Social Studies of Science
http://sss.sagepub.com

Biodiversity Datadiversity

Geoffrey C. Bowker
Social Studies of Science 2000; 30; 643
DOI: 10.1177/030631200030005001

S|S|S

ABSTRACT Biodiversity is a data-intense science, drawing as it does on data from a large number of disciplines in order to build up a coherent picture of the extent and trajectory of life on earth. This paper argues that as sets of heterogeneous databases are made to converge, there is a layering of values into the emergent infrastructure. It is argued that this layering process is relatively irreversible, and that it operates simultaneously at a very concrete level (fields in a database) and at a very abstract one (the coding of the relationship between the disciplines and the production of a general ontology). Finally, it is maintained that science studies as a discipline is able to (and should) make a significant contribution to the design of robust and flexible databases which recognize this performative character of infrastructure.

Keywords archives, interdisciplinarity, metadata, social informatics

Biodiversity Datadiversity

Geoffrey C. Bowker

The form of scientific work which has been most studied by sociologists of science is that which leads from the laboratory to the scientific paper by means of the creation of ever more abstract and manipulable forms of data, which Bruno Latour has dubbed 'immutable mobiles' (Latour, 1987: 227ff.). In this process, there is no need to hold on to data after it has been enshrined in a scientific paper: the paper forms the 'archive' of scientific knowledge (frequently adopting names redolent of this storage ambition, such as the *Archives for Meteorology, Geophysics and Bioclimatology*). The scientific paper, which is the end result of science, contains an argument about an hypothesis (which is proved or disproved) and a set of supporting data which is, saving a controversy, taken on faith by the scientific community (Shapin, 1994: 10–13). The archive of scientific papers can then be indexed both in terms of arguments made and information stored.

However, over the past 20 years we have seen in a number of new and of formerly canonical sciences a partial disarticulation of these two features of scientific work. Increasingly, the database (the information stored) is seen as an end in itself. According to most practitioners, the ideal database should be theory-neutral, but should serve as a common basis for a number of scientific disciplines to progress. Thus one might cite the Human Genome Initiative and other molecular biological projects (Hilgartner, 1995) as archetypical of a new kind of science in which the database is the end product. The human genome databank will in theory

Social Studies of Science 30/5(October 2000) 643–83

[0306-3127(200010)30:5;643–683;016844]

be used to construct arguments about genetic causation of disease, about migration patterns of early humans, about the evolutionary history of our species; but the process of producing causation is distinct from the process of 'mapping' the genome – the communities, techniques and aims are separate.

This disarticulation, which operates in the context of producing a working archive of knowledge, is not in itself new. To limit ourselves arbitrarily to the past 200 years, a significant percentage of scientific work has been involved with creating such an archive. Napoleon's trip to Egypt (France, 1966) included a boatload of geologists, surveyors and natural historians, and reflected a close connection between the ends of empire and the collection of scientific knowledge. Thus also William Smith's geological survey of Britain or James Cook's travels to Australia. Thomas Richards' *The Imperial Archive* (1996) presents some wonderful analysis of the imperial drive to archive information so as to exercise control (a theme familiar, of course, to readers of Latour). The working archive is a management tool. What is new and interesting is that the working archive is expanding in scale and scope. As Michel Serres (1990) points out, we are now as a species taking on the rôle of managing the planet as a whole – its ecosystems and energy flows. Where Charles Lyell (1832, Volume 2) saw human influence as essentially natural, we now see nature as essentially only possible through human mediation. We are building working archives from the submicroscopic level of genes up through the diversity of viral and bacterial species to large-scale floral and faunal communities and the mapping of atmospheric patterns and the health of the ozone layer. There is an articulation here between information and theory, but the stronger connection is between information and action – with competing models based on the same data producing policy recommendations. In this new and expanded process of scientific archiving, data must be reusable by scientists. It is not possible simply to enshrine one's results in a paper: the scientist must lodge her data in a database which can be easily manipulated by other scientists.

In the relatively new science of biodiversity, this data collection drive is achieving its apogee. There are programmes afoot to map all floral and faunal species on the face of the earth. In principle, each of these maps should contain economic information about how groups of animals or plants fend for themselves in the web of life (http://curator.org/WebOfLife/weboflife.htm) and genetic information (about how they reproduce). In order truly to understand biodiversity, the maps should not only extend out in space but back in time (so that we can predict how a given factor – like a 3°C increase in world temperature – might affect species distribution). At the beginning of the 1990s, Vernon Heywood noted:

> Global synoptic or master species databases are being developed for a diverse array of groups such as viruses, bacteria, protists, fungi, mollusks, 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: 10)

Similarly, NASA's Mission to Earth programme 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 meets 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. (as cited in Elichirigoity, 1999: 14-15)

The UK Systematics Forum publication *The Web of Life* (Forum, 1998: 25) quotes E.O. Wilson's invocation, 'Now it is time to expand laterally to get on with the great Linnaean enterprise and finish mapping the biosphere', and speaks of the need to 'discover and describe the Earth's species, to complete the framework of classification around which biology is organized, and to use information technology to make this knowledge available around the world'. These panoptical dreams weave together work from the very small-scale molecular biological to the large-scale geological, and temporally from the attempt to represent the present to a description of the history of all life on earth. They constitute a relatively direct continuation of the drive for the imperial archive so well described by Richards (1996): where the notional imperial archive sought to catalogue completely the far-flung social and political empire in order to govern it better,[1] biodiversity panopticons seek to catalogue completely the natural empire, for much the same reason (Serres, 1990). Although they work as 'oligopticons' (Latour & Hermant, 1998: 58) – covering, as we shall see, only a thin slice of species and environments – they are created to be, and are manipulated as if they were, panopticons.

The information collection effort that is being mounted worldwide is indeed heroic. Databases from far-flung government agencies, scientific expeditions, amateur collectors are being integrated more-or-less successfully into very large-scale searchable databases. This work is being done largely without the input of social analysts of science. In this paper, I make the argument that we in social studies of science have a significant contribution to make to the process of federating databases in order to create tools for planetary management. In particular, I argue that we can produce means to engage the complexity and historicity of data within the sciences so that social, political and organizational context is interwoven with statistics, classification systems and observational results in a generative fashion. In short, I will argue that we need to historicize our data and its organization so as to create flexible databases that are as rich ontologically as the social and natural worlds they map (as would be required by Ashby's law of requisite variety), and so which might really help us gain long-term purchase on questions of planetary management (Elichirigoity, 1999).

The Scientific Archive

Let us assume for the nonce a perfect world. In this perfect world, we have solved the naming problem – names mean just what we want them to mean and nothing else, and they allow us to tell stories which are just so. Lest the reader relax too much, be assured that all is not well. For even if we can name everything consistently, there are the problems of how to deal with old data and how to ensure that one's data doesn't rot away in some 'information silo' (in Al Gore's memorable phrase) for want of providing enough context. The problem with biodiversity data – as with much environmental data – is that the standard scientific model of doing a study doesn't work well enough. In the standard model, one collects data, publishes a paper or papers and then gradually loses the original dataset. A current locally-generated database, for example, might stay on one's hard drive for a while then make it to a zip disk, then when zip technology is superseded it will probably become for all intents and purposes unreadable until one changes jobs or retires and throws away the disk. There are a thousand variations of this story being repeated worldwide – more generally along the trajectory of notebooks to shelves to boxes to dumpsters.

When it could be argued that it was precisely the rôle of scientific theory as produced in journals to order information – to act as a form of memory bank – this loss of the original data was not too much of a problem. The data was rolled into a theory which not only remembered all its own data (in the sense of accounting for it and rendering it freely reproducible) but potentially remembered data which had not yet been collected (for the concept of 'potential memory', see Bowker, 1997). By this reading, what theory did was produce readings of the world that were ultimately data-independent – if one wanted to descend into data at any point all one had to do was design an experiment to test the theory and the results would follow.

However, two things render this reading of the data/theory relationship untenable. First, it has been shown repeatedly in the science studies literature (the *locus classicus* is Collins, 1985) that scientific papers do not in general offer enough information to allow an experiment or procedure to be repeated. This entails that in a field where old results are continually being reworked, there is a need to preserve the original data in as good a form as possible. Secondly, in the biological sciences in general – and the environmental sciences in particular – the distributed database is becoming a new model form of scientific publication in its own right. The Human Genome Initiative is resulting in the production of a very large collaborative database, for example. In the environmental sciences, where the unit of time for observing changes can be anything from the day to the millennium, there is a great value in having long, continuous data sets. The problem of what data to retain in order to keep a data set 'live' is a metadata problem; and as Rick Ingersoll and his colleagues note: 'the quality of metadata is probably the single most important factor that determines the longevity of environmental data' (Ingersoll et al., 1997: 310).

As Bruno Latour reminds us in *Science in Action* (1987), science is an eminently bureaucratic practice deeply concerned with record-keeping. Disciplines do mixed jobs of keeping track of their own results over time – indeed a key finding of science studies has been that using 'theory' as a way of storing old, and accounting for potential, data can be highly problematic, since replacement theories do not automatically account for all the data held in the outgoing one (the *locus classicus* is Kuhn, 1970). The difficulties become apparent when you move beyond the arrangement and archiving of data within a given science to look at what happens in the efforts of a vast number of sciences (working from the scale of molecular biology on up to that of biogeography or even cosmology) to coordinate data between themselves within the field of biodiversity. In practice, the sciences use many differing 'filing systems' and philosophies of archival practice. There is no automatic update from one field to a cognate one, such that the latest classification system or dating system from the one spreads to the other. Further, it is often a judgement call whether one needs to adopt the latest geological timeline, say, when storing information about ecological communities over time; particularly if one's paper or electronic database is structured in such a way that adopting the new system will be expensive and difficult. Such decisions, however, have continuing effects on the interpretation and use of the resultant data stores.

There has been relatively little work in science studies dealing with the organizational, political and scientific layering of data structures. In his analysis of work in the GE lab, Bernie Carlson (1991) made the important point that frequently political issues are fought out using technical language – a point developed in a different context by Michel Callon in his classic study of electrical cars in France (Callon, 1986). Paul Edwards (forthcoming) has demonstrated the temporal layering of models used in climate science – an assumption written into a 1960s model might well still survive into a 2000s' application. In her rich study of the development of the Internet, Janet Abbate (1999) has shown how the distinction between the material world (hardware; wiring) and the discursive world (code; protocols) breaks down in the instauration of modern infrastructures. The assignation of an attribute to the world of discourse or of materiality is shifting, *post hoc* (Latour makes a similar point in his *We Have Never Been Modern* [1993], though without the empirical base). Together, this range of sources points to a path for understanding information infrastructures in technoscience that I will develop throughout this paper. In sum, they suggest that information infrastructures such as databases should be read both discursively and materially; that they are a site of political and ethical as well as technical work (see Bowker & Star, 1999: Chapter 1); and that there can be no *a priori* attribution of a given question to the technical or the political realms.

