Biological Conservation xxx (2013) xxx–xxx

Contents lists available at ScienceDirect

Biological Conservation

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

Perspective

Best practice for biodiversity data management and publication

Mark J. Costello [a,*], John Wieczorek [b]

[a] Institute of Marine Science, Leigh Marine Laboratory, University of Auckland, New Zealand
[b] Museum of Vertebrate Zoology, University of California, Berkeley, CA 94720, USA

ARTICLE INFO

Article history:
Received 15 June 2013
Received in revised form 16 October 2013
Accepted 26 October 2013
Available online xxxx

Keywords:
Methods
Biodiversity informatics
Conservation
Data standards
Global Biodiversity Information Facility
Ocean Biogeographic Information System
World Register of Marine Species
Species 2000

ABSTRACT

There is increasing pressure from the scientific community, including funding agencies, journals and peers, for authors to publish the biodiversity data used in published articles and other scientific literature. This enables reproducibility of research and creates new opportunities for integrating data between research projects and analysing data in additional ways. The long-term availability of data is especially important in conservation science because field data can be costly to collect. In addition, historic data, especially on threatened species and their associated biota, become more valuable over time. This paper summarises current standards and best practices for the management and publication of biodiversity data. It includes recommendations for citing sources of species determination and standards for formatting species distribution data. Whenever possible, data should be published for inclusion in data access platforms that integrate datasets (e.g. GBIF, GenBank) and so enable new analyses and broader impact. Data centres (e.g. PANGAEA) provide added value in quality checks on data. A minimum standard recommended is that data should be permanently archived in an online, open-access repository with sufficient metadata for potential users to understand how and why they were collected.

Contents

1. Introduction

The purpose of scientific publication is to recognize the work of authors and make it available so that others can learn, repeat, build on, and cite that work (Lawrence, 2008). This purpose is best achieved if the associated data are also published. Data publication, that is making data available without conditions on their use, is preferable to 'data-sharing' because it ensures that data are permanently available for future research, and because publication is a meritorious scientific activity (Costello and Vanden, 2006; Costello, 2009; Whitlock, 2011; Costello et al., 2013a; 2013b; 2013c). Furthermore, publication is a well-understood process, and clarifies potential concerns over intellectual property rights; including where data are concerned (reviewed by Reichman and Uhlir, 2003; Hagedorn et al., 2013).

Making primary data available is essential for scientific analyses to be reproducible and independently verified. When combined with additional data, it may reveal new insights that lead to further advances in the field (Costello, 2009). Nevertheless, most

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

0006-3207/$ - see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.biocon.2013.10.018

ecological data are not accessible after their analyses have been published (Reichman et al., 2011; Hampton et al., 2013). Organising data so others can understand it is a chore. There can also be issues related to cost, intellectual property rights, and data ownership. However, overcoming these issues and publishing data is the right thing to do for science, and can lead to increased visibility of the researcher's work, increased citations, and increased invitations to collaborate (Costello, 2009). Increasingly, research funding agencies and their evaluators of applications expect or require data to be publicly available. Thus scientists that demonstrate their good citizenship in this way may have more funding success.

There is a shared responsibility for authors, editors, and referees to ensure that data are published along with analyses, and a wide range of national and international science and publishing polices recommend this (reviewed in Costello, 2009). While print media have shied away from publishing primary data in recent decades to save on printing and postage costs, the advent of online appendices (or supplementary material) and other online repositories have reduced the costs of publishing data and thus have removed one of the major impediments to making data available upon publication of a study (Costello, 2009). However, how the data are published has implications for their discovery, re-use, and permanent availability. An increasing number of journals in the fields of biology and ecology are now formally adopting a Joint Data Archiving Policy (Anon, 2013). Some additional recommended practices are proposed here to ensure biodiversity data are (a) of good quality, (b) easily understood, and (c) easily and permanently accessible. These recommendations are directed to scientists whose data may include observations and related sample data (e.g., when, where, what method) and associated environmental (e.g., temperature, salinity, altitude), ecological (e.g., habitat, associated species, host), photographic, sound, video, and other biological data (e.g., body size, sex, age) from field and laboratory studies.

The need for biodiversity data to be easily and permanently accessible is particularly important for conservation. Collecting data on the occurrence of species of conservation concern is especially difficult, and thus costly, particularly for species that are low in abundance, geographically rare, and that avoid people due to hunting. Perhaps half of all species have distribution data in the main world species database, namely the Global Biodiversity Information Facility (GBIF) (Costello et al., 2013b). This makes supporting species' Red List assessments with empirical data challenging. Considering the concerns over species extinctions, it is critical that past and recent biodiversity data are readily available to researchers and policy makers to enable the best possible conservation decisions.

