Mapping Biodiversity

Geoffrey C. Bowker
Department of Communication
University of California, San Diego.

Keywords: **Biodiversity, GIS, Information Integration; Scientific Communication**

Acknowledgements: I wish to thank Francis Harvey, Nick Chrisman, Leigh Star for their comments on an earlier version of this paper and the inspiration of their work; and the participants of the Intersections: Society, Technology, and Geographic Thought conference held in Kentucky for providing such rich and generous discussions. I also wish to thank the anonymous reviewers for their constructive criticism.

Abstract

It is argued that there is no simple metadata solution to the problems of integrating information from multiple scientific disciplines in biodiversity and thus to problems of adopting GIS for producing maps of biodiversity. To the contrary, it is maintained that within this field, GIS needs to take account of and represent an irreducible ontological diversity in the many biodiversity databases being produced.

Introduction

In this paper, I will discuss a set of intriguing problems that have arisen as scientists from many different disciplines have begun working together to attempt to build up a unified picture of global biodiversity useful for both basic scientific research and for science policy. In order to produce such a map, there must be some basic agreement about data standards, classification systems and spatiotemporal frameworks. I shall argue in this paper that such agreements are hard won and frequently involve necessary but difficult trade-offs between the needs of different scientific fields and governmental agencies.

Biodiversity is a subject of intense concern in the world scientific and policy communities. Heywood noted at the beginning of the 1990s that numerous attempts are being made to produce very large federated databases of plant and animal life:

> Global synoptic or master species databases are being developed for a diverse array of groups such as viruses, bacteria, protists, fungi, molluscs, arthropods, vascular plants, fossils etc. and the SPECIES 2000 programme of IUBS, CODATA and IUMS has proposed that many of these should be joined to create a federated system that could eventually lead to the creation of a synoptic database of all the world's known organisms…(Heywood, 1991: 53).

Similarly, NASA's Mission to Earth program is trying to "'document the physical, chemical, and biological processes responsible for the evolution of Earth on all time scales'" (Elichirigoity, 1999: 14) and Michael and Wayne Moneyhan write of the United Nation's Global Resource Information Database:

> 'Traditional access to environmental data – in shelves of reports and proceedings as well as in fast-aging maps and charts – no longer meet the demands of planners faced with a world in which the nature of environmental change is infinitely complex. With the development of computers that can handle large quantities of data, a global database is now possible'. (cited in Elichirigoity, 1999: 14-15)

I will not in this paper be concerned with representational issues of how to turn a given database into a readable map but with issues of data quality. In particular, I will address the issue of integration: can information derived from multiple scientific disciplines and governmental and non-governmental agencies (each with their own work practices, tools and sets of information standards) be integrated into large federated databases from which maps can be generated?

The chief current response to the problem of data integration is to produce metadata - 'data about data' (Dempsey and Heery, 1998: 145) – standards. This has been generally highly successful within the field of GIS. I shall argue in this paper that there are specific kinds of problems of data reconciliation which metadata standards cannot solve – in particular (and in turn) issues of naming entities and issues of interdisciplinary communication. Naming controversies are normal practice *within* a discipline; they become more problematic when one is trying to create a way of integrating information between several disciplines for two reasons:

- It is practically extremely difficult to automatically update fundamental name changes across a scattered set of layered databases (indeed it is hard enough to do so within any given field of practice – for example, medicine (Bowker and Star, 1999);

- It is difficult to hold a naming decision open when trying to export results from one discipline to another: in general what gets exported are decisions about names, not descriptions of controversies.

The exigencies of interdisciplinary communication, as we shall see, raise similarly thorny issues:

- Different disciplines may have good reasons for adopting incompatible naming conventions;
- Different disciplines may have irreconcilable standards of spatio-temporal scale.

These problems are not intractable. However, I shall argue that we need more than just good metadata standards in order to address them.

The issue of integration has been raised within GIS in the past few years in work by Nicholas Chrisman and Francis Harvey, amongst others (Chrisman, 1998, Chrisman, 1999, Chrisman, 1997, Harvey, 1997); (Shepherd, 1991). Thus Harvey and Chrisman (1998: 1688) write that:

> The negotiation of differences between disciplines and institutions with entrenched perspectives on the world and their mandate can be quite difficult.

They describe the multiple agencies involved in the attempts to standardize wetlands definitions and classifications, using the Cowardin system – Chrisman (1999) describes a case where, comparing four agencies' coverage of a single county: "only 3.5% of the area determined to be wetland in at least one of the sources was similarly classified by all four". The related issue of what happens to data as it gets from the field to the scientific article has been extensively explored in the science studies literature (Goodwin, 1994; Latour, 1993) There has in this area been little analysis, however, of what happens as one moves from raw data to databases – and whether decisions taken in this process have continuing effects on the interpretation and use of the resultant data stores (and in particular, for our purposes, the production of maps of biodiversity). I shall argue that issues of how data is worked into storable form are integrally issues about how groups of scientists communicate with each other and with governmental and other agencies. The paper is in two parts, discussing the question of information integration first by looking at naming practices within the disciplines of biodiversity and then at interdisciplinary communication in the field.

Naming

There are many things, as Douglas Adams (Adams and Lloyd, 1990) reminds us in <u>The Deeper Meaning of Liff</u>, that we would like to talk about but just don't have the words for. In general, you cannot develop a database without having some means of putting data into pigeonholes of some kind or another – you can't store data or produce a map without using multiple classification systems. Thus a representation of tropical rainforest cover might involve not only a classification of vegetation covers, but also of the various surrogate measures (indirect means of sensing that can be used to infer cover).

A major problem for mapping biodiversity globally is that while classification systems are about kinds of things (flora, fauna, communities etc), the world of biodiversity data is radically singular. Just as species can be endemic to very small areas, so too can data about species. As Heywood says: "It has to be remembered that the vast majority of species described in the literature are 'herbarium' or 'museum' species and their existence as coherent, repeatable population-based phenomena is only suppositional" (Heywood, 1988: 48). Raven et al. remark that this throws into relief the fact that our names for organisms do not contain much information:

> the taxonomic system we use appears to communicate a great deal about the organisms being discussed, whereas in fact it communicates only a little. Since, in the vast majority of instances, only the describer has seen the named organism, no one with whom he is communicating shares his understanding of it (Raven et al., 1971: 1212).

There is often only one of a given species in preserved form, so that in order to check one's own samples one needs an exhaustive global database (in the form of a robust body of literature – of which more anon – or in searchable electronic form) and a means of transporting type specimens from one site to another. So working out what is in one's own collection as a prelude to cataloguing it and putting it into searchable form represents a bootstrapping problem – unless you have described your collection well, others can't describe theirs; but equally you can't describe yours until they have theirs. In a study of the International Classification of Diseases, Bowker and Star (Bowker and Star, 1999: Chapter 2) remarked that this bootstrapping problem is a common feature of the development of global databases.