I will be returning in the conclusion to the phenomenon of *irreversibility* (Callon, 1991) in the development of an information infrastructure (a federation of databases in the case of biodiversity).[2] The central argument of the paper questions a powerful trend in science studies, one which privileges the 'school' or the 'community of practice'. In this line of work, when infrastructural technologies such as classification systems are discussed, it is in terms of the mapping between a given group's ideology or technology and their results: one community, one ideology. Within the field of taxonomy, for example, there has been John Dean's lovely analysis of controversy in taxonomy, with his conclusion (Dean, 1979: 225) that taxonomies are invented, not discovered: 'The different goals of the orthodox and experimental taxonomist are pursued in different institutional locations and are thus linked to sets of professional vested interests'. Malcolm Nicolson (1989), in his discussion of the taxonomy of plant communities in North America (Clements) and Southern Europe (Braun-Blanquet), produces a similar conclusion. He argues convincingly that the larger community units in America cannot simply be attached to the fact that nature is different in America (or else Russia would have evolved a similar system; and Britain would have developed a different one). Rather, he maintains, the social and political work that the classification systems were doing institutionally and (in America) in terms of practical land management questions made the difference. Both of these cases convince, but both deploy a one-to-one mapping between a classification system and its setting. However, it is precisely this setting that cannot be assumed when we look at the development of very-large-scale infrastructures. To the contrary, classification systems are increasingly being yanked out of their institutional and political contexts, and applied in other fields with differing ontologies and associations. Leigh Star has developed the generative terms of 'boundary objects' and 'boundary infrastructure' (Bowker & Star, 1999: 15-16, 292-314) to describe one aspect of this process. But again, with the development of shared boundary objects there is a process of negotiation between the different communities of practice deploying a common infrastructure – each knows very well what the object means for itself. With biodiversity science of the type I address in this paper, there is a layering of boundary objects many levels deep into an increasingly coherent infrastructure. I will argue that this creates a form of irreversibility in the infrastructure for two reasons: first because the infrastructure is *performative*; and second because the infrastructure is *diffuse*.

In this paper, then, I build on the science studies work surveyed above to develop an understanding of the development of large-scale information infrastructures in terms of the converging and conflicting information needs of multiple communities of practice (Lave & Wenger, 1991). I use this analysis to show how the field of science studies can make a useful contribution to the world of biodiversity informatics. I identify three major dimensions of the development of very-large-scale databases in biodiversity science – naming entities; retaining the context of information as it passes beyond one's disciplinary bounds on the one hand, or as time passes within one's own discipline on the other; and integrating information held by multiple disciplines or in multiple sites. I draw heavily from the extant literature in the field of biodiversity informatics for my development of these dimensions. However, there is in general very little purchase in this field on how to think about them. There is currently widespread belief in 'technical fixes' for inherently social, organizational and philosophical problems – such as curing the ills of incompatible datasets through developing metadata standards (for example: Michener et al., 1997; Michener et al., 1998). Further, there is a disjunction between the policy and the informatics discourses. Major works on the politics of environmental discourse (such as Hajer, 1995, and Connelly & Smith, 1999) do not mention environmental informatics at any point: they write as if there is no layer at all between science and politics. In a concluding section I will summarize and expand on the case for a contribution by science studies both to the ongoing data management effort and to the discourse on biodiversity policy. I will argue that this is important work, since the field of data management is one of crucial importance to the interface between theory and practice – in our case scientific discussions of biodiversity and planetary management.

Naming

Things That Are Hard to Classify

As Douglas Adams and John Lloyd remind us in *The Deeper Meaning of Liff* (1990), there are many things that we would like to talk about but just don't currently have the words in the language for.[3] 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 without a classification system. (I say 'in general' because there are systems which claim to do classification work after data entry – such as systems that produce automatic, dynamic classifications of large bodies of free text data. For our purposes, though, the problem only recurses with these systems, since the free text itself needs to have some kind of regular naming system – or set of naming systems – for it to be amenable to useful automatic classification.)

A major problem in the world of biodiversity is that while classification systems are about kinds of things (flora, fauna, communities, and so on), the world of biodiversity data is radically singular. 'Species' and 'kinds' are generally really just specimens. 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, 1995: 48).

Just as species can be endemic to very small areas, so too can data about species. Peter Raven and his colleagues remark that this throws into relief the fact that our names for organisms do not contain much information:

> [T]he 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' (specimens which define the species) 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 & Star, 1999: 118) remarked that this bootstrapping problem is a common feature of the development of global databases.[4]

There are many entities or communities that in principle one would like to be able to classify so as to get good comparative data, but which do not prove easily amenable to classification – and so have led to the proliferation of local data stores, or to no data collections at all. This sets up a reverse bootstrapping process, whereby things which cannot be described easily and well get ignored, and so receive an ever decreasing amount of attention. This has been a problem for those working with fossils, for example. Fossils are notoriously difficult to deal with. They are preserved in several different fashions, and each mode of preservation favours its own set of body parts:

> The number of currently recognized compression species is much too high due to underestimation of individual variation and poor knowledge of the leaf polymorphism. Many fewer species have been distinguished in permineralized material. (Galtier, 1986: 12)

This means that it is often very unclear whether one is dealing with a new species descended from one below it (following the general rule that going down means going back in time), or whether one is dealing with a variant leaf form of the same species. It is often not clear how to match parts of plants (leaves and stems and flowers), let alone seeds recovered from a midden, with any given leaf (compare the descriptions of the various historical reconfigurations of body parts in the Burgess Shale by Gould, 1989). However, fossil information is extremely valuable in, for example, the mapping of oil fields – you can track down unconformities in the subsoil through characterization of fossil communities. Norman Hughes notes the reverse bootstrapping:

> The use of fossils has in recent years come to be regarded as cumbersome and unproductive; the work is said to abound with unimaginative complacency, with the obscurity of esoteric terminology, and with lack of compatibility of treatment of different groups including even that between the fossils of plants and animals. As a result, much effort has been directed by geologists in solving their stratal problems towards employing any other available physical or chemical phenomena, and thus to avoiding altogether the 'expensive' and supposedly ineffective use of paleontologists and their fossils.
>
> Paleontologists, who almost all continue to believe that their fossils and the distribution of these form the only viable method of discriminating diverse and confusing strata, are striving to present their fossils more ingeniously and to win back the confidence of the geologists. (Hughes, 1991: 39)

His response has been to produce a 'paleontologic data-handling code' which does not force one to choose a taxon descriptor for a given fossil. It is based, he claims:

> . . . first on equal treatment of all observational records, to avoid the loss of primary information by burial of records within taxa, which for fossils are inevitably subjective, and which render the original records uncheckable. . . . All observational records are made of paleotaxa (of similar scope to species but immutable) or as comparison records formalized to require an instant author decision on the degree of similarity. As the traditional generic name of fossils conveys no precise information, a binomial is completed instead by means of a 'timeslot' name based always on a timescale division of the Global Stratigraphic Scale of the International Union of Geological Sciences. (ibid.: 42)

This solution in turn leads to its own problems – first with respect to ensuring adoption (no easy matter since there are numerous groups using fossil data, often with their own well entrenched schemes), and secondly with respect to the stability of the global stratigraphic scale and of dating techniques.[5] This has led to the situation where geologists from some petroleum companies have abandoned the scientific literature and developed their own coding for fossil remains. David Hawksworth and Frank Bisby (1988: 16) describe the use of 'arbitrary coding systems' by some prospectors. And, as Michael Boulter and his colleagues point out (Boulter et al., 1991: 238), even when the agreed-upon naming procedure is followed, the fossil genera might be buried in, say, the *Journal of the Geological Society of London*, where no database manager nor neobotanist would look for a new genus. This *Journal* is, of course, well known to geologists, but it is not generally known by systematists; indeed, in practice, neobotanists have more than enough journals to work with already without having to search through the literature in any number of cognate fields for previous occurrences of a given species or genus: it is neither incompetence nor idleness that keeps the disciplines as apart as they are, but rather the cost of labour.

Electronic and organizational fixes are constantly being proposed, but they run up against the differing needs and standards of different disciplines. In the same paper, the authors note that there are just too many conflicting uses for fossils:

> Many of the difficulties in palaeobotany and its sub-discipline, palynology, occur as a result of the sometimes conflicting aims of botanical and geological researchers handling these data. (Boulter et al., 1991: 232)

The point here is that the bootstrapping work of getting any classification system off the ground is, as Hughes (1991: 39) notes, always a battle of 'winning confidence' and 'ingenious presentation'; and that, without this work, no technical fix to the data-storage problem will ever get established.

But it is not just the world of data which is radically singular: so is the world itself. A second class of things which it is hard to classify is entities without clear boundaries, like soils. Soils enter into some biodiversity databases as related to particular kinds of communities of flora and fauna: their classification is necessary for the generation of theories about the relationship between the geological and the biological. And yet different agencies within a nation often adopt different soil classifications, rendering the pooling of data complex. In an excellent article discussing 'the monumental (or dinosauric) *Soil Classification, A Comprehensive System – 7th Approximation*' (published in 1960 by the Soil Survey Staff of the US Department of Agriculture), Bennison Gray (1980) discusses the problems of classifying something that does not break up into natural units. Different sets of researchers will have very different intuitive definitions of what soil is. For many agriculturists, soil is just something you can or cannot grow crops in; for ecologists, soil can include hard rock (to which lichen can cling) and be defined as that part of the surface of the earth subject to weathering influences – a comprehensive but itself imprecise definition. Soil, Gray notes, is three-dimensional – but there it is an indexical question how far down or across you go to determine a unit (or pedon): in the prairie states of the United States there is a lot of uniformity in the topsoil; in California with its complex geological history, much less so. There are major national variations in classification in the USA, Australia, Belgium, Canada, Holland, England, France, Germany, New Zealand, Russia and Scandinavia: 'And, because none of these countries are tropical, soils of the tropics are not given detailed coverage in any of the nationally oriented taxonomies' (Gray, 1980: 141) – an exact analogy to a complaint launched by tropical countries against the early International Classification of Diseases, which was based on a list of causes of death in Paris (Bowker & Star, 1999: 114). Richard Huggett (1997: 178) remarks that:

> soil classification schemes are multifarious, nationalistic and use confusingly different nomenclature. Old systems were based on geography and genesis. . . . Newer systems give more emphasis to measurable soil properties that either reflect the genesis of the soil or else affect its evolution.

The local variations in classification are a variant on a major theme in geological history: because each region of the earth has its own complex story written into its soil and subsoil, a 'genetic' classification by origin which, since the mid- to late-19th century has been the canonical form of universal classifications in many fields (see Tort, 1989),[6] cannot so much uncover types as the singularity of every story. . . . One immediately sees the difficulty of trying to merge information on, say, soil type and species limits in different countries or across different agencies within a country – it is extremely difficult to normalize data across conflicting classification schemes, and the agencies or countries involved often have good reasons (bureaucratic, to do with the cost of data handling; educational, to do with norms at local institutions; or theoretical) for continuing to use their own schema. What is the status of such problems? They seem trivial to many scientists and informaticians, and yet they never in practice go away. Local data cultures constantly recreate themselves, in much the same way that 'races' and 'nations' do in the political world.

Along the same vein as soils, landscape topology has proven very difficult to classify. Huggett (1997: 228) notes that: '. . . many landform classifications are based purely on topographic form, and ignore geomorphic process' – they are not genetic but descriptive. He also notes that the Davisian 'geographical cycle' in the late 19th 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 through the stages of youth, maturity, and old age. However, these terms, borrowed from biology, are misleading and much censured. (ibid.: 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. However, it is impossible to characterize geodiversity over time without there being agreement on a classification of topography.