2. Before submission for publication

Most papers about biodiversity include information on one or more species. There are two basic aspects to reporting useful species data. First, the scientific names used should be valid or accepted according to the appropriate code of nomenclature. Second, the organisms represented should be identified as accurately as possible, ideally according to a specified treatment or circumscription. When using only a reference guide, a full determination may not be possible. In all cases, it is important to check names against an appropriate authority. Steinke and Hanner (2011) provide detailed recommendations to ensure the accuracy of species identification and accompanying information, such as who collected and identified the specimens; Franz et al. (2008) discuss the use of taxonomic concepts in biodiversity research, and Frankham et al. (2012) in conservation. The Darwin Core standard is in widespread use among biodiversity initiatives for data sharing, and provides well-defined fields to record information about names, ranks, identifications, and taxonomic concepts (Wieczorek et al., 2009, 2012). Species names are subject to variations in spelling, synonymy and alternative genus combinations (Costello et al., 2013a; Costello et al., 2013b). There are numerous sources of species names on the internet, and as within the published literature, many will include misspellings and synonyms because they have not been quality controlled by experts. Other species names come from secondary sources that collect validated names from taxonomic databases and may therefore be out of date. Even the best sources may have errors. Thus, at an early stage in any study, three steps are recommended:

1. State how the species were identified and cite the identification references. Even if an expert has identified the specimens, both acknowledge them and cite the literature they used. It is possible that a particular species according to one source will be found to be a different species in the future, or under a different taxonomic opinion.
2. Check all species names against authoritative taxonomic checklists (Table 1), such as the Catalogue of Life (Bisby et al., 2012), AlgaeBase (Guiry and Guiry, 2013), and World Register of Marine Species (Appeltans et al., 2013). Such lists are never quite up to date, but nomenclatures of names for animals, fungi (Robert et al., 2005) and plants are being developed that will contain all proposed names (Table 1). Although these nomenclatures may include names that are synonyms, they may not be recognised as such. At least the spellings and authorities of these names are more likely to be correct.
3. Deposit voucher specimens in a well-curated, publicly accessible natural history collection whenever possible and appropriate. These specimens enable the identifications to be checked by other researchers and provide material for future taxonomic research. Of course, type specimens of species new to science must be lodged in museums or herbaria as a prerequisite for publication.

Some journals stipulate that a species name should include the author and year it was described when first mentioned in the paper. This is highly recommended because there are instances of the same names (homonyms) for distinct species, usually across animals, plants and/or fungi, but also among unresolved names within the same code of nomenclature. For example, *Viola montana* L. is a distinct species of plant from *Viola montana* Juzepzuk (Roskov, pers. comm.). Another example is that the name of a cherry tree, *Prunus padus,* may refer to up to any one of six different species of plant unless its authority (i.e., the author(s) who named it first) is provided (Pankhurst, 2009): (1) *P. padus* L. is an accepted name; (2) *P. padus* sensu Ledeb. and (3) *P. padus* sensu Nakai are provisionally accepted names of probably new species but their authors misinterpreted them as *P. padus* L.; (4) *P. padus* Brandis is a synonym for *Prunus cornuta* (Wall. ex Royle) Steud.; and at the infraspecies level, (5) *P. padus* var. *japonica* Miq. is a synonym for *Prunus grayana* Maxim, and (6) *P. padus* var. *cornuta* (Wall. ex Royle) A. Henry is a synonym for *P. cornuta* var. *cornuta* (Wall. ex Royle) A. Henry.

Including a species' author and year is not a substitute for citing the identification guide because it does not clarify what publication was actually used in the species identification process. It is also not uncommon for some regional species guides (especially popular guides) to include only some of the species likely to be found in the region and not indicate that there may be other similar species, and sometimes have labelled images and descriptions with the wrong species name.

3. Data formatting

Regardless of how a data set is published, it must contain sufficient information about the data to make them intelligible to users.

Please cite this article in press as: Costello, M.J., Wieczorek, J. Best practice for biodiversity data management and publication. Biol. Conserv. (2013), http://dx.doi.org/10.1016/j.biocon.2013.10.018

Table 1
Resources useful for checking species names, publishing biodiversity data, and metadata standards.

