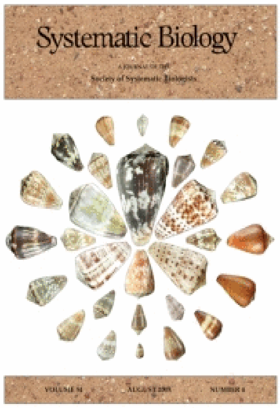

Systematic Biology

Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713658732

The Web and the Structure of Taxonomy

H. C. J. Godfray [a]; B. R. Clark [a]; I. J. Kitching [b]; S. J. Mayo [c]; M. J. Scoble [b]
[a] Department of Zoology, University of Oxford, Oxford, UK
[b] Department of Entomology, Natural History Museum, London, UK
[c] Royal Botanic Gardens, Kew, Richmond, UK

First Published on: 01 December 2007
To cite this Article: Godfray, H. C. J., Clark, B. R., Kitching, I. J., Mayo, S. J. and Scoble, M. J. (2007) 'The Web and the Structure of Taxonomy', Systematic Biology, 56:6, 943 - 955
To link to this article: DOI: 10.1080/10635150701777521
URL: http://dx.doi.org/10.1080/10635150701777521

Syst. Biol. 56(6):943–955, 2007
Copyright © Society of Systematic Biologists
ISSN: 1063-5157 print / 1076-836X online
DOI: 10.1080/10635150701777521

The Web and the Structure of Taxonomy

H. C. J. GODFRAY,[1] B. R. CLARK,[1] I. J. KITCHING,[2] S. J. MAYO,[3] AND M. J. SCOBLE[2]

[1]*Department of Zoology, University of Oxford, South Park Road, Oxford, OX1 3PS, UK; E-mail: charles.godfray@zoo.ox.ac.uk (H.C.J.G.)*
[2]*Department of Entomology, Natural History Museum, Cromwell Road, London SW7 5BD, UK*
[3]*Royal Botanic Gardens, Kew, Richmond TW9 3AE, UK*

Abstract.—An easily accessible taxonomic knowledge base is critically important for all biodiversity-related sciences. At present, taxonomic information is organized and regulated by a system of rules and conventions that date back to the introduction of binomial nomenclature by Linnaeus. The taxonomy of any particular group of organisms comprises the sum information in the taxonomic literature, supported by designated type specimens in major collections. In this article, the way modern means of disseminating information will change the practice of taxonomy, in particular the Internet, is explored. Basic taxonomic information, such as specimen-level data, location of types, and name catalogues are already available, at least for some groups, on the Web. Specialist taxonomic databases, key-construction programs, and other software useful for systematists are also increasingly available. There has also been a move towards Web-publishing of taxonomic hypotheses, though as yet this is not fully permitted by the Codes of Nomenclature. A further and more radical move would be to transfer taxonomy completely to the Web. A possible model of this is discussed, as well as a pilot project, the "CATE" initiative, which seeks to explore the advantages and disadvantages of such a move. It is argued that taxonomy needs to forge better links with its user-communities to maintain its funding base, and that an important part of this is making the products of its research more accessible through the Internet. [Taxonomy; Internet; e-Science; Web revision.]

The two main products of contemporary taxonomy are hypotheses of species boundaries and hypotheses of the phylogenetic relationships among species. The latter may be formulated as explicit cladograms and phylogenetic trees or implicitly in the classification of species into genera and higher taxonomic groups (Wiley, 1981; Forey et al., 1992). It is the triumph of Linnaean taxonomy that it has produced a system of classification and nomenclature that is universally accepted and that at least in theory provides the "end-users" of taxonomy with a stable means of naming and referring to different species and other taxa. In this article, we explore how taxonomy is currently organized and presented to the wider biological community. We argue that this happens chiefly using a means of disseminating information that would have been recognized by Linnaeus 250 years ago. We shall then ask how modern means of sharing data, especially the Internet, might be used to make taxonomy more efficient and to improve its links with its end-users. We do this by reviewing different ways in which taxonomy has already made use of the Web or may do so in the future. One possibility is converting taxonomy to a largely Web-based exercise, and we discuss in more detail a particular project, the CATE initiative, that seeks to explore the costs and benefits of such an approach. We conclude by speculating on how this area might develop in the next few decades.

It is common to talk about science in particular, and society in general, entering a new "information age" driven by the inexorable spread of connectivity brought about by the Web and associated technologies (Benkler, 2006). In biology, the uptake of the Web has been spurred by the "bioinformatics crisis" caused by the explosion of data in molecular genetics and associated fields. The crisis was solved by making genomic sequence and other "-omic" data available (largely freely) on the Web, and by developing a series of data handling and data mining tools to make it accessible and straightforward to analyze. But we prefer to think of this as the second bioinformatics crisis, the first being that faced by Linnaeus in the mid-18th century. At that time a particular plant or animal "species" could have many names, some very long ("phrase names"), and with no agreed standard. The genius of Linnaeus and his successors was to solve this problem using the technology available in a day of purely printed publication and slow communications via an unreliable post (Blunt, 1984; Koerner, 1999). In retrospect, the success of Linnean taxonomy was probably the result of its functionality, its relative democracy—anyone could contribute to the taxonomic project—and the conflict resolution mechanisms that Linnaeus initiated but were later developed and enshrined in the still-evolving Zoological and Botanical Codes of Nomenclature (International Commission for Zoological Nomenclature, 1999; McNeill et al., 2006).

The dawning of the information age comes at a time when taxonomy in the Linnaean tradition finds itself particularly under pressure. Throughout most of the 19th and first half of the 20th century, taxonomy and comparative physiology were the twin pillars of classical biology. This changed in the 1950s as the field of biology began a phase of radical expansion. Taxonomy was squeezed from the whole-organism end of biology as ecology, evolution, and the environmental sciences blossomed, but especially from the subcellular end due to the massive growth of molecular biology. Though funding to biology also increased, it could not keep pace with this expansion, and resources flowing into taxonomy declined in real terms (House of Lords, 2002). But worse, taxonomy became perceived by many as unfashionable and even unscientific, purely descriptive rather than hypothesis driven, and almost bureaucratic in the importance it attached to the rules of nomenclature (Krebs, 1992).

In the second half of the 20th century, taxonomy reinvented itself as a modern science. This reinvention had two main components. First was the proposal (now almost universally accepted) that classification should reflect phylogeny, determined using the ideas initially developed by Hennig (1966) and generally referred to today as *cladistics*. Species and other taxa were now considered to be hypotheses defined by clear criteria and tested by data (Hennig, 1996). Evolution had been at the heart of taxonomy since Darwin, but Hennig made the connection more formal and more systematic. Second, the molecular revolution moved out of the specialist genetics laboratory to provide fast, cheap sequence data. Sequencing generates a huge number of new characters with which to build phylogenies and this has led to the construction of very many new trees. It has revolutionized prokaryote taxonomy, which is now almost completely sequence based (Oren, 2004; Oren and Stackebrandt, 2002). For multicellular eukaryotes we are now in a position where robust hypotheses for the relationships between all animal phyla and all plant families are likely to be in place within a decade (Soltis et al., 2005).

In this article we ask whether there needs to be a further component to the reinvention of taxonomy: Is the way in which taxonomic hypotheses are published and presented to the broader scientific community appropriate for a 21st century science? Does it maximize the efficiency of taxonomy? We believe these questions are important not only for the future of the subject but also because there is a greater need now for the products of taxonomic research than ever before. The case for the importance of phylogenetic information for the broader biological community has long been understood. But the ever greater appreciation of the impending threats of climate change, habitat loss, and other motors of global change has reinvigorated biodiversity science and produced a huge demand for the resources to measure and study the diversity of plants and animals on earth (Janzen, 2004). We believe taxonomy must evolve new means of organizing and disseminating information to meet these challenges, and that this will inevitably involve using the Web as the medium of data access and exchange. We are concerned that without these changes Linnaean taxonomy will become increasingly irrelevant to biodiversity science and replaced by other means of quantifying and organizing biodiversity.

