Introduction

M. F. Watson, C. H. C. Lyal and C. A. Pendry

We are living in an age where biodiversity is being lost at an unprecedented rate, with the well-documented problems of habitat destruction being compounded by the largely unknown future effects of climate change. High-quality, accurate and reliable biodiversity data are needed by biologists, conservationists and environmental modellers to understand and assess the ecosystems in which they work, to produce effective conservation strategies, and to feed computer-generated models, which predict what environments and habitats we might face in the future. Descriptive taxonomy – the collection, characterisation, description and naming of biological organisms – has been a cornerstone for the provision of authoritative data for these purposes, in addition to its general primary purpose of developing a taxonomic understanding of the biota. The chapters in this book explore changes to the traditional methodology of descriptive taxonomy, how new technologies are being embraced and how new requirements are being met.

Floras and Faunas, and their equivalents in other biological kingdoms, present data on organisms occurring within a geographical region. In addition to monographs, which focus on a taxon at the global scale, they have been a major output from taxonomists. They establish a regional classification bringing order to past works, provide tools for identification and inventories through keys and descriptions, and record data on distribution ranges, ecology and conservation status. These works have long been acknowledged as critical for understanding the organisms of a region and,

Descriptive Taxonomy: The Foundation of Biodiversity Research, eds M. F. Watson, C. H. C. Lyal and C. A. Pendry. Published by Cambridge University Press.

in recent years, as fundamental to the formulation of conservation strategies. Electronic data management and dissemination has the potential to transform floristic and faunistic projects radically and diversify the users of these data. Online dissemination enables unparalleled access to, and offers novel means of presenting, data in a variety of formats tailored to specific user requirements.

This book is based on the theme of the Sixth Biennial Conference of the Systematics Association, held at Royal Botanic Garden Edinburgh, 28–31 August 2007. It brings together scientists working on floristic, faunistic and mycological projects, including those producing field guides, and representatives from other user groups. The intent is to discuss novel areas of research and to help create a more outward looking, integrated approach, which better serves the needs of the end users. The book is arranged in five parts: the first addressing the issues of the current and potential users of descriptive biodiversity data and the relevance to diverse audiences; the second looks at the outputs and impacts of descriptive taxonomic products through a series of regional case studies; the third takes a closer look at field guides and applications of floristic and faunistic works; the fourth assesses the influence of new technologies on the gathering and management of data and collections in the field; and the fifth covers the impact of technological advancements in data location, dissemination and integration, and the new field of DNA barcoding.

Part I: The widening audience

Traditionally Floras, Faunas, Mycotas and the like have been intended for natural historians, in the wider sense, and for taxonomists. Their aim, as pointed out in several papers in this section, has primarily been to enable users to identify organisms within a geographical area and, secondarily, to act as a published repository for a wide range of related information, including taxonomic, systematic, morphological, behavioural, biological, ecological and phenological. They have ranged from the scholarly and academic to the illustration-rich and text-light field guide to simple lists of species with minimal additional information. Each of these products has its place, but this place is not always recognised by the users. The first three chapters in this section explore how the contents of Floras and Faunas have been and can be used by conservationists, ecologists and others to address questions that were perhaps outside the intent of the original compilers. Each of them also makes recommendations for future developments in floristic and faunistic studies, in terms of content, increasing use of digital tools and in collaboration across disciplines. The development of new methodologies in compilation, and highlighting valuable content types, are valuable for the preparation of Floras and Faunas, but must be considered in the context of publication; unless a product is to

be completely digital (and see later chapters for examples of this idea) it will need to be published. As discussed in Chapter 4, while publishers are considering digital products in this area, they are for their main market of paper-based publications constrained by market requirements for content and delivery style, and those engaged in the preparation of any new Fauna or Flora needs to take this into account.