A third class of things which it is hard to classify is things which are of indeterminate theoretic status. Thus the concept of 'communities' in ecology. In the first half of the 20th century, Frederic Clements and Henry Allan Gleason took opposing views on the existence of natural communities (Journet, 1991). Clements said that there were natural large 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. But 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). Gleason rejected the overall stability of the climate (ibid.: 457). What Derek Ager calls 'the new catastrophism' (Ager, 1993) in geology resonates with a move in many fields over the past 20 to 30 years to stress disruption to cycles and to question any stabilities – accordingly 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, so-called 'GAP analysis', relies on a pragmatic definition of community, which includes many *ad hoc* assumptions, recognized as such by its developers (Edwards et al., 1995). GAP analysis involves taking aerial photographs of a given area, then marking off 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: section 4-2; see also the account by Tim Clark and Ron Westrum, 1977, of the difficulty of counting black-footed ferrets in Wyoming).

So we have covered here three kinds of thing that are hard to classify: entities where the data itself is singular and scattered; entities with fuzzy boundaries and which are singular; and entities of contested theoretical status. 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. 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 may mean 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 (Tort, 1989) against a stable temporal backdrop can be relatively easily named, but when the pace and texture of time are variable, then the naming mechanisms tend to break down. This can be combined with the first generalization in the argument – namely, 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. Where it takes on depth is that it takes us very quickly from data structures into constraints on the historiography emerging from the sciences of biodiversity. If certain kinds of entities and certain kinds of context are being excluded from entering into the databases we are creating, and those entities and contexts 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 structures. In other words, 'raw data' enfolds just as many layers of organizational and social decision making as the scientific texts that we more generally analyze.

As classification systems get enfolded into infrastructures and then picked up by people working in distant fields for their own ends, we run up quickly against the limits of rawness. In a well-defined field, the junior scientist will learn the canon of literature, and will be told (be it in lectures or in 'war stories' in the lab: see Orr, 1996) about where the uncertainties in a particular kind of classification or measurement lie. It is just not possible, using current educational and social practices, to propagate the informal knowledge of classifications which might permit their sensitive use in practice throughout the multiple disciplines which seek to draw on varied databases for theorizing biodiversity.

Things That Do Not Get Classified

We can develop this point further by looking at things which do not get classified. There are certain kinds of plants, animals and systems which are charismatic. These in turn create a set of others, entities which are often just as important, but which are overlooked because they do not lead to spectacular science or good funding opportunities.

The use of the word 'charismatic' is adopted immediately from the biodiversity literature – though the concept itself traces back to Weberian sociology (Thorpe & Shapin, 2000: 548-51, 576-80). Edwards and his colleagues (Edwards et al., 1995: 1-1) talk about the problem of getting too much information about 'charismatic megafauna'. Certain species are more likely to get attention from policymakers and the public than others – many more care about the fate of the cuddly panda, the fierce tiger or indeed the frequently drunk and scratchy koala bear, than about the fate of a given species of seaweed (or sea vegetable, to use the more recent, kinder, popular coinage). And this attention has very direct consequences. On the one hand, scientists are more likely to get funding for studying and working out ways of protecting these charismatic species, rather than others; and on the other hand, people are more likely to become scientists with a view to studying such entities – another feedback loop which skews our knowledge of the world.

Further, they are more likely to get funding for using exotic, expensive tools – a participant at a symposium in 1988 noted:

> Something called Stearn's Law has entered the literature, following a remark of mine at a conference in 1968. This states that at a given time the perceived taxonomic value of a character is directly related to the cost of the apparatus and the difficulty of using it. (Hawksworth, 1991: 176)

Readers of *Science in Action* (Latour, 1987) will find this no surprise. An unfortunate corollary which is central to current non-naming problems within biodiversity is that taxonomic and systematic work which, particularly in the morphological tradition, uses non-charismatic technology, has consistently lost out to more 'exciting' areas of research which do not try to provide consistent names – arguably that necessary prerequisite to good biodiversity policy. Rendering more stark this picture of things which do not get classified is the problem of the 'disappearing taxonomist'. Thus in the taxacom listserv, devoted to questions of systematics, a thread was opened in January 1999 around the question: 'besides the self-regret and crying, how to be a taxonomist in a primarily non-taxonomic world' (TAXACOM, 29 January 1999); Quentin Wheeler and Joel Cracraft complain that 'faculty positions in systematics have been replaced by non-taxonomic, often molecular, ones. . .' (Wheeler & Cracraft, 1997: 444); and, echoing remarks from the 1950s (Vernon, 1993), Charles Gunn and his colleagues bemoan 'the increasing migration of taxonomists away from the herbarium and into the laboratory' (Gunn et al., 1991: 19). This is paradoxical indeed; since, as Hawksworth and Bisby remark after deploring the shortage of taxonomists:

> The demand for systematics is high and increasing in conservation, environmental monitoring, agriculture and allied sciences, biotechnology, geological prospecting, and education and training. The requirements of consumers for stability in names, reference frameworks, relevant taxonomic concepts, user friendly products, and communication need to be addressed by systematists if they are to meet this demand. (Hawksworth & Bisby, 1988: 3)

So on the one hand we have increasing demands for stable classification systems so as to meet the data demands of a burgeoning set of scientists and entrepreneurs, and on the other we have taxonomy as low-status work not attractive to incoming students.[7] The general problem remains that across the sciences, the activity of naming is mundane and low status, even though it is an activity central to the development of good databases. Insofar as it remains – or becomes even more – so, there is little prospect of an abundance of taxonomists describing the plenum of nature irregardless of the relative status of problems within taxonomy. To the contrary, the skewing of our data collection and management efforts promises to continue unabated.

We are, then, dealing with a finite group of taxonomists striving to generate a world of data. Heywood characterizes those species that tend to attract most favoured entity status:

> . . . our knowledge of biodiversity in terms of named organisms is highly skewed in favor of certain groups such as birds, mammals and flowering

> plants, rather than invertebrates, fungi and microorganisms, except when the latter are of vital importance to humankind in terms of economics or health. (Heywood, 1997: 9)

So in general we tend to protect and study things which are about our size or larger, and things which are spectacular in one way or another. In the plant kingdom, vascular plants get the nod, as Cyrille de Klemm notes:

> An analysis of plant protection legislation in 29 European countries shows that in most cases the lists of protected taxa are relatively short, seldom exceeding one hundred entries, are almost always entirely composed of vascular plants and are largely dominated by spectacular species which are attractive to collectors and to the public. (de Klemm, 1990: 30)

Very small things often get the least attention – Rita Colwell (1997: 282) points out that there are more viruses than bacteria in the open ocean. These viruses do not receive scientific attention commensurate with their numbers – which leads to us building causal narratives about life and earth that features, as the heroic figures, entities of the same general size and lineage as humans.

As a codicil, though, it should be noted that complementary to the heroic is the strange. Judith Winston tells the sad story that she has found it much easier to get funding to study Antarctic bryozoans than those in local coastal waters off the North East United States. She says that this skewing continues to this day, such that:

> . . . many of the most urbanized coastal areas, the very areas in which human impact has been greatest, have never been surveyed. The practice continues today – if systematics has any glamour at all, it is only when carried out in the most exotic locale. (Winston, 1992: 157)

So what gets studied is the exotic other. And if the choice is between two others, then the more exotic will be chosen (cf. Boyle & Lenne, 1997: 33).

Terry Erwin, famous for his early estimates of the global numbers of species, notes a kind of feedback loop common to data collection efforts, one which places technoscientific practice at the operational centre (just as human action is at the narrative core) of our description of nature:

> The *described* species of beetles, about 400,000+ . . . , comprise about 25% of all described species on Earth. This dominance of beetle taxa . . . in the literature has resulted in Coleoptera being perceived as Earth's most speciose taxon. Thus, it has garnered further taxonomic attention from young taxonomists which in turn has resulted in more species of beetles being described than in other groups. Beetles are relatively easy to collect, prepare, and describe, significantly adding to their popularity. Such unevenness in taxonomic effort may or may not give us a false picture of true relative insect diversities. (Erwin, 1997: 29)

Erwin makes two points here: first that we tend to study what has already been studied, which is a version of the 'Matthew Effect'; and second that beetles get studied because they are easy to study.[8] Brenchley and Harper (1998: 318) provide a graphic illustration of the connection between apparent species richness and intensity of study of fossils from various epochs – showing that the best correlation for 'apparent species richness' of an epoch can be found by looking at the number of 'paleontological interest units' (the number of paleontologists studying that epoch).[9] It can of course be argued that there are more people studying the recent periods because there is a larger available area to study. However, in general, figures on apparent species richness which at face value encourage the narrative of simple-to-complex in the development of life reflect just as much the technoscientific fact of the number of paleontologists working on a time span. Once again, the point that I am making here is that the results that we write into (and hence frequently read out of) our federated databases are often integrally results about the world; projections of the current state of technoscience on to the world (see Latour, 1993); and mythic narratives of human heroism.

Things that people don't generally like or need get less attention – both in the real world and in naming. In botanical nomenclature, when plants are considered of particular value, then their names can be protected against the ravages of the inquiring taxonomist who finds a prior description of the species under another name. This is an information retrieval issue: protection is put in so that searches through the literature for such plants will not become unduly complicated. However, weeds – even when economically important – frequently do not get this protection, as Gunn and his colleagues complain:

> In cases where names of economically important species are involved, the Code provides for conservation in order to preserve current usage. Judging from the success of past species conservation or rejection proposals, it is not always clear as to what constitutes an economically important species. Some weeds of significant agricultural importance have not qualified for name conservation. With the expanding potential of gene transfer for crop improvement involving more distantly related taxa, more species will become useful to agriculture. Communication about such species depends on a stable nomenclature. . . (Gunn et al., 1991: 18)

Humans are the measure of all things in the information world that we create. When entities have the misfortune to be small and generally disliked, then they will certainly not get the attention that others do. Such is the fate of the parasite. Mike Claridge notes that the geographical isolation speciation concept basic to much evolutionary theory, does not apply in its normal form to parasites, since for parasites speciation can easily occur within a given geographical location by the choice of a different host; and that in general, taxonomic literature has been anthropocentric, using the story of human evolution as a 'reference point' (Claridge, 1988). The exclusion of the parasite has the kind of double effect noted above – database tools are developed which can very easily process data that fits the theory that speciation is due primarily to geographical isolation caused by tectonic and other events, rather than by competition – so you can swap techniques from one species to another. Conversely, little work is being done to store data that represents other forms of speciation. As the new databases consolidate, the parasite gets harder and harder to represent. At the other end of the production process, examples are chosen in textbooks which reinforce our world as one with the form of speciation particular to charismatic species.

Developing this latter point, Robert O'Hara, in a fine paper about narrative representation and the study of evolutionary history, notes that:

> Conclusions about evolutionary processes that are based on the structure of trees that have been, by selective simplification, brought into alignment with pre-existing nomenclature probably say less about evolution than they do about the narrative character of the preexisting nomenclature. (O'Hara, 1992: 153)

He points out subtle factors affecting the representation of cladograms (trees representing hypotheses about phylogenetic descent or patterns of character traits among groups of species – depending on one's theory), such as the fact that you tend to choose a representation which includes 'clades' (related groups of species) which have already been named, thus creating what he calls a 'grooving effect' (ibid.: 151).