DISTRIBUTED TAXONOMY

How is the taxonomy of a particular group of plants or animals organized today? Hypotheses about species identities and other taxa are contained in a scientific literature that may span 250 years and may be written in any language. Species are described with reference to type specimens (reference specimens that act as namebearers for species). The Zoological and Botanical Codes of Nomenclature specify that type specimens should be deposited in major museums or herbaria and are expected to be available for future consultation and study. A sizeable portion of the taxonomic literature about a particular group does not concern its biology—but instead deals with nomenclatural issues (Godfray, 2002). These include the correct application of a Linnaean binomial, resolution of conflicts such as the application of multiple names to one taxon, the interpretation of historical literature that was often written to standards that would be unacceptable today, and the curation of type specimens. The Codes of Nomenclature govern nomenclatural acts and provide a means of resolving conflicts. The taxonomy of any particular group is thus the sum of publications scattered throughout the literature. A major part of being a specialist in a group is the ability to navigate this information space and to understand the location of different sources of information and specimens (Scoble, 2004).

This approach has some huge advantages, not least the fact that it has a proven track record of working. The current distributed taxonomy does not need to be organized or administered but simply accumulates as various workers publish on a particular group (though the information base in libraries and the specimen base in museums, herbaria, and private collections must be paid for and maintained).

It also has some disadvantages. First, the scattered nature of taxonomy leads to such a complicated information space that few nonspecialists can engage directly with the primary taxonomic literature. This may not matter if there is an intervening layer of "processed taxonomy," a regional flora or synoptic revision deliberately written for a general audience that can be used by the nontaxonomist. But these are frequently missing, especially for the tropical fauna and flora that are increasingly the focus of biodiversity studies. The provision of processed taxonomy has never been a formal part of the taxonomic project, and is not covered by any of the Codes. In botany, there has long been a tradition of producing regional floras, though even here the costs of paper publication often mean they are written using a terse, specialized vocabulary with few illustrations and hence are more useful to other taxonomists than the wider community (Frodin, 2001). For some popular groups worldwide (chiefly birds), and for many more groups in more restricted geographic regions, field guides and similar publications provide the processed taxonomy used by working ecologists, environmental scientists, and enthusiastic amateurs. But with the exception of birds, and possibly butterflies, there is probably no major group of organisms for which the resources exist for nonspecialists to identify them anywhere on earth. Even when comprehensive taxonomic monographs do exist, they can become out of date almost before they are published as new species are discovered, classifications are revised, and distributions change. The absence of an up-to-date and accessible taxonomy is a serious shortcoming to an expanding user base and a real impediment to better tropical ecology and to biodiversity monitoring and assessment generally (Mace, 2004). It also undermines arguments for investing more resources in taxonomy: if further taxonomic work leads to an increase in the size of the accumulated fragmented taxonomy with no synthesis for the end-user, it will fare badly in the intense competition for science funding.

A related issue is that even if a nontaxonomist does get to grips with the fragmented nature of taxonomy, they may still be confronted by alternative taxonomic hypotheses. One of the benefits of cladistics in taxonomy was to reinforce the idea that all taxonomic concepts are hypotheses subject to testing and refutation (Kluge, 1997). But this can be unhelpful to end-users who, while accepting the provisional nature of any taxa, still require guidance to navigate their way through the diversity of opinions. One solution to this would be for the systematists involved in a particular group to provide a current consensus taxonomy for the wider community that would evolve in the light of further research. This is what happens de facto when a flora or synoptic revision is published, but there is currently no provision for a dynamic consensus taxonomy to be produced so that taxon concepts can be traced as the consensus evolves and changes.

We referred above to the “taxonomic housekeeping,” which has been an essential part of maintaining the stability and coherence of the subject. Without taxonomists being willing to spend the time interpreting the historic literature, and curating the specimen base, the Nomenclatural Codes could not be operated. But can this process be made more efficient? Should the complete corpus of historical literature remain part of the distributed taxonomy of a group to be consulted forever, or should it be possible to draw a line and reset the taxonomy of a group with a set of names and descriptions upon which future systematic research can be based without the need always to go back to 18th- and 19th-century literature? The arguments for a more frequent use of resetting include the inefficiency of having to invest time in what might be called archaeological bibliography and the difficulty of persuading science funders that this activity is a high priority. Some groups are also in such a taxonomic mess that they are very unattractive for systematists to study. Especially in Europe many species were named numerous times by 18th-, 19th-, and even 20th-century taxonomists and resolving the ensuing confusion is often a lifetime’s work (see, for example, Graham, 1969). It is ironic that there are some insect groups where the taxonomy is better known in tropical America than in Europe because of the dead hand of the past. But against this is the argument that for many taxa, the historical literature is sufficiently well-reviewed that its consultation is not too onerous, and that the rate-limiting steps in taxonomic revisions are activities such as specimen study, type location, and examination that would not be affected by such a resetting. Furthermore, although resetting a taxonomy may reduce the need for nomenclatural housekeeping, it may discourage study of that significant component of the older literature housing nuggets of biological information. Those familiar with the observations recorded by Linnaeus, his predecessors, and successors will fear the neglect of an important source of information. They may consider that finding a balance between purposeful scholarship and an obsession about nomenclature is well worth the effort. The authors of this article are not of one view on the relative merits of the two arguments.

Irrespective of these decisions, can we indeed look forward to a time when most past problems will have been resolved, and when taxonomy, if carried out in accordance with the modern Nomenclatural Codes, will require only a fraction of the effort that currently needs to be devoted to housekeeping? Two developments may frustrate this. First, taxonomy is rightly moving from an activity purely carried out in developed countries to a global exercise. This is particularly appropriate as most biodiversity hotspots are in the developing world (Myers et al., 2000; Grenyer et al., 2006; and see http://www.biodiversityhotspots.org/xp/Hotspots/). But many taxonomists in the developing world frequently lack access to the library facilities that enable them to navigate fully the fragmented taxonomy of a group (Wheeler et al., 2004; Carvalho et al., 2005). As a consequence, they either cannot contribute to taxonomy, or do so in ways that risk producing problems that will become the housekeeping tasks of future generations of taxonomists (for example, by creating synonyms). Second, non-professional taxonomists have always made a major contribution to the field, and one that has been overwhelmingly positive. Yet in some popular groups, for example the larger Lepidoptera, reptiles, orchids, and succulent plants, there are those with strong views and access to privately published journals who are able to describe large number of taxa using standards of evidence at variance with the broad consensus of opinion. Such people have an inalienable right to put forward these nonstandard hypotheses, but a way needs to be found to prevent the resulting plurality of views from hindering the uptake and use of taxonomy by the broader biological community.

TAXONOMY ON THE WEB

Here we review some of the different ways taxonomists have so far made use of the Web. We suspect that the large majority of taxonomists now use the Web in one form or another, and we make no pretense at attempting a comprehensive review. Instead we want, perhaps appropriately, to try to classify the different manners in which the Web has been used so far, in particular by their chief audience and by the form of the output. We concentrate on the taxonomy of living eukaryotes and, as in the rest of the article, say little about fossils, prokaryotes, or viruses.

Fundamental Taxonomic Data on the Web

To do good taxonomy, a researcher needs specimens to study, a catalogue of previous taxon hypotheses, and access to the relevant literature. Aids to all of these have been developed on the Web (Scoble, 2004). These resources are the raw material of taxonomy typically made available for the use of other taxonomists.

Most major museums and herbaria have projects underway to provide catalogues of the specimens they hold on the Web (see the online appendix, available at http://SystematicBiology.org, section 1, for examples of relevant Web sites). The magnitude of the task varies with the size of the collection: for some, the entomological holdings of the major European and North American museums for example, full digitization is surely unlikely ever to be completed. The value of specimen-level data depends on several factors, in particular the accuracy with which specimens are identified, and the amount of associated data stored with them. Typically the date and collecting locality are recorded, but the latter is much more useful if it is expressed in a standardized, machine-readable form (for example, a set of geographical coordinates, http://www.opengeospatial.org/standards/). Of course, adapting traditional locality labels to such standards increases the time and expense of data transcription. For many groups prioritization of the digitization of specimen-level data is both inevitable and sensible. Most museums and herbaria have, or have plans to produce, digitized catalogues of their holdings of type material, the most important single category of specimens to other taxonomists. Other museums have targeted particular taxa for what they can say about conservation priorities or for changes in distribution, for example, brought about by climate change.

