HPLS (2020) 42:40
https://doi.org/10.1007/s40656-020-00337-8

EDITORIAL

The history and philosophy of taxonomy as an information science

Catherine Kendig[1] · Joeri Witteveen[2]

© Springer Nature Switzerland AG 2020

1 Taxonomy in the information age

We undeniably live in an information age—as, indeed, did those who lived before us. After all, as the cultural historian Robert Darnton pointed out: 'every age was an age of information, each in its own way' (Darnton 2000: 1). Darnton was referring to the news media, but his insight surely also applies to the sciences. The practices of acquiring, storing, labeling, organizing, retrieving, mobilizing, and integrating data about the natural world has always been an enabling aspect of scientific work. Natural history and its descendant discipline of biological taxonomy are prime examples of sciences dedicated to creating and managing systems of ordering data.

In some sense, the idea of biological taxonomy as an information science is commonplace. Perhaps it is because of its self-evidence that the information science perspective on taxonomy has not been a major theme in the history and philosophy of science. The botanist Vernon Heywood once pointed out that historians of biology, in their 'preoccupation with the development of the sciences of botany and zoology … [have] diverted attention from the role of taxonomy as an information science' (Heywood 1985: 11). More specifically, he argued that historians had failed

Catherine Kendig and Joeri Witteveen have contributed equally to this paper and are listed in alphabetical order.

✉ Catherine Kendig
kendig@msu.edu

Joeri Witteveen
jw@ind.ku.dk

[1] Department of Philosophy, Michigan State University, 368 Farm Lane, 503 South Kedzie Hall, East Lansing Michigan, USA

[2] Department of Science Education, Section for History and Philosophy of Science, University of Copenhagen, Rådmansgade 64, Copenhagen, Denmark

Published online: 31 August 2020

to appreciate how principles and practices that can be traced to Linnaeus constituted 'a change in the nature of taxonomy from a local or limited folk communication system and later a codified folk taxonomy to a formal system of information science [that] marked a watershed in the history of biology' (ibid.).

A similar observation could be made about twentieth-century philosophy of biology, which mostly skipped over practical and epistemic questions about information management in taxonomy. The taxonomic themes that featured in the emerging philosophy of biology literature in the second half of the twentieth century were predominantly metaphysical in orientation. This is illustrated by what has become known as the 'essentialism story': an account about the essentialist nature of pre-Darwinian taxonomy that used to be accepted by many historians and philosophers, and which stimulated efforts to document and interpret shifts in the metaphysical understanding of species and (natural) classification (Richards 2010; Winsor 2003; Wilkins 2009). Although contemporary debates in the philosophy of taxonomy have moved on, much discussion continues to focus on conceptual and metaphysical issues surrounding the nature of species and the principles of classification. Discussions centring on whether species are individuals, classes, or kinds have sprung up as predictably as perennials. Raucous debates have arisen even with the aim of accommodating the diversity of views: is monism, pluralism, or eliminativism about the species category the best position to take? In addition to these, our disciplines continue to interrogate what is the nature of these different approaches to classification: are they representational or inferential roles of different approaches to classification (evolutionary taxonomy, phenetics, phylogenetic systematics)? While there is still much to learn from these discussions—in which we both actively participate—our aim with this topical collection has been to seek different entrypoints and address underexposed themes in the history and philosophy of taxonomy. We believe that approaching taxonomy as an information science prompts new questions and can open up new philosophical vistas worth exploring.

A twenty-first century information science turn in the history and philosophy of taxonomy is already underway. In scientific practice and in daily life it is hard to escape the imaginaries of Big Data and the constant threats of being 'flooded with data'. In the life sciences, these developments are often associated with the so-called bioinformatics crisis that can hopefully be contained by a new, interdisciplinary breed of bioinformaticians. These new concepts, narratives, and developments surrounding the centrality of data and information systems in the biological and biomedical sciences have raised important philosophical questions about their challenges and implications. But historical perspectives are just as necessary to judge what makes our information age different from those that preceded us. Indeed, as the British zoologist Charles Godfray has often pointed out, the piles of data that are being generated in contemporary systematic biology have led to a *second* bioinformatics crisis, the first being the one that confronted Linnaeus in the mid-18th century (Godfray 2007).

