	Subsequent usage & variant of *Gasterosteus*
5	Schöpf 1788:168	3	Species	*Sallatrix*	4	5	Subsequent usage & variant of *G. saltatrix*
6	**Lacépède 1802:435**	**6**	**Genus**	***Pomatomus***	…	**6**	**Protonym of genus *Pomatomus***
7	**Lacépède 1802:436**	**7**	**Species**	***Skib***	**6**	**7**	**Protonym of species *P. skib***
8	**Cuvier 1816:346**	**8**	**Genus**	***Temnodon***	…	**8**	**Protonym of genus *Temnodon***
9	Günther 1860:479	8	Genus	*Temnodon*	…	9	Subsequent usage of *Temnodon*
10	Günther 1860:479	3	Species	*Saltator*	9	10	Subsequent usage & variant of *G. saltatrix*
11	Günther 1860:479	7	Species	*Skib*	-	10	Synonym treatment of *P. skib*
12	Bean 1903:445	6	Genus	*Pomatomus*	…	12	Subsequent usage of *Pomatomus*
13	Bean 1903:445	3	Species	Saltatrix	12	13	Subsequent usage & combination of *G. saltatrix*

and transformations of taxon names across different spellings, synonymies and classifications. For example, the species *Gasterosteus saltatrix* Linnaeus, 1766, has also been spelled *Sallatrix* in at least one Reference, spelled *saltator* in at least 16 References, and the species epithet (by whichever spelling) has been variously combined with the genus names *Pomatomus* Lacépède, 1702, *Temnodon* Cuvier, 1816 and *Cheilodipterus* Lacépède, 1801. Moreover, the GNUB index also records the fact that at least twelve other species have been treated as a junior synonyms of *saltatrix*, and these species have been variously combined with at least ten different genus names. Thus, through GNUB we can see that the species originally established by Linnaeus 1766 as *Gasterosteus saltatrix* has been variously referred to in literature by at least 28 different text-string scientific names (inclusive of both homotypic and heterotypic synonyms, suite of GNUB and GNI services, content in otherwise disconnected datasets can be cross-linked despite heterogeneous taxon names.

A successful proof of concept

Largely through support from two separate NSF grants (DBI-1062441; DBI-0956415), the GNA has been developed into a highly successful proof of concept. The most visible representation is the ZooBank registry (http://zoobank.org). ZooBank was first proposed as an official online nomenclatural registry for zoology, under the auspices of the International Commission for Zoological Nomenclature (ICZN) by Polaszek et al. 2005. It was first launched as an early prototype on 1 January 2008 to commemorate the 250th anniversary of the official start of all zoological nomenclature (Pyle et al. 2008; Pyle and Michel 2008; Pyle and Michel 2010; Rosenberg et al. 2012). ZooBank was later reconceived as a service operating on top of GNUB (Pyle and Michel 2009), and the new GNUB-based ZooBank was publicly launched on September 4, 2012, coinciding with the amendment to the ICZN Code supporting electronic publication (ICZN 2012a; 2012b). The amended Code requires all electronically published works in Zoology be registered in ZooBank, thus representing the first mandatory electronic/online registration requirement for any major Code of Nomenclature (the bacteriological Code includes a paper-based registration system [Tindall 2009], and the Code for algae, fungi and plants includes a registration system for fungal names that went into effect on 1 January 2013).

Prior to 2012, ZooBank registrations grew steadily from approximately 100 registrations per month in 2008-2010, to approximately 500 registrations per month in 2011–2012. After the new GNUB-based implementation of ZooBank was launched in September 2012, registrations increased almost ***ten-fold***, to nearly 5,000 per month. The vast majority of these regiwstrations are prospective – that is, for works and names that are newly established. Retrospective content for ZooBank will be added through the bulk importation of existing databases, and through harvesting protonyms from BHL and other sources. Commensurate with the rise in registrations has been an increase in the ZooBank user-base. From 2008–2012, the ZooBank user base grew steadily to a little over 100 active users. In less than a year since the GNUB-based ZooBank was launched, the user base has also grown nearly ***ten-fold***, to over 1,000 users (and it continues to grow). As successful as the new GNUB-based ZooBank has been, it is important to emphasize that ZooBank is only one example of a service that GNUB can facilitate. In addition to ZooBank as a model for GNUB-based registration systems in other nomenclatural domains, there are many other services that GNUB can facilitate.

Whereas name-usages within static References are indexed directly as TNUs, these are mapped to records in external and/or dynamic data sources through a simple identifier cross-link feature in GNUB. This feature, which currently includes nearly half a million links from records in more than 200 external databases to over 320,000 GNUB records, enables much more than simply linking GNUB records to external databases; specifically, it allows external databases to be linked to ***each other***.

For example, GNUB includes links to over 111,000 registered names in ZooBank, nearly 140,000 records (taxonomic serial numbers) in ITIS, and nearly 70,000 genus-group and species-group name records in the Catalog of Fishes (CoF). Besides allowing these three datasets to link directly to GNUB (and vice-versa), the Identifier cross-link service also enables direct cross-links between each of these otherwise disconnected datasets (in this case, 67,000 linked records between ZooBank and CoF; 26,555 linked records between ZooBank and ITIS, and 26,467 linked records between ITIS and CoF). Because of this cross-linking feature, new names registered in ZooBank could be presented to ITIS and CoF for inclusion in their databases, and corrections made to errors in CoF could be propagated to both ZooBank and ITIS. By establishing cross-links between equivalent records in different database systems, we not only expand the ability for end-users to directly access records in the different systems, but we also create novel opportunities for proactive collaboration between different systems with overlapping content. While other systems include support for similar features (e.g., the EoL "Partner Links", and the NCBI Taxonomy "LinkOut"), the GNA provides a single shared platform for all cross-links, such that anytime a record is indexed in GNUB, it is automatically cross-linked to all other data systems that are indexed in GNUB.

This cross-linking service is not limited to taxon names. For example, GNUB includes links to over 3,300 journals registered in ZooBank and over 3,200 journals scanned in the Biodiversity Heritage Library (BHL). Through the BHL "OpenURL" service, over 50,000 ZooBank species pages (as well as nearly 100,000 other TNU records) now have direct access to the corresponding page image in BHL. Likewise, because GNUB is linked to over 34,000 authors in the Authors of Plant Names (APN) directory, and nearly 21,000 authors registered in ZooBank, we can compare authorship trends in both domains (e.g., fewer than 1% of all authors have published new scientific names for both plants and animals).

This same cross-linking service applies to records in more than 200 different external databases (not all databases have been fully indexed yet). As such, GNUB can serve as a universal hub to cross-link records (not only Authors, References, and TNUs/Protonyms; but virtually any other data object as well), which will facilitate collaboration and data exchange (as in the names-linking example), enhance web services to infer and establish other links (as in the BHL page example), and to allow analysis of patterns that had not previously been possible (e.g., patterns of authorship over time, such as those as used by Costello et al. 2012). This cross-linking service developed for GNUB represents an important step towards empowering the collective utility of biological datasets on a global scale.

Several other services and APIs were developed for searching, dereferencing, editing, and inserting Agents, References, and TNUs (particularly Protonyms), with a variety of output formats (e.g., HTML, JSON). The most recent service is called "GNIE" (originally an acronym for "Global Names Index Export", and retained despite its expanded utility), which accepts an identifier for a Protonym and returns a set of all scientific names indexed in GNUB that have been used to represent the same taxon, including all homotypic synonyms (spelling variants, alternate genus combinations, etc.) and heterotypic synonyms (names that have been treated as either junior or senior synonyms of the indicated Protonym). Documentation for all of these services is included on the ZooBank API page (http://zoobank.org/Api).