I have defined these kinds of non-naming process elsewhere (Bowker, 1994b: 162-66) as a form of *convergence*. What this means, in this case, is that a set of data structures and information retrieval models are set up so that a particular, skewed view of the world can be easily represented. With these structures and models in place, it is easier to get funding and support for research which reproduces this view – your work will be understood more easily, you can make good use of material from cognate areas, and so forth. Thus the world that is explored scientifically becomes more and more closely tied to the world that can be represented by one's theories and in one's databases: and this world is ever more readily recognized as 'the real world' – especially when measures can be taken to save only entities which have been named and studied (here again we see the performative aspect of the infrastructure). Taking together the things that are hard to classify (from the last section) and do not get classified (from this one), we see the first threads of a pattern to the convergence that I have described.

In general, what is not classified gets rendered invisible. Years ago, Jacques Derrida (1980) wrote of the importance of looking for the 'other' categories – things that are bundled out of a given discourse as not significant. He argued that these 'others' (what is excluded) often indicate the very problem that a totalizing philosophical discourse cannot deal with. Thus he wrote of the difficulty of speech-act theorists of dealing with the category of humour. The apparently heterogeneous set of others that we have uncovered in this section tells a consistent story. The negative telling is that things that do not get classified are not considered of economic, aesthetic or philosophical importance – for example, weeds not affecting crops,[10] and non-charismatic species. The positive telling is that our databases provide a very good representation of our political economy, broadly conceived: that which we can use through our current modes of interaction with nature and other cultures is well mirrored in our data structures. What gets excluded as the 'other' is anything which does not support those modes of interaction. Put this way – doubtless far too broadly – we see that there may well be hidden systems to the exclusions. There are regularities to the ways of knowing and being that we fail to give a name to in the throes of our current 'archive fever' (Derrida, 1995). It is clear, in general, that as we create worlds of electronic information which reflect our political economy in all its contradictions, it should be no surprise if the policies that get read out of these worlds should help us shape the world in the image of that political economy – again in all its contradictions.

Things That Get Classified in Multiple Ways

A banner headline in *The Independent* newspaper (London, 23 November 1998) reads: 'Scientists Reclassify All Plants' (see APG, 1998, for the publication which led to this item). The headline is precisely correct, but it can conjure up the highly inaccurate image of scientists wading through collections in pressbooks all over the worlds and relabelling specimens. In theory, there is now a new classification of the flowering plants; in practice, reclassification is a long, slow process, and there is no simple path from the molecular sequencing techniques referred to in the body of the paper to the changing of plant classifications in national and local legislatures, in nurseries throughout the world, and so forth. Even – perhaps especially – in the world of electronic databases we are moving into, there is no touch of a button that would allow us to bring in any new system. To the contrary, when a given database of plants, of the ecology of a given area, of paleontology (and the like) is designed, it necessarily draws on a contemporary classification – and will be very rarely updated (and very difficult to update) should the classification change. Very few disciplines or agencies are sufficiently motivated to keep up with the latest classification schemes of flora or fauna in their publications or records: it is an enormously expensive and time-consuming procedure to do so. Changes in name introduced as taxonomic theory develops can have large-scale economic consequences: it could cost tens of millions of dollars to relabel packets of tomato seeds, revisit regulations, and so forth; one commentator noted that 'single name changes can cost the horticultural trade millions of dollars, and . . . nurserymen would go out of business if they took the matter seriously' (Brandenburg, 1991: 30). Even where a single set of names is adopted, there are problems of synonymy: the venerable *Index Kewensis* has been estimated to offer 950,000 names for 250,000 flowering plant species (Lucas, 1993: 11; cf. Hawksworth & Mibey, 1997: 57, on synonymy among fungi). The result is a tower of Babel, where numerous outdated classifications present themselves to the scientific researcher with equal force: a choice must be made if the associated data are to have any value. This may seem to be a contingent point – perhaps, some day, the various user communities will catch up with the latest schemata, particularly if computer scientists can use the latest database technology to make this easier. But the point is that this is never the case – it is the same problem as that of third-world countries catching up with first-world technology; or of poor community groups bridging the great divide.

The move to register all names, to agree on model data structures and formats for botanical databases so as to facilitate biodiversity management (Heywood, 1997: 12) is just as urgent and just as overly optimistic as the calls of Alphonse de Candolle (1867) in the 19th century for a rational system of nomenclature. Such 'incomplete utopian projects' (Gregory, 1999) are so pervasive in the history of naming and record-keeping that they should be regarded as standard rather than abnormal. Currently, as with all such projects, there is a continuation of the proliferation of names for plants in different organizational contexts:

> The net result of instability in names has been an increasing tendency for consumer organizations to issue their own standard lists of names fixing species names, at least for set periods of time. Examples are the CITES convention concerning trade in furs and plants. . ., the International Seed Testing Association, and FAO/WHO lists of organisms of quarantine importance. (Hawksworth & Mibey, 1997: 19) [11]

The so-called 'bleeding edge' (one step beyond the cutting edge and one step behind vapourware),[12] always offers integration, democracy and apple pie: that is one of its features. Within information science, so much theory is written as if the little practical details were irrelevant, if not now, then soon; one of the chief arguments of this paper is that they will always be analytically and theoretically pertinent. It is in the empirical regularities of these details that we can trace the interweaving of knowledge and power.

What's in a Name?

We have in this section on naming looked in turn at things that are hard to classify, things that do not get classified, and things that are classified in multiple ways. Within each category, we have seen organizational and political concerns rubbing up against scientific and philosophical questions. It is the database managers who do most of the articulation work between the research, organizational and political spheres. They are the people who have to construct a system which works well enough, given the current state of the art and technology of naming and storing data. Yet in the field of science studies we have in general focused attention on what scientists do with data, rather than on the mode of data production and storage. The names we choose for our biodiverse world tell organizational stories and contain within them the seeds of historical narratives: it is within the database that the nature/society hybrid so well described by Latour (1993) and Haraway (1997) is born.

Retaining Context

The Problem of Context

The observation that we are dealing both by necessity and by choice with data stores as fundamental to biodiversity science does not of course mean that theory drops out of the equation. Indeed, as shown in my earlier section on classifying, classification systems are theoretically shaped in their inception (by reflecting a set of theoretical beliefs) and shape theory in their deployment in databases (by making it easy to follow certain theoretical paths and hard-to-impossible to follow others). As the layering of the classification systems into articulated frameworks for particular disciplines/subdisciplines becomes more concrete (ever more inscribed into infrastructure), so it becomes more abstract and general – the ideological commitments which survive the layering process are those with the widest application.[13]

What we are frequently left with is a set of legacy data housed in legacy systems – for example, botanical information stored in an hierarchical database and using an outmoded classification scheme. Clearly it is often just not worth the effort of first massaging the data into a more contemporary classification and then (as database people say) 'migrating it across' into a new format. This is particularly true since, as for organisms, so for data: the vast swathes of data held in mundane format on old systems receive neither the theoretical attention nor the funding of esoteric data sets held in expensive new equipment. Thus, for example, Judy Weedman (1998) has written of the problem of getting computer scientists interested in solving the kind of problems that environmentalists handle: these problems are just not seen as theoretically interesting, and so will not advance one's career. Therefore, as in many other fields, the work of the computer scientist is often only of peripheral interest to the domain scientist; and few domain scientists can take the career risk of moving into computer science in order to address their own issues.

In this section, I will try to lay out some basic ways of understanding legacy data. I will treat the issues arising by looking in turn at the Scylla and the Charybdis of long-term data management: keeping too much of the past, and not keeping enough of the past. These two continually chase each other's tails. Both resolve practically into the single observation that if a legacy data store does not retain its own context as a formally separable set of entries, then it is useless. More deeply, they raise theoretical questions of just what do we know about the past, and how do we know it.

Not Keeping Enough of the Past

Carl Bowser (1986) has published a fine description of the problems of dealing with old data – from what is now an NSF-funded Long Term Ecological Research (LTER) centre in Wisconsin.[14] In 1898, Edward A. Birge, Chancey Juday and their associates established a Limnology group at the University of Wisconsin at Madison. Subsequently, research at Trout Lake Station was initiated in 1925. This is a region with one of the densest concentrations of lakes in the world, and with a number of accessible bogs and wetlands. Bowser relates that there were seven decades of data available to modern researchers – though most was concentrated in the period 1925 to 1941. He describes four major problems with the older data. First was the issue of sampling the lake water. Although it was recognized as early as 1924 that there were seasonal variations in acidity and other key measurements, the lake waters were only sampled once per year, in the summer. Thus, to make meaningful comparisons, modern studies had to use only July-through-September data. Second, there were terminological changes: what had been a distinction between 'bound' and 'free' CO_2 became a distinction between 'alkalinity' and 'acidity'. This was uncovered through a literature search of the old data. Third, a more difficult problem was that the pH measures would be different depending on whether they were taken in the laboratory on return from the field trip, or in the field – loss of CO_2 in samples over a few hours changes the measurements. This information was nowhere mentioned in the published reports, but fortunately Bowser and his colleagues were able to locate a retired limnologist who remembered the procedure. Fourth, they found that there was a shift in the kind of measuring techniques used for alkalinity – from methyl orange to electrometric methods – which caused a break in values which could be compensated for, once the change was recognized.

These four kinds of problem are archetypal. At the time of data collection, one might:

- change measurement techniques without proper records (new alkalinity measures);
- use current terms localized in place or time to describe the data ('bound' and 'free');
- not record key data which is seen as part of normal behaviour in a given community of practice (measurement time); and
- not realize that additional data is needed (measurement date in our case).

The first of these is trivial on the surface – though dealing with it is extremely difficult. It represents part of the normal work of producing metadata standards: one tries to script standards in such a way that changes are automatically recorded. There is the standard practical problem here (cf. Bowker & Star, 1999: 65; Fagot-Largeault, 1989: Chapter 3, for the medical analogue) that the person filling in the forms is keen to be doing other things than ensuring the perfection of the record – since he or she knows about the new techniques and will not be led into making any mistakes with the dataset, and has better things to do than through an altruistic gesture make things easier for notional future users.[15] The second and third problems are more difficult and deeper. They require an act of imagination on the part of the record-keeper to place themselves in the position of any possible future reader. They might assume that the average user of the data will know such and such a term through their formal education or through apprenticeship – but as we shall see, with biodiversity and other forms of interdisciplinary data, much less can be assumed than one might suppose. In essence, the record-keeper is being asked to abstract the record set from the historical flow of time – to provide enough information so that a limnologist from Mars (who presumably has been out of work for several million years) can come along and from the dataset and a sufficient command of English interpret the data. Measurements taken as recently as in the 19th century, using nationally accepted standards, cause problems today, as a query to the TAXACOM listserv (22 March 1999) indicates:

> Are conversion tables from fathoms to meters (or linear nonmetric to metric) for different countries available anywhere on the WWW? For example, a Danish fathom is 1.883 m whereas the Imperial (British or American) fathom is only 1.829 m, but if there are web sites with this kind of information, I am asking the wrong questions as I have not been able to find them. Are there similar differences for other European countries? For example, what system does France follow? Germany? – Would this be the Bavarian system? At the smaller end of the scale, a Danish inch is (still) 26.112 mm and a Danish line (1/12 of a Danish inch) is 2.176 mm, whereas an Imperial inch is 25.4 mm and the line is 2.116 mm. These nonmetric measurements, especially fathoms ('orgyiar' [*sic*] in the Latin of Malmgren, 1867) and lines, occur in many older descriptions and station lists from the 1800s.

