Biological Conservation xxx (2013) xxx–xxx

Contents lists available at ScienceDirect

Biological Conservation

journal homepage: www.elsevier.com/locate/biocon

Strategies for the sustainability of online open-access biodiversity databases

Mark J. Costello [a,*], Ward Appeltans [c,l], Nicolas Bailly [j], Walter G. Berendsohn [b], Yde de Jong [d], Martin Edwards [h], Rainer Froese [g], Falk Huettmann [k], Wouter Los [d], Jan Mees [c,e], Hendrik Segers [f], Frank A. Bisby [i,1]

[a] Institute of Marine Science, Leigh Marine Laboratory, University of Auckland, New Zealand
[b] Botanic Garden and Botanical Museum Berlin-Dahlem, Freie Universität Berlin, Koenigin-Luise-Str. 6-8, 14195 Berlin, Germany
[c] Flanders Marine Institute (VLIZ), Wandelaarkaai 7, B-8400 Ostend, Belgium
[d] Faculty of Science, University of Amsterdam, Science Park 904, 1098 XH Amsterdam, Netherlands
[e] Ghent University, Department of Biology, Marine Biology Research Group, Krijgslaan 281/S8, 9000 Gent, Belgium
[f] Belgian Biodiversity Platform, Freshwater Laboratory, Royal Belgian Institute of Natural Sciences, Vautierstraat 29, B-1000 Brussels, Belgium
[g] Helmholtz-Centre for Ocean Research, Duesternbrooker Weg 20, 24105 Kiel, Germany
[h] Sir Alister Hardy Foundation for Ocean Science, Citadel Hill, The Hoe, Plymouth PL1 2PB, UK
[i] University of Reading, UK
[j] WorldFish – FishBase Information and Research Group, Aquatic Biodiversity Informatics Office, Los Baños, 4031 Laguna, Philippines
[k] EWHALE Lab, University of Alaska-Fairbanks (UAF), Fairbanks, AK 99775, United States
[l] Intergovernmental Oceanographic Commission of UNESCO, IOC Project Office for IODE, Wandelaarkaai 7/61, 8400 Oostende, Belgium

ARTICLE INFO

Article history:
Received 3 March 2013
Received in revised form 25 July 2013
Accepted 30 July 2013
Available online xxxx

Keywords:
Species
Taxonomy
Publication
Citation
Funding
Intellectual property

ABSTRACT

Scientists can ensure that high quality research information is readily available on the Internet so society is not dependant on less authoritative sources. Many scientific projects and initiatives published information on species and biodiversity on the World Wide Web without users needing to pay for it. However, these resources often stagnated when project funding expired. Based on a large pool of experiences worldwide, this article discusses what measures will help such data resources develop beyond the project lifetime.

Biodiversity data, just as data in many other disciplines, are often not generated automatically by machines or sensors. Data on for example species are based on human observations and interpretation. This requires continuous data curation to keep these up to date. Creators of online biodiversity databases should consider whether they have the resources to make their database of such value that other scientists and/or institutions would continue to finance its existence. To that end, it may be prudent to engage such partners in the development of the resource from an early stage. Managers of existing biodiversity databases should reflect on the factors being important for sustainability. These include the extent, scope, quality and uniqueness of database content; track record of development; support from scientists; support from institutions, and clarity of Intellectual Property Rights. Science funders should give special attention to the development of scholarly databases with expert-validated content. The science community has to become aware of the efforts of scientists in contributing to open-access databases, including by citing these resources in the Reference lists of publications that use them. Science culture must thus adapt its practices to support online databases as scholarly publications.

To sustain such databases, we recommend they should (a) become integrated into larger collaborative databases or information systems with a consequently larger user community and pool of funding opportunities, and (b) be owned and curated by a science organisation, society, or institution with a suitable mandate. Good governance and proactive communication with contributors is important to maintain the team enthusiasm that launched the resource. Experience shows that 'bigger is better' in terms of database size because the resource will have more content, more potential and known uses and users of its content, more contributors, be more prestigious to contribute to, and have more funding options. Furthermore, most successful biodiversity databases are managed by a partnership of individuals and organisations.

* Corresponding author. Tel.: +64 94145031.
E-mail address: m.costello@auckland.ac.nz (M.J. Costello).

[1] Authors are alphabetical except for the first (lead) author and the last (deceased).