Although these services were used extensively in the development of ZooBank (Figure 9), funding from NSF supported the development and implementation of these and other as-yet undocumented services in various stages of development and testing on two other database systems as well: Bishop Museum natural science databases (http://nsdb.bishopmuseum.org/), and the "Explorer's Log" (http://www.explorers-log.com/). Bishop Museum manages several major specimen and occurrence databases (Plants, Insects, Birds, Mammals, Amphibians and Reptiles, Fishes, Marine Invertebrates, Mollusks, and Pacific Center for Molecular Biodiversity, as well as several regional checklist databases), and we are currently in the process of building support for GNUB as the taxonomic authority against which Bishop Museum specimen databases are indexed. The "Explorer's Log" (http://www.explorers-log.com) is a feature-rich suite of web-based applications designed to support field-based data collection for organismal occurrence records (including specimens and associated tissue samples, multi-media documentation such as images, videos, audio recordings and telemetry data, visual observations, and literature-based occurrence records) and associated data. This system has now completely replaced its internal taxonomy tables and utilizes GNUB services to assign taxonomic identifications to organism occurrences. The purpose of developing these prototype services was to demonstrate the ability for external data management systems to utilize GNUB data and web services directly to support broader biological datasets, without the need to re-invent a custom taxonomic authority system (as is currently done for most biological data management systems).

In addition to these services designed primarily to support external systems, several services designed for internal GNUB use were also developed. These include a user/login account management system, a robust record de-duplication resolution system, a prototype data reconciliation tool (currently optimized for Agents and journal titles) used for bulk data imports, multi-lingual support, a tool for visualizing the publication timeline history for authors, a suite of database statistics visualization services, and services to facilitate data contribution and management by publishers and editors of journals.

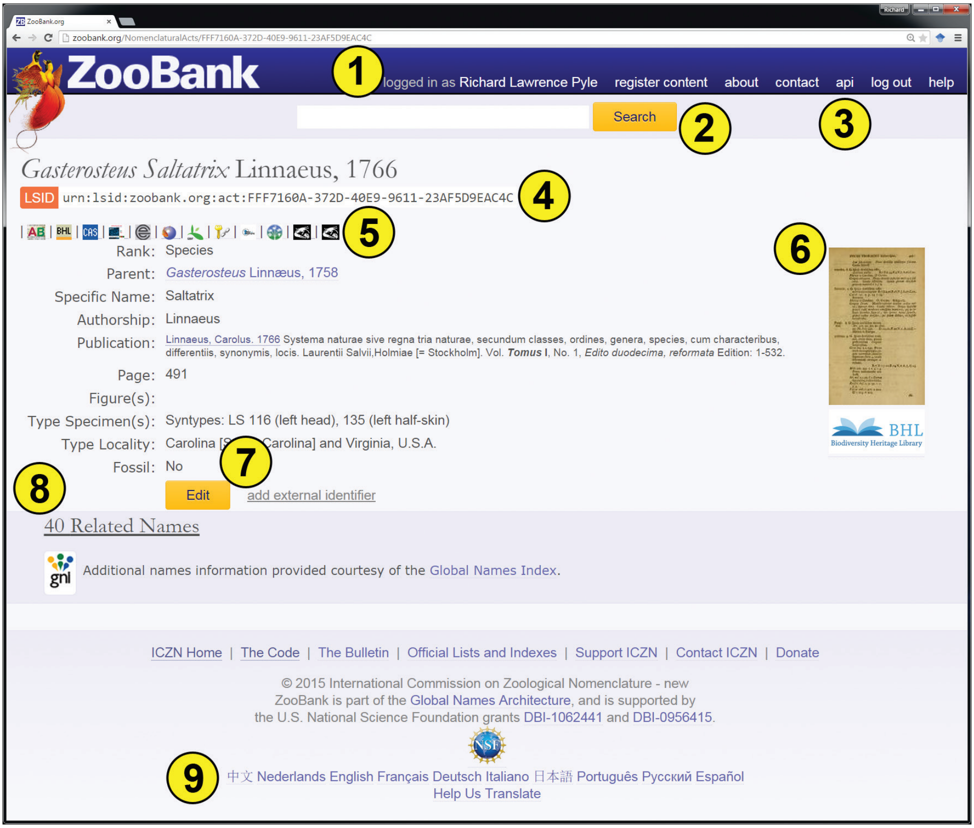


Figure 9. An example ZooBank page, illustrating several GNUB services: **1** user authentication **2** "fuzzy" searching of GNUB content **3** APIs and services **4** ZooBank registration **5** External Identifier cross-linking **6** BHL page linking **7** record editing capabilities **8** similar/related name discovery (via GNI's name searching service); and **9** multi-lingual support. Not shown are services to manage user accounts, de-duplicate records, prototype reconciliation tools, services for journal publishers, and visualization tools for author publication history and other statistics.

An integrative infrastructure for biodiversity data

Unlike most existing biodiversity data initiatives, the components of the GNA are not primarily intended to provide novel information; rather, the GNA includes an index of core facts (and associated data services) that are shared across all of biology. Nothing in the GNA is original or novel content; it merely represents a structured way of organizing information to facilitate broader data integration among other databases that do contain original information. Thus, the GNA does not compete with other data resources; but rather serves as a core infrastructure for cross-linking (and thereby empowering) other biological data sources.

Although the GNA is primarily intended to provide a cross-linking service between existing databases, the data model is sufficiently robust and complete that it can fulfill the primary needs of representing nomenclature, taxonomy and classification for groups that are not otherwise represented by existing databases. Thus, while its primary function is to integrate biodiversity data across multiple disparate systems, the GNA is capable of filling the gaps in taxonomic coverage for groups of organisms not already well-represented in the broader biodiversity data landscape.

Salvaging the global biodiversity library

Biodiversity is Earth's greatest Library, representing the culmination of information that has been written and re-written, edited and re-edited, over the past four billion years. We are like Kindergartners running through the Library of Congress: surrounded by vast amounts of incredibly valuable information – the genomic equivalents of the works of Homer, Shakespeare, and blueprints for a nuclear power plant and 95% efficient conversion of sunlight energy to stored chemical energy, but we are currently only able to interpret this information at the equivalent of "See Spot Run". Someday soon (within the next few decades) we will have the ability to truly understand the information in the Biodiversity Library. As we face the 6th Great Extinction event, we recognize that the Biodiversity Library is burning, so the information will be gone before we have a chance to understand its true value. Whenever a species goes extinct, it's like burning the last copy of a book. Taxonomists are the Librarians, and have perhaps the most important job of all: building the digital equivalent of the "card catalog" for the Biodiversity Library.

This audacious task was begun in the 1750s by Carl Linnaeus, and was dramatically extended by Charles Davies Sherborn 150 years later. With the advent of modern electronic information management, we are poised to achieve the vision of these two pillars of science; but we are in a race against the destruction of what we seek to document. We are the first generation in human history to understand our own impact to biodiversity, and we are very likely the last generation in a position to do anything about it. The Global Biodiversity Library is burning, and we must tirelessly continue to document the richness of form and function in nature before it is lost forever.