Part II: The products of descriptive taxonomy

While many Floras and Faunas are nationally based there is, perhaps increasingly, a requirement for regional studies also. In many cases these can be built on previous work, although bringing that work together can have its challenges. A first step is to determine just what species are known from a region, and how they are distributed. This deceptively simple task has posed significant problems in some projects, and is still an issue for some users. For example, transnational control of pests, or management of genetic resources under the Nagoya Protocol require that all parties are using the same names for the same organisms. An attempt to address the problem of harmonising names and constructing a list is discussed in Chapter 5. Building on earlier experience, the management of the project was of critical importance, and particularly guiding and facilitating the work of contributors across Europe. The importance of sustaining activity after funding for a particular project comes to an end is stressed in Chapter 5, and is also a theme of Chapter 6. The development of the work that compiled the *Flora Europaea* through Euro+Med PlantBase is described. Again, there is a mixture of technological solutions and facilitating and encouraging the work of a cadre of taxonomists to update, discover and deliver information. The next two chapters in this section also cover regional approaches with a digital approach. Chapter 7 focuses on making available in digital format the great quantity of work that has been done on African Floras. Currently inaccessible to many, digitising the legacy literature in a controlled fashion and delivering it in a simple manner has made access very much simpler to a wide audience. The approach used XML markup – something that is returned to in Chapter 19 – to effectively turn the literature content into a database that could be queried with a set of predetermined questions, prioritised because of the likely use of the content. This has taken already published regional Floras to a new readership and, importantly, laid the groundwork for future work on African plant diversity. It is an example of reuse of published work – an area that has to be a priority given the huge volume of study available in libraries but with relatively limited accessibility. Chapter 8 discusses a different approach, to making information available, although this time there are several strands to the information: literature, nomenclature and taxonomy, distributions and associations, and biographical about mycologists. Two of

these strands are available in multiple languages, enabling the system to function in the context of national and regional websites elsewhere. As with other case studies in this section, the use of digital media has increased access and flexibility of use of resources, and enabled use of the resources in a number of different contexts.

The final chapter in this section, Chapter 9 on the Zooplankton Identification Manual for North European Seas (ZIMNES) project, takes the digital argument a little further and into a different domain, that of marine environments. In this case the compilation is for a digital system that features a web-based identification guide, but which also will be available in PDF format. Users are expected to encompass both specialists and non-specialists, a trend that is continued in the next section.

Part III: Outputs and impact: field guides and applications

While the more academic 'end' of Floras and Faunas is aimed at professionals and overlaps with other forms of academic taxonomic literature, there is a large constituency that needs Floristic and Faunistic treatments that are explicitly intended to assist in identification in the field – field guides. Such tools need to be very strongly tailored to their users and the ways in which those users access information. The most common type of field guide is a 'standard' handy-sized volume with illustrations and text. The precise format is of importance to users, and this is addressed in some detail in Chapters 10, 11 and 12. With regard to the images presented, care must be taken to ensure that the illustrations truly capture the organism in the manner in which people are likely to observe it. Chapter 10 provides a case study in which the execution of the guide relied not so much on formulaic 'traditional' images but much more on images presented to aid identification of the living organism. This approach is perhaps antithetical to the concept of an all-inclusive Fauna or Flora as being discussed in the first section of the book, but equally valid and with a very clear user base. Chapters 11 and 12 consider field manuals for quite restricted groups of users, in the context of particular projects and particularly for assessment and collecting purposes, and associated with training. In these cases the developers of the guides have been able to work closely with the users, getting feedback from them and being able to incorporate this into redesign. In the case study in Chapter 11 the developers focussed, as with the dragonfly work in Chapter 10, on the aspects of the organisms that will be most apparent to the user. In the case of plants this largely excluded reproductive structures, as the majority of times that people come into contact with plants, perhaps particularly in tropical work, the plants are not in a reproductive phase. The project discussed in Chapter 12 addressed this problem in a slightly different manner, by stressing

phenological data in the products, so that the users could time their work in the ecosystems covered to maximise the likelihood of finding identifiable material, something that was particularly important for the Seed Bank collectors. The emphasis of working with the users is taken a further step in Chapter 13, which is explicitly about training and recognition of the most effective 'shortest route' to effective field identification through selection of characters and hands-on experience with collecting and preparing plants. This feeds back to the discussions in earlier chapters on character selection, and also provides insights into how field guides are actually used, and the types of learning reinforcement that can assist users.

The final chapter in this section takes us in a different direction. As is discussed in Chapter 22, the use of DNA in identifications is a growing trend. In most cases, however, this is done in the laboratory and some time after field collection. However, there are situations where rapid field identification of critical species is required, and DNA might be the most effective way of achieving this. In Chapter 14 we have an example of such a scenario and tools developed to facilitate this. Two tools are discussed: comparison of a specimen with a three-dimensional image displayed by a wearable computer, and a portable DNA laboratory for use in the field. In this case study, the concept of a field guide has been translated into a small suite of high-tech. tools that are tightly focussed on meeting the needs of particular users in a particular set of circumstances – perhaps the opposite end of the spectrum from the large-scale scholastic tomes.