I have written elsewhere (Bowker, 1999) of problems of integrating names within the field of botany. There are two major sets of difficulties in this field (and also, by extension, in zoology). Firstly, classification systems as developed by systematists reflect the latest theory of a taxon's genetic history: they can be relatively unstable over time. However there is strong pressure to keep taxon names the same. Thus a nursery owner doesn't want to have to relabel all her tomato plants because the tomato has moved to a fresh genus: the old name is sufficient to her purpose. It is estimated that a change on such an important crop could cost many millions of dollars. Further, a given orchid might find itself no longer protected by the legislation if the systematists decide that its name should be changed (the law is written to protect a species by name). A classic response to the need to maintain multiple sets of incompatible naming conventions is to produce tables of synonyms: but this is practically extremely difficult and sometimes theoretically impossible when one is trying to integrate data across multiple agencies each of which has good reasons for sticking to their own naming systems. Secondly, historically national lists have historically grown up somewhat independently of each other. For example, the recent genus change of tomatoes is due to the work of reconciling Chinese and European national floras. The work of reconciling national lists is not simple: one must go back to the original type specimens held in herbaria in order to judge whether two species which have received different names because they are geographically separated should be seen as one. And then one has to convince users of one of the lists to accept the priority of the new name. Insofar as the data being collected was to be used by a local scientific community, the different lists were not a problem: now they are to be shared internationally by the global database community.

However, it is not just the names of taxa that causes a problem for the integration of data into global databases. There is also the question of what kinds of frames to wrap the data in so as ultimately to produce useful maps. For example, in order to produce generalizations about biodiversity, it would be useful to be able to attach plant or animal communities to particular landscapes. However, landscape topography has proven very difficult to classify. Huggett notes that: "… many landform classifications are based purely on topographic form, and ignore geomorphic process". (Huggett, 1997: 228) – in our terms they are not genetic but descriptive. He notes that the Davisian 'geographical cycle' in the late nineteenth century was the first modern theory of landscape evolution:

> It assumed that uplift takes place quickly. The raw topography is then gradually worn down by exogenic processes, without further complications from tectonic movements. Furthermore, slopes within landscapes decline through time (though few field studies have substantiated this claim). So, topography is reduced, little by little, to an extensive flat region close to base level – a peneplain. The reduction process creates a time sequence of landforms that progresses though the stages of youth, maturity, and old age. However, these terms, borrowed from biology, are misleading and much censured". (Huggett, 1997: 230).

In more recent thinking, steady states in the landscape are the exception rather than the rule (ibid.: 236), and indeed there is an argument that landscapes themselves are evolving over time – evolutionary geomorphologists argue that rather than an "'endless' progression of erosion cycles", there have been several geomorphological revolutions (ibid.: 258). This ties in directly to biodiversity arguments, since the argument has been made that there is a growing geodiversity subtending our current relative biodiversity peak (ibid.: 299) - itself offset by the current extinction wave. If, as this theory asserts, we are developing new kinds of landscape features over time, then, we cannot retrospectively assign current classifications to maps of the earth in the past. Specifically grass, a fairly recent ecological actor, radically changed sedimentation rates. As Chrisman (Chrisman, 1997: 55) notes, time is a major underworked area in GIS. We run into the problem of time here: for the level of granularity of many biodiversity maps, it is hard to find any one classification system that is stable over eras of geological time and disciplinary history … .

A third set of classification difficulties arises for things which don't have easy boundaries, and are of indeterminate theoretic status. Soils, for example, enter into some biodiversity databases as related to particular kinds of communities of flora and fauna. And yet different agencies within a nation often adopt different soil classifications, rendering the pooling of data complex. In a superb article discussing: "the monumental (or dinosauric) *Soil Classification, A Comprehensive System - 7th Approximation*" published in 1960 by the Soil Survey Staff of the U.S. Department of Agriculture, Bennison Gray ((Gray, 1980) discusses the problems of classifying something that does not break up into natural units. Different sets of researchers will have very different disciplinary perspectives on what soil is. Similarly, there are difficulties with the concept of communities in ecology – a central organizing feature of GAP analysis. The first issue in mapping communities is whether or not the theoretical construct has any validity. In the first half of the twentieth century, Clements and Gleason took opposing views on the existence of natural communities (Journet, 1991). Clements said that there were large natural units of vegetation, such as a forest. He saw the community as a 'superorganism'. (ibid., 452) These communities go through predictable stages of growth and development – called 'series'. There are natural climax communities optimally suited to a given (stable) climate regimen. Gleason saw the groups of plants as being less an organism than a coincidence: "the implication of Gleason's arguments is that what an ecologist calls a community may be only an artificial boundary drawn around a collection of individuals". (ibid., 455). He rejected the overall stability of the climate. (ibid., 457). In the new catastrophism (Ager, 1993), there are many who have come to accept the position that stability, the prerequisite for many definitions of communities, just does not obtain sufficiently to make this a useful concept. However, some of the more exciting work in biodiversity research, GAP (Gap Analyis Program), relies on a pragmatic definition of community, which includes many ad hoc assumptions, recognized as such by its developers (Edwards et al., 1995: passim). Gap analysis has had a major impact on the fields of biogeography and biodiversity: it provides a kind of area coverage that is inconceivable with more fine-grained analyses (such as the problematic All Taxon Biodiversity Inventory method – which comprises exhaustive analysis of small areas). The technique involves taking satellite images of a given area, then classifying these into homogeneous units imputed with labels for natural vegetation communities. Vegetation communities are then used as surrogates for animal communities. The difficulties with the analysis include that aerial photographs can only be used to characterize vegetation communities with about 73% success; vegetation communities are difficult to define anyway; even if they are successfully defined and located, they are not necessarily good surrogates for animal communities (free ranging predators might not distinguish between any two vegetation communities) - bats with point distribution (in caves) cannot be picked out; and in general structural features of the landscape are not mapped, so that if a species requires a given type of setting within a floristic community, the setting has to be assumed to exist (Edwards et al., 1995: 4-2). Gap analyses, then, depends on being able to provide good definitions of particular communities of flora and fauna – and on being able to use the latter as a surrogate for the former. If communities cannot be easily named (fuzzy borders; problems of the level of granularity and so forth) then the technique itself must be questioned.