The DiGIR project takes Web-based specimen catalogs one step further by allowing institutions to expose their data in a standard way allowing use of generic software tools that query multiple collections at once. Several taxon-specific Web sites have been established that act as portals to this global data set (online appendix section 2).

In addition to the raw specimen data, collection-level metadata on, for example, the number of specimens of a given species housed within a particular museum can provide an invaluable source of information for the taxonomist. Collecting and presenting on the Web this type of data can be done far more quickly than specimen-level digitization (for example, see the www.biocase.org project, which included both elements).

Knowing where a type or other specimen is housed is just the first step in using it to do new taxonomy. But at least for some groups the Web can be used to avoid the need to borrow the specimen or make a journey to the institute where it is housed. Modern high-quality photography combined with high-bandwidth Internet connections allow high-resolution images of specimens to be universally shared (see collected papers in Häuser et al., 2005). This is done most simply for two-dimensional or near two-dimensional specimens. Herbaria sheets lend themselves very naturally to this approach and a number of large projects are already nearing completion. For example, the African Plants Initiative (API) has digitized the types of 191,000 of African plant species (with a goal of 250,000) that are now available for a modest fee via the Web (online appendix section 3). As we shall discuss below, Lepidoptera also lend themselves easily to Web display as their most important taxonomic characters are wing pattern, which is revealed in a single plane in cabinet-set specimens, and the structure of the genitalia, the latter often mounted on a microscope slide. Increasingly Web-visualization will become easier for more three-dimensional structures (Giles, 2005a). Already modern computerized scanning techniques can produce images of small specimens with great depth of focus, allowing a much clearer view of the organism than could be obtained by looking at it under a microscope. A three-dimensional specimen can also be scanned and then examined virtually by computer, rotating and zooming in where needed. This is time-consuming to set up, and then expensive in storage space and bandwidth, but at least the last two constraints are rapidly becoming less important.

Before any serious work can be done on a group of organisms, a catalogue of the names associated with the taxon is essential (Scoble, 1999). Producing a catalogue of names is one of the most important things a systematist can do to facilitate further research on a group. Catalogues and lists are very simply transferred to the Web and numerous projects exist to provide such resources on the Web (online appendix section 4). Catalogues vary greatly in the amount of information they contain, some being little more than a list of names, whereas others carry considerable information about synonymy, literature citation, location of type material, and even distribution and ecology. They also differ in completeness, some reflecting a systematic attempt to locate all names applied to a taxon, whereas others are more an accumulation of information as it becomes available. They may also differ in geographical scope, ranging from regional—the Fauna and Flora Europaea projects, for example, to the global—Species 2000 and ITIS, which aggregate taxonomic databases to produce the annual Catalogue of Life, containing the names of approximately half of all known organisms.

In addition to specimens and a catalogue of names, the taxonomist requires access to the literature, which traditionally has meant a library. One of the recent sea changes in nontaxonomic biology and the rest of science is that the more recent primary literature is now almost universally available in electronic form, although not necessarily free of charge. Scientists in major research institutions now seldom visit the library but obtain the literature purely through their computer terminals. Even for scientists without such strong institutional support, the ease of obtaining literature has improved thanks to the simplicity with which pdf-reprints can be shared via e-mail or downloaded anonymously via the Web.

Taxonomy is different for two reasons. First, as discussed above, papers published decades if not centuries in the past are still of value to the subject. With the exception of a limited number of major historical findings, few papers cited in the ecological sciences are more than a couple of decades old, whereas in molecular biology the citation half-life is only a few years. All major ecological and molecular biological journals have been stored in digital form for the last 20 years and hence the vast majority of works that need to be consulted are at least potentially available on the Web. The taxonomic literature ages much more slowly and hence a much smaller fraction of relevant material is in digital form. Moreover, the taxonomic literature is large and diffuse and a much greater fraction is potentially relevant compared with most sciences. A taxonomist cannot just ignore a paper published in an obscure journal in an unfamiliar language as another scientist might be tempted to. A description anywhere is a taxonomic hypothesis that needs to be considered.

There are a number of large projects that aim to address this problem in the next decade or so (online appendix section 5). The Biodiversity Heritage Library initiative is specifically aimed at digitizing much of the heritage literature—a large proportion of it taxonomic, while big projects such as the Google Books Library Project, which plans to digitize the entire holdings of major collections such as, among others, the libraries of Oxford, Harvard, and Stanford universities will collaterally include many relevant taxonomic works. But it is not yet clear the extent to which material still in copyright will be made available on the Web, and under what conditions (Agosti and Johnson, 2002; Agosti, 2006), or the degree to which digitization projects will be able to capture all the numerous and various ways in which taxonomic hypotheses are published.

We have concentrated here on revisionary taxonomy, but much of the raw data used in phylogenetic studies is already freely and universally available via sequence databases. There is a growing movement throughout the sciences to make more data available in standardized and machine-readable form. To do this for anything more complicated than the simplest type of data requires ontology tables and metadata standards. These are already being developed for some types of data that may become increasingly useful for taxonomists (genomic and EST data for example). Metadata standards for the character matrix information used in constructing phylogenies from morphological data already exist (Thacker, 2003; online appendix section 6), but exchange of data sets is hampered by differences in the terminology used to describe characters and states (Paterson et al., 2004).

Raw Data Syntheses

As will have become apparent from the brief review above, there are already a large number of sources of basic taxonomic data available on the Web, and this has naturally led to the need for a means to index and locate them. We discuss here a few of the projects that seek to do this on the Web. However, some of the most powerful tools for taxonomy on the Web are the major search engines, Google in particular, which are already a prime means of locating new and relevant sites (see also the discussion of mash-ups, below). However, specialized synthesis sites and portals may be able to provide more targeted search tools and summaries for the taxonomist (online appendix section 7).

Before the Web existed, taxonomic abstracting services provided some of the services we now expect from online databases. For animal taxonomists the Zoological Record provided a record of species descriptions that achieved high although not perfect coverage. This Zoological Record continues today as a commercial subscription service (http://scientific.thomson.com/products/zr/) that can be accessed via the Web. The possibility of mandatory registration of new taxa has been repeatedly discussed over the years, but no model that has been able to gain wide support in either the botanical or zoological communities has yet emerged. One proposal for plant name registration was rejected in 1999, though the International Plant Name Index mentioned above fulfils many of the functions of a mandatory scheme. A new proposal for animal names, ZooBank (Polaszek et al., 2005), backed by the International Commission on Zoological Nomenclature (ICZN), is currently being developed (http://www.zoobank.org/query.htm). One of the difficulties of previous proposals has been finding a successful administrative and funding model, and it will be interesting to see whether ZooBank's championing by the ICZN and the involvement of the Zoological Record with its large historical database will allow it to fare better.

There have also been a series of projects to synthesize information in disparate databases to produce catalogues of all species on earth. An example of this is the Species 2000 project, a federated network of 26 databases, each with a series of standard nomenclatural fields. The chief advantage of a project such as Species 2000 is that it provides a snapshot of the progress that has been made in enumerating life on earth. It also provides a universal portal allowing access to the contributing databases. The chief restriction is that for the professional taxonomist working on a particular group, it provides little in addition to what is in the contributing databases, whereas to the nontaxonomist, a name with relatively little associated information is of limited use.

Taxonomic Tools

All aspects of taxonomy are becoming increasingly computer dependent, and in addition to providing a means of sharing raw data, the Web also represents a source of the software needed by practicing taxonomists. Perhaps the most widely available resources are those shared by taxonomists with other biologists, particularly those working in evolutionary and molecular biology: these include software for storing and manipulating sequence data, and for constructing and analyzing phylogenies. Some of this software is proprietary but much is freely downloadable (online appendix section 8).

Nearly all revisionary taxonomy involves the manipulation of large amounts of data, a task that has become much easier with the widespread availability of standard spreadsheet and database programs. However, taxonomic data have their own particular structures and peculiarities, and handling these can be assisted with databases and data-handling routines explicitly designed for this market. Similarly, a very common task for taxonomists is the construction of identification keys. The traditional dichotomous key was a clever means of organizing information before the age of computers, but now identification can often be made easier by allowing multiple entry points into the key and for the user to choose which characters to score (Pankhurst, 1991; Farr, 2006). Both proprietary and free-to-download software for constructing computerized keys is now available over the Web. Lucid (http://www.lucidcentral.org/) is probably the current brand-leader, but Delta (http://delta-intkey.com/) is also very widely used.