Species nomenclature		
Algae	AlgaeBase	www.algaebase.org
All species	Species 2000, Catalogue of Life	www.sp2000.org
Animal names	ZooBank	www.zoobank.org
Flowering plants	The International Plant Names Index (IPNI)	www.ipni.org
Fungi	MycoBank	www.mycobank.org
Marine species	World Register of Marine Species	www.marinespecies.org
Specialist data publishers		
Biological images	Morphbank	www.morphbank.net
Data associated with published papers	Dryad	www.datadryad.org
Environmental data	PANGAEA	www.pangaea.de
Fossils	Paleobiology Database	paleodb.org
Genetic	GenBank	www.ncbi.nlm.nih.gov/genbank
Marine species distribution	Ocean Biogeographic Information System	www.iobis.org
Phylogenetic knowledge	TreeBASE	treebase.org
Species distribution data	Global Biodiversity Information Facility	data.gbif.org
Vegetation plots	VegBank	vegbank.org
Vertebrate specimen collections	VertNet	vertnet.org
Protein structure	Protein Data Bank	www.wwpdb.org
Metadata standards		
Geographic Metadata	ISO 19115:2003	www.iso.org/iso/catalogue_detail.htm?csnumber=26020
Ecological Metadata Language Andelman et al. (2013)	The Knowledge Network for Biodiversity	knb.ecoinformatics.org/software/eml
Dublin Core Metadata Initiative	DCMI Specifications	dublincore.org/specifications/
Darwin Core Archives	Global Biodiversity Information Facility	www.gbif.org/informatics/standards-and-tools/publishing-data/data-standards/darwin-core-archives/
Darwin Core Standards http://www.tdwg.org/standards/450/	Mapping Darwin Core to Old Versions	http://rs.tdwg.org/dwc/terms/history/versions/index.htm#dwcobis

Several standards exist for data-set-level metadata including Dublin Core, Ecological Metadata Language and ISO 19115 (Table 1). Cross-mapping between standards is possible. These standards all generally include information on who, what, where, when, why and how the data were collected, a summary of what the data set contains, and contact details for further information. It is useful to keep some guidelines in mind in preparing data for publication (Table 2) (Cook et al., 2001; Borer et al., 2009). It is helpful to readers to provide a recommended citation for the data set in the conventional form of author (or editor)-year-title format, to which will be added the address of the data archive. Michener et al. (1997) and Hook et al. (2010) provide guidelines for geospatial data. In addition to the actual data, if the data analysis process ('work-flow') was documented, as is increasingly the case using the statistical software R, then include this file with the data. This will enable other researchers to replicate the analysis accurately, perhaps with additional or improved data.

4. Where to publish data

4.1. Occurrence data

Though data will vary greatly between studies, the place and time of collection or observation of species are commonly included. These data fall into the category of Darwin Core "Occurrence" data (Wieczorek et al., 2009, 2012). GBIF and its participants publish over 400 million occurrence records of specimens and observations nearly 12,000 data sets, covering over 1.4 million confirmed species supplied by 700 organisations from over 40 countries (Table 1). This is achieved by organising the data into a standard format using Darwin Core Archives (Wieczorek et al., 2012; Table 1) to facilitate automatic data integration into an aggregate database that can be searched online. Data may be published into GBIF through its regional nodes (including national biodiversity data centres), PANGAEA, and thematic nodes such as VertNet (Constable et al., 2010; Table 1) and the Ocean Biogeographic Information System (OBIS) (Costello et al., 2007; Table 1).

4.2. Beyond occurrence data

At present, the data schema used by GBIF is based on species occurrences in samples or collections, such as checklists, plus extensions that can elaborate on these core aspects of the data. If the content of the data fall outside the capabilities of a particular data aggregator such as GBIF, they can be published through less structured repositories such as PANGAEA, Dryad, and/or archives managed by national data centres (Table 1).

Remy-Zephir et al. (2012) provide an example of data publication and archiving in Dryad of maps suitable for Geographical Information Systems associated with a paper in *Biological Conservation* about habitat mapping in a marine reserve (Leleu et al., 2012). Similarly, data from papers in this journal related to ant diversity in forests (Bihn et al., 2008a,b), nest predation in songbirds (Remeš et al., 2012a,b), and monitor lizards (Luke et al., 2013a,b) have been archived in Dryad.

There are other kinds of biodiversity data, including data for fossils, genetics, and images, among others. These have their own community databases, such as GenBank, the Paleobiology Database (Alroy 2013), Protein Data Bank (Berman et al., 2000), TreeBase (Piel et al., 2013), MorphBank, and VegBank (Peet et al., 2012) (Table 1). Other databases may emerge for particular kinds of data and, if they satisfy the requirement of making data openly and permanently accessible online, they may be more appropriate for data publishing than the above-mentioned options.

4.3. Data centres

An important benefit of publishing through data centres (e.g., PANGAEA, OBIS) is that their staff will check the format of the data set before making it available online. However, if the data centre

Please cite this article in press as: Costello, M.J., Wieczorek, J. Best practice for biodiversity data management and publication. Biol. Conserv. (2013), http://dx.doi.org/10.1016/j.biocon.2013.10.018

Table 2
Guidelines to aid preparation of data for publication. See Cook et al. (2001), Borer et al. (2009), and Hook et al. (2010) for more details.