Although the aim of this collection is to clear a path for new discussions of taxonomy from an information science-informed point of view, we continue where others in the history, philosophy, and sociology of science have already trod. Historians of science have in recent years drawn attention to the continuities underneath the

disruptions that are due to present-day data-intensification. For example, Bettina Dietz has explicitly described natural history in general and Linnaean botany in particular as an information system. Its systematicity functioned in what she describes as 'a collaborative and contributive way' (Dietz 2012). Her work on the cultures of collaboration and co-production of eighteenth-century botanical compendia reveals the iterative, networked nature of information accumulation as a knowledge-making process. Likewise, Isabelle Charmantier and Staffan Müller-Wille have shown that Linnaeus's use of botanical paper slips and other paper tools constituted a novel, more flexible and dynamic means of processing information than the traditional approach of using bound manuscripts (Charmantier and Müller-Wille 2014). These studies give substance to Godfray's remark that Linnaeus already faced a threat of information-overload that compelled him to devise new methods, tools and technologies to manage expanding flows of data. They make us aware that concerns about being swamped by data preceded our recent experience of digital Big Data.

More generally, historical perpectives on taxonomy as an information science help to avoid that 'the hype of the present moment outshouts extrapolations from the past and into the future' (Daston 2017: 11). Of course, this is not to say that we should not give a voice to practices, tools, technologies, social conditions, and institutional contexts that have reformed and reconfigured past information practices. But in taxonomy perhaps more than in other sciences, contemporary practices need to heed past uses of data and maintain a bridge to bygone approaches to scientific knowledge capture and coding. The essays in this collection aim to show that problems of data and information have played important roles in the curating of taxonomic information whether it be from paper slips and constant collaborative correspondence, to public archives, crowd-sourcing and debates surrounding Genbank and the Barcode of Life Data system. We believe that an appreciation of biological taxonomy as an information science raises many questions about the philosophical, theoretical, material, and practical aspects of the use and revision of biological nomenclatures in different local and global communities of scientists and citizen-scientists. In particular, conceiving of taxonomy as an information science directs attention to the temporalities of managing an accumulating data about classified entities that are themselves subject to revision, to the means by which revision is accomplished, and to the semantic, material, and collaborative contexts that mediate the execution of revisions.

2 Taxonomic practice in the spotlight

Attending to taxonomic practices allows one to discover implicit norms in taxonomic information processing activities that remain hidden from the more abstract theoretical or metaphysical treatments. Saying that taxonomy is ineliminably normative does not imply that it is in some way non-naturalist or worryingly subjective. The normativity of taxonomic information is investigated through different routes that have been largely left out of interdisciplinary discussions even between practitioners and philosophers or have been abandoned to a large undifferentiated bucket of *just* pragmatic concerns. These normative concerns have not only historical and

sociological bases, they have metaphysical, epistemological and scientific impacts. Attending to these normative concerns prompts new lines of inquiry: What are the best principles and practices for keeping track of named groups? How are names made to 'stick' to fluid entities? On what basis do taxonomists decide whether it is worth naming a group? And how do names, labels, and the kinds they refer to affect taxonomic theory and practice once they are brought into circulation? Our aim in conceiving of biological taxonomy as an information science is to focus attention on the nexus of material, explanatory, ontological, and normative concepts underpinning current and historic taxonomic practices, and on how these have coped with and shaped the kinds of data captured and generated.

In so doing, this work can be understood as part of a long-standing conceptual project in integrated HPS, philosophy of science in practice (PSP), and science and technology studies (STS). In all three, scholarship on the epistemic and metaphysical dimensions of scientific practice center on the interplay of theory, practice, theoretician, practitioner as knower in the generation of knowledge. Recently, much of this has centred on reconfiguring the object of philosophy of science from a discipline which has been conceived of largely as a theory-driven exploration to one that includes both theory and practice whilst also troubling the grounds on which the long-assumed theory-practice dichotomy rests (Ankeny et al. 2011; Soler et al. 2014; Kendig 2016a). The importance of an integrated historical and philosophical focus on the role of exploratory investigation, the epistemological role of experiment, modelling and model integration, and knowledge-making activities within science and engineering has generated a wealth of scholarship. Attention to careful and critically discussed case studies (rather than wide unempirical generalizations) provides the basis for understanding. For instance, Leonelli's (2016) work on data practices and data journeys has been conducive in raising awareness of how information content in the plant and biomedical sciences is shaped.