In addition to computerized tools to allow taxonomists to carry out their traditional tasks, new tools are also required to allow novel activities to be carried out on the Web. We discussed above some of the raw data resources that were becoming availably digitally, and also projects designed to link them and produce global taxonomic and biodiversity resources. This interconnectivity can be greatly facilitated if databases are built with this goal in mind, and organize their information using standardized data structures. It is the goal of Biodiversity Information Standards (previously known as the Taxonomic Database Working Group [TDWG], http://www.tdwg.org/) to provide the data standards allowing this interoperability. Currently, the biodiversity informatics community is developing data standards for common taxonomic entities such as taxonomic names and concepts, specimen level data, descriptions, and literature.

Another important class of tools is the Globally Unique Identifier (GUID; Leach et al., 2005). These are digital codes that, when used with an appropriate infrastructure, can be used to identify and retrieve information about a variety of objects and concepts. In taxonomy, such identifiers could be applied to specimens, papers, illustrations, and taxa. They are particularly valuable because logical connections between different identifiers can be simply constructed in a machine readable way. At its simplest this might just say that specimen x is determined as species y but by using a GUID objects are connected to a potentially complex, global web of related (meta-)data that software can access without the need for human intervention (Page, 2006). The Global Biodiversity Information Facility (GBIF; http://www.gbif.org) is currently piloting a scheme to use a particular GUID technology, LifeScience Identifiers (LSIDs; Clark et al., 2004; Taxonomic Databases Working Group, 2006) for uniquely identifying data in biodiversity and systematics.

Web Publication

The simplest way that taxonomy can move to the Web is for the current publication model to remain unchanged but that in addition to (or instead of) the production of a paper the results are published on line. As discussed above, all mainstream science journals are published in digital form, and projects exist to digitize the past biodiversity literature. Recently, some primarily digital taxonomic journals have been started (for example, *ZooTaxa*, Zhang, 2006; online appendix section 9), but there is also a huge number of small, low-circulation journals that are not available digitally and have no plans to move to the Web. Some of these journals are essentially private publishing ventures.

The current Zoological Code allows "publication" in Web-only journals provided a small number of copies are stored in major libraries. These copies may be in the form of CD-ROMs, which is a controversial provision that has been attacked by some taxonomists. The chief worry about allowing Web-only publication seems to be that the content might be ephemeral and not easily available in the future. In fact, at the moment it is possible to access the vast majority of pages previously mounted on the Web as they have been archived by organizations such as The Internet Archive (http://www.archive.org). If such projects continue and the costs of data storage keep coming down, as seems very likely, these concerns are unlikely to arise. Alternatively, were Web publication in taxonomy to increase, the libraries of our major museum and herbaria might be persuaded to take on the responsibility of preserving and archiving the digital taxonomic record, in exactly the same way they currently maintain their paper-based libraries today.

Moving to Web-only publication for the majority of the more technical taxonomic publications has some very great advantages. Taxonomy is expensive to publish because papers are often long, and often require copious illustrations. Much, though by no means all, of the difficulties nonspecialists experience when consulting the basic taxonomic literature is caused by the terse vocabulary and underillustration that the cost of paper publishing imposes. Web publication gives taxonomists the space they need to be understood.

Purely digital papers also allow features to be added that would not be possible in traditional publications. Sound recordings are very useful in the taxonomy of some groups, as is video. Illustrations can be provided at much higher resolution than in journals, and facilities such as user-driven zooming and magnification incorporated (Häuser et al., 2005). Enhanced keys of the type mentioned in the last section can also be included. More technically, parts of a digital publication (such as the diagnosis, for example) can be marked-up in a way that allows information in the publication to be incorporated automatically into broader digital resources. Metadata about the contents can also be incorporated, and these can be scanned by search engines to make it easier for potential users to become aware of or locate the publication (Lagoze and Van de Sompel, 2001).

For the Web to be an attractive publication medium for taxonomists, it needs to be citable, accessible, and to have the same or higher prestige as paper publications. E-publication is in its infancy, but projects such as the Open Archive Initiative have already developed protocols and standards that allow the efficient dissemination of metadata from online archives of digital content. Such protocols also enable the number of Web citations to be calculated (for example, Thompson Scientific's Web Citation Index, or the not-for-profit Citseer), and one can envisage more sophisticated indices being developed that record the use of different taxonomic concepts. We think that it will be possible to develop Web-based methods of assessing the use of taxonomy that are fairer to the field than current scientific paper based–citation counts.

Who should pay for Web publication? At present most taxonomic journals are supported by library and individual subscriptions. This model could continue with Web publication; typically the costs per page of publishing on the Web are much smaller than for print, though journals have also found their income to fall as individuals drop personal subscriptions when they have easy access to digital institutional subscriptions. The alternative is to make the Web publication freely available to everyone at the point of access and move the costs of production to the author. There is intense debate in other areas of biology about the long-term feasibility and stability of the author-pays financial model, especially in the biomedical sciences where some of the major research sponsors are demanding open-access publication (Suber, 2005; Wren, 2005). How open-access publication might work in taxonomy is not yet clear.

Wikis—Collaborative Authoring

A very different model for how taxonomy might develop on the Web is through the idea of community-based projects such as wikis. The changes made to a wiki are unmoderated, can be made by anyone, often without the need for registration, and are not checked for accuracy and quality. Without doubt the most famous site is Wikipedia (online appendix section 10), an encyclopedia of everything, which currently (early 2007) contains more than 1.5 million articles in English and perhaps 3 million in all languages. Within broadly defined limits, users are free to add entries and to edit or overwrite entries written by others. Naturally, the quality of different articles varies, though a large community of people now uses it very extensively (see, e.g., Benkler, 2006). In 2005, the journal *Nature* controversially compared Wikipedia entries favorably to those in the Web-version of the venerable *Encyclopaedia Britannica* (Giles, 2005b), though the latter argued vigorously that the comparison was unfair (Encyclopedia Britannica, Inc., 2006).

Wikipedia itself contains many entries to species and other taxa but the organization behind it, the Wikimedia Foundation, has gone further and in 2005 set up Wikispecies, a "free directory of life" consisting of articles organized by the Linnaean hierarchy. Currently it contains over 82,000 entries, though many of these are simply names ("stubs"). However, if the popularity of its parent organization is replicated here, the number of entries and their richness of information are likely to increase rapidly.

The wiki model has both strengths and weaknesses, as exemplified by Wikipedia itself. The strengths are its democratic and open nature, and the benefits of harnessing a huge community of people to produce a substantial volume of product. Its weaknesses are its susceptibility to sabotage, and the lack of quality and authoritativeness of some of its entries. Wikipedia has had to restrict editing of some entries, though obviously taxonomic articles are unlikely to be quite as politically sensitive. Also, Wikipedia has failed to attract many experts to provide entries, perhaps because of the lack of any editorial moderation. Partly because of this, a new project called Citizendium that places experts in a privileged position to review the work of nonexperts has been set up. We suspect the Wikispecies model is likely to prove popular and will grow in size, but whether it provides taxonomic resources of value to biodiversity scientists and serious amateurs is less clear.

Mash-Ups—Federating Data

The idea of having a Web page for every species (Wilson, 2003) is an enticing vision underlying Wikispecies, Tree-of-Life (http://www.tolWeb.org/tree/) and, before its apparent demise, the All Species Foundation (http://www.all-species.org/). The challenges of creating this resource are enormous, but could it be done automatically, by searching the Web for all the information available on a particular taxon? Sites produced in this way are called mash-ups (Butler, 2006). The iSpecies project (http://darwin.zoology.gla.ac.uk/rpage/ispecies/) is a simple test of how this might be done; code is written to interrogate biodiversity, image, sequence, and literature databases, making use of standard application programming interfaces (APIs) that they expose to other software. The user enters a taxon name and the program returns a list of links to the resources indexed by that name. It is already quite impressive the number of links that are found for even quite obscure taxa, and it is certain that this will increase in the future. What iSpecies demonstrates is the enormous potential that could be achieved if all databases from natural history museums and herbariums were linked in a searchable distributed network.