However flawed the GAP maps of vegetation might be they are still much better than the current alternative in some areas – no maps at all. Here we run directly up against issues of communities of practice: if one is producing a map for conservation purposes, one might be less concerned with theoretical validity than with getting some representation out that can do the political work that is required. This practical attitude is openly adopted in GAP research. Sismondo remarks on a similar issue in the use of island biogeographic (IB) modeling for the demarcation of protected biodiversity areas. He notes arguments that even though IB does not fit the real world well: "it could be useful in the design of nature preserves, by predicting the number of species preserved in a given conservation area" (Sismondo, 1999: 245). The argument runs that that alternatives to its use – such as detailed study of endangered organisms and habitats –

may be too expensive and time-consuming. I can but cite his conclusion: "Too simple an application of the language of truth and falsity does not do justice to the complex representational relationships that theoretical models have. … rightness or wrongness depends upon what the community decides gets to count as being right or wrong in particular circumstances, what the community wants the object to do" (ibid., p.258).

The data format of a biodiversity database invariably reflects the set of temporalities and spatialities that are attached to the objects (taxa, species, genes, communities …) being represented. We have seen above how these objects themselves do not nest neatly inside each other along either of these axes. There is no clear absolute nesting of species inside genera inside families – it depends on what you call a species, and on what you do with border problems like hybrids, which do not fit neatly into any classification scheme. In a fully lawlike world, it would be possible to create a synoptic database that held all the information from all of these partial objects: we are begging the ontological question if we attempt to render our data in a completely flat, stable database.

Equally, there is no universal spatial or temporal nesting. One cannot just nest chunks of space one inside the other from the level of the entire globe down to the $1m^2$ patches of the traditional ecologist. For example, a 'standard' biogeographical model (Brenchley and Harper, 1998: 273) of the world shows six major provinces, each province being centered on a major landmass. A recent panbiogeographic model of the world (Cox, 1998: 825) shows five major provinces, each centered on an ocean. The panbiogeographic model comes out of a team based in New Zealand, an island deep in the Pacific – and emphasizes the role of islands (by way, on recent theory, of terranes – pieces of continental plates that travel relatively independently, such as Point Reyes in California and so carry floral and faunal islands across the ocean from one landmass to another). The locus of object/space/time production is not an accidental feature here (cf Bowker, 1994b). To the contrary, we have seen that the classification systems used tend to reflect their origins (the number of angiosperm species being an index of European plant life, for example); in general in systematics, Heywood writes:

> There are currently five or six different species concepts in use and no agreement between the different practitioners on how to develop a coherent theory of systematics at the species level. … In addition, species concepts differ from group to group and there are often national or regional differences in the way in which the species category is deployed". (Heywood, 1997: 9).

The same particularism can be noted temporally with respect to the presentism in much discourse about climate in climate change theory (the value that in many texts is placed on holding the world climate to current parameters).

The importance of site indicates a fact about biodiversity science of central importance for its mapping: it is the science of the radically singular, and so its maps will always enfold – in complex ways – traces of their community and site of production. The underlying question is what diversity there is in this world now – not what diversity there may be in earth-like planets under different sets of conditions. However, for many involved in biodiversity science, the totalization that is sought is a predictive, lawlike knowledge that will allow for an understanding of the wellsprings of biodiversity. Maps of the fossil record recapitulate these problems. Koch (Koch, 1998: 199) cites Ager on the taxonomic barrier whereby the Austro-Hungarian and British Empires can be traced through fossil collections in Vienna and London respectively – this led to different sets of synonyms and thus to different apparent biogeographic regions. The political empire writ large on the natural world (cf Richards, 1996 on the imperial drive to total knowledge)! Similarly within the US, he cites Sohl's analysis of apparent difference in species between the Texas and the Tennessee and Mississippi areas being: "artifacts of too stringent a taxonomy and an evident belief on the part of some workers that gastropods had very narrow dispersal limits'". (ibid.: 199) Similarly there are people – as Sohl points out – who have " 'a tendency to either work on Cretaceous or on Tertiary assemblages but seldom both'", thus emphasizing the differences between the eras; Erwin says the same for the Paleozoic and the Mesozoic (ibid.: 199-200)

So we have covered here three kinds of thing that are hard to classify and map: entities where the data itself is singular and scattered; entities which are not amenable to genetic classifications and are singular, and entities (or spatiotemporal regions) which may or may not exist. If we take the first two cases, the first generalization to emerge is the relatively obvious one that when either data or the world is singular, then naming schemes tend to break down and comparable maps are hard to design. More interesting is a corollary – that once they do break down, these things do not get represented in databases, or if represented are represented in incompatible forms across different sites. This in turn means that less attention is paid to such entities. If we take the last two cases, the obvious generalization is that things amenable to genetic classifications against a stable temporal backdrop can be relatively easily named, but when time itself is a dynamic variable then the naming mechanisms tend to break down. This can be combined with the first generalization in the argument that things that are singular either in space (one unit of space is not equivalent to another) or in time (the history of the earth is too secular) cannot be easily named. As I write this, I am aware that this is on one reading a remarkably obvious statement. Of course singulars are harder to name consistently across disciplines – unique id numbers are relatively easy to generate, but each discipline would have its own numbering system. Where my generalization takes on depth is that it takes us very quickly from data structures subtending maps into intellectual developments in the history of the sciences of biodiversity. If certain kinds of entities are being excluded from entering into the databases and maps we are creating, and if those entities share the feature that they are singular in space and time, then we are producing a set of models of the world which – despite its frequent historicity – is constraining us generally to converge on descriptions of the world in terms of repeatable entities: not because the world is so but because this is the nature of our data models.

Interdisciplinary Communication

Problems associated with naming relate to what gets mapped and what does not. The naming issues evoked in the last section scale up in the world of biodiversity mapping to issues of interdisciplinary communication. One can have consistent sets of names in subdisciplines but no good way to bring them all (literally) onto the same map; and one can have single datasets produced for two or more groups who integrate the dataset into entirely different contexts. These are the two issues that I discuss in this section. We will see that a fundamental challenge for large-scale biodiversity maps is that the data exist in a folded space and time, torqued by the practices and problematics of the local scientific communities that produce them. There is no possible registration mechanism for overlaying the data from one community onto that of another. In this section, I will lay out this problematic, and in the conclusion I shall argue that it is not an insuperable barrier to communication across scientific disciplines.

Biodiversity and Ecology vs Biodiversity and Systematics

There has been a move in a number of sciences over the past few hundred years to analyze basic scientific units in terms of information storage and transmission (Bowker, 1994a): be this genes, or quarks or species. This move is a point of articulation for two divergent information collection strategies in biodiversity research. By the simplest definition, biodiversity is just about the number of species that there are – one assigns a unique identifier to each species and then counts the number of species in a given unit area to get a biodiversity score. Many practitioners of biodiversity science have argued that such a measure does not give an indication of 'true' biodiversity. They point, for example, to the case where one has a dozen species of rats and one of pandas (Vane-Wright et al., 1991: 237) – note the charismatic megafauna being pitched against the ever unpopular rat. The argument is made that in terms of information held in gene stock the panda well outweighs the several rat species, which contain a set of overlapping genes. Formally, this means taking diversity as "a measure of information in a hierarchical classification" (ibid.) – with the implication that one wants to save ancestor species and species with few close relatives for preference over more recently evolved species with many siblings. As Barrowclough indicates in a very well worked through example, this line of argument can lead to a series of non-obvious choices; in an example he discusses : “some of the taxa in the coastal forest of southeastern Brazil have a sister group relationship to other species throughout the Amazon basin and hence are in some sense equivalent to that entire avifauna”. (Barrowclough, 1992: 137).