Conclusions

Throughout most of the history of modern taxonomy and nomenclature, the basic tasks performed by taxonomists have remained remarkably unchanged. Technology has allowed some improvements, with modern electronic information dissemination representing the most significant advancement. At this point in history, biodiversity data are being digitized at an impressive rate, but in most cases the data remain in "silos", with limited interconnectivity. As such, the accumulated digitized data cannot be used to its full potential. The most effective way to integrate disparate biodiversity datasets is through scientific names, but for many reasons, text-string names alone are not effective for this purpose. The Global Names Architecture (GNA) has been developed to provide core indexes and cross-linking services, to help leverage otherwise disparate biodiversity data. ZooBank, the official online registry of zoological nomenclature, represents only one example of how the GNA can improve interconnectivity among biodiversity data. Going forward, the priority should be to continue digitizing data, and to develop robust cross-links among existing biodiversity datasets. Global biodiversity is precious – perhaps Earth's most valuable resource – yet we have only begun to catalog its contents. With global climate change and accelerating rates of extinction, it is more important than ever to extend the work of Charles Davies Sherborn to apply to all known and yet-to-be-discovered taxa.

Sherborn himself seemed to understand the challenges of his task, many of which remain true today. In the Epilogue of *Index Animalium*, (Section 2, Part 29, pp. vi-vii), he wrote:

Now my work is finished, it may well be to glance at the difficulties met with during compilation… This want of every book and every edition has been a serious hindrance and loss of time to me while working for over forty years in the British Museum (Natural History) and though I have acquired over a thousand volumes for the libraries, gaps still remain to be filled… On the whole one has met with a generous response, but the amused smile, real apathy, or the remark 'we have no money' … have been encountered…

He was also acutely aware of the nature of evolving technology:

And now that rotography has superseded photography as regards cost, a rare tract can be reproduced in a few hours and placed on its proper shelf in any Library for a few shillings.

But most important of all, Sherborn understood the grandeur of his quest, and knew full well that it was far greater than his own personal contributions:

In conclusion I may add that the whole of my papers, Books of Reference and apparatus will remain at the Museum for my continuator and I trust that arrangements will be made for the permanent indexing of even current literature as the only true method of economizing the time of the working zoologist.

It is our responsibility as modern biologists to harness the power of new technology to continue is this all-important task of documenting biodiversity.

Acknowledgements

I wish to thank Donat Agosti, Rod Page, Ross Mounce, and especially Michael P. Taylor (http://orcid. org/0000-0002-1003-5675) for their excellent and constructive reviews of this manuscript.

References

Bates HW (1863) The naturalist on the River Amazons, a record of adventures, habits of animals, sketches of Brazilian and Indian life and aspects of nature under the Equator during eleven years of travel (volume 1). J. Murray, London [1892], 351 pp. doi: 10.5962/bhl.title.21335

Bean TH (1903) Catalogue of the fishes of New York. Bulletin New York State Museum 60: 1–784. doi: 10.5962/bhl.title.23134

Chapman AD (2009) Numbers of Living Species in Australia and the World (2nd edition). Australian Biological Resources Study, Australia. http://www.environment.gov.au/biodiversity/abrs/publications/other/species-numbers/2009/index.html

Chmielewski JG, Krayesky D (2013) General Botany Laboratory Manual. AuthorHouse, Bloomington, Indiana, 300 pp.

Costello MJ, Wilson S, Houlding B (2012) Predicting total global species richness using rates of species description and estimates of taxonomic effort. Systematic Biology 61(5): 871–883. doi: 10.1093/sysbio/syr080

Cuvier G (1816) Le règne animal distribué d'après son organisation, pour servir de base à l'histoire naturelle des animaux et d'introduction à l'anatomie comparée. Les reptiles, les poissons, les mollusques et les annélides [Vol. 2, Edition 1]. Déterville, Paris, 532 pp. doi: 10.5962/bhl.title.49223

Fontaine B, van Achterberg K, Alonso-Zarazaga MA, Araujo R, Asche M, Aspöck H, Audisio P, Aukema B, Bailly N, Balsamo M, Bank RA, Belfiore C, Bogdanowicz W, Boxshall G, Burckhardt D, Chylarecki P, Deharveng L, Dubois A, Enghoff H, Fochetti R, Fontaine C, Gargominy O, Lopez MSG, Goujet D, Harvey MS, Heller K-G, van Helsdingen P, Hoch H, de Jong Y, Karsholt O, Los W, Magowski W, Massard JA, McInnes SJ, Mendes LF, Mey E, Michelsen V, Minelli A, Nafria JMN, van Nieukerken EJ, Pape T, de Prins W, Ramos M, Ricci C, Roselaar C, Rota E, Segers H, Timm T, van Tol J, Bouchet P (2012) New species in the Old World: Europe as a frontier in biodiversity exploration, a test bed for 21st century taxonomy. PLoS ONE 7: e36881. doi: 10.1371/journal.pone.0036881

Günther ACLG (1860) Catalogue of the acanthopterygian fishes in the collection of the British Museum [volume 2]. Squamipinnes, Cirrhitidae, Triglidae, Trachinidae, Sciaenidae, Polynemidae, Sphyraenidae, Trichiuridae, Scombridae, Carangidae, Xiphiidae. Taylor and Francis, London, 548 pp. doi: 10.5962/bhl.title.8321

Hedgpeth JW (1961) Taxonomy: Man's Oldest Profession. University of the Pacific Research Lecture, Stockton (California), 18 pp.

Heidorn PB (2008) Shedding light on the dark data in the long tail of science. Library Trends 57: 280–299. doi: 10.1353/lib.0.0036

ICZN (1999) International Commission on Zoological Nomenclature [4th edition]. The International Trust for Zoological Nomenclature, London, 306 pp.

ICZN (2012a) 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. doi: 10.3897/zookeys.219.3944

ICZN (2012b) Amendment of Articles 8, 9, 10, 21 and 78 of the International Code of Zoological Nomenclature to expand and refine methods of publication. Zootaxa 3450: 1–7.

IISE (2010) International Institute for Species Exploration. SOS – State of Observed Species Report. http://species.asu.edu/SOS_2010

Jordan DS, Evermann BW (1903) Descriptions of new genera and species of fishes from the Hawaiian Islands. Bulletin of the US Fish Commission 22(1902): 161–208.

de Lacépède BGÉ (1801) Histoire naturelle des poissons [Vol. 3]. Plasson, Paris, 558 pp. doi: 10.5962/bhl.title.6882

de Lacépède BGÉ (1802) Histoire naturelle des poissons [Vol. 4]. Plasson, Paris, 728 pp. doi: 10.5962/bhl.title.6882

Lapage SP, Sneath PHA, Lessel EF, Skerman VBD, Seeliger HPR, Clark WA (1990) International code of nomenclature of Bacteria. ASM Press, Washington, DC, 232 pp.

Linnaeus C (1736) Fundamenta Botanica, quae Majorum Operum Prodromi instar Theoriam Scientiae Botanices by breves Aphorismos tradunt. Salomon Schouten, Amsterdam, 40 pp.

Linnaeus C (1753) Species plantarum, exhibentes plantas rite cognitas, ad genera relatas, cum differentiis specificis, nominibus trivialibus, synonymis selectis, locis natalibus, secundum systema sexuale digestas. Laurentii Salvii, Holmiae [= Stockholm], 560 pp.

Linnaeus C (1758) Systema naturae per regna tria naturae, secundum classes, ordines, genera, species, cum characteribus, differentiis, synonymis, locis. Laurentii Salvii, Holmiae [= Stockholm], 824 pp.

