Costs and benefits of Web access to museum data

Brooke's[1] News & Comment alluded to the concerns museum curators have about unrestricted Web access to biological specimen data – the objective of several competing international bioinformatic networks. All parties agree that museum specimen data have many scientific and commercial applications, but the costs and benefits of open Web access remain to be reconciled.

First, relatively few specimens have been computerized (e.g. only 13% of 2 350 000 skins in the three largest ornithological museums in Tring, New York, USA and Washington, USA). At prevailing costs (US$1.50–US$2.00/specimen in Washington), computerizing the remaining 87% of the specimens would cost between US$3 200 000 and US$4 200 000; however, accurate georeferencing might double or triple the cost. External pressure for computerization is increasing, but who will pay? Museum funding continues to decline and many institutions survive on starvation budgets; thus computerization is a lower priority than other urgent infrastructure needs. And, if such funds were available, wouldn't they be better spent on new biodiversity surveys?

Brooke mused that museums 'should accept that, by making their data available, they are providing an immensely valuable service to the world community and use that to leverage more core funding from national governments'. If past experience is a guide, the ardent courtship of museums and promises of financial support by bioinformatics advocates would probably vanish as soon as the specimen data have been extracted.

The fundamental problem is not free transfer of museum data, but to whom data are transferred and for what purpose. Free access to museum specimens and data has long been provided to bona fide researchers. However, before major museums will permit sensitive data fields (e.g. collector, date, specific locality and reproductive data) for entire collections to be posted on the Web, bioinformatics networks must provide legally binding safeguards and provisions. These safeguards include: (1) intellectual property rights; (2) commercial use and licensing; (3) appropriate acknowledgement in the publication; (4) endangered species data; and (5) privacy rights of collectors and researchers. At present, the major networks rely on unsecured disclaimers to prevent data miners from systematically downloading, repackaging and selling museum data, from transferring data to third parties, and from using compilations of raw museum data to leverage grants and contracts in direct competition with museums. Reliance on the honor system to protect the intellectual property rights of museums and researchers is naïve. Moreover, it is virtually impossible to compel Web database users to acknowledge data sources appropriately (e.g. Ref. 2). Unfiltered data might also be used to locate and exploit commercially valuable or endangered species, and might provide the means for anticollecting zealots to interfere with ongoing research programs. Once released, the digital genie can never be put back in the bottle.

Until flaws in network safeguards are fixed, museum curators can better facilitate free access to museum resources, and protect the interests of museums and researchers, by handling data requests on a case-by-case basis – the same as requests for specimen use.

Gary R. Graves

Dept of Vertebrate Zoology, National Museum of Natural History, Smithsonian Institution, Washington, DC 20560, USA (graves.gary@nmnh.si.edu)

References

1 Brooke, M. de L. (2000) Why museums matter. *Trends Ecol. Evol.* 15, 136–137

2 Peterson, A.T. *et al.* (1999) Conservation of ecological niches in evolutionary time. *Science* 285, 1265–1267

I agree entirely with Brooke in his recent *TREE* article[1] that 'museums matter' both in science and education. Surely, the rich resources, both intellectual and material, within these institutes provide ample evidence to refute any perception of museums being metaphorically 'dusty places'. Indeed, I wonder if such perceptions exist beyond the minds of those insiders mildly paranoid about how they might be viewed by outsiders with limited imagination. Nevertheless, these perceptions appeared to spawn an interesting meeting.

Museums and other physical 'memory institutions' (http://www.ariadne.ac.uk/issue22/dempsey), such as libraries, are fast populating digital space on the Internet, with the result that, in the future, information about collections is likely to be far more accessible. In the context of Brooke's article[1], and against the background of a fast evolving situation, it is worth carefully considering two important points on databasing museum specimens.

First, we should be aware of the limits of many data in natural history museum collections. Although these institutes certainly hold material of great value to understanding changes in the distribution of organisms, much of the information is likely to be of a more qualitative than quantitative kind. This is partly because museums have traditionally acquired specimens through 'collecting', rather than the less biased procedure of 'sampling', and partly because much geographical information on specimen labels (particularly for older material) is broader than the more precise point (coordinate) data.

Second, the value of providing wide access to label data is dependent on the accuracy of the identification of the specimen with which those data are associated. In species-rich groups, such as insects, which are often represented in collections by large numbers of specimens, accuracy of identification is frequently a serious problem. Well ordered collections often disguise significant levels of misidentification, notably in those taxa where morphological preparation is required for confident determination or where modern revisionary work has yet to be undertaken. Misidentifications can lead to false portrayals of distribution.

These points are by no means intended to devalue the wealth of distributional information contained in natural history museums. Rather, they suggest that as curators (and particularly managers) express their enthusiasm for computerizing collections data, there is a crucial need to control the momentum by scrutinizing data standards and protocols. Brooke[1] cites the European Natural History Specimen Information Network (ENHSIN) – this thematic network, which is supported by the European Community's Improvement of Human Potential programme, has been initiated to address such matters across European institutions holding natural history collections. Recommendations to help standardize data fields are needed; software enabling interoperability across databases requires selection or development; conflicting issues demand resolution – notably balancing the defence of intellectual property rights against a desire to provide open access to data; and deep thought is required to meet or anticipate the needs of users.

Brooke notes that the many initiatives that have arisen, in response to moves towards what he calls a 'world museum', compete to some extent; indeed, at times the field does look rather crowded. However, although the Global Biodiversity Information Facility (GBIF) is a 'top-down' global initiative, ENHSIN deals explicitly with specimen (unit) data[2] in European collections and is, as such, a 'bottom-up' initiative (http://www.nhm.ac.uk/ science/enhsin/). In this it is also complementary to BioCISE (http://www.bgbm.fu-berlin.de/BioCISE/), an EC Concerted Action broadly covering 'metadata' associated with natural history collections.

Malcolm J. Scoble

Dept of Entomology, The Natural History Museum, Cromwell Road, London, UK SW7 5BD, and Network Coordinator of ENHSIN (m.scoble@nhm.ac.uk)

References

1 Brooke, M. de L. (2000) Why museums matter. *Trends Ecol. Evol.* 15, 136–137

2 Berendsohn, W.G. *et al.* (1999) A comprehensive reference model for biological collections and surveys. *Taxon* 48, 511–562

In his stimulating recent News & Comment, Brooke[1] described the ever increasing possibility of accessing collection data via the Web. He probably got a bit carried away when he wrote, 'Traditionally, the scientist has accessed such…data by touring the museums…But this approach promises to be superseded.' I have toured many European and American museums while searching for distributional data of blennioid fish from the Mediterranean Sea and the eastern Atlantic (50 species). Not only were approximately 5% of the species misidentified (which in some cases would have led to totally wrong