This problem is trivially easy to solve in theory; but very difficult to solve in practice when multiplied over a vast number of data sources. Of course, complete transparency of old data is not possible. Indeed, what is being demanded of the dataset is precisely something which over 20 years' research in science studies has shown cannot be asked of the scientific paper – to stand outside of time. The fourth problem is positively brutal – the measurements that are made now are necessarily constrained by current theory, and there is no way of making allowances for future possible theories.

Bowser (1986: 174) writes that overall: 'The availability of "data" has generally not been the problem with historic data, it is the unequivocal documentation of techniques that has concerned us most'. What is needed is a record of processes as well as a record of facts. However, in principle, processes and facts cannot be disentangled, so we are never going to have a perfect dataset wrapped in complete metadata. Moreover, the processes that we need to record so as to ensure the viability of data in the long term do not constitute an easily enumerable set: they include information about how classification systems are arrived at; what the local coding culture is; what techniques were used, and so forth – social, organizational and technical processes all make a difference. The difficulty of so doing is rendered greater by the non-discussion within the scientific literature of coding culture or of organizational processes: this is, as we shall see in the conclusion, an area where science studies can make a useful contribution to database design.

Time and Space as Charged Containers

Any data-coding scheme bears traces of its own past, frequently buried deeply enough that it might not be apparent for a contemporary user – particularly from a cognate discipline and untrained in the vagaries of that scheme. An archeological effort is needed in order to uncover deep-rooted biases. A classic example here is Stuart Max Walters' study (1986) of the European bias in angiosperm classification. Walters demonstrated that Linnaeus used the folk classifications available to him at the time, and that these classifications in turn favoured describing more genera in a family for economically important plants such as *Umbelliferae* (the carrot family) than for chickweed. Inversely, there are many more species per genus in the economically important grass genera (*Graminae*) than in sedge (*Cyperaceae*). He notes that a New Zealander starting *ex nihilo* would order the same set of species quite differently. Bowker and Star (1999: 114) have noted this European bias in the International Classification of Diseases (ICD) – a bias originating in the Parisian origins of the current classification and which, through the history of the deployment of the ICD, has led to a number of complaints from tropical countries, feeling that their disease entities were under-represented. Walters goes on to criticize evolutionary hypotheses that have been made based on number of species in a genus and number of genera (maintaining that the oldest genera have the most divergence) as being an artefact of the original folk classifications (Walters, 1986: 538). This example of traffic between data structure and narrative is not a pathological case to be rooted out. To the contrary, it is very much business as usual in our classification schemes: one of the true tasks for metadata development is how to represent such archeological richness in our data coding.

Models using any given dataset frequently contain hidden traces of their own past. This issue has been explored by Paul Edwards (forthcoming) in a study of circulation models used in climate research: he has shown inheritance from the 1960s to the present of tenuous assumptions in some models. Models developed today can suffer from the same kind of 'presentism' that plagues much historical writing. Such data or modeling problems are frequently well known within any one given discipline: there has generally been some kind of apprenticeship process whereby the new botanist learns the history of Linnaean classifications, or the climate modeler learns of the equable climate paradox (Wing, 1997).[16] However, things become more complex in a field like biodiversity science, where many others might be using one's data or drawing conclusions from one's models. Any given complex model of a given biogeographic region at some previous geological epoch might be drawing on a large number of models and datasets, each with their own relatively recondite legacies.

Consider, for example, the various bases that are available for providing temporal scales for the events on the earth. A number of different 'clocks' have been suggested over the past 200 years. One person – for example Bishop Ussher – might count the number of generations reckoned in the bible, and multiply them by average generation span to date time since the creation of the earth. Charles Lyell might reckon events from sedimentary evidence – using this to put lower limits on the age of the earth that far exceeded the length of time indicated by the generational clock. Then Lord Kelvin might use the pearl of all the sciences – physics – to show that, on the assumption of a molten earth at origin and the current indirectly measurable inner heat of the earth, one only had about 40,000 years to date events in – not the 400,000 or 400 million indicated by geology (Burchfield, 1990). Which brings us into this century, when there have been a series of other dating conflicts. One which has caused a lot of controversy in the past ten years has been using the rate of mutation of genomes to give a 'clock' for species development. Following the theory that mitochondrial DNA was passed down only from the mother (recently shown to be only approximately correct, as many such dogmas are), and that it mutated at a constant rate (recently shown to be probably incorrect; and certainly requiring ranking of different parts of mitochondrial DNA as better or worse timekeepers), one could argue for a 'mitochondrial Eve' for various species – producing a clock which throws key evolutionary events much further back in time than would be indicated by paleontological evidence in the form of fossil remains (Strauss, 1999).

Two observations emerge from the difficulties of reconciling clocks over different scientific disciplines over time. The first is that there is a strong power component in the debates – she who controls the clocks controls knowledge. Lyellian geology was developed very directly in opposition to literal creationist narratives, so the struggle for the clock was simultaneously an assertion of the rights of the scientist to speak to questions of the history of the earth. Kelvin's attack on the Lyellian clock was equally an assertion of the power of physics as a deductive science over geology (Burchfield, 1990). More recent debates have been equally about whether to prefer molecular biological data over field data: for example, Ernst Mayr (1988: 9) decries the 'unpleasantly arrogant manner' of some molecular biologists talking to morphologists.

The second observation is that if you want to create a biodiversity database using a commonly accepted standard timeline then you are going to have serious problems. In some fields there are standards-setting bodies at work which relatively regularly produce consensus clocks. Within geology, one might decide to use the Geological Timescale produced by the Geological Society of America in 1989, or the timescale more recently produced by Felix Gradstein and James Ogg (1996) – or a mixture, as in a UNESCO project where contributors of paleotectonic maps were directed to use a timescale as follows:

> . . . early Permian to Holocene ages are from Gradstein and Ogg, 1996, A Phanerozoic Time Scale: *Episodes*, 19:3-5. Late Carboniferous ages are from Harland et al., 1990, *A geologic time scale*: Cambridge University Press.[17]

The reason for choosing a mixture is that one timescale might better reflect a given set of measurements than another (keeping an extinction event within a certain epoch, for example, rather than having it straddle two). This is not just a problem for reconciling clocks – it occurs at each level of data integration across disciplines and across agencies.

With much biodiversity data, there is no point at which you can say that such and such is a bedrock standard: it's triangulation all the way down. A given geological timescale might make sense because of a certain reading of the fossil record, itself coordinating between various kinds of surrogate measures, taphonomic theories and current understandings of paleoclimates. . . . In a wonderfully named article, 'Absolute Ages Aren't Exactly', Paul Renne and his colleagues point out that there are a number of decay constants in play for leading indicators of absolute age, different disciplines producing their own incompatible absolute scales:

> The decay constants used in the nuclear physics and chemistry literature are based on counting experiments of α, ß, or χ-radiation activity, whereas the values used by geochronologists also include, in some cases, the results of geologic intercalibrations or laboratory accumulation experiments. For ^{40}K, the two communities' values for the total decay constant differ by 2.1% simply because of different choices in filtering the same set of activity data. For ^{87}Rb, the value used by geochronologists, based on laboratory in-growth experiments, differs by nearly 3% from the value used in the nuclear physics and chemistry literature, a difference well beyond stated errors of the two values. (Renne et al., 1998: 1840)

They note that some key standards – such as of ^{40}Ar/^{39}Ar decay – have never received international review, and comment that where there have been reviews of other standards, there has been poor interdisciplinary communication.

Equally, there is no clear spatial or temporal nesting. We have seen that in the construction of biodiversity-relevant databases, there are two generally competing models: the ecological and the systematics; and that the ecological model is spatially organized, and tries to break the world up into chunks (or partial objects, since they are only constituted as objects within a given discourse) which are appropriate units of biological activity. The appropriate temporality for any of these partial objects in space is singular. The systematics model is temporally organized, and tries to represent the complete history of life on this planet – from the ticking of the genetic clock in mitochondrial DNA up to the enfolding of the history of all living beings in a coherent cladistic classification. Temporally, one great divide is between contingent historical time lines (stressing catastrophes like meteor showers, floods and the like: see Ager, 1993) and cyclical times (beloved of climatologists: see Lamb, 1995). Objects in either of these temporalities can be unsettled by objects in the other. Thus there are cyclical stories about meteors (the movement of the solar system taking us into and out of dense patches on a cyclical basis) and secular stories about climate change (regular decrease in carbon dioxide in the atmosphere over time up to the present). Another temporal great divide, with equally complex sets of objects, is between stasis or equilibrium (the eternal present) and progressive change. A third temporal divide is between viewing human development as part of a natural series, or of a *sui generis* social series. Thus paleontologist Elisabeth Vrba ties the development of a large brain in humans to heterochrony (the speeding up and slowing down of growth phases in the embryo) itself embedded in a pulse of global cooling – large juvenilized bodies being associated with cold temperatures and consequent vegetatively open environments (Vrba, 1994: 354). Anthropologists, on the contrary, argue about whether carnivory or eating tubers is at the base (Pennisi, 1999). The natural scientist embeds a new temporality of development in a natural cycle (the ice age); the anthropologist delineates an archetypically social order (the cooked *vs* the raw). It is an open question whether the two sets of registers and associated temporalities are contradictory: within biodiversity data there are many partially nested sets of such open questions.

As far as databases go, the point is that there is no uniform way of separating off the data objects (which themselves enfold complex histories) from their spatial and temporal packaging, and inserting them in some other panoptical Cartesian space and time. As you nest cycles one inside the other, you find secular change irrupting into the story; as you nest secular narratives, cycles emerge. This is a problem across the divide. Furthermore, even within each divide, there is no simple nesting. The work of flattening out all the narrative sciences into a single narrative timeline is a *productive* effort that articulates data formats with relative power relationships between disciplines – through the mediation of classification systems and data standards.

Similarly 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 & Harper, 1998: 273) shows the world split up into several regions (nearctic, neotropical, Ethiopian, oriental, palaearctic and Australasian) based on continental masses and plate tectonics. A panbiogeographic model (Cox, 1998) comes out of a team based in New Zealand, an island deep in the Pacific – and emphasizes the rôle of islands (by way, on recent theory, of terranes – pieces of continental plates that travel relatively independently, such as Point Reyes in California). The key blocks are the Atlantic, Pacific, Indian, Arctic and Southern oceans. The locus of object/space/time production is not an accidental feature here (cf. Bowker, 1994b: Chapter 2, on geophysical theory and site). 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 (1997: 9) 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.

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 central fact about biodiversity science that I have already touched on: it is the science of the radically singular. 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. The fossil record recapitulates these problems. Carl Koch (1998) 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! Similarly within the USA, Koch cites Norman F. 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' (Koch, 1998: 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 (ibid.: 199-200); Terry Erwin says the same for the Paleozoic and the Mesozoic.