However, the value of this approach depends on the reliability of the underlying taxonomic nomenclature, as well as the taxonomic accuracy of the original studies that the mash-up collates. Taxonomic names are not always unique identifiers for taxonomic hypotheses, and consequently software that uses a Linnean binomial to search other resources may aggregate data about different taxonomic concepts (Berendsohn, 1995). It will work best for well-known, easily distinguishable species, and poorly for those groups where either the taxonomy is in flux or specimens difficult to identify. Some of these problems can be addressed by the use of Globally Unique Identifiers such as the LifeSciences Identifiers (LSID) technology discussed above, which makes it possible to identify data in an explicit and machine-readable way. The Universal Biodiversity Indexer and Organiser (http://www.ubio.org) is a project that has annotated almost nine million taxonomic names with LSIDs. Software clients can access the data associated with these LSIDs using a protocol called SOAP (Box et al., 2000).

There is an argument, termed the wisdom of crowds (Surowiecki, 2004), that says that the truth emerges as an average of the opinions of many. Ask a large number of people the number of baked beans in a big jar and the average is often remarkably close to the true value. But this seems to us a poor analogy for how taxonomic questions might be settled and although we see mash-ups as a very valuable tool for exploring potential information on the Web (especially as they can be set up quickly), their use will be enhanced by better authoritative Web taxonomies rather than being a substitute for them.

Unitary Taxonomies

Another radical use of the Web in taxonomy would be to build a Web-based environment in which all aspects of the taxonomic project could be pursued. New versions of the Codes could insist that for a nomenclatural act to be valid it would have to published within this environment. Such an approach has been called a unitary taxonomy (Godfray, 2002) and although it would retain much of taxonomy as currently practiced, the fact that all the resources relevant to the systematics of a particular group would be present at a single site would lead to a very different information landscape compared with the current distributed taxonomy.

There are different ways a unitary taxonomy could be constructed, and here we briefly outline one particular model (Godfray, 2002). Consider a taxon containing a significant number of species distributed throughout the globe. The first Web revision of such a taxon would consist of standardized descriptions and illustrations of all the species and higher taxa it contained, much as would a major paper-based revision today (or rather a Web-enhanced but otherwise traditional Web revision). The first Web revision would also contain synonymies and other alternative taxonomic hypotheses from which the compilers of the consensus taxonomy might dissent. But unlike a standard revision today, once the first Web revision was completed and published, it would fully circumscribe the set of names and concepts that would need in the future to be referred to by systematists working on the group. The information space for the taxon would be clearly delimited and accessible to all, no matter how restricted their access was to traditional libraries. No longer would there be a risk of nomenclatural instability due to the discovery of an overlooked or misinterpreted specimen. Clearly the first Web revision would be a major step in the taxonomy of a group and important to be of as high a quality as possible. One could envisage an extensive period of Web-based refereeing by the community of people working on the taxa, and then the equivalent of an editorial board, a "taxon committee," deciding among competing views.

How would taxonomy proceed within the unitary environment? Suppose a systematist describes a new taxon, or proposes some other change to the taxonomy of the group. He or she would enter it into the Web site and it would then go through a period of Web-based review by the community. At the end of this, the editorial committee, based on the community responses, would decide whether it became part of the current Web revision, or whether it was provisionally excluded. The current Web revision would thus represent a provisional, consensus taxonomy. The end-users of taxonomy, the ecologists, environmentalists, other biologists, and amateurs, would have an unambiguous recommendation that they could use in their work (with the possibility of investigating alternative hypotheses should they want). But importantly all currently rejected hypotheses would be maintained on the Web site for further consideration, and they may, if the evidence base changes, become part of a future consensus taxonomy. Strong differences of opinion within the community may still be present despite the consensus presented for wider consumption. The point is for the user community to be protected from the confusion of a plethora of different hypotheses, but not to be misled by any implication that the consensus is unchanging or that it is intended to mask possible underlying alternative views. Taxonomists rightly guard their independence and the plurality of opinion possible with a distributed taxonomy. All the benefits of this could be retained in a unitary taxonomy environment, with the one proviso that a community provisional consensus would also be provided to give a product of maximum use to the consumers (and funders) of taxonomy. Again, confidence in the fair determination of a consensus taxonomy would require a representative taxon committee operating under clear and transparent guidelines; the International Commissions might be asked to include overseeing this process as part of their roles.

What are the advantages and disadvantages of this model for doing taxonomy? Consider the advantages first. As with any Web-published revision, taxonomists would be liberated from the constraints of paper publication; they could produce as long a description as necessary, with as many illustrations as required. Video, sound, or information in any other digitizable media could be included. The information would be available (we believe for free), to anyone, anywhere, no matter how poor their local library facilities, provided, of course, they had an Internet connection. Even the latter constraint might be relaxed as the full taxonomy could be downloaded onto a CD or other medium, at the cost of some functionality.

There are further advantages to unitary taxonomies that would not apply to a Web-based but otherwise distributed taxonomy. If the complete nonspecimen resources for the taxonomy of a group were available at a single site, it would make it immeasurably easier for systematists in developing countries to make valuable contributions to a group's taxonomy. (For that matter it would greatly help the large number of taxonomists who live in developed countries but who do not have easy access to the best specialist libraries.) By ensuring that all new taxa were described in one place, to at least a minimum standard, and refereed in a uniform manner, it would help prevent the nomenclatural confusions whose resolution occupies so much of the time of contemporary taxonomists. It would also allow smart use of modern digital information technology. All this, we think, would allow taxonomists to be more productive and more efficient.

We believe that it is important for unitary taxonomies to be integrated within the global network of biodiversity databases. We mentioned above Globally Unique Identifiers (GUIDs; Leach et al., 2005): every taxon concept (in the sense of Berendsohn, 1995) in a unitary taxonomy would have a GUID. As the current Web taxonomy replaces the previous one, any alteration in the taxon concept attached to a particular name would result in a new GUID for the new hypothesis and a logical statement connecting the two. A biologist writing a paper who today might refer to a Linnean binomial could then supplement it with a GUID that denotes the particular taxonomic concept he or she is using. In the future, readers could understand immediately in what sense the binomial was being used, and then employ the unitary taxonomy to tell them how that taxon is to be interpreted today. We would hope that journal editors might begin to insist on such designations, just as they would today demand a GenBank (http://www.ncbi.nlm.nih.gov/Genbank/) accession number for a quoted DNA sequence. Much more sophisticated uses of GUIDs are possible; any item of information including a physical specimen can be assigned a GUID and linked logically to the taxon in a unitary taxonomy.

GUIDs are also valuable because they expose information in a standard way to other software. The information in a unitary taxonomy could be harvested by other systems; for example, providing databases of specimens, images, names, and publications. One can envisage a community of taxonomic and taxonomy-related information systems that exchange information with each other and seamlessly provide different types of resources to different categories of end-users (Paterson et al., 2006). GUIDs in unitary taxonomies may help underline the importance of the subject by providing metrics for taxonomic use (frequency of citation of GUIDs). Taxonomy is notoriously badly served by current science research bibliometrics, possibly affecting its funding base.

What about the disadvantages? The most frequent criticism of unitary taxonomy is that it is top-down and authoritarian, at variance with the bottom-up and individualistic tradition of taxonomy (Thiele and Yeates, 2002). Insisting that all properly formulated (i.e., Code compliant) taxonomic hypotheses are maintained on the Web site might go some way to allay this, but nevertheless the designation of a consensus taxonomy (even with the caveats about checks and balances discussed above) is new and raises understandable concerns. It is surely true that major treatments, such as monographs, are typically treated de facto as a consensus by most users, even those from the community working on that group, but that is different from a formal designation of a consensus. Taxonomy needs to ask itself whether the possible dangers of designating a consensus taxonomy are outweighed by the advantages to the end-users of taxonomy, and whether it will stimulate a larger and more vocal constituency of end-users who will then argue for more resources and funding for the subject.

