

## 7 EPISTEMIC CONFLICT

### CONFLICTS

For all the benefits that composite classifications offer the data-rich biodiversity world, such an approach has not been embraced full-scale by all taxonomists and bioinformaticians. And this is for good reason, both technical and intellectual. Stephen Thorpe's general critique of the management classification approach on the Taxacom listserv can set the stage for these critiques:

Biological classification is a mixture of scientific fact . . . and subjective opinion. . . . Both these factors taken together doesn't make life very easy, and it is all in perpetual flux. . . . However, I don't think that the issue can be "managed" in quite the way that is envisaged by some. . . . My primary governing principle is that, subject to monophyly, classification is primarily a filing system to make information management easier. So, it doesn't really matter which classification is followed, PROVIDING that it is explicitly stated which one. The problem with adopting a particular classification for a large group (like the "Protista") is that advances in taxonomy happen on much smaller subgroups, so if you blindly follow one particular broad classification, then you cannot accommodate the advances very easily. Hence, I think you have to simply treat matters on a case-by-case basis, and just choose and specify a sensible classification for that particular case (and change it, if necessary, if something more convincing is published). To try to come up with a single "officially endorsed" classification would simply be to ignore the subjectivity and fallibility of taxonomy. (2009)

In general, critiques come from two directions. On the one hand, a composite classification presents technical complications, including issues such as proliferating data error, assessing data quality, and the like. These technical concerns, while vital to the integrity of the system, are not altogether new in the realm of database work. Thorpe's view certainly seems reasonable in this respect, especially given the fact that such top-down approaches to classification are antithetical to the way taxonomy has functioned for hundreds of years (Godfray 2002). Contemporary biodiversity practice, however, does need to attend to the ever-increasing rates of data production, and large-scale questions about environmental issues and mass extinctions are driving the need for these more-integrated consensus approaches (Guralnick and Hill 2009). Additionally, maintaining numerous independent taxonomic systems is not necessarily a pragmatic approach. This said, while the world checklist system and consensus-oriented view might work at the production level, "for advanced uses, it's sh\*t," as one prominent taxonomist poetically proclaimed during one of our conversations. Finding the balance between generalized and expert system design is easier said than performed in practice.

The more serious critiques, at least as they relate to the main narrative of this book, concern the epistemic conflicts that occur, since, when multiple taxonomies are integrated into a composite space, they lose nuance and are manipulated in ways that contradict local scientific practice. The reality is that contributed taxonomies are "qualitatively transformed" as they are ingested into the compiled system (Remsen 2010). As databases enter the Catalogue of Life, the editors do their best to maintain the integrity of the original taxonomic structure, but there are times when "adjustments may need to be decided upon by the editors on where and how to insert it, to make it as consistent as possible [with the rest of the Catalogue of Life], while not losing the essential taxonomic information it has been created to provide" (Species 2000 2016a). Let us assume for one moment that we are contributing a taxonomy to the Catalogue. And let us further assume that conflicts arise between our contributed piece and the Catalogue's existing structure. One can make an argument that, however trained and expert an editorial board might be, the best person to make a determination about

the integrity of the contributed taxonomy is the original creator. Losing their intellectual agency and authority is of great concern to contributing taxonomists.

To concretize how this conflict might transpire, refer to figures 7.1 and 7.2, which compare the taxonomic hierarchy from both the combined Species 2000 and ITIS Catalogue of Life (figure 7.1) and the stand-alone Integrated Taxonomic Information System (figure 7.2), to see how this kind of editing might take place within database environments. The ITIS taxonomic hierarchy is far more detailed in its composition, particularly because it contains a more finely articulated higher-level taxonomic ranking system (inclusive of subkingdom, subphylum, infraphylum, superorder, etc.). While some might say that this information isn't vital to a species-level articulation, other might note that the existence of higher-level taxa is important for contextual reasons. In comparison, looking at the Catalogue of Life's taxonomic tree, you can see that the Catalogue editors chose the