There is an interesting convergence between the world and its information at this point. Both in terms of databases on computers and the world as database, scientists are seeking for the minimum dataset which is needed in order to preserve biodiversity – whether that dataset be the genes held in organisms or the bits held in computers. In both cases this is not seen as the ideal outcome; to the contrary, it is seen as a practical choice – given that we are destroying biodiversity, and given that we do not have enough systematists… . The estimated completion of the Flora Neotropica (“a published inventory of the plants of the entire New World tropics” - http://www.nybg.org/bsci/ofn/infoengl.html), begun in 1968, is the year 2397 – this is not commensurate with the rate at which the environment is changing (Heywood et al., 1995: 5-19). This matches the famous nineteenth century information problems that the Oxford English dictionary was being produced slower than the language was changing; and that the American census was being carried out oftener than its results could be processed – leading perhaps to a modulation of Ross Ashby’s law of requisite variety: a law of requisite pace. Stork notes that we seem to be in the middle of an extinction crisis and yet only about 1000 species: “are recorded as having become extinct in recent years (since 1600)” (Stork, 1997: 45). He notes that it is hard to estimate loss, since in any case most species are only known through a single example – their holotype. Indeed, to continue the convergence theme, we could note that many of these holotypes are only known through books or other publications, since the original specimen either was not collected or has been lost: these are called lectoholotypes. In general it can be argued that such a convergence leads, through the deployment of a common set of metaphors and methods, to a close resonance between the world and its information (see Bowker, 1998): the two are conjured into the same form. For our immediate purposes, the assertion that biodiversity should be seen as an information issue entails that strategies for both data collection and habitat management are intimately wrapped up in ontological questions about what kind of a thing biodiversity is.

On the one hand, then, we have a set of information collection strategies twinned with biodiversity protection strategies based on the view that species are information units in a genealogical hierarchy. Eldredge (Eldredge, 1992: 1) contrasts this to an *ecological* perspective, which he ties back to an economic hierarchy - thus going back to the common roots of ecology and economics in the Greek word for household, *oikos* (Williams, 1983: 110). The distinction works as follows. Ecological diversity reflects: “the number of different sorts of organisms present in a local ecosystem” (Eldredge, 1992: 2). Now a species does not operate as an economic unit – indeed a given species is generally a member of a number of ecosystems. Ecological diversity (the number of species in a given community) is orthogonal to biodiversity. Eldredge argues that species are part of the Linnaean hierarchy, whereas local ecosystems are part of interacting economic systems. The former are extended in time and genealogical – they “act as reservoirs of genetic information” (Eldredge, 1992: 5), whereas the latter are extended in space, their temporal hallmark being “moment-by-moment interactions”. At base, then, Eldredge is arguing that we are dealing with two entirely different ways of being in the world: demes (subgroups of a particular species living together) obtain spatially and are part of ecological diversity whereas species obtain temporally and are part of biodiversity. Similarly, within biogeography, Grehan (Grehan, 1994: 461) makes the assertion – challenged by Cox (Cox, 1998: 821) that there are: “dichotomies of ecology v. history, and dispersal v. vicariance”.

By this persuasive pitching of ecology against history, the complexity of integrating biodiversity data across multiple disciplines is increased. The opposition has opened a rift in the data collection efforts which merits deep consideration – it poses the question of what we should be mapping. Thus Wheeler and Cracraft inveigh against the concept of the All Taxon Biodiversity Inventory (ATBI), which has been an influential model in recent years. In an ATBI, an area is marked off and a group of taxonomists and parataxonomists work on inventorying all the species in that area. The authors argue that:

> While periodic collecting at known sites is a prerequisite for documentation of the status and trends of biodiversity, the very notion of long-term study at a few anointed sites is inherently an ecological approach while the resolution of fundamental questions about biodiversity require answers grounded in a systematic biological approach. (Wheeler and Cracraft, 1997: 439)

What they mean here is that the biodiversity information question is not to be answered by surveying a few communities in depth but by doing the systematics work necessary to developing strategies for retaining the most genetic information. Ecological data has traditionally been collected at very small units of extent – plots of $\leq 1\ m^2$ over relatively short periods of time (Michener et al., 1997: 330). It is not enough just to scale up and integrate over the multiple disciplines that might contribute biodiversity data (a difficult problem in its own right, as we have seen) – that data comes in two major incompatible flavors with maps that are at very different levels of scale. Different scientific approaches are both vying for the scarce resources to carry out data collection.

Biodiversity Science vs Biodiversity Politics

We have just seen that there is a problem with integrating data across a range of disciplines which have two fundamentally incommensurable ontologies. A second integration problem is that data which is collected is being integrated into two discourses – a scientific and a political discourse - which operate in two different (overlapping but sometimes analytically distinct) sets of relations.

A comparison of maps of systematics collections against species richness indicates one broad stroke of the problem: broadly speaking, species are a third world commodity; information about species is a first world commodity.

