1 Introduction

Changing roles and perceptions in European natural history collections: from idiosyncrasy to infrastructure

Malcolm J. Scoble

The history of collecting specimens from nature has been a long one, although just how long is uncertain. While illustrations from a mid-fourteenth century Flemish manuscript show people chasing butterflies (see Salmon, 2000), it is unclear as to whether the specimens were ever caught or stored. Later, however, particularly in the seventeenth and eighteenth centuries, it became fashionable to keep collections – creating a 'Cabinet of Curiosities' was a popular leisure activity. In those times, such an undertaking was mainly a gentleman's pursuit, although women, too, had their cabinets – a famous example belonged to Queen Ludovica Ulrica of Sweden, the contents of which were arranged by Carl Linnaeus, the father of taxonomy, and on which he based descriptions of many species.

A recent estimate (Duckworth, Genoways and Rose, 1993) suggests that there are now around 2.5 billion specimens in natural history collections across the world, although how close this figure is to reality requires further examination given the wide range suggested by L. Speers (pers. comm.) of 1.5 to 3 billion. Whatever the true figure, a substantial proportion of these specimens certainly are housed in European institutions, and, given the history of the region, it is of no surprise that Europe should hold the best sample of organisms through time and in terms of geographical coverage.

The most extensive personal collections were built in the nineteenth century, a time of great exploration, and one when many European countries were colonial powers with associated personnel spread across the globe. Hence many collections were made by administrators, and members of the church and armed services – largely as a leisure activity. Thus the results of 'hobby collecting', unstrategic and individualistic an activity as, very largely, it

was, constitute a significant component of the holdings in many of our national, regional and local museums. Certainly, the earliest collections are part of our cultural heritage. But what we term, subjectively, 'historical' specimens together comprise just a small element of collections as they are composed today. The value of our current holdings, therefore, has been transformed from that in the somewhat idiosyncratically formed individual cabinets of curiosities of the past, to the better integrated and collectively curated systems found in our great museums and herbariums of today. Such integration is far more purposeful and functional, for curated natural history collections provide the raw material on which identifications of taxa, the construction of classifications, and the description of new species rest.

These primary roles suggest that collections are typically associated with the work of taxonomists. Their contents, while uneven, represent the best, and often the only, material information we have about the distribution of organisms over the last few hundred years. Unfortunately, many early collections have little in the way of label data. But before we disregard the apparently naked and bereft specimens of which they are composed, it should be noted that associated observations made by collectors or observers are sometimes capable of being illuminated by tracing them to literature and archives. In 1699, the London apothecary and collector James Petiver wrote: 'The greater silver-streaked Fritillary [the Silver Washed Fritillary butterfly, *Argynis paphia* (Linnaeus)]. I once found this in the physick-garden Chelsea [the Chelsea Physick Garden, London]'. This observation (Petiver, 1699) represents the first British record for the species (Plant, 1987: 124) and can be traced to a particular locality.

Viewed in a more integrated way, collections made many years ago are part of a seamless resource, for all collections of organisms are historic to the extent that they represent samples from the past, albeit often biased in time and space. In terms of European human history, Linnaeus's (eighteenth century) collection is old; but, in the context of the history of life, the temporal difference between its content and specimens collected in the year 2003 barely registers. And while collecting material for interest often meets with disapproval today, and frequently is banned or restricted by law, there is a certain irony in the fact that much of our understanding about classification, species diversity and biogeography has been derived from material composed significantly of the accumulations of those who often collected for personal interest and pleasure. When seen as a whole, collections can constructively, and appropriately, be perceived as an infrastructure of relevance to science and socio-environmental requirements.