A second potential disadvantage is that whereas the current distributed model of taxonomy requires little maintenance and upkeep other than maintaining a library and a vibrant publishing sector, the unitary model requires a Web site to be maintained and for there to be an overseeing taxon committee. The obvious candidates for running these Web sites are our major museums and herbaria. But few if any of these institutions have spare financial capacity, and they are already being requested to do new things in addition to traditional research and curation. New money will certainly be required, and it is inevitable (as well as sensible) that should unitary taxonomies to be adopted it would be slowly, with different groups transferring to the Web at different times and rates. More positively, the burden of supporting unitary taxonomy sites could be shared around the globe, and in particular the many excellent taxonomists in developing countries could contribute as equal partners to the taxonomic project to a far greater degree than they do today, hampered as they are by poor access to the taxonomic literature. Note that ultimately there will be savings in terms of journal subscriptions, which might even be sufficient to pay for the unitary model, though this is unlikely to be felt in the short term. An injection of funds to provide the "activation energy" would still be essential.

Nothing in the model of unitary taxonomies described above changes the underlying specimen-based approach to systematics as currently practiced. Type specimens would still be designated, and new systematic research would still normally require the examination of types. Borrowing types, or visiting institutions to view types, will still often be a rate-limiting step in taxonomy, and one that will disenfranchise many developing-world taxonomists who cannot afford to travel or live in states to which institutions in developed countries are unwilling to lend. By providing many more illustrations, Web-based taxonomy can help by making clear which specimens need to be consulted physically. What is more, as discussed in the section on raw taxonomic resources on the Web, modern imaging techniques can be used to mount on the Web images of specimens and associated morphological preparations that may be easier to interpret than by traditional physical examination.

THE CATE PROJECT

It is inconceivable that the more radical facets of a unitary taxonomy would be acceptable to the taxonomic community, in particular its mandatory aspects, without a demonstration that the taxonomy of a significant group could be mounted at a single site. Likewise, it remains to be shown that a viable Web environment can be constructed in which taxonomy can be pursued. We also suspect that the funding for unitary taxonomies, as well as the essential commitment by major institutions, would demand proof of principle. The goals of the CATE project are to develop one-stop taxonomic Web resources for two significant (>1000 species) groups of plants and animals, and to develop the software resources to allow new taxonomic hypotheses to be advanced within the site. In addition, it is hoped that the project will flag up any unanticipated problems and issues, as well as demonstrating to the generators and users of taxonomy the benefits of simplified access to systematic resources. Although the taxon treatments that will be mounted will be the equivalent of a first Web revision, it will of course have none of the special nomenclatural privileges of unitary taxonomies proposed above, and similarly any new taxonomic hypotheses would need to comply with the Codes as currently formulated. Yet we hope it will be a test case for the feasibility, advantages, and disadvantages of moving away from distributed taxonomies.

The project acronym stands for Creating a Taxonomic E-science and is funded by the UK's Natural Environment Research Council's E-Science Program. It is a consortium led by the Natural History Museum (NHM), London, and including the Royal Botanic Gardens, Kew (RBGK), and Oxford University. The NHM is responsible for the animal exemplar group, the Sphingidae (hawk moths, Lepidoptera); RBGK is responsible for the plant exemplar group, part of the Araceae (aroids); whereas software development is led by Oxford. The project began in early 2006 with material to be displayed on the Web as it is produced (the first substantial upload will take place in early 2007; online appendix section 11).

The two exemplar taxonomic groups were chosen to be of a feasible size to tackle with the resources available in the project, and to be groups for which a relatively mature taxonomy already existed (Mayo et al., 1997; Kitching and Cadiou, 2000). A plant and animal group were chosen to reflect the two different International Codes and their associated taxonomic traditions. The two groups differ considerably in the kind of material normally included in a standard revision. For the Sphingidae (as for most Lepidoptera), there is not the tradition of producing dichotomous keys, and instead diagnosis and identification typically relies on a text description, illustration of wing patterns, and illustrations of male and female genitalia. In contrast, aroid taxonomy requires a text description and a dichotomous key, ideally with illustrations of flower and leaf morphology (Mayo et al., 1997; Croat, 2004).

The Web revision for each taxon will consist of a series of "pages" for each taxonomic concept. A species page for a hawk moth will consist of a series of photographs of male and female specimens, both uppersides and undersides, and photographs of the genitalia. There will be a text description that concentrates on diagnostic features. The location and designation of type material will be described, as well as proposed synonymy. At least within the current project the distribution of the species will be given at the country/major region level using the standardized codes developed by Biodiversity Information Standards (Brummitt et al., 2001). There will be a section on biology that will initially include, where known, food plant and phenology, and any other major feature of the taxon. We hope to supplement this data with links to external sources of data, such as the project to barcode the Sphingidae of the world (http://www.lepbarcoding.org). The species pages for the aroids will also consist of a text description and diagnosis, and illustrations where appropriate of herbarium sheets, living specimens, and drawings of diagnostic features. Information on habitat, altitudinal range, and flowering phenology will also be provided, as will proposed synonymy and the location of type specimens. All species will be keyed using interactive keys. Where necessary, the author of the species page can add comments describing the basis for his or her decisions.

The initial goal of the project is to mount a first Web revision, though this will largely reflect the current taxonomy of each group rather than incorporating new research. This taxonomy will then be available for review and revision by the hawkmoth and aroid communities and, indeed, anyone else. Changes, such as new species and synonymies, will take the form of submitted contributions in the form of pages, which will then be made available on the site for a period of review and comment. An editorial committee independent of the authors will then make the final decision as to what enters the consensus, a process akin to the standard peer review system common to all the sciences. The next edition of the Web taxonomy will include both the consensus taxonomy and all alternative hypotheses. All previous editions of the taxonomy will be archived and accessible on the Web site.

The structure of each Web site consists of an underlying database and a software application that makes the data available as Web pages. Users are, by default, directed to the consensus classification, but are able to view alternative hypotheses, proposed changes and previous Web taxonomies where they exist, aspects of the Web site that should be particularly attractive to specialist users.

Users will be able to register and propose changes to the current consensus classification. Changes could include adding new information to an existing taxon page or defining new taxonomic hypotheses. Taxonomists will be able to develop a proposal using resources available within the site, and, when complete, submit it for Web review. Before successful submission, the software will ensure that a set of minimum criteria are fulfilled; for example, a new species requires a diagnosis, designation of type specimen, etc.

Once a proposal has been submitted to the Web revision, it is available for other users to review, in an open process. After a sufficient period of review, the editors will examine the submission and the reviews, and decide whether the changes should be incorporated into the consensus classification and provide an explanation of why the new taxonomic hypothesis was or was not included.

One of the goals of unitary taxonomies is to increase the engagement of the broader biological community with the taxonomic process, and to produce a biologically richer information source than is currently associated with a traditional taxonomic revision. With this aim in mind, users will be able to add information to the species page, whether it is a new image, a fact about the biology of the species, a new distributional record, or indeed anything else. These wiki-like contributions will be accessible from but not part of the formal taxon page. The information they contain may become part of the formal taxon page when it is revised, and after the normal refereeing process. If this happens then the original contribution is referenced, as one would a traditional scientific paper. As with other wikis, people can also comment on contributions.

Finally, the Web pages will link to other resources available on the Web. The common reference may be either the Linnaean binomial or the formally assigned LifeScience Identifier (LSID; Clark et al., 2004) if the other resource chooses to use them. In addition to specific links made by the author of each taxonomic hypotheses, further Web resources can be identified by automated search procedures on the model of the iSpecies project (aggregating results from other resources such as search engines) or by interfacing with major biodiversity clearing houses, such as the GBIF project, as they develop. Of course, it is important to distinguish the refereed contents of the taxon Web pages from the material whose relevance to the classification presented in the unitary taxonomy is dependent on the accurate application of the Linnean name or other identifier to the data.