Part IV: The influence of technology on data gathering in the field

The previous section took us to the field use of Faunas and Floras. The other side of the same equation is the gathering of data in the field to support an understanding of biodiversity, and how the collections made are managed and used. Traditionally collectors used to operate in a fairly simplistic manner – go to a place, search for organisms, kill them, bring them back and then sort out the information they provide back at the laboratory. This methodology is increasingly outdated. Of course, most collectors learned particular methods to locate target groups, but with advances in technology these have, in some cases, become both more refined and standardised. This is a theme of Chapters 15 and 16, where a range of techniques is addressed, and Chapter 17 where the implementation of novel technologies in a very challenging environment for discovering biodiversity – marine systems – are discussed. Other technological solutions are discussed in all chapters in this section, in particular digital capture of information in the field. Particular attention is paid to digital photography in all chapters, ideally linked to global positioning system recording. This is linked to the selection of just what to collect, and what to

preserve in collections; increasingly it is important to understand the collecting biases that led to what is preserved and accommodate that in analyses – issues that had previously been discussed in terms of contents of Floras and Faunas in Part III. It is also important to collect efficiently, so that images and samples for DNA analysis are all that is required, rather than entire specimens. Data capture in the field may be more efficient than retrospective searching for data, and the ways in which such data capture can be brought into a more efficient workflow are touched on in Chapter 16 and explored further in Chapter 18.

Part V: New technologies: their current use and future potential

The impact of novel technologies unsurprisingly is a component of many chapters in the volume, and is taken further in this section. Chapter 19 focuses on the transformation of texts to XML documents, which allows direct querying of their content, as well as extraction and reuse. The example discussed in Chapter 7 gives some idea of the power of this approach, but in Chapter 19 the concept is taken further. Such a system would allow users to effectively compile bespoke Faunas and Floras from published works. While the concept of bringing together information from different sources and uniting it sounds, if not simple, at least practical, it brings with it a host of problems. The issue of agreeing which names to use for species across different countries has been discussed in Chapter 5, but this is even more of an issue when one is attempting to bring together information not only from different countries but also from different authors and compiled at different times. Names are, if not invariant, at least possible to track through variations of spelling as taxonomists change their systematic position to reflect new understanding of the species' relationships. However, there is no such easy link to the species concept employed by the authors, especially if we are attempting to semi-automate the process. Chapter 20 explores this issue through the application of globally unique identifiers, an increasingly important element in the biodiversity informatics aspect of floristic and faunistic work. An indication of the rate of change in this area is in numbers; when the chapter was written the Global Biodiversity Information Facility mediated some 140 million occurrence records linked to names, a figure that, in 2014, has now risen to 416 242 316. Chapter 21 also calls for rapid change, pointing out that descriptive taxonomy is a slow labour intensive process, and identifying the need for tools to accelerate the process. In Chapter 12 a methodology was discussed that produced an average of three field guides per month; in Chapter 21 further tools are discussed for use in a wider context. Since the chapter was written in 2010 a great deal has changed and tools that are discussed as necessary are now appearing. For example, the development of a smooth online

workflow for the whole life cycle of a manuscript, from writing to submission to peer review to publication and dissemination, is now a reality in the *Biodiversity Data Journal*.

The final chapter in the book brings us back to species and species concepts, but again seeking the most rapid and effective way of assessing what is in an area and, thus, what is to be covered in faunistic and floristic work. In areas that are very poorly known, and for taxa that pose taxonomic problems, a traditional solely morphological approach is extremely time-consuming. The addition of the use of DNA barcodes to the process, again tied to an efficient workflow, demonstrably both speeds the process and, tied to morphological examination, can provide a far more accurate and detailed understanding of the biota.

Overall, the chapters in this volume demonstrate the wide range of uses of descriptive taxonomy and how to overcome some of the problems in applying it beyond how the original authors anticipated their publications would be used. Novel techniques are discussed, both to access legacy material and also to collect, prepare and publish new information. The importance of understanding the purpose of each part of the process and how it contributes to meeting the needs of the user is highlighted and, most significantly, workflows to maximise efficient discovery, compilation and delivery are presented.