Linnaeus C (1766) Systema naturae sive regna tria naturae, secundum classes, ordines, genera, species, cum characteribus, differentiis, synonymis, locis. Edito duodecima, reformata. Laurentii Salvii, Holmiae, 532 pp. doi: 10.5962/bhl.title.37256

McNeill J, Barrie FR, Buck WR, Demoulin V, Greuter W, Hawksworth DL, Herendeen PS, Knapp S, Marhold K, Prado J, Prud'homme Van Reine WF, Smith GF, Wiersema JH, Turland NJ (2012) International code of nomenclature for algae, fungi, and plants (Melbourne Code), Adopted by the Eighteenth International Botanical Congress Melbourne, Australia, July 2011. Koeltz Scientific Books, Koenigstein, 240 pp.

Mora C, Tittensor DP, Sina Adl S, Simpson AGB, Worm B (2011) How many species are there on Earth and in the Ocean? PLoS Biology 9(8): e1001127. doi: 10.1371/journal.pbio.1001127

Patterson DJ, Cooper J, Kirk PM, Pyle RL, Remsen DP (2010) Names are key to the big new biology. Trends in Ecology and Evolution 25: 686–691. doi: 10.1016/j.tree.2010.09.004

Polaszek A, Agosti D, Alonso-Zarazaga M, Beccaloni G, Bjørn PdP, Bouchet P, Brothers DJ, Earl of Cranbrook, Evenhuis N, Godfray HCJ, Johnson NF, Krell F-T, Lipscomb D, Lyal CHC, Mace GM, Mawatari S, Miller SE, Minelli A, Morris S, Ng PKL, Patterson DJ, Pyle RL, Robinson N, Rogo L, Taverne J, Thompson FC, van Tol J, Wheeler QD, Wilson EO (2005) Commentary: a universal register for animal names. Nature 437: 477. doi: 10.1038/437477a

Prévost D'Exiles AF (1746-1789) Histoire générale des voyages, ou nouvelle collection de toutes les relations de voyages par mer et par terre qui ont été publiées jusqu'à présent dans les différentes langues de toutes les nations connues. Didot, Paris, 24 pp.

Prévost D'Exiles AF (1746) Histoire générale des voyages. Paris.

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

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

Pyle RL, Michel E (2008) ZooBank: Developing a nomenclatural tool for unifying 250 years of biological information. Zootaxa 1950: 39–50. http://www.mapress.com/zootaxa/2008/f/zt01950p050.pdf

Pyle RL, Michel E (2009) Unifying nomenclature: ZooBank and Global Names Usage Bank. Bulletin of Zoological Nomenclature 66: 298.

Pyle RL, Michel E (2010) ZooBank: reviewing the first year and preparing for the next 250. In: Polaszek A (Ed.) Systema Naturae 250. The Linnaean Ark. CRC Press, Boca Raton, FL, 173–184. http://www.routledge.com/books/details/9781420095012/

Ramsbottom J (1938) Linnaeus and the species concept. Proceedings of the Linnaean Society of London 150: 192–220. doi: 10.1111/j.1095-8312.1938.tb00181

Regan CT (1909) The species of three-spined sticklebacks (*Gastrosteus*). Annals and Magazine of Natural History 8(4): 435–437. doi: 10.1080/00222930908692694

Regan CT (1926) Organic evolution. Report of the ninety-third meeting of the British Association for the Advancement of Sciience, Office of the British Association, London, 75–86.

Rosenberg G, Krell F-T, Pyle RL (2012) Nomenclature: Call to register new species in ZooBank. Nature 491(7422): 40. doi: 10.1038/491040b

Thessen AE, Patterson DJ (2011) Data issues in the life sciences. ZooKeys 150: 15–51. doi: 10.3897/zookeys.150.1766

Tindall BJ (2009) Registration of names: the bacteriological experience. Bulletin of Zoological Nomenclature 66(2): 117–119. http://www.bionomenclature.net/documents/BZN_66_2_Registration_of_names_article.pdf

Trontelj P, Fišer C (2009) Cryptic species should not be trivialized. Systematics and Biodiversity 7: 1–23. doi: 10.1017/s1477200008002909

Wheeler Q, Meier R (2000) Species concepts and phylogenetic theory: A debate. Columbia University Press, New York, 256 pp.

Wilkins JS (2009) Defining Species: A Sourcebook from Antiquity to Today. Peter Lang International Academic Publishers, New York, 238 pp.

ZooKeys 550: 283–298 (2016)
doi: 10.3897/zookeys.550.10042
http://zookeys.pensoft.net

RESEARCH ARTICLE

A peer-reviewed open-access journal
ZooKeys
Launched to accelerate biodiversity research

Manual for proposing a Part of the List of Available Names (LAN) in Zoology

Miguel A. Alonso-Zarazaga[1], Philippe Bouchet[2],
Richard L. Pyle[3], Nikita Kluge[4], Daphne Gail Fautin[5]

1 *Depto. de Biodiversidad y Biología Evolutiva, Museo Nacional de Ciencias Naturales, Jose Gutierrez Abascal 2, E-28006 Madrid, SPAIN* **2** *Laboratoire de Biologie des Invertébrés Marins et Malacologie, Muséum National d`Histoire Naturelle, 55, rue Buffon, 75005 Paris, France* **3** *Natural Sciences, Bishop Museum, Honolulu, 1525 Bernice Street, 96817 Honolulu, USA* **4** *Saint-Petersburg State University, Saint-Petersburg, Russia* **5** *Ecology and Evolutionary Biology, and Natural History Museum, University of Kansas, 1200 Sunnyside Avenue, 66045 Lawrence, Kansas, USA*

Corresponding author: *Miguel A. Alonso-Zarazaga* (zarazaga@mncn.csic.es)

Academic editor: *E. Michel* | Received 25 May 2015 | Accepted 25 May 2015 | Published 7 January 2016

http://zoobank.org/24689AE4-C77C-488F-8E27-E2B1D1618E80

Citation: Alonso-Zarazaga MA, Bouchet P, Pyle RL, Kluge N, Fautin DG (2016) Manual for proposing a Part of the List of Available Names (LAN) in Zoology. In: Michel E (Ed.) Anchoring Biodiversity Information: From Sherborn to the 21st century and beyond. ZooKeys 550: 283–298. doi: 10.3897/zookeys.550.10042

1. Introduction

Article 79 of the Fourth Edition of the International Code of Zoological Nomenclature (henceforth Code) describes an official List of Available Names in Zoology (henceforth LAN), consisting of a series of "Parts" (of defined taxonomic and temporal scope), compiled by relevant experts. The LAN represents a comprehensive inventory of names available under the Code. The aim of this manual is to define a procedure for implementing Article 79, with format suggestions for zoologists aiming to create a Part of the LAN for family-group, genus-group, or species-group names in zoological nomenclature. Because the LAN may serve as an important basis for retrospective content in ZooBank, the structure outlined here is designed to allow easy importation to ZooBank.

A Part ultimately adopted for the LAN will contain nomenclaturally available names but not necessarily all those within the scope of the Part: the comprehensiveness of the candidate Part is at the discretion of the experts proposing the Part. They may choose to include all nomenclaturally available names or use the proposal of a Part to pare away *nomina dubia* so they lose “status in zoological nomenclature despite any previous availability” (to quote Articles 10.7 and 79.4.3; that this was the intention of the framers of Article 79 is clear from the Preface to the Code). Nonetheless, we advocate that the proposing body include an inventory of all known names deemed to be available so it will be obvious that names not advocated for inclusion in the Part have not simply been overlooked. Because a candidate Part of the LAN is for an entire taxon at the specified rank and for the specified period, it must include the names of both living and fossil representatives of the taxon.