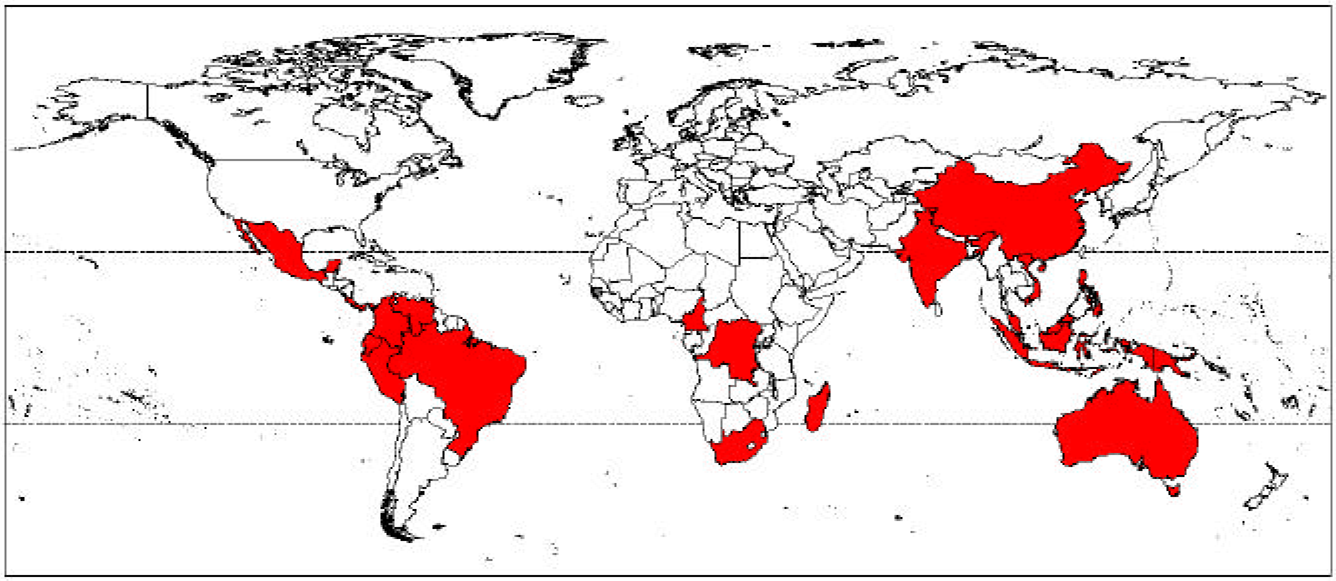

Figure 2. Countries with highest species diversity
These are the 20 countries with the highest species diversity as estimated by the WCMC national biodiversity index. Data from an earlier version of Table 1 (multiple sources). See Text (and notes) to Table 1 for explanation.

Figure 1 Countries with highest species diversity

(ftp://ftp.wcmc.org.uk/products/wcmc.publications/1.sourcebook - World Conservation Monitoring Centre, Biodiversity Data Sourcebook, Figure 2)

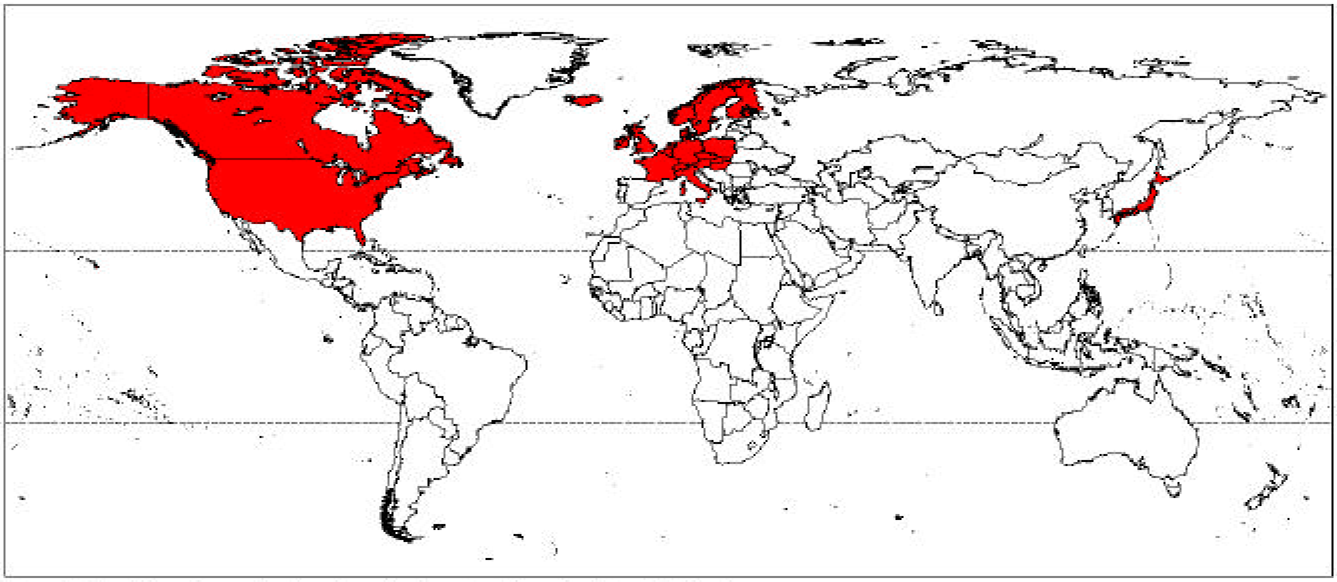

Figure 8. Countries with most systematics collections in relation to national biodiversity
Data from Tables 8 and 1 (multiple sources). Levels of biodiversity assessed by WCMC national biodiversity index; see Text (and Notes) to Table 1 for explanation.

Figure 2 Countries with most systematics collections (ftp://ftp.wcmc.org.uk/products/wcmc.publications/1.sourcebook - World Conservation Monitoring Centre, Biodiversity Data Sourcebook, Figure 8)

At the top level, there is a question of equity here. Underdeveloping countries (to use John Berger's (Berger and Mohr, 1975: passim) phrase) are becoming reluctant to share information and specimens, for good economic reasons. Thus Roblin bemoans the current difficulty of shifting microbial strains across borders:

> The days when one could walk into any country with interesting habitats for microbial diversity and walk out with one's pockets full of interesting samples appear to be over. Countries containing such habitats now are aware that they may harbor microorganisms with commercial potential. (Roblin, 1997: 472).

There have been innumerable cases in the past decade of access closing to information which was once seen at the prerogative of the scientific or imperial elite – some examples at different levels of granularity include the controversy about the genetic sampling of the complete human population of Iceland (Enserink, 1998); the Human Genome Diversity project with its attempt to collect gene samples from isolated communities (Hayden, 1998); the repatriation of the Neodata tropical fish database to Brazil.

These equity questions are not easily solved however – equity entails commensurability, and in many cases the economic and information systems that are being integrated are fundamentally incommensurable. It is impossible to give fair recompense for information if one cannot determine its owner. First of all, there is frequently a different understanding of 'ownership' – as Posey writes, intellectual property rights:

> (1) …are intended to benefit society through the granting of exclusive rights to 'natural' and 'juridical' persons or individuals, not collective entities such as indigenous peoples. As the Bellagio Declaration puts it: Contemporary intellectual property law is constructed around the notion of the author as an individual, solitary and original creator, and it is for this figure that its protection are reserved… they cannot protect information that does not result from a specific historic act of 'discovery'.
>
> (2) Indigenous knowledge is transgenerational and community shared. Knowledge may come from ancestor spirits, vision quests, or orally transmitted lineage groups… (Posey, 1997: 86).