Recognized by	Rank	Classification
<ul style="list-style-type: none"> <li>• <b>Species 2000 &amp; ITIS Catalogue of Life: April 2013</b></li> <li><a href="#">View in classification</a></li> </ul>	Species	<ul style="list-style-type: none"> <li><a href="#">Animalia</a> ±</li> <li><a href="#">Chordata</a> ±</li> <li><a href="#">Mammalia</a> ±</li> <li><a href="#">Carnivora</a> ±</li> <li><a href="#">Ursidae</a> ±</li> <li><a href="#">Ursus</a> ±</li> <li><b><i>Ursus arctos</i> Linnaeus, 1758</b></li> <li><a href="#">Ursus arctos alascensis</a> Merriam, 1896</li> <li><a href="#">Ursus arctos arctos</a> Linnaeus, 1758</li> <li><a href="#">Ursus arctos beringianus</a> Middendorff, 1851</li> <li><a href="#">Ursus arctos californicus</a> Merriam, 1896</li> <li><a href="#">Ursus arctos collaris</a> F. G. Cuvier, 1824</li> <li><a href="#">Ursus arctos crowtheri</a> Schinz, 1844</li> <li><a href="#">Ursus arctos dalli</a> Merriam, 1896</li> <li><a href="#">Ursus arctos gyas</a> Merriam, 1902</li> <li><a href="#">Ursus arctos horribilis</a> Ord, 1815</li> <li><a href="#">Ursus arctos isabellinus</a> Horsfield, 1826</li> <li>6 more... <a href="#">show full tree...</a></li> <li><a href="#">Ursus americanus</a> Pallas, 1780 ±</li> <li><a href="#">Ursus maritimus</a> Phipps, 1774</li> <li><a href="#">Ursus thibetanus</a> G. [Baron] Cuvier, 1823 ±</li> </ul>

**Figure 7.1**

Two different curated taxonomies displayed by the Encyclopedia of Life for the species *Ursus arctos* (Encyclopedia of Life 2017). CC-BY 4.0, Catalog of Life, used by permission.

<p><b>Integrated Taxonomic Information System (ITIS)</b>  <a href="#">View in classification</a></p>	Species	Animalia +
		Bilateria +
		Deuterostomia +
		Chordata +
		Vertebrata +
		Gnathostomata +
		Tetrapoda +
		Mammalia Linnaeus, 1758 +
		Theria Parker and Haswell, 1897 +
		Eutheria Gill, 1872 +
		Carnivora Bowdich, 1821 +
		Caniformia Kretzoi, 1938 +
		Ursidae Fischer de Waldheim, 1817 +
		Ursus Linnaeus, 1758 +
		<b>Ursus arctos Linnaeus, 1758</b>
		<a href="#">Ursus arctos alascensis Merriam, 1896</a>
		<a href="#">Ursus arctos arctos Linnaeus, 1758</a>
		<a href="#">Ursus arctos beringianus Middendorff, 1851</a>
		<a href="#">Ursus arctos californicus Merriam, 1896</a>
		<a href="#">Ursus arctos collaris F. G. Cuvier, 1824</a>
		<a href="#">Ursus arctos crowtheri Schinz, 1844</a>
		<a href="#">Ursus arctos dalli Merriam, 1896</a>
		<a href="#">Ursus arctos gyas Merriam, 1902</a>
		<a href="#">Ursus arctos horribilis Ord, 1815</a>
		<a href="#">Ursus arctos isabellinus Horsfield, 1826</a>
		6 more... <a href="#">show full tree...</a>
		<a href="#">Ursus americanus Pallas, 1780 +</a>
		<a href="#">Ursus maritimus Phipps, 1774</a>
		<a href="#">Ursus thibetanus G. [Baron] Cuvier, 1823 +</a>

**Figure 7.2**

The (default) classification hierarchy for the species provided by the Catalogue of Life. (Bottom) The classification hierarchy provided by ITIS. CC-BY 4.0, Catalog of Life, used by permission.

*genus* node as the connection point for the ITIS database. Note the lack of an author and publication designation after the species epithet, *Ursus*, at the genus level in the Species 2000 taxonomy. Additionally, the Species 2000/Catalogue of Life hierarchy retains the ITIS taxa detail below the genus node connection point (beginning at *Ursus arctos* Linnaeus 1758), whereas above that connection point, the Catalogue’s management classification backbone maintains authority. The act of choosing one part of the ITIS classification over any other transforms the ITIS structure in fundamental ways, divorcing *Ursus arctos* from the upper-level backbone context in its original location in the ITIS database. In all, this example is fairly generic, but other decisions can be far more detrimental.