In the proposal for adding a Part to the LAN, an unavailable name corresponding to a later available one should be included in the Remarks section of the available name. Unavailable names that have not subsequently been made available can be added at the end of the candidate Part, along with information explaining them. The Commission and reviewers of the candidate Part will thereby have a list of such names and an understanding of why they are not available. Moreover, these names can be discussed during the periods required by Article 79 for input by the zoological community, when change in their status can be advocated by members of the community interested in the taxon under consideration.

2. Code authority

The consideration of candidate Parts of the LAN is specified by some articles of the Code as well as the Bylaws and Constitution of the International Commission on Zoological Nomenclature (henceforth the Commission). Parallel with this are provisions for the treatment of the Official Lists and Indexes. The relevant portions are reproduced below.

2.1. List of Available Names, Official Lists and Indexes

2.1.1. In the Preface of the Code

“We can anticipate that zoologists and other users of scientific names will before long require still further changes in the Code, perhaps especially concerning procedures for the listing of existing names and the registration of new ones. With regard to the former, extensive databases are now appearing in quick succession and are being consolidated by such enterprises as Species 2000, and this fourth edition of the Code has already taken a significant step through the provisions for the development and adoption of List(s) of Available Names in Zoology.”

“Progress is made in this edition to establish a mechanism to facilitate access to previously established names, and to achieve certainty that searches made for names are complete, by enabling international groups of specialists to compile lists of extant and known available names in major taxonomic fields, and to have these lists adopted by the Commission. Names not in a relevant adopted List would not be available. A similar policy has already been adopted for all genera and species in microbiology, where neither past nor new names are available unless they have been officially recorded."

"Lists of Available Names

15. The Commission is empowered, with safeguards, to adopt lists of names in major taxonomic fields. Names within the scope of such an adopted list but not listed in it will be treated as unavailable. Lists may only be adopted by the Commission which have been proposed by international bodies, and only after publication of the proposals, wide consultation with specialist committees and others, and taking into account public comment."

2.1.2. In Articles of the Code

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].

57.2. Primary homonyms. Identical species-group names established for different nominal taxa when originally combined with the same generic name (see also Articles 11.9.3.2 and 57.8.1) are primary homonyms [Art. 53.3] and the junior name is permanently invalid (but see Article 23.9.5) except when:

- 57.2.1. its use as a valid name (a nomen protectum) is maintained under the conditions specified in Article 23.9, or
- 57.2.2. it is conserved by the Commission under Article 81, or
- 57.2.3. it, but not its senior homonym, is included in a relevant adopted Part of the List of Available Names in Zoology (see Article 79.4.3).
- 78.2.1. The Commission may, under procedures specified in Article 79, establish a List of Available Names in Zoology and may adopt Parts of the List (for the status of names in the List of Available Names in Zoology, and the name-bearing types of the nominal taxa the names denote, see Article 79.4).

78.4. Other duties. The Commission shall

- 78.4.2. enter in the relevant Official Lists and Indexes the names and works that have been the subject of rulings by the Commission in its Opinions (including Official Corrections);

Article 79. List of Available Names in Zoology. An international body of zoologists (such as an International Congress, an international society, or a consortium of national or regional societies, or a Scientific Member of the International Union of Biological Sciences) in consultation with the Commission may propose that the Commission adopt for a major taxonomic field (or related fields) a Part of the List of Available Names in Zoology. The Commission will consider the proposal and may adopt the Part subject to the proposing body and the Commission meeting the requirements of this Article.

79.1. Form of the proposal. The proposal to the Commission shall be made in the form of the Part proposed for adoption and shall
- 79.1.1. specify the scope of the proposal, such as the taxonomic field, ranks, and time period covered, (e.g. Amphibia, Names of the Species Group established before 31 December 1995 [full date, i.e. day, month, year]);
- 79.1.2. for each name to be listed, give the bibliographic reference to the work in which it is established, its authorship, its date of publication and its status (including its precedence if this is different from its priority);
- 79.1.3. for each name to be listed, give details of the name-bearing type of the nominal taxon it denotes; in the case of a species-group name, if the details of how the type specimen(s) may be recognized are not known, state whether the name is based on a holotype, syntypes, lectotype or neotype and the place(s) of deposition (if any) recorded in the type fixation (but no lectotype or neotype designation can be made for the purposes of listing alone [Arts. 74.7, 75.3]);
- 79.1.4. for any name to be listed which has been the subject of a Commission ruling [Arts. 80, 81], give the relevant Opinion and the status of the name as ruled therein; and
- 79.1.5. if applicable, specify how homonymy with names beyond the scope of the proposal has been resolved.

79.2. Requirements concerning notification, consultation and voting by the Commission.
- 79.2.1. Upon being advised by an international body of zoologists that it intends to propose a Part of the List, the Commission shall appoint by its Council an *ad hoc* committee [Constitution Art. 10] to consult with the proposers.
- 79.2.2. Upon receipt of a proposal the Commission shall
  - 79.2.2.1. publish a notice of the proposal in the Bulletin of Zoological Nomenclature giving details of the proposing body, proposed scope of the Part and a source from which copies (on paper or otherwise) of the proposed Part may be obtained by zoologists, and inviting comments from zoologists during the following twelve months;
  - 79.2.2.2. submit the notice for publication in journals publishing taxonomic work in the taxonomic field covered by the proposal;
  - 79.2.2.3. refer the proposal to its *ad hoc* committee for it to receive comments, consult with the proposers and others and, not less than two years from the date of publication of the notice referred to in Article 79.2.2.1, consider either a revised proposal or a recommendation that the proposal be abandoned;
  - 79.2.2.4. ensure that the revised proposal does not contain any name established less than five years before the submission of the initial proposal;

79.2.2.5. following receipt of the revised proposal from its *ad hoc* committee, publish notice of it and invite comments on the revised proposal in the same manner as for the initial proposal [Arts. 79.2.2.1, 79.2.2.2];

79.2.2.6. take into account comments received (if any) and comments of the proposers thereon, and vote to adopt the Part proposed or to abandon the proposal, under procedures prescribed in the Constitution [Art. 12] and the Bylaws of the Commission for voting under its plenary power.

79.3. Effective date of Parts and their accessibility. The Commission shall publish a notice in the Bulletin of Zoological Nomenclature of a decision to adopt any Part of the List of Available Names in Zoology as soon as possible after the decision is taken.

79.3.1. Before publishing the notice of adoption, the Commission shall satisfy itself that the Part newly adopted is accessible either by purchase or gratis and shall include that information in the notice.

79.3.2. Any Part of the List of Available Names in Zoology adopted by the Commission becomes effective from the date of publication in the Bulletin of Zoological Nomenclature of a notice of the decision of the Commission to adopt it.

79.3.3. The notice shall specify the title under which the Part of the List adopted by the Commission shall be known and its scope (including the taxonomic field and dates covered).

79.4. Status of names, spellings, dates of availability, and types specified in the List of Available Names in Zoology.

79.4.1. A name occurring in an adopted Part of the List of Available Names in Zoology is deemed be an available name and to have the spelling, date, and authorship recorded in the List (despite any evidence to the contrary).

79.4.2. A nominal taxon denoted by a name occurring in an adopted Part of the List of Available Names in Zoology is deemed to have the name-bearing type recorded therein (despite any evidence to the contrary).

79.4.3. No unlisted name within the scope (taxonomic field, ranks, and time period covered) of an adopted Part of the List of Available Names in Zoology has any status in zoological nomenclature despite any previous availability.

Recommendation 79A. Citation of previously available names. If for taxonomic and historical purposes an author desires to cite a name that is no longer available because it is not included in the relevant Part of the List of Available Names in Zoology adopted by the Commission, it should be made clear that it no longer has a status in zoological nomenclature.