This volume includes a series of papers that are concerned essentially with the latest conceptual transformation to apply to collections. As noted already, the first transformation was represented by the shift from keeping personal cabinets of curiosities to the development of a series of better-integrated *physical* resources within single organisations. The second is about gaining even greater value to information in collections through common *electronic*

access to those data via the Internet. Since collections are widely distributed across European institutions, the information content is also distributed. Before discussing this second development, I explore, briefly, the kinds of collections relevant to the European Natural History Specimen Information Network (*ENHSIN*) and other networks.

Collections from different centuries:

Butterflies from the collection of James Petiver (now part of the Sir Hans Sloane Collection, housed in the Natural History Museum, London; 17th C)

An original drawer of butterfly specimens from the collection of Carl Clerck (Linnaeus's contemporary), mid-18th C. (now housed in the Naturhistoriska Riksmuseet, Stockholm)

A modern drawer of moth specimens from the collection of the Department of Entomology, Natural History Museum, London.

Natural history collections

Institutions traditionally housing preserved collections are museums and herbariums. Such organisations typically contain specimens of extant macro-species and also of extinct species – fossils. Many natural history collections also house minerals – inanimate artefacts of nature – and there exist other institutions that keep newer kinds of natural history specimens with data to which we also need to provide access. Examples of the latter are cultures of living material, notably micro-organisms, and collections of frozen tissue samples. As the poverty of our knowledge of, for example, soil and water micro-organisms and their role in ecological functionality is appreciated, the value of collections of them is likely to grow. The same is true of tissue collections, the importance of which will increase with expanding requirements for molecular sequence data. Other important examples of specimens of living material are to be found in botanical gardens, and their zoological counterparts, and in seed banks. The value of such living collections includes not just the whole specimens, but also the underlying genetic resources. These latter collections can only grow with increasing environmental pressures.

Natural history collections typically have much associated archival information. Those most closely linked to specimens take the form of simple observations, and collections databases need to be able to handle such observations. Frequently, prodigious amounts of taxonomic data exist on archives of index cards, each card typically containing information about a species (name, author, date, original reference etc.). Such a card is reasonably treated as a specimen, and access to the source data on the card archive across the Internet is desirable for many potential users (Beccaloni *et al*., this volume).

The ENHSIN(etwork)

Viewed as a whole resource, natural history collections have always been scattered across the world. Their users (often staff working within institutions housing collections) have traditionally networked in an informal way and exchanged or lent each other specimens. It is rarely practicable, however, for all relevant collections to be visited for a particular study. Improving data access across European (and other) institutions is realistic in only two ways. One is to amalgamate collections on a large scale, so that nations or regions have just a few, centralised physical facilities. The other is to provide Internet access to the distributed data. Nation states are unlikely to support the amalgamation of collections, in whole or in part, even within their own countries: costs would probably prove too great, and associations between collections and states and, indeed, particular institutions, are often very strong. Furthermore, amalgamation would be impossible in the case of the many living specimens existing in botanical gardens.

Although some rationalisation of collections has certainly occurred, the most realistic means of enabling common access to their information content is via the Internet. For certain users, access of this kind may be all that is

required. For specialists involved in creating the content stored in collections databases, preliminary access will help them to source material and to decide which collections to visit or what specimens to borrow.

The project featured in this publication is a *network* – specifically the European Natural History Specimen Information Network. It is a network in two senses. First, it is a prototype *electronic* network of distributed specimen databases made available through the Internet via a common access system. Second, it is a network of *people and institutions* across Europe. Furthermore, since the *ENHSIN* pilot is being made operational and greatly expanded by the Biological Collection Access Service for Europe (*BioCASE*), which also falls within Framework Programme V of the European Union, the initial network of seven partner institutions has been expanded across Europe. There can be little doubt that if united through a more effective network, European collections will be much more useful and have a more visible presence. The value of having the capacity to search collection databases across a wide range of European natural history organisations will rise in a non-linear way to the number of them connected to the system.