We curated this collection on purpose from a wide range of interweaving disciplinary approaches in integrated HPS, PSP, and STS. Philosophers, historians, sociologists, and biologists wrestle between and among these disciplines focusing on the normative challenges surrounding the procuring and ordering of data and the collaborative nature of taxonomic naming. With regard to biological taxonomy, we, Catherine Kendig and Joeri Witteveen, the two co-editors of this collection of essays, have individually made excursions into practice-based history and philosophy of taxonomy. For Kendig, this has included research on the comparability and translatability of the varied names and descriptions of parts contained within different synthetic biology repositories. She has critically assessed the problems arising with naming and tracking of parts within and across repositories, but also how comparisons across different databases might be facilitated using computational models that capture information and meta-information as similarity measures for different biological ontologies (Kendig 2016a, Kendig and Bartley 2019). In these, Kendig focuses on the inextricability of epistemological and ontological activities required for data-categorizations in synthetic biology. In other work, she has also tracked historic revisions of kind-making or 'kinding' practices used to shape and re-tune concepts of species, the concept of homology, and homologizing practices in order to analyze their role in taxonomic naming (Kendig 2016b). Witteveen's research in

the history and philosophy of taxonomy has centered on the nature and development of systems of taxonomic naming and reference. For example, he has shown how key principles that underpin contemporary naming practices were articulated in the nineteenth century in a context of marked changes in taxonomic tools, methods and practioners, intense institutional politics, and lengthy debates about the logic of reference (Witteveen 2016, 2018). These historical studies are both informed by and inform Witteveen's research on the philosophical foundations of contemporary principles and procedures of naming taxa (Witteveen 2015).

3 Contributions to the topical collection

The contributions to this topical collection reflect on the foundations and challenges of capturing, storing, organizing, retrieving and communicating taxonomic information from a diversity of historical, philosophical, sociological and scientific vantage points. Uniting this diversity is a common focus on taxonomic practice based on illustrative case studies. Each provides a new frame through which investigation can be pursued and within which the problem-space is reconfigured. In keeping with the theme of this collection, we have classified the contributions according to their common themes and interests: sourcing, naming, and ordering of taxonomic information. Though we realize that this is a highly contrived classification and are aware that some papers could be cross-listed, we think that this organization can help to bring out some shared questions and concerns.

4 Sourcing

The first three papers in the collection engage with the social conditions of procuring and communicating taxonomic information. This theme is explored by studying the relations and interactions between communities of experts, amateurs, and citicizen scientists (Mahr and Dickel), the conditions under which consensus can be obtained within a community of practitioners (Sterner, Witteveen and Franz), and by fleshing out the history of the collaborative nature of curating names through the ongoing comparative work of botanists (Dietz). This eclectic mix of contributions from across sociology, history and philosophy of science engages with questions that cross-cut the social spaces within which taxonomic knowledge is generated and used. Overarching questions that are raised here include: what counts as sourcing?; who are considered a reliable sources (and why)?; and what role does society (broadly and narrowly construed) contribute to the capture of information and communication of data? The contributors discuss sourcing of collaborators and stakeholders, sourcing data and through databases, as well as through collaborations.

Dana Mahr and Sascha Dickel provide a compelling case troubling the unquestioned distinction between citizen-based Big Data research and expert-led Big Data research by focusing on the amateur-led biogeographical birdwatching research, like the Audubon Christmas Bird Count in the 19th and 20th century. Although more recent approaches to citizen science emphasize a crowd-sourcing model where

citizens are 'invited' to participate while it is the professional scientists that do the 'serious' scientific work, this expert-lay person divide obscures much epistemic work on the part of the citizen groups. Mahr and Dickel compare the recent 'invited' incarnations of unequal co-creators of knowledge in current narratives of citizen science with earlier more complementary approaches of 'uninvited', autonomous, and people-driven science. They show how this latter model has and can be supported through a set of political and normative values.