And so are the dangers of universal and consensus classification systems: they normalize data in ways that obfuscate different epistemic points of view. Sterner, Witteveen, and Franz address this fact in their article, “Coordinating Dissent as an Alternative to Consensus Classification: Insights from Systematics for Bio-ontologies.” The authors identify three types of consensus systems that have “advantages in certain contexts and for certain ends” (2020, 2). The first is the Definitional Consensus Principle (DCP), a classification system intended to be internally coherent (in the sense Jonathan Furner notes) and one that represents consensus about the definitions of entities included in the classification. DCP is further bifurcated into two types: (1) one that adheres to ontological realism as articulated by Barry Smith and Werner Ceusters (2010) and is termed Realist Interpretation (DCP/R), and (2) one that is based on local context, termed a Contextual Interpretation (DCP/C). DCP/R maintains that to limit the production of multiple competing biological (computational) ontologies, centralized reference systems should be created that represent “settled science” (Barry Smith and Ceusters 2010, 1) as it expresses some true external reality (realist). Such a metaphysical approach would presume a uniform (philosophical) ontological understanding of universals, types, and properties (inclusive of relations) that should then be formulated into bio-computational ontologies. DCP/C, on the other hand, attends to classificatory consensus at a local level that represents that particular group’s (and potentially *only* that group’s) metaphysical or epistemic principles that, might, “for example, characterize what counts as mature or settled science” in that context (Sterner, Witteveen, and Franz 2020, 10).

In response to DCP positions, Sterner, Witteveen and Franz (2020, 7–20) indicate a consensus system should focus not on the ontological or epistemic aspects of classification (that is, to produce a taxonomy that is presumed to be *correct* in both of these senses), but rather on aspects meant to coordinate names, primarily, with the underlying taxonomic structure acknowledged as a mechanism for communication. In some ways, this approach is aligned with that of the Catalogue of Life or the GBIF Backbone taxonomy, with the exception that systems such as the Catalogue are often touted as representing, at least at the sub-consensus level, a “correct” or

“current” taxonomic arrangement. A coordinated consensus principle (CCP) approach eschews this rhetoric of correctness entirely and claims only that the names and the type it refers to—collectively known as the taxonomic concept—are valid in this space. Such a system should make no claims as to how these taxon concepts might then relate to the articulation of a species concept or larger taxonomic ecology—this interpretive work should be left to the local and contextual classification, as they always have been.

Adding to the complexity of these epistemic conflicts is that taxonomic opinions are not only in a constant state of change but they also do so quite frequently. This taxonomic contingency will perpetually be in tension with the urge to control change in “officially endorsed” systems designed for stability such as the Catalogue. Such ontogenic concerns have been of increasing interest in the information studies community (Tennis 2002, 2012, 2015), concerned as we are with how to theorize and manage the lateral transformations of classification systems. Such change is important if we are to keep them relevant to our organizational and access concerns. In the biodiversity world, change becomes an ever-more-present issue as consensus structures have emerged on the scene. For all their virtues, consensus structures have not been designed to track internal structural changes. And in biodiversity space, structural change is always produced by a shift or update to some theoretical condition.

As Tennis notes, ontogeny is the examination of “the life of a subject over time—the subject’s scheme history” (2012, 1351). Ontogeny, or ontogenesis, is a surprisingly appropriate term for the development of classifications in our context, given its general usage in the field of biology to describe the mechanism and development of individual organisms. Ernst Haeckel, in *Generelle Morphologie der Organismen* (1866), defines four general concepts: ontologie, phylogenie, ontogenesis, phylogenese. Løvtrup summarizes Haeckel’s concepts in this way: “‘Ontogenie’: the history of the development of the individuals. ‘Phylogenie’: the history of the development (evolution) of the taxa.” Ontogenie refers to the development of individual organisms—as in the progressive and regressive development of one organism over its life span from its embryonic stage to death (Løvtrup 1987, 201). Phylogenie, on the other hand, refers to the

historical reconstruction of taxa as represented in phylogenetic classifications. In the context of biodiversity classification, it is imperative that we are able to trace both the changes that taxon concepts undergo and the changes that might take place with regard to how these are activated within classification systems and related as species with various hierarchies.