0006-3207/$ - see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.biocon.2013.07.042

1. Introduction

The Internet has rapidly become the first place most people look for information. They will find information on almost anything, but still in a somewhat anarchic form. It behoves the scientific community to ensure that authoritative information is available, can be easily distinguished from less scholarly or up-to-date resources, and is as comprehensive as possible (Costello et al., 2006). As the print media developed, so did peer-review and editorial control, where individual scientists published their work through resources whose quality is primarily controlled by the scientific community. A similar approach is desirable for online resources, including taxonomic and ecological databases (Costello and Vanden Berghe, 2006; Costello, 2009a; Huettmann, 2005, 2009; Costello et al., 2013a,b). Thus online resources should be led, edited and authored by well-qualified experts (Zuckerberg et al., 2011; Cushman and Heuttmann, 2010). As in the print media, scientists need to discipline their efforts to create authoritative, collaborative, online information resources and databases overseen by the scientific community. Print publications have the benefit of one-off production costs, revenue earned from sales, and archiving in libraries, but they also entail significant costs for scientific institutions to maintain subscriptions, library premises and staff. Particularly in areas of scientific research, they may become out of date quickly. In contrast, web resources can be regularly updated and upgraded at little cost, but they require regular input and quality control efforts that have continuous costs for maintenance. Furthermore, instead of earning revenue from sales or user fees, both contributors and users expect their access to be free and open. Criticisms from users who have unrealistic expectations that online resources should be even more comprehensive and accurate than the print literature can undermine the support for these same resources. Thus, once project funding to establish a web resource has expired, it can become difficult to maintain it (Graham et al., 2004; Merali and Giles, 2005; Costello, 2009b; Drew, 2011; Brewer et al., 2012). This is especially a problem for stand-alone web resources and short-term projects that are serving data based on human validation, and thus dependent on curation to maintain quality and up to date service. Although the lack of long-term funding support for such open-access databases is a great worry to amongst scientists, the options for such funding, and the factors facilitating a sustainable business model, have not been discussed in the literature. Here we seek to address this from our experiences in developing a variety of biodiversity databases since the 1990s, and observations of related initiatives world-wide. We provide examples of how some leading global, online biodiversity databases have been developed, managed, and governed; describe the challenges and costs in their maintenance; and the importance of clarity in Intellectual Property Rights for database succession planning. We conclude with a summary of management options and recommendations to maximise the sustainability of biodiversity databases.

2. Biodiversity databases

Some biodiversity databases collect primary data and information, such as measurements from instruments, records of biological specimens and observations, and/or may contain expert judgements on species concepts and classifications, or interpretations of data and current knowledge. Much of the content may have been derived from the published literature, as is the case in many scientific papers and books. Other web sites, such as iSpecies and the Encyclopedia of Life (EoL), aggregate content from such databases. Here we are primarily concerned with primary resources validated by experts in the subject. These experts choose to contribute their time to making their accumulated data and knowledge publicly available because they see this as an important service to science and society (Fig. 1). These resources are thus scholarly publications and should be so recognised (Box 1). However, this is often not practised by their institutes, funders or publishers. For example, the manner of citation of online resources in journals is variable (Box 2).

Box 1 Databases as publications.

It is important that we consider these online resources as scientific publications (Huettmann, 2007a; Costello, 2009a). All the authors and editors involved should be named and their contributions citable. The scientist's names and standing lend authority to the quality of the content. Knowing the authors can also indicate their bias, and as they may support particular schools of thought or advocate particular approaches to issues related to their science. Citations provide the recognition that scientists require for their career development, and that their employer organisations and funding agencies may require to demonstrate their productivity and excellence. In science, authors and editors do not typically get paid for publishing their knowledge, but it is the practice to cite their work, and it influences their reputation, employability and promotion. As Latour (1987) explained in a still debated interpretation of what is science, the capital of the scientists is not money but rather recognition that is gained by citations of their work, which is measured today only on traditional publications (printed or electronic). Citation also has implications for the permanency and archiving of versions of the database, because science requires accessibility to a resource as it was cited, not only the latest version.

At least in the biology and ecology disciplines, the scientific community appears still to be reluctant when it comes to citing online databases and generally prefer to cite traditional references. For example, some of the species databases from the FADA project have been made available as classical scholarly publications, in addition to as a web resource. Whereas both are equally sound scientifically, having passed peer review, the latter is being regularly updated and should therefore be the more useful resource. Nevertheless, whereas one of them (Segers, 2007) has been cited 71 and 102 times as a paper publication, not a single citation of any of the 14 FADA online databases could be traced (Science Citation Index Expanded 3rd February 2012 and Google Scholar search 7th February 2012 respectively). Similarly, the editors of WoRMS were publishing a series of synthesis papers in a journal, but of the first five papers so published, not one cited the WoRMS database in their references, and only two mentioned it in their text. Thus, even the scientists developing scholarly archived web resources need to be reminded to cite them; and the same applies to editors and publishing houses.

Please cite this article in press as: Costello, M.J., et al. Strategies for the sustainability of online open-access biodiversity databases. Biol. Conserv. (2013), http://dx.doi.org/10.1016/j.biocon.2013.07.042