Beckett Sterner, Joeri Witteveen and Nico Franz provide a critical examination of the (practical) social epistemology of bioinformatics, focusing on the principles, aims and assumptions that inform contemporary computational architectures for data discovery and integration. They observe that the design of bio-ontologies in the life sciences tends to be rooted in a 'definitional consensus principle.' In order for ontologies to serve their communicative and integrative purposes, scientists need to come up with definitions of the classificatory entities that feature in the ontology. Recognizing that the practical and philosophical aspects of formulating 'good' definitions has been a topic of debate, the definitional consensus principle itself has been treated as self-evident: how else can we facilitate the accumulation and integration of knowledge in the data-intensive life sciences other than by agreeing on the meaning of the terms that the growing piles of data can be associated with? Sterner et al. argue that there is an alternative solution, which has deep roots in biological systematics. They present biological systematics as an area in the life sciences that is based on an a 'coordinative consensus principle'. A formal classificatory system for expressing a body of data should be grounded in a consensus standard for coordinating the application of names of classified entities.

Bettina Dietz reveals the basis for synonyms, their use, and the collaborative nature of their curation relying on the work and networks that the Dutch botanist Johannes Burman worked within. Dietz describes the project of listing and evaluating synonymous names for the same specimen or organism. The comparison of plant names and the use of synonyms as well as the discrimatory work of trying to decide which names were synonyms for which specimens was a project that was central to making and storing textual and visual knowledge. This work was fundamentally collaborative as it relied upon herbariums, private gardens, and the comparative work of botanists. Dietz shows how this information curating was collaborative both in terms of the reliance of ongoing corrections to synonyms and publications, but also in the ongoing revisions of texts that were inherited and continued to be revised by later generations of botanists. These practices relied on an international community of botanists rather than the exclusive identification by one solitary botanist.

5 Naming

Connecting with the sourcing papers that consider the location and provenance of informational resources, the naming papers enter the discourse by inviting discussion over the principles and practices of applying names to taxa. On the face of things, naming may appear to be the least interesting aspect of taxonomy. It is not science proper, but merely constitutes a means of facilitating the storage and

retrieval of scientific information. The second duo of papers in this collection challenge this common attitude to naming and by showing that decisions that need to be made in applying names are rich in normative and epistemic issues that affect taxonomic practice as a whole. Topics discussed here include the role of non-Code compliant names or 'grey names' (Minelli) and the evolving possibilities of being in error about the application of names (Witteveen and Müller-Wille).

Alessandro Minelli explores complications that grey nomenclature brings to the formal use of naming conventions dictated by Linnaean nomenclature and the International Codes of Nomenclature. He analyzes not only the formal practices of naming but also how these come apart from the interests and goals of those using taxonomy. Focusing on intentional contraventions of these naming conventions—in terms of not following gender agreement in Latin, the use of numbers to index barcoded genetically understood taxa, and the role of the individual doing the naming—all lead to potentially problematic taxonomic ambiguity. This ambiguity exists not only in terms of the names but also in the kinds the names are intended to apply, e.g. specimen, taxa. Minelli proposes a means by which non-Code compliant grey names may continue to be used, outlining a set of normative rules grey names must adhere to in order for the information contained within them to be of international use.

Joeri Witteveen and Staffan Müller-Wille consider the question what it takes to be in error about the application of a name to a given taxon. Their study is prompted by a recent highly publicized case of the (alleged) error Linnaeus made in the naming of the Asian elephants (*Elephas maximus*). Linnaeus, it has been argued, made the error of applying the name for the Asian elephants to what we now know to be the African elephants (*Loxodonta africana*), because he was mistaken about the taxonomic identity of a specimen he used as name-bearer. Witteveen and Müller-Wille demonstrate that this case study is considerably less clean-cut than it would at first appear, and show that it raises interesting historical and philosophical questions about precisely what sort of error Linnaeus made. In answering these questions, they argue that the very possibility of being in error about the identity (and hence naming) of taxa has a history. And in so doing, they identify key differences between the (implicit) philosophy of Linnaeus's approach to naming and contemporary 'Linnaean' codes of nomenclature.

6 Ordering

The last two papers in the collection enter into conversations surrounding data and taxonomy that complement those from the sourcing and naming approaches. These papers consider how order is created (Tamborini) and what information content ordered taxa contain, if any (Reydon).