Watson-Verran (for example in Watson-Verran and Turnbull, 1995) has discussed similar issues arising from different ontologies of ownership between white Australians and aborigines; Turnbull (Turnbull, 1989) discusses the use of traditional Aboriginal representations of the land in court cases. Hayden (Hayden, 1998) discusses the difficulty of locating the 'owners' of information about herbs sold in markets in Mexico – frequently the traders are peripatetic, buying the herbs from a number of different sources; and they gain information about their medical use partly from their local contacts and partly from others passing through markets buying their herbs and telling them their medical use. Who, in this case, should be reimbursed for giving Monsanto information about an herb that leads to lucrative drug development? In principle, there needs to be an ethnography of ownership (a longitudinal qualitative study of community knowledge) prior to each particular determination; in practice this just does not happen, and the requirement to respect intellectual property rights is honored by formally designating the trader as the fount of knowledge (Hayden, 1998). As Posey, Watson Verran and Hayden indicate, there is currently no standard, workable organizational interface permitting the fair exchange of information across cultural and economic divides. Even making the bold assumption of good will on all sides, then, there is continuing de facto information imperialism, causing a net data drain out of the Third World into Western databanks. And yet information from many countries must be integrated in order to carry out biodiversity research and develop reasonable policies for the planet as a whole, since environmental questions as a whole do not respect national borders.

These equity issues speak to the difficulties of gathering together information for some notional database housed, more likely than not, in North America or Europe. Political questions do not go away once the hurdle of access is cleared. Indeed they are continually raised, through the multiple uses of biodiversity databases, in issues of data algorithms and granularity of descriptions. Edwards et al, for example, discuss the use of vegetation as a surrogate for animal species presence in Gap Analysis – they use vegetation because it can be classified from aerial photographs. In ground checks, they found that this led to more errors of commission than omission in the locating of animal species, but argued:

> Given that Gap Analysis is a tool for predicting geographic distributions of terrestrial vertebrates for use in conservation planning, we argue that commission is preferred over omission. (Edwards et al., 1995: 4-10)

Such generous errors are frequently made in estimations of the number of species in the world (Paul, 1998: 3). Thus, Stork argues that molecular based species counts, which give higher estimates of numbers than counts using morphospecies concepts, are sometimes used as a political club (Stork, 1997: 60) – and incidentally this contributes to the supplanting of morphology by cladistics (and so is experienced by many scientists as part of their disciplinary struggle for survival). Similar biases occur in estimates of the number of extinctions that are occurring. This most certainly does not imply that biodiversity problems are not of crucial and pressing importance. To the contrary. It does indicate the difficulties of using data in multiple ways. Relatedly, Klemm notes that one cannot legislate for the ways in which people will use the data in a public database – he describes one problem with contradictory implications as follows:

> A difficult problem has always been to decide whether or not the location of endangered, rare or protected plants should be kept secret. Keeping the location secret avoids unscrupulous collection, vandalism or willful destruction by

> landowners fearing restrictions to development. On the other hand publicizing the location avoids inadvertent destruction in good faith. (Klemm et al., 1990: 28)

Biodiversity work is integrally scientific and political, so it should not be surprising that its data shares the same features.

The Problems of Integration

In this section we have seen how there are two kinds of integration ideally going on in biodiversity work – between ecological and systematics data and between scientific and political discourse. We have also seen that both kinds of integration cannot in principle be smoothly accomplished. Ecological and systematics data cannot be rendered equal just by standardizing over a set of weights and measures; and scientific data cannot be collected without making politically charged decisions about its representation in a database. I have included both of these under the same general rubric because analytically much the same processes are occurring in both cases: the forging of a dynamic uncompromise between agonistic groups in the very creation and structuring of biodiversity databases. The databases being developed today do not impose a hegemonic solution: they unfurl within them, at the level of data structure and data processing algorithms, the contradictions folded into their creation.

Conclusion: Time, Space and Biodiversity Databases

In the introduction, I evoked the problem of the integration of biodiversity data. I have explored two broad themes: how objects are named (and what is not named); and how do the intended (and unintended) users of the databases communicate. On the one hand, I have argued throughout that what we need to know about data in a database is far more than the measurement standards that were used, and on the other hand, I have argued that atomic elements of a database such as measurement standards, contain complex local histories folded (Deleuze, 1996: passim) into them, histories which must be understood if the data is to persist and the maps incorporating the results of multiple conflicting databases to be drawn.

Each particular discipline associated with biodiversity has its own incompletely articulated series of objects, times and spaces. These objects each enfold an organizational history and subtend a particular temporality or spatiality. They frequently are incompletely articulated with other objects, temporalities and spatialities – often legacy versions when drawing on non-proximate disciplines. All this sounds like an unholy mess if one wants to produce a consistent, long-term database of biodiversity relevant information the world over. At the very least it suggests that global panoptica are not the way to go in biodiversity data!

Failure to name and standardize should not be read as a product of consistent contingent failures of nomenclatural bodies and data standards committees. On the contrary, it is deeply reflective of the nature of disciplinary research in biodiversity related sciences. These sciences deal in objects, spaces and times that cannot be readily normalized one against the other. To get an understanding of biodiversity data, we need to move away from a goal of producing a global panopticon, which will always be unattainable. We should instead begin to look at the practices that are productive of local orderings and alignments of datasets.

Each (sub)discipline is acting as an effective spokesperson (Latour, 1987: Chapter 3) for the objects plus spatial and temporal units it produces. What we have seen throughout this paper is that the ordering of data across multiple disciplines is not simply a question of finding a commonly accepted set of spatial and temporal units and naming conventions – though this is the way that it is often portrayed in the literature (Michener et al., 1998; Michener, Brunt et al., 1997; Dempsey and Heery, 1998). To the contrary, these ordering issues lead us very quickly on the one hand into deep historiographical questions and on the other to questions of institutional politics. If we are going to develop decent biodiversity maps to inform policy then we need databases held together through more than just good metadata practice. In a biodiverse world we need to be think through ways of manipulating ontologically diverse data. It is surely a rich challenge to the GIS community to devise forms of representation which integrate without traducing the multiple data diversities of the field of biodiversity.

References:

ADAMS, D. and LLOYD, J., 1990, *The Deeper Meaning of Liff* (New York: Harmony Books).

AGER, D. V., 1993, *The new catastrophism : the importance of the rare event in geological history* (Cambridge: Cambridge University Press).

BARROWCLOUGH, G. F., 1992, Systematics, Biodiversity and Conservation Biology, In *Systematics, Ecology, and the Biodiversity Crisis*(Ed, Eldredge, N.) New York: Columbia University Press, pp. 121-143.

BERGER, J. and MOHR, J., 1975, *A seventh man : a book of images and words about the experience of migrant workers in Europe* (Harmondsworth ; Baltimore: Penguin).

BOWKER, G. C., 1994a, Information Mythology: the world of/as information, In *Information Acumen: the understanding and use of knowledge in modern business*(Ed, Bud-Frierman, L.) London: Routledge, pp. 231-247.

BOWKER, G. C., 1994b, *Science on the Run: Information Management and Industrial Geophysics at Schlumberger, 1920-1940* (Cambridge, MA: MIT Press).