One can picture an ideal data-storage system which would be aware of all the standards that had gone into the naming of its categories and the partitioning of its intervals. In this system, when the American Geological Society changed its timescale, paleoecological and paleoclimatological databases could easily be adjusted to fit the new standard. What's wrong with this picture? First of all, standards get deeply entrenched into infrastructures so that, as with the climate models referred to above, it can be very unclear to current users that such and such a standard has been implemented. Second, even if the paleoecologist hears about, say, the new timescale from the mitochondrial clock, he might not be at all willing to change his database to fit in with the new absolute timescale because such a change would throw off a series of other correlations (which may or may not be based in turn on their own sets of entrenched data). Finally, it is not only impossible to separate off data structures from the internal theory of the discipline(s) whose data is being stored: it would not be a good idea anyway. For as we have seen with the case of the warring chronologies, there is a lot of good content and, at the same time, a lot of good political work being done at the level of deciding database issues.

The caution here is that with the process of the mutual imbrication of data standards, each of which covers its own context (its own past) in archeological layers or through radical renaming (see Bowker, 1997), there are certain secular processes of information convergence which entail that some kinds of narrative get progressively excluded from the picture in the same process which creates ever more robust (though always necessarily incomplete) mutual bootstrapping of the rest. If we want to create data structures that are more open – though none will ever be completely so – then we need to do two things. The first is a very practical step: we need to retain the context of development of a given database in reasonable detail; the political and social and scientific contexts of a set of names and data structures are all of interest. I emphasize 'reasonable detail' here: a perfect archival system is a chimera. And the second step is to admit that the goal of metadata standards should not be to produce a convergent unity. We need to open a discourse – where there is no effective discourse now – about the varying temporalities, spatialities and materialities that we might represent in our databases, with a view to designing for maximum flexibility and allowing, as possible, for an emergent polyphony and polychrony. Raw data is both an oxymoron and a bad idea; to the contrary, data should be cooked with care.

Information Integration

Biodiversity and Ecology vs *Biodiversity and Systematics*

In a number of sciences over the past few hundred years, there has been a move to analyze basic scientific units in terms of information storage and transmission (Bowker, 1994a; Keller, 1995) – be they 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 quotient. 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) – 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 several rat species (providing one remains!), which contain a set of overlapping genes. Formally, this means taking diversity as 'a measure of information in a hierarchical classification' (ibid.: 237) – 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 George 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*, begun in 1968, is the year 2397 – this is not commensurate with the rate at which the environment is changing (Heywood, 1995).[18] It is hard to estimate species loss, since in any case most species are only known through a single example – their holotype. 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. Niles Eldredge argues that this is in contrast 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, 1992b: 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', whereas the latter are extended in space, their temporal hallmark being 'moment-by-moment interactions' (ibid.: 5). 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, John Grehan makes the assertion – challenged by Cox (1998) – that there are 'dichotomies of ecology v. history, and dispersal v. vicariance [the separation or division of a group of organisms by a geographic barrier]' (Grehan, 1994: 461).

If we accept Grehan's pitching of ecology against history, then the complexity of integrating biodiversity data across multiple disciplines is increased since it leads to a rift in data collection efforts which merits deep consideration. 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. Wheeler and Cracraft (1997: 439) 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.

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 levels of scale – plots of $\leq 1\text{m}^2$ 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 flavours, with different scientific approaches 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.[19] At the top level, there is a question of equity here. 'Under-developing countries', to use John Berger's phrase (Berger & Mohr, 1975: 233ff.), are becoming reluctant to share information and specimens, for good economic reasons. Thus Richard Roblin (1997: 472) 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.

There have been innumerable cases in the past decade of access closing to information which was once seen as the prerogative of the scientific or imperial élite.

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 Darrell Posey (1997: 86) writes, intellectual property rights:

> . . . 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 is reserved . . . [T]hey [the laws] cannot protect information that does not result from a specific historic act of 'discovery'.
> Indigenous knowledge is transgenerational and community shared. Knowledge may come from ancestor spirits, vision quests, or orally transmitted lineage groups. . .

Helen Watson-Verran (for example, in Watson-Verran & Turnbull, 1995) has discussed similar issues arising from different ontologies of ownership between white Australians and aborigines. Corinne Hayden (1998) discusses the difficulty in biodiversity-prospecting 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 prior to each particular determination; in practice, this just does not happen, and the requirement to respect intellectual property rights is honoured by formally designating the trader as the fount of knowledge. As Posey, Watson-Verran and Hayden all 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 flow of raw data out of the Third World into Western databanks – where it is converted (at times; rarely betimes) into economically valuable information and knowledge, and then sold back. 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 and his colleagues, 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 locating 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 Nigel 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 molecular biology. Similar errors 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, however, indicate the difficulties of using data in multiple ways. Relatedly, Cyrille de 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 wilful destruction by landowners fearing restrictions to development. On the other hand publicizing the location avoids inadvertent destruction in good faith. (de Klemm, 1990: 28)

Thus the representation (the map of biodiversity) might well affect the territory.

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 knowledge production and planetary management. 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. 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: Science Studies and Biodiversity Databases

I have elaborated – through an analysis in turn of naming practices, context description and information integration – three major broad dimensions of metadata information: how objects are named (and what is not named); how much information is given about data collection procedures and initial data use; and who are the intended (and unintended) users of the databases. 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 histories folded (Deleuze, 1996) into them, histories which must be understood if the data are to persist.

To summarize. Each particular discipline associated with biodiversity has its own incompletely articulated series of objects. 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. If one wants to produce a consistent, long-term database of biodiversity-relevant information the world over, all this sounds like an unholy mess. At the very least it suggests that global panopticons are not the way to go in biodiversity data.

In my introduction, I spoke about the problem of irreversibility and infrastructure. We see this irreversibility starkly in the field of biodiversity research. Picture a powerful biodiversity database that enables policy-makers with limited resources to save all and only those species which it describes. Two points emerge. First is that the database itself will ultimately shape the world in its image: it will be *performative*. If we are only saving what we are counting, and if our counts are skewed in many different ways, then we are creating a new world in which those counts become more and more normalized. The second point is that once this effort has been made, there is (at present) no possible reverse engineering – recreating a lost species. Now clearly there is not one single such database, nor is there such a narrowly-constrained policy that it can save all and only – nature and science are both far too messy. However, there are, as I have shown, some emergent regularities in the sets of databases we are creating – in terms of naming (atomistic, human-centred, economically charged) and in terms of spatiotemporal units (the struggle between the disciplines writ large on to the world of nature).

I have evoked a strategy for rendering the irreversibility that we choose as robust and useful as possible through deep historicization of our datasets. We cannot retain everything about a set of data (this would be bureaucracy gone wild), but it is practical to suggest ways of categorizing and formalizing historical perceptions of data (ways with, of course, their own reflexive problems) which would allow us to retain the right sort of information about the past, at the same time as we create the future world and its knowledge. For this to be done, it is vital to dissolve the current disjunct between database (as technical storage medium) and policy (as way of acting in the world). The production of the database is productive of the new world we are creating.

The political possibility of an international consensus on the definition of biodiversity and the organization of a unified data-collection effort is slight (witness the problems this last century with attempts to create single universal classifications of diseases or job types). Even if it succeeds, there will still be coding cultures specific to given locations and particular disciplines. 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, the failure deeply reflects the nature of disciplinary research in the sciences related to biodiversity. These sciences deal in objects, spaces and times that cannot be readily normalized one against the other. It may be theoretically possible to produce political agreements that would create a single integrated database, but no field at all has been able to make those agreements – the field of medicine, for example, has been attempting to produce universal classifications for over a hundred years without success (Bowker & Star, 1999). We should begin to look at the machines that produce local orderings and alignments of datasets (oligopticons).

I have claimed that these sets of local orderings, in turn, produce irreversibility within our regime of knowledge/power. We have seen this at the level of discipline, where more powerful disciplines get to define the space, time and ontology of objects in the databases – and so to facilitate their own agendas preferentially over other disciplines, which have to make do as best they can with ill-fitting data objects. We have also seen this at the level of producing the worlds itself – since kinds of entities, temporalities and spatialities which are excluded from the databases are by extension excluded from management policies. They may proliferate just fine (the databases are not simply deterministic), but they may also fall between cracks (or between the lines of code) in the biodiversity model being used to develop policy for a given region.

I have not, in general, questioned the mania to name which is rife in the circles whose work I have described. There is no absolutely compelling connection between the observation that many of the world's species are dying and the attempt to catalogue the world before they do. If your house is on fire, you do not necessarily stop to inventory the contents before diving out the window. However, as Jack Goody (1977) and others have observed, list-keeping is at the heart of our body politic. It is also, by extension, at the heart of our scientific strategies. Right or wrong, it is what we do.

We have seen throughout this paper 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., 1997, 1998; Dempsey & 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 communication patterns both between various scientific disciplines and between those disciplines and legal and political bodies. If we are going to develop decent biodiversity policies then we need databases held together through good metadata practice. The field of science studies, which has developed very rich languages for describing ontological diversity, can make a significant contribution to this important work. In a biodiverse world we need to be able to manipulate ontologically diverse data.

Notes

Thanks to Leigh Star, Francis Harvey, Nick Chrisman, Fernando Elichirigoity and David Stockwell for their insights. I am most grateful to Mike Lynch, Sergio Sismondo and David Edge for their insight and encouragement, and to the anonymous reviewers for their comments. This paper was made possible by an NSF Professional Development Grant.

1. In Richards' words (1996: 152): 'The museum is no longer the privileged archive of culture; the archive, the sum total of what can or cannot be said or done, has become the very form of the modern state'.
2. Cf. Veyne (1971) on irreversibility and specificity in history.
3. For example, their neologism 'naugetuck' – 'A plastic packet containing shampoo, mustard etc., which is impossible to open except by biting off the corners' (Adams & Lloyd, 1990: 72).
4. Compare Berg & Bowker (1997) on the need to discipline local users of a universal tool.
5. We shall return to this below (665-67).
6. Cf. Rudwick (1985) on the Devonian controversy in geology.
7. I will use the words 'taxonomy' and 'systematics' in accord with the following definition of the products of taxonomy as being:

   > [A] taxonomic information system comprising classification, nomenclature, descriptions, and identification aids. Systematics is then taxonomy *plus* the biological interrelations – breeding systems and genetics, phylogeny and evolutionary processes, biogeography, and synecology (participation in communities). (Hawksworth & Bisby, 1988: 9)

   On the absolute decline in taxonomy in recent years in undergraduate courses, and on the ageing workforce in taxonomy, see Gaston & May (1992). There have been attempts to meet this problem in the past ten years – such as the NSF PEET (Partnership for Expanding Expertise in Taxonomy) and discussion of the use of parataxonomists in biodiversity data collection, technically assisted with the development of interactive keys for plant or animal identification.
8. Cf Kohler (1994) on the use of drosophila; and the reasons for the adoption of *c elegans* as a model organism (Star & Ruhleder, 1996).
9. Compare Henri Dumont on rotifers: 'About 2000 "species" (by which I mean taxonomic bonomens or trinomens) of Rotifers are now known throughout the world; ca. 1350 of these occur in Europe … . This is no coincidence, and reflects the distribution of rotiferologists (in Scandinavia, Germany and Great Britain) rather than that of Rotifers' (Dumont, 1983: 20).