If successful, a unitary taxonomy site on the CATE model should help achieve the following goals. (i) To make it easier for taxonomists to do research and for their research to be more visible. (ii) To maintain the pluralistic model of current distributed taxonomy by ensuring alternative hypotheses about taxa are mounted for future examination, but with the advantage that they are accessible within a single site on the Internet. (iii) To create an enhanced product for the end-users of taxonomy through a dynamically updatable consensus taxonomy, along with the electronic means of navigating among and tracing references between different hypotheses and editions. (iv) To increase the constituency of people assisting the taxonomic enterprise by encouraging wiki-style contributions (with appropriate recognition and reference). (v) To connect taxonomy further with the broader biological community by links to other major resources and by automated search of the Web. (vi) To incorporate an online peer-review system to enhance the scientific authority of the Web site. (vii) To provide a flexible system for online revisionary taxonomy that can be customized through open-standards and protocols and rendered applicable to a diverse spread of taxa.

CONCLUSIONS

Though we believe very strongly in the importance of taxonomy for biodiversity scientists trying to study and protect life on earth, we are also concerned about the threats facing the subject. To confront these challenges we think taxonomy has to evolve to become more obviously relevant to the environmental sciences, and that the only realistic means to achieve this goal, both in terms of building taxonomy and providing access to it, is to render it primarily Internet based. We do not claim that the model of Web-based, unitary taxonomy that we have described above is the only way to proceed, but we would argue that it and related ideas deserve serious consideration and discussion. We finish by listing some of the threats that face the subject, and how unitary taxonomies may help allay them.

The first threat we perceive is that taxonomy is thought of as poor value for money by science funders (Krebs, 1992; House of Lords, 2002), that the output of taxonomy, the current distributed taxonomy of a group that needs further interpretation before it can be used by nonspecialists, is not seen to be a priority, especially for those diverse, species-rich groups whose classification and identification is essential for many areas of ecology and biodiversity science. By centralizing resources for taxonomy at one site a unitary taxonomy will be more efficient, whereas the production of a consensus taxonomy will provide an important product for the people who use taxonomic research. We believe this can be achieved without breaking with the pluralist tradition of taxonomy where every worker has the right to advance a hypothesis about a group. We are also optimistic that if taxonomy provides better products for its end-users this will increase the number of people arguing for more funds for the subject.

Arguments for why more resources should be directed to taxonomy frequently point out the importance of establishing species boundaries and identities (e.g., Bionet, 2007). Without a species being correctly defined or identified, control of an important human disease agent or vector fails, or the correct biological control agent of an agricultural pest is not found, or the source of a useful plant secondary compound is overlooked. Identifying species and determining species boundaries has been one of the most important traditional tasks of taxonomists and will continue to be so. But taxonomists should beware of making this a major justification for supporting the subject. Even considering the complexities of species limits discussed in other contributions to this symposium, today it is often easier for a biologist to develop molecular tools to identify species or to determine the amount of genetic differentiation between two populations than to use morphological approaches to answer the same questions. With the relentless advance of molecular techniques, this will become even simpler in the future. Indeed, for questions of medical or economic importance, it is pretty much inconceivable today that molecular methods will not be applied. Of course, it may be taxonomists wielding the pipette but our point is that they no longer necessarily have a special expertise or monopoly in this field. To maintain the role of taxonomy in addressing these questions, its practitioners must offer more than could be achieved by the ordinary biologist. Maintaining unitary taxonomy Web sites that collate and curate the extraordinarily rich information base on a taxon is one way for the subject to continue to engage with the rest of biology and beyond.

Biodiversity scientists who need to identify large assemblages of species must, for nearly all groups, either work with taxonomists or use the fruits of their labor. For groups where specialists are scarce or unavailable, and where the taxonomic literature is difficult or impossible for nonspecialists to use, there is a significant impediment to research (Gotelli, 2004). Today there is a way round this impediment. Short diagnostic DNA sequences, DNA barcodes, can be used to distinguish amongst large assemblages of species. The current barcoding initiative (Hebert et al., 2003, and see http://barcoding.si.edu/) based on the mitochondrial *co1* gene has attracted a variety of criticisms (Meyer and Paulay, 2005; Brower, 2006; Waugh, 2007) and may not work for all groups (including the not insignificant Plantae) or be capable of distinguishing amongst all species pairs. Nevertheless it works for many and can be developed further to include new markers that are likely to make it much more efficient. There seems to us a real danger that it will prove easier for ecologists and other environmental scientists to switch from traditional taxonomy to a "DNA taxonomy" where individuals are assigned to clusters in sequence space rather than to base their work on traditionally described species. Some enthusiasts for DNA taxonomy (though not the leaders of the barcoding initiative; see Schindel and Miller, 2005) even welcome and promote this as the only future, on the grounds it will make taxonomy what they think of as modern and efficient. Our view is that it would be a major retrograde step, disconnecting future taxonomy from the wealth of knowledge that has been built up and indexed under the Linnaean system (see also Vogler and Monaghan, 2007). It would also lead to a schism in taxonomy, at least until DNA sequencing became cheap enough to be universal. However, we believe that a Web-based, unitary taxonomy has the capacity to include all kinds of data, whether morphological, behavioral, molecular, or of any other kind, and thus render unnecessary and counterproductive this and possible future schisms.

REFERENCES

Agosti, D. 2006. Biodiversity data are out of local taxonomists' reach. Nature 439:392.

Agosti, D., and N. F. Johnson. 2002. Taxonomists need better access to published data. Nature 417:222.

Beccaloni, G. W., M. J. Scoble, G. S. Robinson, and B. Pitkin, eds. 2003. The Global Lepidoptera Names Index (LepIndex). World Wide Web electronic publication. http://www.nhm.ac.uk/entomology/lepindex [accessed 6 January 2007].

Benkler, Y. 2006. The wealth of networks. Yale University Press, New Haven, Connecticut.

Berendsohn, W. G. 1995. The concept of "potential taxa" in databases. Taxon 22:207–212.

Bionet. 2007. http://www.bionet-intl.org/opencms/opencms/case Studies/default.jsp [accesseed May 16 2007].

Blunt, W. 2002. The complete naturalist: A life of Linnaeus. Frances Lincoln, London.

Box, D., D., Ehnebuske, G. Kakivaya, A. Layman, N. Mendelsohn, H. Frystyk Nielsen, S. Thatte, and D. Winer. 2000. Simple Object Access Protocol (SOAP) 1.1. World Wide Web electronic publication. http://www.w3.org/TR/2000/NOTE-SOAP-20000508/ [accessed 2 April 2007].

Brower, A. V. Z. 2006. Problems with DNA barcodes for species delimitation: "Ten species" of *Astraptes fulgerator* reassessed (Lepidoptera: Hesperiidae). Syst. Biodiv. 4:127–132.

Brummitt, R. K., F. Pando, S. Hollis, and N. A. Brummitt. 2001. World geographical scheme for recording plant distributions, edition 2, *in* Plant Taxonomic Database Standards, number 2, edition 2. Hurst Institute for Botanical Documentation, Carnegie Mellon University, Pittsburg, Pennsylvania.

Butler, D. 2006. Mashups mix data into global service. Nature 439:6–7.

Carvalho, M. C., F. A. Bockmann, D. S. Amorim, M. Vivo, M. Toledo-Piza, N. A. Menezes, J. L. Figueiredo, R. M. C. Castro, A. C. Gill, J. D. McEachran, L. J. V. Compagno, R. C. Schelly, R. Britz, J. G. Lundberg, R. P. Vari, and G. Nelson. 2005. Revisiting the taxonomic impediment. Science. 307:353.

Clark, T., S. Martin, and T. Liefeld. 2004. Globally distributed object identification for biological knowledgebases. Brief. Bioinformatics 5:59–70.

Croat, T. B. 2004. Revision of *Dieffenbachia* (Araceae) of Mexico, Central America, and the West Indies. Ann. Mo. Bot. Gard. 91:668–772.

Encyclopedia Britannica, Inc. 2006. Fatally flawed: Refuting the recent study on encyclopedic accuracy by the journal Nature. World Wide Web electronic publication. http://corporate.britannica.com/britannica_nature_response.pdf [accessed 2 April 2007].

Farr, D. F. 2006. On-line keys: More than just paper on the web. Taxon 55:589–596.

Forey, P. L., C. J. Humphries, I. J. Kitching, R. W. Scotland, D. J. Siebert, and D. M. Williams. 1992. Formal classification *in* Cladistics: A practical course in systematics. Systematics Association Publications, volume 10. Oxford University Press, Oxford, UK.