Tracing this change over time is incredibly difficult in the space of a composite classification, particularly because change is happening at two levels, both of which are nearly impossible to track. On one hand, we have the change that is occurring within local taxonomies that are contributed to the backbone. On the other hand, we have the changes that are occurring by way of editorial intervention in the Catalogue itself. These are both crucial pieces of information for taxonomic experts, and one of the primary reasons that consensus systems often fall short of their intended use. As we've discussed, contributed taxonomies are snipped out of their original context for inclusion into the Catalogue, but the minute editorial changes that are taking place are difficult, if not impossible, to reconstruct. For experts, working with such a moving target is not a tenable situation. The position of a species is also continually redefined in the Catalogue's space, if not by its absolute position, then by its relative position. As taxonomies are incorporated into the Catalogue, they are set alongside other contributed taxonomies. So, by virtue of these juxtapositions, the relative position of one subject to another is constantly transformed. Understanding these relative relationships is essential to understanding the internal logic of any given taxonomy. In *The Organization of Knowledge and the System of the Sciences*, Bliss speaks of the order of disciplines in the Dewey Decimal Classification schedule and how the placement of topics may cause confusion: "Methodology (112) stands between Ontology and Cosmology and far from Logic, of which it is usually regarded as an extension" (1933, 218). Olson also points to the definitional qualities of classificatory structures: "The hierarchy thus created structured knowledge by putting every subject in its place. It creates a context for each subject within this hierarchical arrangement" (2002, 22). Within biological taxonomies, a species' relative position with regard to other species is of great consequence, both theoretically and methodologically. To change a position is to change a species' overall context.

## EXPERTISE, NUANCE, AND MAINTENANCE

Another pressing issue of concern is that these composite classifications extend beyond the boundaries of the scientific domain itself, meaning that their impact beyond their intended audience must be considered. This is a major point of friction within the biodiversity community. Increasing numbers of scientists and historians of science worry that such taxonomic backbones, and the corresponding epistemic and ontological assumptions they convey, undermine the authority of taxonomic opinion more broadly speaking. If we return to our case of the classification of the Australian dingo, we can see how the use of these systems by nonspecialists would present some serious material problems. On the face of it, the Catalogue inherits the classification of the dingo as represented in ITIS, where the dingo is categorized as *Canis lupus dingo*, a bona fide species of its own. However, there is nothing to say that some other consensus classification might (perhaps wrongly or correctly) classify the dingo as a feral wild dog, under the species *Canis familiaris dingo*. Under the circumstances of the latter, policy makers might rule the dingo unworthy of protection and order the mass eradication of the species on the basis of it being a pest. The crux of this matter is that, while the specialist may be able to understand the nuances of the dingo's classification, in the space of a composite classification, there is no indication that this conflict exists within the taxonomic community to a nonspecialist user. And herein lies one of the dangers of the consensus system: if systems claim to be validated and managed by experts, the prevailing assumption is likely to be that the entire system is, in fact, accurate. And given that "accuracy" proper is not an appropriate way to view taxonomic hypotheses, since there are many viable ways of constructing a classification, this nuance is important.

A larger extension of this problem is how to guide users through the taxonomy in a way that they are able to navigate the space with a critical eye that can spot issues related to quality and completeness. Some of these problems are obvious and clearly present themselves as errors, whereas others are so nuanced that they are never questioned. This is one way that power is systemic—it is obfuscated behind technics. Classifications of this



magnitude are assembled from many moving parts, and at any of these junctures, the introduction of error is not only a distinct possibility, it may be even likely. Given the size of the Catalogue, the introduction of error is not often the fault of any one person, but is rather a consequence of rote data manipulation or local taxonomic practices. Problems arise when users are not versed in understanding how systems function, or what they are and are not meant to express. As Franz and Sterner indicate, as errors proliferate in these consensus systems, trust in them dwindles (2018). “Error,” however, ranges from small issues, such as name misspellings, to more serious issues such as species miscategorization.

Robert Mesibov (2018), a biochemist and active taxonomist in the realm of biodiversity informatics, conveys a peculiar and fascinating example that illustrates the proliferation of erroneous taxon records within database environments. During an audit of the GBIF database, Mesibov points to his discovery of the invalid species name, *Not Chan*, 2016, that led him to examine the extent of similar problems in taxonomic databases. The “Not problem” shows how markers of unknown identifications in databases (represented by terms such as, not found, not given, not known, not identified, not listed, and so on) find their way into aggregated spaces despite their invalid nature. Subsequent examination led Mesibov to find similar issues within the Catalogue and Encyclopedia of Life. The result in this case is that multiple and similar kinds of nomenclatural and concept error were collocated across the global data landscape. Mesibov asserts that consensus backbones such as those produced by the Catalogue and GBIF provide a space in which these errors can “come to light” (Page 2018)—errors that would have otherwise been retained at the local level. A global view affords a perspective that a fragmented data landscape cannot. But even still, the errors proliferate until these discrepancies are located and fixed—and more often than not, it is impossible to track down all the systems that have integrated this error into their infrastructures.