79.5. Power of the Commission to amend the status of a name occurring in the List of Available Names in Zoology. If there are exceptional circumstances and only when an entry in the List of Available Names in Zoology is a cause of confusion, the Commission may amend the entry by use of its plenary power [Art. 81] and publish its ruling in an Opinion [Art. 80.2].

79.5.1. From the date of the publication in the Bulletin of Zoological Nomenclature of the amended entry the relevant name has the status, spelling, date of availability, and authorship, and the nominal taxon it denotes has the name-bearing type, as shown in the amended entry.

79.5.2. The requirement that amendments to the status of names occurring in the List may be made only by the Commission using its plenary power does not prevent an author from designating a type species for a nominal genus-group taxon published before 1931, if one has not already been fixed, or from designating a lectotype [Art. 74] from syntypes recorded in the List of Available Names in Zoology, or a neotype when circumstances exist that require neotype designation [Art. 75]. Such subsequent fixations may be inserted by the Commission in the List.

Recommendation 79B. Request to authors designating lectotypes or neotypes for names in the List of Available Names in Zoology. Authors are requested to inform the Commission of lectotype or neotype designations made by them for the nominal taxa of names in the List of Available Names in Zoology as soon as possible after publication.

79.6. Power of the Commission to add omitted names to the List of Available Names in Zoology. If the Commission determines that there is a previously available name within the scope of an adopted Part of the List of Available Names in Zoology that has been omitted from the List, in exceptional circumstances the Commission may by use of the plenary power add an appropriate entry to that Part of the List and record this in an Opinion. The availability of the name thereby becomes restored.

Article 80. Status of actions of the Commission. As a consequence of actions required of it by the Code, the Commission may publish Declarations, Opinions, the Official Lists and Official Indexes, and may adopt and publish Parts of the List of Available Names in Zoology. The status of these published acts, and of names and works in the Official Lists and Official Indexes, is specified in this Article.

80.4. Corrections of errors or omissions in Opinions. Official Corrections to errors and omissions (such as a bibliographic error, lapsus calami, or an omission in placing a conserved or suppressed name on an Official List or Index) may be published by the Commission without further vote unless the error or omission negates the ruling or its consequences. If the ruling is negated by the error or omission, the Commission shall reconsider the matter and publish a further Opinion.

80.6. Status of works, names and nomenclatural acts in Official Lists. The Commission publishes the effects of its Opinions on individual names and works in the Official Lists and Official Indexes. In the case of names and works in the Official Lists:

80.6.1. A name entered in an Official List is an available name.

80.6.2. The status of a name entered in an Official List is subject to the ruling(s) in any relevant Opinion(s), including any Official Correction of an Opinion [Art. 80.4]; all other aspects of its status derive from the normal application of the Code. However, if such a name is given a different status in the List of Available Names in Zoology the latter status is deemed to be correct [Art. 80.8].

80.6.3. A name may be placed in an Official List without any additional qualification.

80.6.4. If a name entered in an Official List is thought to be a synonym of another available name (whether in an Official List or not), their relative precedence is determined by the normal application of the Code unless the Commission rules or has ruled otherwise.

80.6.5. A name or nomenclatural act occurring in a work entered in the Official List of Works Approved as Available for Zoological Nomenclature is subject to the provisions of the Code and to any limitation imposed by the Commission on the use of that work in zoological nomenclature.

80.7. Status of works, names and nomenclatural acts in Official Indexes. The Commission publishes the effects of its Opinions on individual names and works in the Official Lists and Official Indexes. In the case of names and works in the Official Indexes:

80.7.1. A work, name or nomenclatural act entered in an Official Index has the status attributed to it in the relevant ruling(s).

80.7.2. A name or nomenclatural act occurring in a work entered in the Official Index has no availability or validity in zoological nomenclature, unless the Commission by use of its plenary power rules otherwise. However, such a work may be used as a source of information relevant to zoological nomenclature unless the Commission has ruled that the work is to be treated as unpublished.

80.8. Contradictory status accorded by the Commission to names in the List of Available Names in Zoology and in the Official Lists. In the event of contradictory status being accorded by the Commission to a name included in the List of Available Names in Zoology, in an Official List, or in an Opinion, the status accorded in the List of Available Names in Zoology is deemed to be correct unless the Commission has ruled otherwise [Art. 79.5].

2.1.3. In the Glossary

adoption, n. Of a Part of the List of Available Names in Zoology: the acceptance of the Part by the Commission as specified in Article 79.

Index, Official. See Official Index.

List of Available Names in Zoology, n. The cumulative term for those parts of the List of Available Names in Zoology which have been adopted by the Commission under Article 79.

List, Official. See Official List.

Official Index, n. An abbreviated title for any of the four Indexes, maintained and published by the Commission, citing works or names that have been rejected by rulings of the Commission. For the status of names cited in the Indexes, and of names and nomenclatural acts in works cited in the Indexes, see Article 80.7. The full titles of the Indexes are:

Official Index of Rejected and Invalid Works in Zoological Nomenclature.
Official Index of Rejected and Invalid Family-Group Names in Zoology.
Official Index of Rejected and Invalid Generic Names in Zoology.
Official Index of Rejected and Invalid Specific Names in Zoology.

Official List, n. An abbreviated title for any of the four Lists, maintained and published by the Commission, citing available works or names that have been ruled upon in the Opinions of the Commission. For the status of works, names, and nomenclatural acts in the Lists see Article 80.6. The full titles of the Lists are:

Official List of Works Approved as Available for Zoological Nomenclature.
Official List of Family-Group Names in Zoology.
Official List of Generic Names in Zoology.
Official List of Specific Names in Zoology.
(See also List of Available Names in Zoology).

Part of the List of Available Names in Zoology, n. (q.v.). A list, adopted by the Commission under Article 79, of available names in a major taxonomic field.

proposal, n. (1) An action, whether successful or unsuccessful, to establish a nominal taxon or name or to carry out a nomenclatural act (q.v.). (2) An application to the Commission under Article 79 for the adoption of a Part of the List of Available Names in Zoology.

rejected work. Any work included by the Commission in the Official Index of Rejected and Invalid Works in Zoological Nomenclature.

2.2. Official Lists in the Bylaws

Mention in the Bylaws, which is under the heading "The Secretariat", states:

23. The duties of the Secretariat are:
    (b) To prepare and edit for publication the Bulletin of Zoological Nomenclature, successive instalments of the official lists and indexes (Constitution Art. 14c), and editions of the Code, Constitution and Bylaws.

2.3. Official Lists in the Constitution

Mentions in the Constitution, which are under the heading of Art. 14 "Editorial duties of the Commission", state:

14.3. Maintenance of Official Lists and Indexes. The Commission shall compile and maintain the under mentioned Lists and Indexes:
14.3.1. Official List of Family-Group Names in Zoology;
14.3.2. Official List of Generic Names in Zoology;
14.3.3. Official List of Specific Names in Zoology;
14.3.4. Official Index of Rejected and Invalid Family-Group Names in Zoology;
14.3.5. Official Index of Rejected and Invalid Generic Names in Zoology;
14.3.6. Official Index of Rejected and Invalid Specific Names in Zoology;
14.3.7. Official List of Works Approved as Available for Zoological Nomenclature;
14.3.8. Official Index of Rejected and Invalid Works in Zoological Nomenclature.
14.4. *List of Available Names in Zoology*. The Commission may consider, adopt and publish notices concerning the List of Available Names in Zoology (or Parts thereof) as prescribed in Article 79 of the Code.