PART I

The widening audience

1

Floras yesterday, today and tomorrow

A. G. Miller, M. Hall, M. F. Watson, S. G. Knees, C. A. Pendry and M. R. Pullan

1.1 Floras past

Frodin (2001) masterfully traces the history of Floras from their origins to their present form, offering both historical analysis and insightful critique. The current paper will draw substantially upon this authoritative work, *Guide to standard Floras of the world*, to frame the discussion of the role and purpose of the modern Flora. It does so by examining some of the important historical influences on our understanding of a Flora's purpose. Frodin roughly characterises the historical purpose of descriptive Floras as being twofold. Broadly speaking, the first purpose of Floras is for *identification* of plant species, achieved (hopefully) by the provision of keys, descriptions and illustrations. The second (often opposed) purpose of Floras is *archival* or *encyclopaedic*, with the Flora acting as a repository of comparative, descriptive, taxonomic data, such as extensive synonymy and specimen information.

Frodin (2001) considers the *Flore française* of J. B. de Lamarck and A. P. de Candolle (1805–1815) as the first modern Flora. This pioneering work was, in effect, the third edition of Lamarck's *Flore françoise* (1778), a publication that is highly significant in a historical discussion of Floras for two principal reasons: it was the first floristic work to contain Lamarck's analytical, dichotomous keys for species identification, and (of particular significance for this paper) because it contains some of the earliest, explicit views of the purposes of Floras. For Lamarck, Floras

Descriptive Taxonomy: The Foundation of Biodiversity Research, eds M. F. Watson, C. H. C. Lyal and C. A. Pendry. Published by Cambridge University Press.

were not scholarly works for consumption by other taxonomists. In his opinion, there was no purpose in Floras being elegant, encyclopaedic, library volumes. Rather, their purpose was to be a practical tool for the identification of plant species. Lamarck's opinion on the purpose of descriptive Floras translated into his concise descriptive style, which he employed in *Flore françoise*, along with the consistent use of dichotomous keys, and notes on distribution and habit. Lamarck's use of concise descriptions in vernacular language, the use of vernacular names and the absence of much synonymy and specimen citation was part of its orientation to more popular, practical (as well as scholarly) use.

This concise, pragmatic approach to Flora writing is also characteristic of a series of colonial Floras produced by Kew in response to Britain's desire to document the plant wealth in its colonies. J. D. Hooker's *Flora of British India* (1875–1896) is a significant example of a concise, utilitarian Flora. Before the *Flora of British India* the only Flora ostensibly covering India up to the mid nineteenth century was that of Roxburgh, published complete only in 1832. There were other impressive works such as those of Wight (1831) (exceptional in being heavily illustrated), but Hooker's is notable for taking a radical and sweeping approach, both to species concepts and nomenclature (postulating the 'Kew Rule'). In the preface, Hooker (1875) makes it very clear that in his *Flora of British India* conciseness is paramount:

> An exhaustive Flora would be a work of many years and many volumes; and it is as a handbook to what is already known, and a pioneer to more complete works, that the present is put forward ... the adoption of as concise a style and phraseology as is consistent with clearness, and the avoidance of repetition in the descriptions and remarks on each species will enable me to compress the whole into a portable form.

Not only are concise descriptions clearer, they enable the Flora to be more compact. Hooker was clearly aware that a more compact Flora is produced more quickly and cheaply. However, a problem with this concise approach was that, although the species are usually grouped with shared characters, the descriptions are often not particularly diagnostic and frequently not usable by themselves to separate similar species. In many cases, accurate identification is only achievable with reference to a comprehensive, well-curated herbarium. This is compounded by the fact that only synoptic (not dichotomous) keys are included. Nevertheless, the *Flora of British India* was a revelation for a descriptive Flora in English. Although much of the taxonomy has been subsequently refined, it is a good example of the benefits of producing a usable product 'warts and all'. In the absence of an alternative (until very recently) it has been much used and copied by later workers.

A concise, pragmatic approach to writing Floras was also championed by another British colonial botanist, George Bentham. At the beginning of his *Flora hongkongensis*, Bentham (1872) espoused the principles of writing a concise Flora, which were

subsequently repeated in a number of the other great colonial Floras such as *Flora australiensis* (1863–1878) and Bentham's *Outlines of elementary botany as introductory to local Floras* (1861). These principles are expressed as aphorisms – for brevity, these can be summarised by the following points (Frodin, 2001):

- the principle object of a Flora is to afford the means to identify any plant growing in its area;
- a Flora should have good descriptions;
- a Flora should be concise and accurate;
- a Flora should not be overloaded with technical terms;
- plants should be arranged in a natural system; and
- artificial keys should be provided.