Frodin, D. G. 2001. Guide to standard floras of the world, 2nd edition. Cambridge University Press, Cambridge, UK.

Giles, J. 2005a. Remote-control microscope to ease work in taxonomy. Nature 433:673.

Giles, J. 2005b. Internet encyclopaedias go head to head. Nature 438:900–901.

Godfray, H. C. J. 2002. Challenges for taxonomy. Nature 417:17–19.

Gotelli, N. J. 2004. A taxonomic wish-list for community ecology. Philos. Trans. R. Soc. B 359:585–597.

Graham, M. W. R. d. V. 1969. The Pteromalidae of north-western Europe (Hymenoptera: Chalcidoidea). Bull. Br. Mus. (Nat. Hist.) Entomol. 16:908.

Grenyer, R., C. D. L. Orme, S. F. Jackson, G. H. Thomas, R. G. Davies, T. J. Davies, K. E. Jones, V. A. Olson, R. S. Ridgely, P. C. Rasmussen, T. S. Ding, P. M. Bennett, T. M. Blackburn, K. J. Gaston, J. L. Gittleman, and I. P. F. Owens. 2006. Global distribution and conservation of rare and threatened vertebrates. Nature 444:93–96. 31

Häuser, C. L., A. Steiner, J. Holstein, and M. J. Scoble, eds. 2005. Digital imaging of biological type specimens: A manual of best practice. Staatliches Museum für Naturkunde, Stuttgart.

Hebert, P. D. N., A. Holstein, S. L. Ball, and J. R. DeWaard. 2003. Biological identifications through DNA barcodes. Proc. R. Soc. Lond. B 270:313–321.

Hennig, W. 1966. Phylogenetic systematics. University of Illinois Press, Urbana, Illinois.

House of Lords 2002. Report of Science & Technology Committee. http://www.publications.parliament.uk/pa/ld200102/ldselect/ldsctech/118/11801.htm [accessed 16 May 2007].

International Commission on Zoological Nomenclature. 1999. International Code on Zoological Nomenclature, 4th edition. International Trust for Zoological Nomenclature, London.

Janzen, D. H. 2004. Now is the time. Philos. Trans. R. Soc. B 359:731–732.

Kitching, I. J., and J. -M. Cadiou. 2000. Hawkmoths of the world: An annotated and illustrated revisionary checklist (Lepidoptera: Sphingidae). The Natural History Museum, London & Cornell University Press, Ithaca, New York.

Kluge, A. G. 1997. Testability and the refutation and corroboration of cladistic hypotheses. Cladistics 13:81–96.

Koerner, L. 1999. Linnaeus: Nature and Nation. Harvard University Press, Cambridge, Massachusetts.

Krebs, J. R. 1992. Evolution and biodiversity: The new taxonomy. Natural Environment Research Council Report. Swindon, UK.

Lagoze, C., and H. Van de Sompel. 2001. The open archives initiative: building a low-barrier interoperability framework. Pages 54–62 *in* International Conference on Digital Libraries. Proceedings of the 1st ACM/IEEE-CS Joint Conference on Digital Libraries. Roanoke, Virginia.

Leach, P., M. Mealling, and R. Salz. 2005. IETF RFC 4122. A Universally Unique IDentifier (UUID) URN Namespace. World Wide Web electronic publication. http://www.ietf.org/rfc/rfc4122.txt [accessed 6 January 2007].

Mace, G. M. 2004. The role of taxonomy in species conservation. Philos. Trans. R. Soc. Lond. B 359:711–719.

Mayo, S. J., J. Bogner, and P. C. Boyce. 1997. The genera of Araceae. Royal Botanical Gardens, Kew.

McNeill, J., F. R. Barrie, H. M. Burdet, V. Demoulin, D. L. Hawksworth, K. Marhold, D. H. Nicolson, J. Prado, P. C. Silva, J. E. Skog,

J. H. Wiersema, and N. J. Turland. 2006. International Code of Botanical Nomenclature (Vienna Code). A.R.G. Gantner Verlag, Ruggell, Liechtenstein.

Meyer, C. P., and G. Paulay. 2005. DNA barcoding: Error rates based on comprehensive sampling. PLoS Biol. 3:e422.

Myers, N., R. A. Mittermeier, C. G. Mittermeier, G. A. B. da Fonseca, and J. Kent. 2000. Biodiversity hotspots for conservation priorities. Nature 403:853–858.

Oren, A. 2004. Prokaryote diversity and taxonomy: Current status and future challenges. Philos. Trans. R. Soc. Lond. B 359:623–638.

Oren, A., and E. Stackebrandt. 2002. Prokaryote taxonomy online: Challenges ahead. Nature. 419:15.

Page, R. D. M. 2006. Taxonomic names, metadata, and the semantic web. Biodivers. Informatics 3:1–15.

Pankhurst, R. J. 1991. Practical taxonomic computing. Cambridge University Press, Cambridge, UK.

Paterson, T., J. B. Kennedy, M. R. Pullan, A. Cannon, K. Armstrong, M. F. Watson, C. Raguenaud, S. M. McDonald, and G. Russell. 2004. A universal character model and ontology of defined terms for taxonomic description. Pages 63–78 *in* Data integration in the life sciences (E. Rahm, ed.). Lecture Notes in Computational Science, 2994. Springer, Berlin/Heidelberg.

Patterson, D. J., D. Remsen, W. A. Marino, and C. Norton. 2006. Taxonomic indexing—Extending the role of taxonomy. Syst. Biol. 55:367–373.

Polaszek, A., et al. 2005. A universal register for animal names. Nature 437:477.

Pullan, M. R., M. F. Watson, J. B. Kennedy, C. Raguenaud, and R. Hyam. 2000. The Prometheus Taxonomic Model: A practical approach to representing multiple taxonomies. Taxon 49:55–75.

Schindel, D. E., and S. E. Miller. 2005. DNA barcoding a useful tool for taxonomists. Nature 435:17.

Scoble, M. J., ed. 1999. Geometrid moths of the world: A catalogue. CSIRO Publishing, Collingwood, Victoria.

Scoble, M. J. 2004. Unitary or unified taxonomy? Philos. Trans. R. Soc. Lond. B 359:699–710.

Soltis, D. E., P. S. Soltis, P. K. Endress, and M. W. Chase. 2005. Phylogeny and evolution of angiosperms. Sinauer Associates, Sunderland, Massachusetts

Suber, P. 2005. Open access, impact, and demand. BMJ 330:1097–8.

Surowiecki, J. 2004. The wisdom of crowds: Why the many are smarter than the few and how collective wisdom shapes business, economies, societies and nations. Little Brown, New York.

Taxonomic Databases Working Group. 2006. GUID-1 Workshop Report. World Wide Web electronic publication. http://wiki.tdwg.org/twiki/bin/viewfile/TIP/TipDocuments?rev=;filename=GUID1 Report.pdf [accessed 30 March 2007].

Thacker, P. D. 2003. Morphology: The shape of things to come. BioSci. 53:544–549

Thiele, K., and D. Yeates. 2002. Tension arises from duality at the heart of taxonomy. Nature 419:337.

Vogler, A. P., and M. T. Monaghan. 2007. Recent advances in DNA taxonomy. J. Zool. Syst. Evol. Res. 45:1–10.

Waugh, J. 2007. DNA barcoding in animal species: Progress, potential and pitfalls. BioEssays 29:188–197.

Wheeler, G. D., P. H. Raven, and E. O. Wilson. 2004. Taxonomy: Impediment or expedient? Science 303:285.

Wiley, E. O. 1981. Phylogenetics. The theory and practice of phylogenetic systematics. John Wiley and Sons, New York.

Wilson, E. O. 2003. The encyclopedia of life. Trends Ecol. Evol. 18:77–80.

Wren, J. D. 2005. Open access and openly accessible: A study of scientific publications shared via the Internet. BMJ 330:1128.

Zhang, Z. -Q. 2006. The making of a mega-journal in taxonomy. Zootaxa 1385:67–68.

First submitted 25 January 2007; reviews returned 27 February 2007; final acceptance 1 May 2007

Guest Associate Editor: John Wiens