BOWKER, G. C., 1998, Archival technology in the historical sciences 1800-1997. *History of Technology,* **15,** 69-87.

BOWKER, G. C., 1999, The Game of the Name: Nomenclatural Instability in the History of Botanical Informatics, In *History and Heritage of Science Information Systems*(Eds, Bowden, M. E., Hahn, T. B. and Williams, R. V.) Medford, NJ: ASIS, pp. 74-83.

BOWKER, G. C. and STAR, S. L., 1999, *Sorting Things Out: Classification and its Consequences* (Cambridge, MA: MIT Press).

BRENCHLEY, P. J. and HARPER, D. A. T., 1998, *Palaeoecology : Ecosystems, Environments, and Evolution* (London: Chapman & Hall).

CHRISMAN, N. and HARVEY, F., 1998, Boundary Objects and the Social Construction of GIS Technology. *Environment and Planning A,* **30,** 1683-1694.

CHRISMAN, N., 1997, *Exploring Geographic Information Systems* (New York: John Wiley and Sons).

CHRISMAN, N., 1999, typescript from author, .

COX, C. B., 1998, From generalized tracks to ocean basins - how useful is Panbiogeography? *Journal of Biogeography,* **25,** 813-828.

DELEUZE, G., 1996, *Proust et les signes* (Paris: Presses universitaires de France).

DEMPSEY, L. and HEERY, R., 1998, Metadata: A Current View of Practice and Issues. *The Journal of Documentation,* **54,** 145-172.

EDWARDS, T. C., HOMER, C. G., BASSETT, S. D., FALCONER, A., RAMSEY, R. D. and WIGHT, D. W., 1995, *Utah GAP analysis : An Environmental Information System* (Logan, UT: National Biological Service, Utah Cooperative Fish and Wildlife Research Unit, Utah State University).

ELDREDGE, N., 1992, Where the Twain Meet: Causal Intersections Between the Genealogical and Ecological Realms, In *Systematics, Ecology, and the Biodiversity Crisis*(Ed, Eldredge, N.) New York: Columbia University Press, pp. 59-76.

ELICHIRIGOITY, F., 1999, *Planet Management: Limits to Growth, Computer Simulations, and the Emergence of Global Spaces* (Evanston, IL: Northwestern Unversity Press).

ENSERINK, M., 1998, Opponents Criticize Iceland's Database. *Science,* **282,** 859.

GOODWIN, C., 1994, Professional vision. *American Anthropologist,* **96,** 606-633.

GRAY, B., 1980, Popper and the 7th Approximation: The Problem of Taxonomy. *Dialectica,* **34,** 129-153.

GREHAN, J. H., 1994, The Beginning and End of Dispersal: The Representation of 'Panbiogeography'. *Journal of Biogeography,* **21,** 451-462.

HARVEY, F., 1997, From Geographic Holism to Geographic Information System. *Professional Geographer,* **49,** 77-85.

HAYDEN, C. P., 1998, A Biodiversity Sampler for the Millenium, In *Reproducing Reproduction: Kinship, Power and Technological Innovation*(Eds, Franklin, S. and Ragone, H.) Philadelphia: University of Pennsylvania Press, pp. 173-206.

HEYWOOD, V.H. 1988, The Structure of Systematics, In *Prospects in systematics,* edited by D. L. Hawksworth and Systematics Association. Oxford: Published for the Systematics Association by Clarendon Press, pp.44-56.

HEYWOOD, V. H., 1991, Needs for Stability of Nomenclature in Conservation, In *Improving the Stability of Names, Needs and Options: Proceedings of an International Symposium, Kew, 20-23 February 1991*(Ed, Hawksworth, D. L.) Königstein/Taunus, Germany: Published for the International Association for Plant Taxonomy in conjunction with the International Union of Biological Sciences and the Systematics Association by Koeltz Scientific, pp. 53-58.

HEYWOOD, V. H., 1997, Information Needs in Biodiversity Assessments - from Genes to Ecosystems, In *Biodiversity Information: Needs and Options: Proceedings of the 1996 International Workshop on Biodiversity Information*(Eds, Hawksworth, D. L., Kirk, P. M. and Clarke, S. D.) Oxon, OX, UK ; New York, NY, USA: CAB International, pp. 5-20.

HEYWOOD, V. H., WATSON, R. T. and UNITED NATIONS ENVIRONMENT PROGRAMME, 1995, *Global Biodiversity Assessment* (Cambridge ; New York, NY, USA: Cambridge University Press).

HUGGETT, R. J., 1997, *Environmental Change: The Evolving Ecosphere* (London: Routledge).

JOURNET, D., 1991, Ecological Theories as Cultural Narratives: F.E. Clements's and H.A. Gleason's 'Stories' of Community Succession. *Written Communication*, **8,** 446-472.

KLEMM, C. d., International Union for Conservation of Nature and Natural Resources and World Wide Fund for Nature, 1990, *Wild plant conservation and the law* (Gland, Switzerland: IUCN-The World Conservation Union).

KOCH, C. F., 1998, 'Taxonomic Barriers' and Other Distortions within the Fossil Record, In *The Adequacy of the Fossil Record* (Eds, Donovan, S. K. and Paul, C. R. C.) Chichester, New York: John Wiley, pp. 189-206.

LATOUR, B., 1987, *Science in Action: How to Follow Scientists and Engineers Through Society* (Milton Keynes: Open University Press).

LATOUR, B., 1993, *La Clef de Berlin et autres leçons d'un amateur de sciences* (Paris: La Decouverte).

MICHENER, W. K., BRUNT, J. W., HELLY, J. J., KIRCHNER, T. B. and STAFFORD, S. G., 1997, Nongeospatial Metadata for the Ecological Sciences. *Ecological Applications,* **7,** 330-342.

MICHENER, W. K., PORTER, J. H. and STAFFORD, S. G. (Eds.), 1998, *Data and Information Management in the Ecological Sciences: A Resource Guide* (Albuquerque, NM: LTER Network Office, University of New Mexico).

PAUL, C. R. C., 1998, Adequacy, Completeness and the Fossil Record, In *The Adequacy of the Fossil Record*(Eds, Donovan, S. K. and Paul, C. R. C.) Chichester, New York: John Wiley, pp. 1-22.

POSEY, D. D., 1997, Wider User and Applicatoin of Indigenous Knowledge, Innovations and Practices: Informations Systems and Ethical Concerns, In *Biodiversity Information, Needs and Options: Proceedings of the 1996 International Workshop on Biodiversity Information*(Eds, Hawksworth, D. L., Kirk, P. M. and Clarke, S. D.) Oxon, OX, UK ; New York, NY, USA: CAB International, pp. 69-103.