The structure of the network

The work undertaken in *ENHSIN* falls into four areas, which may be defined broadly as technical, users, management and legal. At the core of the project was the creation of a prototype, demonstrating access to specimen information across seven test databases. The related topic of data quality was addressed, as was that of restrictions imposed by intellectual property issues. A consequence of the latter is to limit the free flow of information across the Internet. Means, and problems, of sustaining effective functioning of *ENHSIN* (and related networks), were also examined. Since the very function of the Network was to provide information access to users, the Partnership undertook a survey of needs of a variety of categories of users.

The ENHSIN prototype

The architecture and components of the *ENHSIN* prototype system are described by Güntsch (this volume). Seven specimen databases have been linked to the prototype, and data on any taxon included can be searched via the Internet by means of the common access system. A significant requirement of the prototype was the need to enable access to databases with structures that are somewhat variable. The basic architecture of the *ENHSIN* pilot should be perceived as an evolving one. In a modified form it is already in use in *BioCASE*, where a large thesaurus of taxonomic and eco-geographical terms is being collated to facilitate data searches by means of the user interface (see Copp, this volume).

Data quality

In our wave of enthusiasm for databasing specimen information, it is essential that the quality of the data is considered. Anecdotal examples given by Bailly (this volume) show that treating specimen databasing purely as a mechanistic exercise without quality control or context is fraught with problems (such as giving a locality record for a marine fish in central Paris – the fish market). The magnitude of the task of 'merely' keyboarding data has already been noted; but the time taken to check data quality is almost infinite.

There are technical means of easing the situation, including the possibility of building some level of data cleansing into software. However, high-quality data are derived from sound research, and should be accepted as one that is time-consuming. More broadly, and, as noted by Bailly, *ENHSIN* should be viewed not only as a means by which digital data can be made accessible, but also as a tool to encourage building the Network by enhancing the quality of the data over the long-term.

Users

The number of users of collections data is growing, and more users will be captured as data sources become richer and as outputs become ever more multifaceted. Some users will simply require a list of institutions where specimens of particular species can be located (i.e. information about what exists in particular collections). Others will need data from many sources delivered via the Internet in the form of spatial outputs, such as distribution maps with some indication of whether gaps in specified areas may be explained by sampling intensity or whether they reflect actual distribution. Taxon and geographical boundaries are indistinct ('fuzzy') and the user community is correspondingly complex. Yet for the value of natural history databases to be realised effectively, and for a significant gain both to the taxonomic community, and to the wider and complex community of users, the development of graphic and highly analysed outputs will be essential.

Although we can only speculate on the breadth of users in the future, the attitudes of existing users of specimen information can, at least, be assessed. In *ENHSIN*, the assessment was based on the results of a questionnaire. The design of the process and results arising from it are reported by Calabuig *et al*. (this volume). All partners commented on the content of the questionnaire before the final version was mailed to a substantial number of recipients. This exercise proved to be a useful basis for considering quite fine detail about user needs, requiring both partners and users to consider the kinds of data associated with specimens. It also enabled the partnership to assess similarities and differences between the needs of the specialist research community and the wider community such as conservation and education bodies.

ENHSIN in context

ENHSIN should be understood as just one component of a programme of work to provide electronic access to information in natural history collections across Europe. Beyond even this broad brief, there is a yet wider vision, which is global in scope (see Berendsohn, this volume). The emphasis being placed on specimen digitisation in collections-rich establishments means that large quantities of data of these kinds are likely to become available in the future. Advances in data standardisation, data modelling, interoperability across distributed databases, and the capacity to 'wrap' databases with somewhat disparate structures has meant that large amounts of data are gradually becoming accessible through the Internet. The European dimension has evolved through a series of projects with complementary aims. Its common aim (Olsvig-Whittaker & Berendsohn, 2000) is the intention to produce:

> *an electronic access system facilitating queries across hundreds of millions of specimens and monitoring or mapping records held by institutions, projects and individual researchers in the EU and partner countries.*