Marco Tamborini investigates an historical case study on odontography. The study of odontography reveals a particularly striking example of the mathematical coding of one aspect of the natural world: teeth. Tamborini provides a detailed account of how the Argentine zoologist, Florentino Ameghino, relied upon mathematical formalizations within odontography that he used to classify animals as a set of shorthand descriptions for fuller anatomical descriptions. Ameghino's

mathematicalization presents a clear case of taxonomy as information science insofar as Ameghino effectively digitized biological information useful for phylogenetic analyses. These formalizations allowed not only Ameghino but also others, such as George Gaylord Simpson and Anne Roe, to produce a quantitative systematic account that ultimately replaced the qualitative morphological approaches that came before it.

Thomas Reydon suggests that although taxa are often thought to represent information about the members of the classes they name, this impression is a mistaken one. The ordering system of taxonomy is not an information storage system and taxa are not information storage units. Reydon notes that this means that they cannot be used to be the grounds for inferences or generalizations over a group of organisms which are members of the taxon. Phylogenetic inference itself rests on the methodological choices that are made by the scientist. Although the idea that taxa are informational units breaks down since the generalizations individual scientists make may not result in the same groupings, taxa may instead be markers that hold information about where they are located in phylogenetic trees.

7 Summary remarks

It is, perhaps, unusual for papers with this breadth of disciplinary perspectives, aims and topics to come together in a single collection. We believe that this broad diversity and the approaches to the topic of taxonomy as an information science shows the potential of embracing different organizing themes as a means of stimulating interaction between historians, philosophers, sociologists and scientists. Our hope is that by bringing the different views in conversation with one another it will lead to new 'information-focused' dialogue. As such, this special collection serves as a discursive template for exploring the interweaving of data narratives surrounding the sourcing, naming, and ordering of taxonomic information from historical, philosophical, and sociological perspectives.

References

Ankeny, R., Chang, H., Boumans, M., & Boon, M. (2011). Introduction: Philosophy of Science in Practice. *European Journal for Philosophy of Science, 1*(3), 303–307.

Charmantier, I., & Müller-Wille, S. (2014). Carl Linnaeus's botanical paper slips. *Intellectual History Review, 24*(2), 215–238.

Darnton, R. (2000). An early information society: News and the media in eighteenth-century Paris. *American Historical Review, 105,* 1–35.

Daston, L. (2017). *Science in the archives*. Chicago: University of Chicago Press.

Dietz, B. (2012). Contributions and co-production: The collaborative culture of linnaean botany. *Annals of Science, 69*(4), 551–569.

Kendig, C. (2016a). What is proof of concept research and how does it generate epistemic and ethical categories for future scientific practice? *Science and Engineering Ethics, 22*(3), 735–753.

Kendig, C. (2016b). Homologizing as kinding. In C. Kendig (Ed.), *Natural kinds and classification in scientific practice* (pp. 106–125). Abingdon and New York: Routledge.

Kendig, C., & Bartley, B. (2019). Synthetic kinds: Kind-making in synthetic biology. In J. R. S. Bursten (Ed.), *Perspectives on classification in synthetic sciences: Unnatural kinds* (pp. 78–96). London: Taylor & Francis.

Leonelli, S. (2016). *Data-centric biology: A philosophical study*. Chicago: University of Chicago Press.

Richards, R. A. (2010). *The species problem: A philosophical analysis*. Cambridge: Cambridge University Press.

Soler, L., Zwart, S., Lynch, M., & Israel-Jost, V. (Eds.). (2014). *Science after the practice turn in the philosophy, history, and social studies of science*. Abingdon and New York: Routledge.

Wilkins, J. S. (2009). *Species: A history of the idea*. Berkeley: University of California Press.

Winsor, M. P. (2003). Non-essentialist methods in pre-Darwinian taxonomy. *Biology and Philosophy, 18*(3), 387–400.

Witteveen, J. (2015). Naming and contingency: The type method of biological taxonomy. *Biology and Philosophy, 30*(4), 569–586.

Witteveen, J. (2016). Suppressing synonymy with a homonym: The wmergence of the nomenclatural type concept in nineteenth century natural history. *Journal of the History of Biology, 49*(1), 135–189.

Witteveen, J. (2018). Objectivity, historicity, taxonomy. *Erkenntnis, 83*(3), 445–463.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.