RAVEN, P. H., BERLIN, B. and BREEDLOVE, D. E., 1971, The Origins of Taxonomy: A review of its historical development shows why taxonomy is unable to do what we expect of it. *Science,* **174,** 1210-1213.

RICHARDS, T., 1996, *The Imperial Archive: Knowledge and the Fantasy of Empire* (London: Verso).

ROBLIN, R., 1997, Resources for Biodiversity in Living Collections and the Challenges of Assessing Microbial Biodiversity, In *Biodiversity II: Understanding and*

Protecting our Biological Resources(Eds, Reaka-Kudla, M. L., Wilson, D. E. and Wilson, E.) Washinton, D.C.: Joseph Henry Press, pp. 467-474.

SHEPHERD, I. D. H., 1991, Information Integration and GIS, In *Geographical Information Systems*, Vol. 1 (Eds, Goodchild, M. F. and Rhind, D. W.) London: Longman, pp. 337-360.

SISMONDO, S., 1999, Island Biogeography and the Multiple Domains of Models. *Biology and Philosophy,* **15,** 239-258.

STORK, N. E., 1997, Measuring Global Biodiversity and Its Decline, In *Biodiversity II: Understanding and Protecting our Biological Resources*(Eds, Reaka-Kudla, M. L., Wilson, D. E. and Wilson, E.) Washinton, D.C.: Joseph Henry Press, pp. 41-68.

TURNBULL, D., 1989, *Maps are Territories: Science is an Atlas* (Geelong , Victoria: Deakin University Press).

VANE-WRIGHT, R. I., HUMPHRIES, C. J. and WILLIAMS, P. H., 1991, What to Protect? - Systematics and the Agony of Choice. *Biological Conservation,* **55,** 235-254.

WATSON-VERRAN, H. and TURNBULL, D., 1995, Science and Other Indigenous Knowledge Systems, In *Handbook of science and technology studies*(Eds, Jasanoff, S., Markle, G., Petersen, J. and Pinch, T.) Thousand Oaks, Calif.: Sage Publications, pp.115-139 .

WHEELER, Q. D. and CRACRAFT, J., 1997, Taxonomic Preparedness: Are We Ready to Meet the Biodiversity Challenge?, In *Biodiversity II: Understanding and*

Protecting our Biological Resources(Eds, Reaka-Kudla, M. L., Wilson, D. E. and Wilson, E.) Washinton, D.C.: Joseph Henry Press, pp. 435-446.

WILLIAMS, R., 1983, *Keywords : a vocabulary of culture and society* (London: Fontana Paperbacks).

Modern biogeography | 273

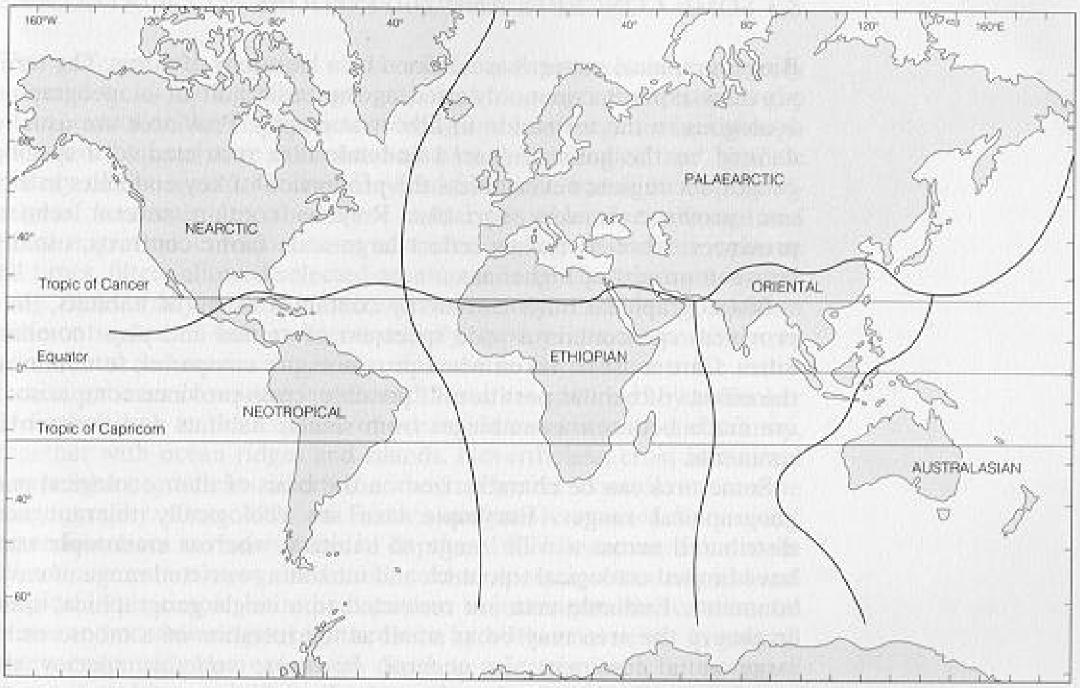

Figure 8.2 Modern biogeography. (Adapted from various sources.)

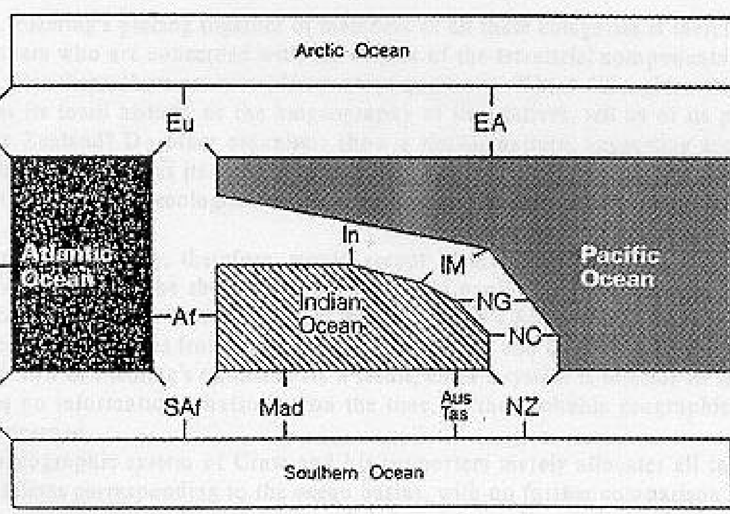

FIG. 5. A new world system of biotic regions based on panbiogeographic analyses. Abbreviations: Af, Africa; Aus/Tas, Australia/Tasmania; CA, Central America; EA, East Asia; EU, Europe; In, India; IM, Indo-Malaya; Mad, Madagascar; NA, North America; NC, New Caledonia; NG, New Guinea; NZ, NeW Zealand; SA, South America; SAf, South Africa. Figure and caption from Craw & Page (1988), p. 181.