Certainly, one way we can assess quality is through the credibility and reputation of the institutions and individuals that provide information to the Catalogue. But even the most credible sources are human, and errors are a fact of life. Such credibility is made visible within systems by linking

back to the contributing GSD or RSD, as well as requiring the “Latest Taxonomic Scrutiny” data field for all contributed data sets to the Catalogue. This field group must include, the “name(s) of the taxonomic expert or editor, who is responsible for the taxonomic concept accepted in the source database and (b) date when the expert or editor [or small team] assessed the record” (Species 2000 2014, 10). Such linkages and attribution are essential in deconstructing the veracity of the Catalogue’s individual entries, especially since the taxonomic editors depend entirely on the expertise—and reputation—of those that contribute this data. Once again, however, the underlying assumption is that the lay user (a) will question the data *at all*, and (b) if they do, they are willing to perform the legwork of verifying a particular source. To the latter point, one need only be employed as a librarian for a few weeks to realize that almost nobody, in fact, verifies sources. Such is the reality in a world where misinformation is running rampant among users who take information at face value.

In addition to these provenance markers, source databases in the Catalogue provide a *confidence rating*, or data set qualifiers, for the taxonomic data, which is certainly a step in the right direction (Species 2000 2017b). This information is then displayed in the main record at the appropriate taxon level (see figure 7.3). Of course, a major issue is that the quality rating itself is provided by the contributing database, and not by independent peer review, so the quality rating is dependent on the contributor’s willingness and ability to acknowledge their own database’s strength and limitations.

Still other data aggregators, such as the Global Biodiversity Information Facility (GBIF), have attempted to articulate more quantitative and decisive mechanisms for quality assessment. Acknowledging these rampant issues, GBIF formed a task force to articulate a series of recommendations on data fitness and efficacy for large-scale modeling use (Anderson et al. 2016). One of the main recommendations by the GBIF Task Force centered on a user’s inability to accurately differentiate between quality and nonquality data:

GBIF.org should serve indicators of precision, quality, and uncertainty of data that can be calculated practically, and preferably “on the fly,” as well as summaries and metrics of completeness of inventories, at scales and for regions

*Acidimicrobium ferrooxidans* Clark & Norris, 1996

<b>Name</b>	<i>Acidimicrobium ferrooxidans</i> Clark & Norris, 1996	
<b>Checklist status</b>	Accepted species	
<b>Classification</b>	<b>Unranked</b>	<i>Biota</i>
	<b>Kingdom</b>	Bacteria Cavalier-Smith, 2002
	<b>Subkingdom</b>	Posibacteria Cavalier-Smith, 2002
	<b>Phylum</b>	Actinobacteria Cavalier-Smith, 2002
	<b>Class</b>	Actinobacteria Cavalier-Smith, 2002
	<b>Subclass</b>	Acidimicrobidae Stackebrandt et al., 1997
	<b>Order</b>	Acidimicrobiales Stackebrandt et al., 1997
	<b>Suborder</b>	Acidimicrobineae Garrity & Holt, 2001
	<b>Family</b>	Acidimicrobiaceae Stackebrandt et al., 1997
	<b>Genus</b>	<i>Acidimicrobium</i> Clark & Norris, 1996
	<b>Species</b>	<i>Acidimicrobium ferrooxidans</i> Clark & Norris, 1996
<b>Source dataset</b>	ITIS: The Integrated Taxonomic Information System 100% ★★★★★	
<b>Link to original resource</b>	<a href="https://www.itis.gov/servlet/SingleRpt/SingleRpt?search_topic=TSN&amp;search_value=958726">https://www.itis.gov/servlet/SingleRpt/SingleRpt?search_topic=TSN&amp;search_value=958726</a>	

**Figure 7.3**

Catalogue of Life taxon record for *Acidimicrobium ferrooxidans* Clark and Norris, 1996. Note the “Source database” field, indicating the ITIS Global source database, the database version date (Sept 2015), the percentage of completeness of the species list this entry is embedded within, and finally, the confidence rating for the quality of the taxonomic checklist (level 5 value). CC-BY 4.0, Catalog of Life, used by permission.

defined by the user. The summaries should display maps and graphs of completeness by region, time-period and taxa. (Anderson et al. 2016, 2)

The task force specifically recommended clear fields to indicate error and uncertainty rates, as well as provide methods for users to visualize data sets to understand the larger contours of that data and “highlight possible inconsistencies and error” (Anderson et al. 2016, 4). A valiant goal, to be sure, but often very difficult to implement.