The task of achieving this vision is demanding, but fortunately there are political and practical reasons why we may be optimistic about progress. Politically, calls for wider and improved access to material held in publicly funded bodies are growing (e.g., in the UK by *Resource*, The Council for Museums, Archives and Libraries, www.resource.gov.uk/index.html), and such calls are being addressed with determination by many institutions. This determination is being supported by considerable enthusiasm in these organisations, as is evident from the growing number of individual or institutional Internet sites providing access to data held in collections. The information available is often a mixture of data on species names, in the form of descriptions of collections, and from labels attached to the very units themselves – the specimens. Although information in collections rests, fundamentally, on the units of which they are composed, a complete database of all European specimens is a distant goal. However, every specimen databased is of immediate value in its own right, and the value of the bank of specimen information will grow with the number of specimens databased.

A further practical means by which data in collections can be made available more quickly than awaiting the long process of complete specimen databasing, is by providing users with descriptive data about whole collections or subsets of them. Berendsohn *et al*. (2000) and Güntsch (2000) explained how such collections metadata can rescue us from an overwhelming delay by allowing users to gain broad access to distributed information about, for example, the particular collections in which target species are represented. More specifically, the authors point out that the relevant metadata items can be applied at the level both of the specimen and of the collection. Moreover, such items belong, for the most part, to a

hierarchical system, whether referring to geography (e.g. state within country) or taxonomy. These two observations mean that the metadata can be arranged into a scaleable system and applied to whole collections down to detailed information about specimens. Such concepts were fundamental to the Biological Collections Identification Service for Europe (*BioCISE*) www.bgbm.fu-berlin.de/BioCISE/database/ and deserve to be understood better by curators and managers in collections-rich institutions. Such an understanding is likely to help resolve differences of opinion as to whether databasing should be focused on specimens or collections. Certainly the nature of the metadata underlines the integrated position of *ENHSIN* in the broader programme of electronic access to collections databases across Europe – and, indeed, the globe.

ENHSIN and the taxonomic system

Viewed in its entirety, the taxonomic system as applied to all organisms is massive and fragmented, being spread across thousands of publications and a growing number of websites. It has been proposed recently (Godfray, 2002) that taxonomy should be viewed as an information science encouraging, as standard, a shift to web-based unitary taxonomies of organisms, including integrated information on keys, descriptions, phylogenies of subgroups, literature and natural history observations. An example of such an integrated site under long-term development for the marine invertebrate Class Echinoidea (sea urchins) is provided at www.nhm.ac.uk/palaeontology/ echinoids/ by A.B. Smith.

The echinoid site deals with taxa rather than specimens, although the information on which it is founded is based significantly on studies of specimens over the long taxonomic history of echinoids. An explicit link between taxon circumscriptions and the specimens on which they are founded is provided by the Prometheus Taxonomic Model (Pullan *et al*., 2000). Prometheus enables users to view multiple classifications of a given taxon (particularly plant taxa, since rules and conventions of the botanical code are adopted in the model), each classification being defined in terms of a particular circumscription. While there is no direct connection between *ENHSIN*, *BioCASE* and Prometheus at present, the emphasis given to specimens by the Prometheus model is indicative of the potential for marrying web-based systems dealing with taxon definitions to sites providing access to the actual specimens on which these definitions are founded.

These two examples emphasise the importance of improving data-quality and access on information relating to specimens and collections as part of a broad yet more unified taxonomic infrastructure.