3. Philosophy

Although it is clear from several Articles of the Code that the LAN and the Official Lists and Indexes are different entities, neither it nor the Bylaws and Constitution are very explicit about the differences and the relationship between them. With the help of the Glossary, we make the following clarifications:

3.1. LAN

An inventory of names in the Family-, Genus- and Species-groups formed by accumulating Parts as stipulated under Article 79 of the Code and adopted by the Commission. The Code does not provide for an equivalent List for Works.

3.2. Official Lists

These four Lists are compiled from the Opinions of the Commission and its use of the Plenary Power. One is related to Works (it is the only way to make available works that otherwise could be doubtfully considered as available). Three others are related to Names and Nomenclatural Acts (one each for Species, Genera, and Families) whose status is ruled by Article 80.6. Their counterparts are the *Official Indexes*, which cite works or names that have been rejected by rulings of the Commission.

3.3. Duties of the Commission

The Commission:

a) shall compile and maintain the *Official Lists* (and *Indexes*) (Constitution Art. 14.3); enter the names and works that have been the subject of rulings by the Commission in its Opinions (including Official Corrections) (Code Art. 78.4.2);
b) may establish a LAN (Code Art. 78.2.1);
c) may adopt Parts of it for a major taxonomic field, when proposed by international bodies of zoologists in consultation with the Commission (Code Art. 79);
d) shall appoint by its Council an *ad hoc* committee to consult with the proposers (Code Art. 79.2.1, Constitution Art. 10); refer the proposal to its *ad hoc* committee for it to receive comments, consult with the proposers and others and, not less than two years from the date of publication of the notice, consider either a revised proposal or a recommendation that the proposal be abandoned (Code Art. 79.2.2.3);
e) shall consider and publish notices concerning the LAN (or Parts thereof) (Constitution Art. 14.4, Code Art. 79):
f) shall invite comments on the revised proposal in the same manner as for the initial proposal (Code Arts. 79.2.2.1, 79.2.2.2, 79.2.2.5);
g) shall take into account comments received (if any) and comments of the proposers thereon, and vote to adopt the Part proposed or to abandon the proposal, under procedures prescribed in the Constitution [Art. 12] and the Bylaws of the Commission for voting under its plenary power.

3.4. Status of names in the LAN

The status of the names in the LAN or its Parts are covered by Article 79.4 of the Code. Names can be amended in (Article 79.5), added to (Article 79.6), or deleted from the LAN by using the Plenary Power and publishing an Opinion. The status of names in the LAN supersede those in the *Official Lists* (cf. Articles 80.6.2, 80.8). Within the scope of period, rank, and taxon for which a Part has been adopted, names not listed are not available (Articles 10.7, 79.4.3).

4. Structure of a proposal

4.1. Introduction

A proposal for a Part of the LAN should have a Title and an inventory or two inventories of nomenclaturally available names within the scope of the candidate Part, each accompanied by the information required in Articles 79.1.2 through 79.1.5. Whether the names are in one inventory or two separate ones, two categories of names must be clearly distinguished:

1) names proposed for adoption as Part of the LAN; and,
2) names proposed not to be included in the Part, with reasons for their lack of inclusion addressed.

The *ad hoc* committee formed under Article 79.2.1 will determine that the candidate Part does not overlap with any other Part of the LAN (accepted or under consideration) and may propose changes in the taxonomic and temporal limits. During the periods of public discussion (Article 79.2.2.1 and 79.2.2.5), any interested zoologist may comment on the lists, or, by addressing a formal request to the *ad hoc* committee, request modifications, ask that names be transferred between categories, or indicate interest in being consulted further (Article 79.2.2.3).

In the following text, curly brackets { } have been used to include terms that should be replaced by those appropriate to the Part of the LAN being presented.

4.2. Title of proposal

Any proposal should be entitled and formatted as follows, as required by Article 79.9.1:

1. **Candidate Part of the List of Available Names in Zoology**
2. **Major taxonomic field**: The name of the taxon, followed by author and date; if needed, the including taxa.
3. **Rank-group**: {Family-, Genus-, Species-} Group Names.
4. **Time period covered**: Before {day month year}. No name established less than 5 years prior to the date of the proposal presentation can be considered (Article 79.2.2.4).
5. **Prepared by**: Authors of the list must provide names, postal addresses, and e-mail addresses.
6. **Presented by**: The international body under the aegis of which this candidate Part of the List was prepared.
7. **Adopted**: Date in which the international bodies presenting the proposal endorsed it (and optionally, place, if in a meeting or congress).
8. **Summary**: An overview of the taxonomic field covered, and possible alternative interpretations if differing taxonomic systems exist (including an overview of related taxa that might be excluded or included in alternative systems). This summary must include the taxonomic criteria followed in compiling the list, citing references or sufficient details to be clear, if the criteria are new or differ considerably from those published. The number of taxa proposed for adoption as Part of the LAN, and the number proposed not to be in the Part (as per § 4.1 above).

4.3. Family-group names

For the purpose of compiling a Part of the LAN, a family-group name is any uninominal, plural noun or nominalized adjective (Arts. 4.1, 11.7.1, 29). It may be formed by adding to the stem of a valid type genus name any suffix and/or ending (Arts. 11.7, 29, 63, 64), even if not standard under the Code (Arts. 11.7, 29); originally it can be attributed to any rank included in the family- group in the system followed by its author or have an unnamed rank within it (Art. 35). In publications issued when the family group did not exist (e.g., prior to 1802) or in those where family group names are not accepted, any rank above genus group may be regarded as belonging to the family group. Names clearly attributed to a family-group rank but not based on a type genus name must also be considered, although only for rejection.

The fields of a candidate Part of the LAN in the Family-Group are:

1. **Stem**: The proposed stem to be followed by standard (and non-standard) endings, according to Article 29 of the Code. The stem must be written in small caps, and capitalized.
2. **Original spelling**: The spelling as originally written, including diacritics, if present. Use small caps, and capitalize. If in the same work a name appears in Latinized and vernacular forms, the Latinized form is preferred. If the name appears in different ranks, the highest rank within the family-group has precedence (Article 24.1) as the origin; others must be mentioned in the Remarks field. Care should be taken not to confuse faulty (or non-standard) Latin with vernacular spelling. Consider also Articles 32.5.3.2 and 35.4.1 for names formed from incorrect subsequent spellings or unjustified emendations (but see Article 35.4.2).
3. **Author**: The name(s) of the author(s) of the taxon.
4. **Year**: The year of effective publication.
5. **Reference**: In the format of the Bulletin of Zoological Nomenclature and the Official Lists and Indexes of Names and Works in Zoology, Supplement 1986-2000.
6. **Rank**: The original explicit rank, if any. If not explicit or the rank is not a standard one (family, tribe; with prefixes: super-, sub-, infra-; prefixes may be combined), state: “Unnamed rank intended to be between {rank_1} and {rank_2}” or “Unnamed rank intended to be immediately above or below {rank}”. Non-standard rank names can be used instead of “Unnamed rank”. If a name given to a phylogenetic clade is also available for nomenclatural purposes under the present Code, it is to be stated as: “Clade child of {clade name_1} and sister to {clade name_2}”.
7. **Type genus**: The valid name of the type genus in **bold font**. If the original spelling was based on an unjustified emendation or on an incorrect subsequent spelling, that must be given in the Remarks field.
8. **Type genus author**: The name(s) of the author(s) of the genus name.
9. **Type genus year**: The year of effective publication of the genus name.
10. **Type genus reference**: In consistent format (see 5 above).