10. Clearly, weeds that affect crops are of economic importance and so are intensively studied.
11. Compare the different fossil-naming schemes attached to petroleum discovery, as discussed above.
12. One can trace a genealogy of many computer programs that go from 'vapourware' (great ideas used for getting funding) to 'ghostware' (programs or applications, such as Memex, that have a great influence in the literature but never existed), without passing through the 'software' phase of the life cycle.
13. This coupling of the highly concrete and the abstract can of course be found in Marx's writing (though without this particular twist); it is also explored by Michel Serres (1993) in his wonderful exploration of the origins of geometry, and elsewhere. It can also be found – though without the references to the layering of infrastructure (so that it sounds like a magic trick) – in the work of Alfred Sohn-Rethel (1975).
14. There are currently 21 such centres, devoted to building up long-term datasets for environmental studies – see *http://lternet.edu/network/sites/index.html*. Paradoxically, each of the centres is on short-term renewable funding.
15. Mike Twidale (personal communication; *http://www.lis.uiuc.edu/~twidale*) is developing the concept of designing for altruism to respond to this problem.
16. Current models of past climates tend to suggest that they were closer to present conditions than the paleontological evidence implies.
17. *http://www.geomin.unibo.it/orgv/igcp/maps.htm*
18. For a discussion of the race between extinction rates and cataloguing rates, see Stork (1997: 45).
19. World Conservation Monitoring Centre, Biodiversity Data Sourcebook, Figures 2 and 8: *ftp://ftp.wcmc.org.uk/products/wcmc.publications/1.sourcebook*

References

Abbate (1999) Janet Abbate, *Inventing the Internet* (Cambridge, MA: MIT Press).

Adams & Lloyd (1990) Douglas Adams and John Lloyd, *The Deeper Meaning of Liff* (New York: Harmony Books).

Ager (1993) Derek Victor Ager, *The New Catastrophism : The Importance of the Rare Event in Geological History* (Cambridge & New York: Cambridge University Press).

APG (1998) The Angiosperm Phylogeny Group [APG], 'An Ordinal Classification for the Families of Flowering Plants', *Annals of the Missouri Botanical Garden* 85/4: 531–53.

Barrowclough (1992) George F. Barrowclough, 'Systematics, Biodiversity and Conservation Biology', in Eldredge (1992a): 121–43.

Berg & Bowker (1997) Marc Berg and Geoffrey C. Bowker, 'The Multiple Bodies of the Medical Record – Towards a Sociology of an Artefact', *The Sociological Quarterly* 38: 513–37.

Berger & Mohr (1975) John Berger and Jean Mohr, *A Seventh Man : A Book of Images and Words about the Experience of Migrant Workers in Europe* (Harmondsworth, Middx & Baltimore, MD: Penguin).

Boulter et al. (1991) Michael C. Boulter, William G. Chaloner and Phil L. Holmes, 'The IOP Plant Fossil Record: Are Fossil Plants a Special Case?', in Hawksworth (1991): 231–42.

Bowker (1994a) Geoffrey C. Bowker, 'Information Mythology: The World of/as Information', in Lisa Bud-Frierman (ed.), *Information Acumen: The Understanding and Use of Knowledge in Modern Business* (London: Routledge): 231–47.

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

Bowker (1997) Geoffrey C. Bowker, 'Lest We Remember: Organizational Forgetting and the Production of Knowledge', *Accounting, Management and Information Technology* 7/3 (July–September): 113–38.

Bowker (1998) Geoffrey C. Bowker, 'Archival Technology in the Historical Sciences 1800–1997', *History of Technology* 15: 69–87.

Bowker & Star (1999) Geoffrey C. Bowker and Susan Leigh Star, *Sorting Things Out: Classification and its Consequences* (Cambridge, MA: MIT Press).

Bowser (1986) Carl J. Bowser, 'Historic Data Sets: Lessons from the Past, Lessons from the Future', in William K. Michener (ed.), *Research Data Management in the Ecological Sciences* (Columbia: University of South Carolina Press): 155–79.

Boyle & Lenne (1997) Tim J.B. Boyle and Jill M. Lenne, 'Defining and Meeting Needs for Information: Agriculture and Forestry Perspective', in Hawksworth et al. (1997): 33–54.

Brandenburg (1991) Willem A. Brandenburg, 'The Need for Stabilized Plant Names in Agriculture and Horticulture', in Hawksworth (1991): 23–31.

Brenchley & Harper (1998) Patrick J. Brenchley and David A.T. Harper, *Palaeoecology: Ecosystems, Environments, and Evolution* (London & New York: Chapman & Hall).

Burchfield (1990) Joe D. Burchfield, *Lord Kelvin and the Age of the Earth* (Chicago, IL: The University of Chicago Press).

Callon (1986) Michel Callon, 'The Sociology of an Actor–Network: The Case of the Electric Vehicle', in Callon, John Law and Arie Rip (eds), *Mapping the Dynamics of Science and Technology: Sociology of Science in the Real World* (London: Macmillan): 19–34.

Callon (1991) Michel Callon, 'Techno-economic Networks and Irreversibility', in John Law (ed.), *A Sociology of Monsters: Essays on Power, Technology and Domination* (London: Routledge): 132–61.

Carlson (1991) W. Bernard Carlson, *Innovation as a Social Process: Elihu Thomson and the Rise of General Electric, 1870–1900* (Cambridge & New York: Cambridge University Press).

Claridge (1988) Michael F. Claridge, 'Species Concepts and Speciation in Parasites', in Hawksworth (1988): 92–111.

Clark & Westrum (1977) Tim Clark and Ron Westrum, 'Paradigms and Ferrets', *Social Studies of Science* 17/1 (February): 3–33.

Collins (1985) H.M. Collins, *Changing Order : Replication and Induction in Scientific Practice* (London & Beverly Hills, CA: Sage Publications).

Colwell (1997) Rita R. Colwell, 'Microbial Biodiversity and Biotechnology', in Reaka-Kudla et al. (1997): 279–87.

Connelly & Smith (1999) James Connelly and Graham Smith, *Politics and the Environment: From Theory to Practice* (London: Routledge).

Cox (1998) C. Barry Cox, 'From Generalized Tracks to Ocean Basins – How Useful is Panbiogeography?', *Journal of Biogeography* 25: 813–28.

Dean (1979) John Dean, 'Controversy over Classification: A Case Study from the History of Botany', in Barry Barnes and Steven Shapin (eds), *Natural Order: Historical Studies of Scientific Culture* (London: Sage): 211–30.

de Candolle (1867) Alphonse de Candolle, *Lois de la nomenclature botanique* (Paris: V. Masson et fils).

de Klemm (1990) Cyrille de Klemm, *Wild Plant Conservation and the Law* (Gland, Switzerland: International Union for Conservation of Nature and Natural Resources [IUCN]).

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

Dempsey & Heery (1998) Lorcan Dempsey and Rachel Heery, 'Metadata: A Current View of Practice and Issues', *The Journal of Documentation* 54/2: 145–72.

Derrida (1980) Jacques Derrida, *La carte postale: de Socrate à Freud et au-delà* (Paris: Flammarion).

Derrida (1995) Jacques Derrida, *Mal d'archive: une impression freudienne* (Paris: Galilée).

Donovan & Paul (1998) Stephen K. Donovan and Christopher R.C. Paul (eds), *The Adequacy of the Fossil Record* (New York: John Wiley).

Dumont (1983) Henri J. Dumont, 'Biogeography of Rotifers', *Hydrobiologia* 104: 19–30.

Edwards (forthcoming) Paul N. Edwards, 'Data-laden Models, Model-Filtered Data: Uncertainty and Politics in Global Climate Science', *Science as Culture*.

Edwards et al. (1995) Thomas C. Edwards, Collin G. Homer, Scott D. Bassett, Allan Falconer, R. Douglas Ramsey and Doug W. Wight, *Utah GAP Analysis: An Environmental Information System* (Logan: National Biological Service, Utah Cooperative Fish and Wildlife Research Unit, Utah State University).

Eldredge (1992a) Niles Eldredge (ed.), *Systematics, Ecology, and the Biodiversity Crisis* (New York: Columbia University Press).

Eldredge (1992b) Niles Eldredge, 'Introduction', in Eldredge (1992a): 1–11.

Eldredge (1992c) Niles Eldredge, 'Where the Twain Meet: Causal Intersections Between the Genealogical and Ecological Realms', in Eldredge (1992a): 59-76.

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

Erwin (1997) Terry Erwin, 'Biodiversity at its Utmost: Tropical Forest Beetles', in Reaka-Kudla et al. (1997): 27–40.

Fagot-Largeault (1989) Anne Fagot-Largeault, *Causes de la Mort: Histoire Naturelle et Facteurs de Risque* (Paris: Librairie Philosophique J. Vrin).

Forum (1998) UK Systematics Forum, *The Web of Life: A Strategy for Systematic Biology in the United Kingdom* (London: UK Systematics Forum).

France (1966) France, Commission des sciences et arts d'Egypte, *Description de l'Egypte, publiée par ordre de Napoléon Bonaparte. Avec la collaboration de deux cents savants et artistes*, préface de Gérard Dacier, maquette de Françoise Guillot (Paris: Éditions d'art A. Guillot).

Galtier (1986) Jean Galtier, 'Taxonomic Problems Due to Preservation: Comparing Compression and Permineralized Taxa', in Robert A. Spicer and Barry A. Thomas (eds), *Systematic and Taxonomic Approaches in Palaeobotany* (Oxford: Clarendon Press; The Systematics Association Special Volume 31): 1–16.

Gaston & May (1992) Kevin J. Gaston and Robert M. May, 'Taxonomy of Taxonomists', *Nature* 356 (26 March): 281–82.

Goody (1977) Jack Goody, *The Domestication of the Savage Mind* (Cambridge & New York: Cambridge University Press).

Gould (1989) Stephen Jay Gould, *Wonderful Life: the Burgess Shale and the Nature of History* (New York: W.W. Norton).

Gradstein & Ogg (1996) Felix Gradstein and James G. Ogg, 'A Phanerozoic Time Scale', *Episodes* 19: 3–5.

Gray (1980) Bennison Gray, 'Popper and the 7th Approximation: The Problem of Taxonomy', *Dialectica* 34/2: 129–53.

Gregory (1999) Judith Gregory, *Changing Patient Care: Creating the Electronic Health Record* (unpublished PhD dissertation, University of California, San Diego, 1999).

Grehan (1994) John H. Grehan, 'The Beginning and End of Dispersal: The Representation of "Panbiogeography"', *Journal of Biogeography* 21: 451–62.

Gunn et al. (1991) Charles R. Gunn, John Harry Wiersema and John W. Kirkbride, 'Agricultural Perspective on Stabilizing Scientific Names of Spermatophytes', in Hawksworth (1991): 13–21.

Hajer (1995) Maarten A. Hajer, *The Politics of Environmental Discourse: Ecological Modernization and the Policy Process* (Oxford: Oxford University Press).

Haraway (1997) Donna Haraway, *Modest_Witness@Second_Millennium.FemaleMan© Meets_OncoMouse™: Feminism and Technoscience* (New York: Routledge).

Harland et al. (1990) W. Brian Harland, R.L. Armstrong, A.V. Cox, L.E. Craig, A.G. Smith and D.V. Smith, *A Geologic Time Scale, 1989* (Cambridge: Cambridge University Press).

Hawksworth (1988) David L. Hawksworth (ed.), *Prospects in Systematics* (Oxford: Clarendon Press; New York: Oxford University Press; for the Systematics Association).