Bentham's aphorisms have much in common with the ethos of the first modern Flora, Lamarck and de Candolle's *Flore française*. Bentham is clear; a Flora is not a repository of knowledge for the taxonomic community, but is a tool for effecting identifications of plants, both by specialists and non-specialists alike.

Frodin (2001) contrasts this floristic approach taken by the so-called 'French School' and 'British School', with that of the 'Central European School'. During much of the nineteenth century the Floras emerging from Central Europe were constructed as encyclopaedias – detailed repositories of information about the plants of an area. This is considered to be a legacy of the Linnaean approach to Flora writing, a legacy that ensured that detailed descriptions and notes, as well as extensive synonymy and specimen citations, were all included. The influential botanical centre of Berlin saw the production of a number of encyclopaedic Floras in late nineteenth and early twentieth centuries; such as Ascherson and Graebner's *Synopsis der mitteleuropäischen Flora* (1896–1939), Reichenbach's *Icones florae Germanicae et Helveticae* (1837–1914) and Hegi's detailed *Illustrierte Flora von Mittel-Europa* (1906–1931).

Indeed, the lack of concise Floras as a whole in Central Europe must partly be a function of the relative importance of monographical work over floristic work in early twentieth-century German botany. Emphasis on large-scale, time-consuming monographical projects, such as Engler and Prantl's *Die natürlichen Pflanzenfamilien* (1849–1893), meant that Floras came to be seen as a secondary consideration. Even the iconic floristic work of this Central European School, *Flora brasiliensis* (Martius et al., 1840–1906), was almost monographical in approach. Frodin (2001) highlights the pervasive significance of this Flora by stating that it 'established the tradition – still with us – of large scale, multi volume, descriptive regional Floras .... They came to be seen as suitable vehicles for submonographical studies .... Most remain more or less encyclopaedic, and as well retain an aura of prestige: a form of institutional "cachet" '.

1.2 Floras present

This brief historical background forces us to question the current purpose of Floras, and we are pushed to consider how effective Floras are in fulfilling their ascribed role. Heywood (1995) notes that in the nineteenth century, world exploration (and exploitation) and building of empires by the European powers led to a need for colonial Floras. Post World War II, the desire for reconciliation and cultural integration led to the creation of the *Flora Europaea* project, which in turn acted as stimulus in other parts of the world (see Chapter 6, Jury, in this volume). Heywood argues that in the last 10–25 years an intensification of Flora writing is attributable, in large part, to the conservation movement, and highlights the need for basic floristic inventory assessments as a basis for resource management, conservation and other biological activities.

Recent, high-profile cases have been made for a strong link between taxonomy and conservation (House of Lords, 1992, 2002; Mace, 2004; Wilson, 2004). Clearly, in the age of a Sixth Great Extinction Crisis (Wilson, 1992), one of a Flora's most important roles is facilitating biodiversity conservation. Although this received wisdom is seemingly self-evident, in a discussion of a Flora's purpose it is important to investigate these purported links between writing a Flora and the conservation of plant species. Such probing will help clarify the needs of Flora users and can feed into discussions of a Flora's design and orientation.

Mace (2004) argues that taxonomy has something of a dual role in its contribution to biological conservation. Firstly, taxonomy must necessarily identify, circumscribe and inventory biological diversity. In any area of vegetation, we must know what exists before we can attempt to preserve its existence (Mace, 2004). In this respect, floristic projects can often be at the forefront of conservation efforts, identifying new species, mapping species distributions and 'Red Listing' plant species that are threatened with extinction (Secretariat of the Convention on Biodiversity, 2002). Floristic taxonomists also have an important role to play in identifying potential sites for conservation, for as Prance (1995) argues 'it is systematists who hold most of the data about centres of endemism and who can locate hotspots accurately'.

This link between taxonomy and conservation can be expanded using the *Flora of the Arabian Peninsula and Socotra Archipelago* (FAPSA; Miller and Cope, 1996–ongoing) as a case study. Flora projects such as FAPSA are often the first point of call for those within a country or region, such as government bodies or research institutes, or international non-governmental organisations (NGOs) (e.g. the World Conservation Monitoring Centre) wishing to obtain authoritative data on plants. Much of this is related to assessing the conservation status of particular areas or species. For example, FAPSA supplied much of the southwest Asian data to two major global conservation initiatives; *Global biodiversity – status of the earth's living resources*