11. **Type genus rank in original description**: Originally described as genus or subgenus of {Genus Author, Year}.
12. **Latinization**: Reference to first latinization for an originally vernacular name, quoting the first author to latinize it and the reference. If original, state "Original". If not so, following Art. 11.7.2, it should be added: "Generally accepted as valid by authors interested in the group concerned and as dating from that first publication in vernacular form". Note that first latinization has not to be standard, according to the present Code.
13. **Qualification**: Include statements about availability and precedence. First Revisers' actions regarding the precedence of names published at the same time must be quoted here, including data on precedence if different from priority (Article 79.1.2). Include information on how homonymy with names beyond the scope of the proposal has been resolved, if applicable (Article 79.1.5). If relevant, statements about the application of Articles 13.2.1 and 40 to the name are expected.
14. **Previous rulings**: Quote in full any ruling (Opinion or Direction) by the Commission on the name and how it is affected by them, including placement on Official Lists or Indexes (Article 79.1.4).
15. **Remarks**: Add any comment to clarify the status of the name or other relevant data.

4.4. Genus-group names

For the purpose of compiling a candidate Part of the LAN, a genus-group name is any uninominal, singular noun or nominalized adjective, regarded to be Latin or latinized, corrected to nominative singular if given in any case other than this because of the requirements of Latin grammar in Latin texts, originally proposed either for a taxon at the rank of genus or subgenus (Article 42.1), or for a taxon attributed to any rank included in the genus-group in the system followed by its author, or having an unnamed rank within it in terms of the Code.

The fields of a candidate Part of the LAN in the Generic Group are:

1. **Name**: Use **bold font**, and capitalize. Do not use diacritics and other marks; incorrect original spellings must be given in the Remarks field.
2. **Author**: The name(s) of the author(s) of the taxon.
3. **Year**: The year of effective publication.
4. **Reference**: In the format of the Bulletin of Zoological Nomenclature and the Official Lists and Indexes of Names and Works in Zoology, Supplement 1986-2000.
5. **Gender**: Masculine, feminine, or neuter.
6. **Kind of type species designation**: By original designation, by monotypy, by absolute tautonymy, by Linnaean tautonymy, by subsequent designation, by subsequent monotypy, or under the plenary power.

7. **Reference if subsequent**: If the type species designation was subsequent or made under the plenary power, the data for this designation must be provided in consistent format.
8. **Rank**: The original explicit rank, if any. If not explicit or the rank is not standard (genus, subgenus), a statement should be included such as: "Unnamed rank intended to be between {rank_1} and {rank_2}" or "Unnamed rank intended to be immediately above/below {rank}". Non-standard rank names can be given instead of "Unnamed rank". If a name given to a phylogenetic clade is also available for nomenclatural purposes, it is to be stated as: "Clade child of {clade name_1} and sister to {clade name_2}".
9. **Type species**: The name of the type species, in its original combination, using **bold font**. If the original spelling was an unjustified emendation or an incorrect subsequent spelling, that must be given in the Remarks field.
10. **Type species author**: The name(s) of the author(s) of the species name.
11. **Type species year**: The year of effective publication of the species name.
12. **Type species reference**: In consistent format (see 4 above).
13. **Type species proposed valid name**: The proposed senior synonym of the type species under the taxonomic system used, in its valid combination [Genus (Subgenus) species Author, Year], if the type species is currently considered to be a junior synonym, using **bold font**.
14. **Qualification**: Include statements about availability and precedence. First Revisers' actions regarding the precedence of names published at the same time must be quoted here, including data on precedence if different from priority (Article 79.1.2). Statements about stems for the formation of family-group names are to be placed here. Include information on how homonymy with names beyond the scope of the proposal has been resolved, if applicable (Article 79.1.5).
15. **Previous rulings**: Quote in full any ruling (Opinion or Direction) by the Commission on the name and how it is affected by them, including placement on Official Lists or Indexes (Article 79.1.4).
16. **Remarks**: Add any comment to clarify the status of the name or other relevant data.

4.5. Species-group names

For the purpose of compiling a Part of the LAN, an available species-group name should be a specific name, regarded to be a Latin or latinized adjective, participle or noun in nominative singular or genitive case (Art. 11.9.1), satisfying the provisions of the Code; it must be originally published in unambiguous combination with a generic name (Art. 11.9.3) and originally proposed either for a taxon at the rank of species or subspecies (Art. 45.1), or published before 1961 for a "variety" or "form" (Art. 10.2) not deemed to be infrasubspecific (Art. 45.6).

The fields of a candidate Part of the LAN in the Specific Group are:

1. **Specific name**: The specific epithet as it should have been correct in the original publication, if any correction of the original spelling is mandatory (Articles 32.5, 33.2.2), using **bold font**. Incorrect original spellings must be given in the Remarks field.
2. **Genus name**: The original genus name as originally spelled, with corrected spelling if originally misspelled, using **bold font**. Incorrect original spellings must be given in the Remarks field.
3. **Author**: The name(s) of the author(s) of the taxon.
4. **Year**: The year of effective publication.
5. **Reference**: In the format of the Bulletin of Zoological Nomenclature and the Official Lists and Indexes of Names and Works in Zoology, Supplement 1986-2000.
6. **Rank**: The original explicit rank, if any. If not explicit or the rank is not a standard one (species, subspecies), a statement should be included such as: "Unnamed rank intended to be between {rank_1} and {rank_2}" or "Unnamed rank intended to be immediately above/below {rank}". Non-standard rank names can be given instead of "Unnamed rank". If a name given to a phylogenetic clade is also available for nomenclatural purposes, it is to be stated as: "Clade child of {clade name_1} and sister to {clade name_2}".
7. **Type specimen(s) data**: State details of how the type specimen(s) may be recognized, including, if relevant, reference to a definition or interpretation such as "designated by" or "figured by", in consistent format. If such details are unknown, it should be stated whether the name is based on a holotype, syntypes, lectotype, or neotype, and the place(s) of deposition (if any) recorded in the type fixation (but no lectotype or neotype designation can be made for the purposes of listing alone [Articles 74.7, 75.3]) (Article 79.1.3). If the type specimen(s) ha(s/ve) been the subject of an Opinion or Declaration to define the taxon, state "as defined by the {neotype, lectotype, holotype}… in {depository}" or "as interpreted by reference to the {neotype, lectotype, holotype}".
8. **Qualification**: Include statements about availability and about validity under the taxonomic system used. First Revisers' actions regarding the precedence of names published at the same time must be quoted here, including data on precedence if different from priority (Article 79.1.2). State also whether the species is the type species of a genus: "specific name of the type species of {Genus Author, Year}". Include information on how homonymy with names beyond the scope of the proposal has been resolved, if applicable (Article 79.1.5).
9. **Previous rulings**: Quote in full any ruling (Opinion or Direction) by the Commission on the name and how it is affected by them, including placement on Official Lists or Indexes (Article 79.1.4).
10. **Remarks**: Add any comment to clarify the status of the name or other relevant data.

5. References

The citations follow the standard in the Bulletin of Zoological Nomenclature.

Authors must furnish a separate list of the references used, formatted as the bibliographic sections of the Bulletin. Each reference should have a day-month-year field (as far as discernable) for publication date to help with priority issues.

Index

Note:
The index covers the text of the chapters and Appendix but not their bibliographies. The suffixes F and T following a page number indicate either a relevant figure or a mention only in a table on that page. In view of the extensive cross-citation, the names of contributing authors have been excluded from the index.