Specimens as primary data: museums and 'open science'

Abstract

In 1977, Eugene Odum advocated a synthetic approach if ecology were to rise above the level of explanation afforded by independent, individual studies [1]. Today, Odum's wish is being fulfilled, and important advances are being made by synthesising data derived from great numbers of studies, either by scaling up temporally or geographically [2]. However, to allow effective, creative, and reproducible integration of ecological and environmental results, the methods and data used need to be made freely accessible and combinable. Only then can integrated ecology become a field where the ideals of 'open science' [3] fully come to fruition. Indeed, although great challenges remain [4,5], open access to ecological data, methods, and analysis is rapidly improving [6,7]. Nonetheless, we here call attention to what we perceive as one important obstacle to open data in biodiversity studies. The 'raw data' in biodiversity research consist not of tabulations of numbers of individuals of species sampled at a particular date and place, but of the properly-labelled specimens themselves; occurrence records associated with specimens are metadata. We feel it is insufficiently appreciated that each assignment of a specimen to a particular taxon (whether a formally described species or a pragmatic 'morphospecies' [8]), is a researcher's interpretation, and therefore not a primary datum. Because the scholarship of biodiversity includes scrutinising earlier work, evaluating what was written before, and adding new information and insight, it should always be possible to return to those specimens. They are the primary evidence for the information presented. The ability of researchers to reexamine the primary data and question the conclusions of previous work is a crucial part of what makes this a scientific activity. Especially in groups where the taxonomy is in flux, this is essential to ensure long-term comparability and vitality of data. Unfortunately, in our experience, the accessibility of specimens sampled during biodiversity studies is problematic for two reasons. First, after publishing their results, many researchers and institutes do not systematically archive the samples of specimens.
that form the basis for the analyses. Specimens are either discarded or only a small reference collection is saved, leaving no way to verify the metadata. Even if specimens are stored, material from separate plots or dates are often pooled to reduce storage space [9], rendering valuable information irretrievable. We therefore suggest that it become accepted policy in ecological research that full, unadulterated collections of all specimens from a study be deposited in a natural history collection. This is common practice in other areas of specimen-based biological research, such as taxonomy and palaeontology. Public natural history collections increasingly make the content of their collections databases available through the Global Biodiversity Information Facility (GBIF), which should facilitate retrieval and verification of specimens as well as reuse of the associated metadata [10]. Moreover, the specimens would then be available for obtaining additional information (such as genetic and morphometric data, and sometimes even information about ecological interactions [11]).

The second reason for the inaccessibility of specimens, however, lies with those same publicly-accessible collections. Natural history museums often appear unable or reluctant to assume a custodian’s role as repositories for bulk samples from ecological studies. This is understandable in view of the traditional focus of natural history museums on systematics and biogeography, which gives rise to a desire to maximise the information density of their holdings by giving priority to previously unrepresented species or localities. Given the universal features of species-abundance distributions, however, biodiversity research will yield samples that are dominated by common and widespread species. Faced with space limitations, and a lack of funding and staff to be able to curate and maintain large ecological collections, museums tend to refuse, cherry-pick, or even dispose of such bulk collections. We argue that by adhering only to their traditional role, natural history museums are missing an opportunity to expand their scientific reach and relevance. Specifically, we advocate that they should act as custodians of biological field samples, including entire collections of samples from biodiversity studies, either by storing the material themselves or by setting up dedicated repositories under their supervision. We also suggest that museums develop collection
management policies that enable the scientific community to access the increasing number of specimens needed to realise the open science concept. For their part,