It is of great import that governments and other funding sources have acknowledged the data maintenance, retrieval, management, and conservation, as an integral part of biodiversity work. If we look to large organizations such as GBIF, the Catalogue, and the Distributed System of Scientific Collections (DiSSCo), the European Union has paved the way for a more sustainable data landscape. GBIF subsidiary partners (global nodes) receive funding from a variety of funders, including individual governments, the

National Science Foundation, the Smithsonian Institution, and multiple ministries of science, education, and the environment, and the like. The reality is that, as the locus of biodiversity data power shifts from local, description-oriented work (individual scientists, research teams, and the like) to aggregative access-oriented data infrastructures, there is a real danger that the expert systematist community could see problems funding their own individual systematic projects. Certainly, GBIF and other organizations understand the problem inherent in this imbalance, and the Alliance for Biodiversity Knowledge (ABK) (Hobern et al. 2019) is one mechanism that has arisen in response to the current competitive research funding model. One goal of the ABK is to limit duplicative efforts on a number of fronts. Rather than everyone competing for limited funds to create many management taxonomies, for example, agreed-upon organizations should be designated to handle certain portions of the data landscape. The ABK also provides avenues for direct communication and participation from the research community to guide the path forward for such informatics-based projects.

However, even with this cooperative agreement, some taxonomists worry that the pendulum has swung too far in the direction of access, to the detriment of individual description-based research endeavors. And while the shift seems relatively benign, with the advent of large-scale infrastructures, some note that prioritizing data precision, stability, and comprehensiveness, rather than the systematic work itself only amplifies the epistemic problems that outfits like the Catalogue perpetuate. In addition, as organizations such as GBIF rise in prominence, they are increasingly becoming the mediators of biodiversity work with national and international governmental agencies. The Catalogue and GBIF are placed in a situation where they must balance both the needs of funders supporting their large-scale retrieval-oriented agenda and the requirements of the taxonomic community, which needs assurance that their taxonomic work will maintain a certain level of integrity in these systems. A fine line to balance, to be sure—and one that will only amplify in the coming years as this trend continues. And while the need to balance these two constituents may not be a bad thing, the reality is that such a position is indicative of the drastic

ways power has shifted from the lab, so to speak, to the global management of accumulated data. As is often colloquially stated, if you want to know where the power is, follow the streams of money.

### **Divergent Traditions and Nameless Taxa**

Perhaps the greatest technical issue for composite spaces is how to manage taxa tokens (names or codes) that do not conform to Linnaean binomial formulations. Genetic markers in the form of DNA barcodes (such as the mitochondrial CO1 gene sequence) have been increasingly useful in constructing phylogenies (Waterton, Ellis, and Wynne 2013; Erickson and Driskell 2012). One result of the increasingly popular approach of phylogenetic inference is the “proliferation of taxonomic categories” (Queiroz and Gauthier 1992, 457). DNA evidence tends to “split” more than “lump” species together into taxon groups. The applications of names to these DNA barcode strings, however, and their reconciliation with existing taxon concepts, is an entirely separate activity, often performed after these phylogenies have been constructed. Increasingly, names are not being applied to this growing cache of genetically labeled information.

Berry van der Hoorn, group leader for biodiversity discovery at Naturalis Biodiversity Center, was kind enough to summarize a case in point (interview 2016). Van der Hoorn’s group was in the process of conducting state-funded research on water quality in the Leiden area of the Netherlands. They were charged with identifying species in various water wells in an effort to better understand shifting species variability in different environments. One related project, in particular, took them to an island in the Caribbean, where they were charged with understanding how species variability changed over time. However, because the research in question was about species *change*, the particular species in question, and their Linnaean names, were not a necessary aspect of this work. Van der Hoorn noted, “Sometimes you don’t even need to know the species name, you just indicate, we found 150 spiders here, 12 spiders there, and that’s enough. Species were identified by ‘OTUs’—operational taxonomic units. And you can use [these OTUs] fairly well as a biodiversity index and you don’t even

need to recognize the species itself.” The same was often the case for their study of water samples in Leiden—the actual species involved were less important than the rates of variability and the rate of change for OTUs. For van der Hoorn, the research questions were ecologically based, formulated within and for very specific conditions. The application of names was secondary to solving these project- and funding-specific research queries.