ENHSIN and modern museums and herbariums

> *'There is only one quality that distinguishes a museum from all other types of entertainment, from scholarly, academic and recreational activities, and that is that we hold collections'* (N. Cossons *in* Fisher, 2002)

The general aims of *ENHSIN* and related data-access projects are highly appropriate to the view that the distinctive feature of museums and herbariums is that they hold collections. Natural history collections are typically used in the way in which they have always been used, that is for research, notably in taxonomy, and education. Yet today, the value of the results derived from the collections seems particularly high, since data about organisms, especially their distribution and association with patterns of land use, are of wider environmental relevance. Collections, although biased in terms of localities, remain the best sample of species distribution that we have. Examining material simply by physically surveying a large collection gives the user some idea of the distribution through time of a given species (or other taxon). The ability to gain instantaneous access to such information across a wide range of European museums via a suitable user interface at a computer screen will be considerably more useful. Imperfect as they are, specimen records, most of which have no geographical co-ordinate data, often provide at least a broad picture of distribution and some idea of patterns of change of this distribution.

Research of all kinds on organisms or minerals, whether applied or not yet applied, depends on the accurate identification of specimens. And means of identification are based largely on work on material stored in museums and herbariums, resulting in the production of identification keys, field guides and monographs on target groups of organisms. Gaining wider access to specimen data speeds the work of the specialists who prepare these products and improve their quality. Of special importance is the need for data about types, reference specimens that bear the names of species. There is a growing interest in access to images of these specimens, a topic that is being explored in several initiatives, including the EU-supported project *ENBI* (European Network of Biodiversity Information Network), www.faunaeur.org/enbi/info.html).

Finally, a thoroughly modern digital approach to disseminating information about museums may, ironically, help save these institutions from what was described by Fisher (2002) as a potential loss of 'confidence in the idea of the museum'. An important component of this malaise, he suggested, is a distaste for the past, and thus a diminution of the value of actual objects as links to the past. It might seem counter-intuitive to think that promoting access to virtual information about museum objects via the Internet could be squared with the idea of the primacy of those objects. Yet alongside museums (memory institutions) as *physical places* that promote the exchange of knowledge, are developing *digital networks*, which serve to connect collections in

a virtual sense (Dempsey, 2000). Paradoxically, an apparent diminution of respect for the objects of the past appears to be occurring with an ever-increasing enthusiasm for access to them in the form of digital information (data or images). Perhaps, far from causing European citizens to ignore objects in collections, the most modern of technologies might encourage them, instead, to be revisited. Digital networks are surely the very route to enable a shared approach to our collective scientific, industrial and cultural heritage. At the same time, virtual access to data in natural history collections is of material value in providing baseline information to those involved in attempting to assess the loss of biological diversity.

References

Berendsohn, W.G., Costello, M.J., Emblow, C., Güntsch, A., Hahn, A., Koenemann, J., Thomas, C., Thomson, N. & White, R. 2000. Concepts for a European portal to biological collections. *In* Berendsohn, W.G. (Ed.), *BioCISE, Resource Identification for a Biological Collection Information Service in Europe*, pp. 59–70. Berlin.

Dempsey, L. 2000. Scientific, industrial, and cultural heritage: a shared approach. www.ariadne.ac.uk/issue22/dempsey

Duckworth, W.D., Genoways, H.H. & Rose, C.L. 1993. *Preserving natural science collections: chronicle of our environmental heritage*. Washington, D.C.:iii+140 pp.

Fisher, M. 2002. Objections to the object: are we losing confidence in the idea of the museum? *The Times Literary Supplement* **5164**: 13–14.

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

Güntsch, A. 2000. The BioCISE collection catalogue. *In* Berendsohn, W.G. (Ed.), *BioCISE, Resource identification for a biological collection information service in Europe*, pp. 53–58. Berlin.

Olsvig-Whittaker, L. & Berendsohn, W.G. 2000. Computerising and networking biological collection data. *In* Berendsohn, W.G. (Ed.), *BioCISE, Resource identification for a biological collection information service in Europe*, pp. 5–12. Berlin.

Petiver, J. 1699. *Musei Petiverani Centuria quarta rariora naturae*.

Plant, C.W. 1987. *The Butterflies of the London Area*. London: 199 pp.

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

Salmon, M.A. 2000. *The Aurelian Legacy – British Butterflies and their Collectors*. Colchester: 432 pp.