Hawksworth (1991) David L. Hawksworth (ed.), *Improving the Stability of Names: Needs and Options*, Proceedings of an International Symposium, Kew, 20–23 February 1991 (Königstein/Taunus, Germany: Koeltz Scientific, for the International Association for

Plant Taxonomy in conjunction with the International Union of Biological Sciences and the Systematics Association).

Hawksworth & Bisby (1988) David L. Hawksworth and Frank A. Bisby, 'Systematics: The Keystone of Biology', in Hawksworth (1988): 3–30.

Hawksworth & Mibey (1997) David L. Hawksworth and Richard K. Mibey, 'Information Needs of Inventory Programmes', in Hawksworth et al. (1997): 55–68.

Hawksworth et al. (1997) David L. Hawksworth, Paul M. Kirk and Stella Dextre Clarke (eds), *Biodiversity Information: Needs and Options*, Proceedings of the 1996 International Workshop on Biodiversity Information (Oxford & New York: CAB International).

Hayden (1998) Corinne P. Hayden, 'A Biodiversity Sampler for the Millennium', in Sarah Franklin and Helen Ragoné (eds), *Reproducing Reproduction: Kinship, Power, and Technological Innovation* (Philadelphia: University of Pennsylvania Press): 173–206.

Heywood (1991) Vernon H. Heywood, 'Needs for Stability of Nomenclature in Conservation', in Hawksworth (1991): 53–58.

Heywood (1995) Vernon H. Heywood (ed.), *Global Biodiversity Assessment* (Cambridge & New York: Cambridge University Press).

Heywood (1997) Vernon H. Heywood, 'Information Needs in Biodiversity Assessments – from Genes to Ecosystems', in Hawksworth et al. (1997): 5–20.

Hilgartner (1995) Stephen Hilgartner, 'Biomolecular Databases – New Communication Regimes for Biology', *Science Communication* 17/2 (December): 240–63.

Huggett (1997) Richard J. Huggett, *Environmental Change: The Evolving Ecosphere* (London: Routledge).

Hughes (1991) Norman F. Hughes, 'Improving Stability of Names: Earth Sciences Attitudes', in Hawksworth (1991): 39–44.

Ingersoll et al. (1997) Rick C. Ingersoll, Tim R. Seastedt and Michael Hartman, 'A Model Information Management System for Ecological Research', *BioScience* 47/5 (May): 310–16.

Journet (1991) Debra Journet, 'Ecological Theories as Cultural Narratives: F.E. Clements' and H.A. Gleason's "Stories" of Community Succession', *Written Communication* 8/4: 446–72.

Keller (1995) Evelyn Fox Keller, *Refiguring Life: Metaphors of Twentieth-Century Biology* (New York: Columbia University Press).

Koch (1998) Carl F. Koch, '"Taxonomic Barriers" and Other Distortions within the Fossil Record', in Donovan & Paul (1998): 189–206.

Kohler (1994) Robert E. Kohler, *Lords of the Fly: Drosophila Genetics and the Experimental Life* (Chicago, IL: The University of Chicago Press).

Kuhn (1970) Thomas S. Kuhn, *The Structure of Scientific Revolutions* (Chicago, IL: The University of Chicago, 2nd edn).

Lamb (1995) Hubert H. Lamb, *Climate, History, and the Modern World* (New York: Routledge).

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

Latour (1993) Bruno Latour, *We Have Never Been Modern* (Cambridge, MA: Harvard University Press).

Latour & Hermant (1998) Bruno Latour and Emilie Hermant, *Paris ville invisible* (Paris: Les Empêcheurs de Penser en rond/La Découverte).

Lave & Wenger (1991) Jean Lave and Etienne Wenger, *Situated Learning: Legitimate Peripheral Participation* (Cambridge: Cambridge University Press).

Lucas (1993) Gren L. Lucas, 'A Worldwide Botanical Reference System', in Frank A. Bisby, George F. Russell and Richard J. Pankhurst (eds), *Designs for Global Plant Species Information Systems* (Oxford: Clarendon Press; New York: Oxford University Press; for the Systematics Association): 9–12.

Lyell (1832) Charles Lyell, *Principles of Geology* (London: Murray).

Malmgren (1867) A.J. Malmgren, *Annulata polychaeta: Spetsbergiæ, Gronlandiæ, Islandiæ et Scandinaviæ. Hactenus cognita* (Helsingfors: Ex Officina Frenckelliana).

Mayr (1988) Ernst Mayr, 'Recent Historical Developments', in Hawksworth (1988): 31–43.

Michener et al. (1997) William K. Michener, James W. Brunt, John J. Helly, Thomas B. Kirchner and Susan G. Stafford, 'Nongeospatial Metadata for the Ecological Sciences', *Ecological Applications* 7/1 (February): 330–42.

Michener et al. (1998) William K. Michener, John H. Porter and Susan G. Stafford (eds), *Data and Information Management in the Ecological Sciences: A Resource Guide* (Albuquerque: LTER Network Office, University of New Mexico).

Nicolson (1989) Malcolm Nicolson, 'National Styles, Divergent Classifications: A Comparative Case Study from the History of French and American Plant Ecology', *Knowledge and Society: Studies in the Sociology of Science Past and Present* 8: 139–86.

O'Hara (1992) Robert J. O'Hara, 'Telling the Tree: Narrative Representation and the Study of Evolutionary History', *Biology and Philosophy* 7: 135–60.

Orr (1996) Julian E. Orr, *Talking about Machines: An Ethnography of a Modern Job* (Ithaca, NY: ILR Press).

Paul (1998) Christopher R.C. Paul, 'Adequacy, Completeness and the Fossil Record', in Donovan & Paul (1998): 1–22.

Pennisi (1999) Elizabeth Pennisi, 'Did Cooked Tubers Spur the Evolution of Big Brains?', *Science* 283 (26 March): 2004–05.

Posey (1997) Darrell D. Posey, 'Wider User and Application of Indigenous Knowledge, Innovations and Practices: Informations Systems and Ethical Concerns', in Hawksworth et al. (1997): 69–103.

Raven et al. (1971) Peter H. Raven, Brent Berlin and Dennis E. Breedlove, 'The Origins of Taxonomy: A Review of its Historical Development Shows Why Taxonomy is Unable to Do What We Expect of It', *Science* 174 (17 December): 1210–13.

Reaka-Kudla et al. (1997) Marjorie L. Reaka-Kudla, Don E. Wilson and Edward O. Wilson (eds), *Biodiversity II: Understanding and Protecting our Biological Resources* (Washington, DC: Joseph Henry Press)

Renne et al. (1998) Paul R. Renne, Danial B. Karner and Kenneth R. Ludwig, 'Absolute Ages Aren't Exactly', *Science* 282 (4 December): 1840–41.

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

Roblin (1997) Richard Roblin, 'Resources for Biodiversity in Living Collections and the Challenges of Assessing Microbial Biodiversity', in Reaka-Kudla et al. (1997): 467–74.

Rudwick (1985) Martin J.S. Rudwick, *The Great Devonian Controversy : The Shaping of Scientific Knowledge among Gentlemanly Specialists* (Chicago, IL: The University of Chicago Press).

Serres (1990) Michel Serres, *Le Contrat Naturel* (Paris: F. Bourin).

Serres (1993) Michel Serres, *Les Origines de la Géométrie* (Paris: Flammarion).

Shapin (1994) Steven Shapin, *A Social History of Truth: Civility and Science in Seventeenth-Century England* (Chicago, IL: The University of Chicago Press).

Sohn-Rethel (1975) Alfred Sohn-Rethel, 'Science as Alienated Consciousness', *Radical Science Journal* 5: 65–101.

Star & Ruhleder (1996) Susan Leigh Star and Karen Ruhleder, 'Steps Toward an Ecology of Infrastructure: Design and Access for Large Information Spaces', *Information Systems Research* 7/1 (March): 111–34.

Stork (1997) Nigel E. Stork, 'Measuring Global Biodiversity and Its Decline', in Reaka-Kudla et al. (1997): 41–68.

Strauss (1999) Evelyn Strauss, 'Can Mitochondrial Clocks Keep Time?', *Science* 283 (5 March): 1435–38.

Thorpe & Shapin (2000) Charles Thorpe and Steven Shapin, 'Who Was J. Robert Oppenheimer? Charisma and Complex Organizations', *Social Studies of Science* 30/4 (August): 545–90.

Tort (1989) Patrick Tort, *La Raison Classificatoire: les Complexes Discursifs – Quinze Etudes* (Paris: Aubier).

Vane-Wright et al. (1991) Richard I. Vane-Wright, Chris J. Humphries and Paul H. Williams, 'What to Protect? – Systematics and the Agony of Choice', *Biological Conservation* 55/3: 235-54.

Vernon (1993) Keith Vernon, 'Desperately Seeking Status: Evolutionary Systematics and the Taxonomists' Search for Respectability 1940–1960', *British Journal for the History of Science* 26: 207–27.

Veyne (1971) Paul Veyne, *Comment on écrit l'histoire; augmenté de Foucault révolutionne l'histoire* (Paris: Éditions du Seuil).

Vrba (1994) Elisabeth Vrba, 'An Hypothesis of Heterochrony in Response to Climatic Cooling and its Relevance to Early Hominid Evolution', in Robert S. Corruccini and Russell L. Ciochon (eds), *Integrative Paths to the Past: Paleoanthropological Advances in Honor of F. Clark Howell* (Englewood Cliffs, NJ: Prentice Hall): 345–76.

Walters (1986) Stuart Max Walters, 'The Name of the Rose: A Review of Ideas on the European Bias in Angiosperm Classification', *The New Phytologist* 104: 527–46.

Watson-Verran & Turnbull (1995) Helen Watson-Verran and David Turnbull, 'Science and Other Indigenous Knowledge Systems', in Sheila Jasanoff, Gerald E. Markle, James C. Petersen and Trevor Pinch (eds), *Handbook of Science and Technology Studies* (Thousand Oaks, CA: Sage Publications/4S): 115–39.

Weedman (1998) Judith Weedman, 'The Structure of Incentive: Design and Client Roles in Application-Oriented Research', *Science, Technology, & Human Values* 23/3 (Summer): 315–45.

Wheeler & Cracraft (1997) Quentin D. Wheeler and Joel Cracraft, 'Taxonomic Preparedness: Are We Ready to Meet the Biodiversity Challenge?', in Reaka-Kudla et al. (1997): 435–46.

Williams (1983) Raymond Williams, *Keywords: A Vocabulary of Culture and Society* (London: Fontana).

Wing (1997) Scott L. Wing, 'Global Warming and Plant Species Richness: A Case Study of the Paleocene/Eocene Boundary', in Reaka-Kudla et al. (1997): 163–85.

Winston (1992) Judith E. Winston, 'Systematics and Marine Conservation', in Eldridge (1992a): 144–68.

Geoffrey C. Bowker is Professor of Communication at the University of California, San Diego. His areas of specialization are biodiversity informatics, the archaeology of scientific collaborations and memory practices in the sciences. He is the author, with Susan Leigh Star, of *Sorting Things Out: Classification and its Consequences* (Cambridge,MA: MIT Press).

Address: Department of Communication, University of California, San Diego, 9500 Gilman Drive, La Jolla, California 92093–0503, USA; fax: +1 858 534 7215; email: bowker@ucsd.edu