In practice, numerous scholars have pointed to this widening divide between traditional and phylogenetic approaches, and the detrimental effects it is having on the adequate accumulation and collocation of scientific knowledge as each proceeds forward invoking and implementing different methodologies. Nico Franz discusses this increasing tendency to avoid translating phylogenies into classifications in his article “On the Lack of Good Scientific Reasons for the Growing Phylogeny/Classification Gap.” Franz notes, “By supplementing a traditional classification with a more precise estimate of phylogeny, one has not yet ‘removed the need to use’ any or all parts of that classification. In the vast majority of cases, the more recent phylogenetic analyses are properly considered revisions of pre-existing hypotheses (however coarse) about the relationships among taxa and the evolutionary histories of character traits” (Franz 2005, 496). The traditional modes of taxonomy that have been built over the last 250 years, including the application of names to taxon groups, are essential to contextualizing and making meaningful the results of phylogenetic analysis. Additionally, unnamed phylogenies become siloed from this cache of knowledge linked to the historiographical record. One camp cannot communicate with other, thereby limiting the ability for systematics as a whole to proceed forward as a coherent unit and to build on the virtues of each approach.

Despite these competing views, there is certainly a synergy between the two camps that can flourish. A project such as the one described by van der Hoorn provides the raw data that taxonomists can use to produce more robust and complete classifications. Systems like the Barcode of Life are intended to serve this very purpose. In the meantime, however, while these increasing caches of barcodes are being produced, this information fails to make its way into aggregated and composite systems such as the Catalogue. GBIF, for example, is currently working on mechanisms by

which genetic barcodes can be appended to their Nub Taxonomy framework. For example, species that are databased can often be recognized at a higher taxonomic level—say, at the order or family level. For example, in Berry van der Hoorn’s described project above, spiders were collected from an area with passive traps. GBIF’s goal in this case might be to append the barcoded data at the highest known taxon level—in this case, the barcoded spider data might be appended to the Araneae order level of the Nub Taxonomy. The assumption is that data imperfectly placed but available is far preferred over perfectly situated data, especially given that the latter can take a great deal of time to assess. Users seeking out this information then have the ability to search through the GBIF portal to locate this information, even if its exact location on the taxonomic scale is unknown.

As Quentin Wheeler stated, “Phylogenetic classifications are optimal for storing and predicting information, but phylogeny divorced from taxonomy is ephemeral and erodes the accuracy and information content of the language of biology” (Wheeler 2004). A fundamental issue, and one that information studies should be closely attuned to, is how the biodiversity world is attending to these divides and attempting to build classification systems that are more inclusive of multiple approaches to classification based on contradictory standards. The ultimate goal of taxonomies such as the Catalogue are to create inclusive spaces that radically commingle divergent taxonomic opinions. This is no easy feat, and only time will tell in what manner all of these problems will work out within the system. That the Catalogue, and other similar structures, are still vying for broad, universal acceptance is secondary to this story in many ways—that they are trying to bridge divides is what is paramount.

However, even while global data are conceptualized as inclusive of multiple forms in museum and biodiversity practice, much of the networked infrastructure that supports this work is incapable of more complex data sets, particularly video and three-dimensional data (because of both the memory required to store these instances and third-party software requirements to display these assets). For example, the media asset management system in use at the Natural History Museum, London, as of 2016—Open Text Media Manager (known internally as MAM)—does

not have the capability to store video or 3D files. As Matthew Woodburn, science data architect for the Natural History Museum's Digital Collections Programme, indicated, "We either have to wait for that capability to be put in place or we need to look for a workaround if it's not going to happen. There are [many] kinds of difficulties on the 3D side of it. Is 3D an image or is it a dataset? Because effectively it's all bits and bytes but it tends to be . . . visualized by a particular piece of software. . . . We can publish it as a dataset, which is not a problem. . . . Then it would be up to them to find a visualization service" (2016). In more than one conversation with museum specialists, a distinction was made between "data that can be used in any particular database" and data "that was unable to be properly ingested into the digitization and digital management program." As such, what constitutes data within an institutional setting at a pragmatic level is often dictated by what formats and types of data any particular system has the ability to accept and integrate.

Lastly, it is one thing to speak about the irreconcilability of Linnaean nomenclature and genetic sequencing data, but it is quite another, more serious, problem to say that the current landscape of biodiversity taxonomy is firmly rooted in a scientific tradition that arose in the context of "the West." This is a fact that has been an undercurrent in some of this book's former chapters—from the exclusion of indigenous information in the classification of the dingo; to biological taxonomy's adoption of Linnaean nomenclature, dependent as it is on Latinate syntactical forms. Further, throughout this text I have spoken about instantiative, instrumental, aesthetic, material, and extensive powers, all of which gain their epistemic strength through the broadscale adoption of colonial reasoning. Speaking with complete honesty, in the many months of fieldwork I undertook to complete this book, the subject of colonial and imperial power as an underlying ethical issue in biodiversity work was the most contentious subject for the many practicing taxonomists I encountered. Many biodiversity workers were keenly aware and conflicted at the imbalance of power their positions represent. Part of this, I believe, has to do with the enormity and pervasiveness of the subject, and the reality that, if one acknowledges it as a problem, then one must also accept the reality that the entire foundation



of taxonomy as a field must reckon with the greatest exertion of power of them all: racism and the imperial and colonial motive to dominate. And although this may be true, the reality is that not only taxonomy but also the whole endeavor of science and the Western traditions that matured around it have to come to terms with this reality. Our natural history museums, after all, are full of specimens and material from foreign lands—much of which was secured during the great colonial expansion of the nineteenth century. If there are “blind spots” for data, chances are they are in areas that have traditionally been exploited and underrepresented on the global stage. I am thinking here of the gap in African data that the Catalogue of Life has been actively trying to rectify. In chapter 8, I take a sharp critical turn and delve more deeply into this issue and frame the work of classification as unescapably intertwined with these historical realities.



This is a section of [doi:10.7551/mitpress/12245.001.0001](https://doi.org/10.7551/mitpress/12245.001.0001)

# Power of Position

## Classification and the Biodiversity Sciences

By: Robert D. Montoya

### Citation:

*Power of Position: Classification and the Biodiversity Sciences*

By: Robert D. Montoya

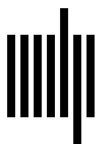
DOI: [10.7551/mitpress/12245.001.0001](https://doi.org/10.7551/mitpress/12245.001.0001)

ISBN (electronic): 9780262369961

Publisher: The MIT Press

Published: 2022

This book is freely available in an open access edition thanks to TOME (Toward an Open Monograph Ecosystem) – a collaboration of the Association of American Universities, the Association of University Presses, and the Association of Research Libraries – and the generous support of Arcadia, a charitable fund of Lisbet Rausing and Peter Baldwin, and the UCLA Library. Learn more at the TOME website, available at: [openmonographs.org](https://openmonographs.org)



The MIT Press

© 2022 Robert D. Montoya

This work is subject to a Creative Commons CC-BY-NC-4.0 license. Subject to such license, all rights are reserved.



This book is freely available in an open access edition thanks to TOME (Toward an Open Monograph Ecosystem)—a collaboration of the Association of American Universities, the Association of University Presses, and the Association of Research Libraries—and the generous support of Arcadia, a charitable fund of Lisbet Rausing and Peter Baldwin, and the UCLA Library. Learn more at the TOME website, available at: [openmonographs.org](http://openmonographs.org).

The MIT Press would like to thank the anonymous peer reviewers who provided comments on drafts of this book. The generous work of academic experts is essential for establishing the authority and quality of our publications. We acknowledge with gratitude the contributions of these otherwise uncredited readers.

This book was set in Adobe Garamond Pro by Westchester Publishing Services.

#### Library of Congress Cataloging-in-Publication Data

Names: Montoya, Robert D., author.

Title: Power of position : classification and the biodiversity sciences / Robert D. Montoya.

Description: Cambridge, Massachusetts : The MIT Press, [2022] | Series: History and foundations of information science | Includes bibliographical references and index.

Identifiers: LCCN 2021033972 | ISBN 9780262045278 (paperback)

Subjects: LCSH: Biology—Classification. | Life sciences—Classification. | Cladistic analysis.

Classification: LCC QH83 .M68 2022 | DDC 570.1/2—dc23/eng/20211221

LC record available at <https://lccn.loc.gov/2021033972>