- The names of files should be informative (e.g., identify kind of data, study, place or date, author's surname)
- Follow conventions and standards where available. Be consistent in style, structure, and content throughout
- Define any abbreviations, acronyms, and key terms used
- Be aware of file encoding. Whenever possible use a common standard file encoding such as ASCII or UTF-8 for text and data files. This will maximize the likelihood that software will render the contents correctly
- Do not mix up the number zero with the letter O
- Absent data values for numeric fields should not be zero. Instead, use a null instead, except for specific applications that require a well-documented default value that is otherwise not suitable as a data value
- Use the appropriate number of decimal places in measurements and other numerical values to reflect their precision. For example, if measurements were precise to the nearest cm, then report the values as integers. If measurements were to the nearest tenth of a cm, then report, for example 1.2 cm rather than 1.23 cm or 1 cm
- Keep all digits in original geographic coordinates. Use a separate measure such as the Darwin Core 'coordinate Precision' to capture the notion of precision. Not doing so can introduce unwanted errors (Chapman and Wieczorek, 2006)
- Use open common standard file formats whenever possible. For example, common preferences for non-relational data (data that could be managed in a simple spreadsheet) are comma-separated value (CSV) and tab-delimited (TXT) files
- When storing sub-sets of data using a flat-file format (rather than relational database files), use distinct files (or spreadsheets) for the different sub-sets. For example, summary statistics should be in a separate file from the raw data. This is to promote consistent formatting within a single document and avoid extra processing to get the data of interest
- Cells in tables should have just one kind of information (e.g., a number and its units should occupy separate fields)
- Each data record (e.g., row of information) must have a unique, and preferably globally unique, identifier within the data set to distinguish it from all other data records. This identifier can be used unambiguously in perpetuity to refer to a record, for example, when someone wants to refer to a record that is particularly important or questionable. To achieve local uniqueness, the identifier can be a distinct combination of attributes of the data, such an abbreviation indicating the sampled location, date, and/or sample number
- Text notes should have a distinct field, and not be interspersed with numerical data
- Rows and columns in Tables should have unambiguous and clear headings, the first column is typically the identifier for what a row is; e.g., a species name in an ecological data set
- Related publications should be cited so readers can see how the data were previously used. Typically, these provide more metadata and examples of its use
- Use the W3C and ISO standard date format (ISO 8601:2004(E)) of year-month-day, e.g. 2013-03-20. This makes dates unambiguous and facilitates sorting
- Specify time zones if appropriate (e.g., 2013-03-01T14:07-0600)
- Use the International Standard of Units (SI units)
- Use international standards such as the Darwin Core (rs.tdwg.org/dwc) for organising data. This will make subsequent publication (e.g., through the Ocean Biogeographic Information System - OBIS or the Global Biodiversity Information Facility - GBIF) more straightforward
- Check that the data set is complete and free of errors using a variety of tests such as the following: (a) review the list of distinct values for a given field to be sure they match expected content; (b) calculate summary statistics such as maximum and minimum for numerical fields and check to see that these match expectations; (c) check that the first, last, and some random rows and columns match the original datasheets or field notes to be sure misalignments did not occur; and (d) map latitude-longitude coordinates to make a visual check that they make sense

Please cite this article in press as: Costello, M.J., Wieczorek, J. Best practice for biodiversity data management and publication. Biol. Conserv. (2013), http://dx.doi.org/10.1016/j.biocon.2013.10.018

does not provide online access, then the requirements for making the data publicly available are not met.

Another advantage of most data centres is that they assign a unique identifier to enable machine tracking of the data set. PANGAEA and Dryad do this by assigning a Digital Object Identifier (DOI), a system widely used to identify papers in journals. DOI are managed in a central registry so the identified object can always be located. Schrag et al. (1995a, 1995b) demonstrated how to link original data to their published research.

5. Data publishing priorities

Publishing through data centres does not prevent the data from being made available through multiple sources, including the web sites of journals, authors, or their institutions. However, these options are not adequate for data archiving. Journal websites are not always ideal repositories for data because many do not provide unrestricted access via the Internet (open access, or OA) and they are not necessarily permanent archives (Santos et al., 2005; Vision, 2010). Thus, we can prioritise options for permanent open-access online archives in which to publish biodiversity data:

1. Through a system that enables integration of the data with other similar data sets (e.g., GBIF, VertNet, OBIS, GenBank).
2. In a repository where staff do quality checks on the data (e.g., PANGAEA, World Data Centres, National Oceanographic Data Centres).
3. In other repositories (e.g., Dryad, some institutional repositories).

Note, Option 1 above relates to particular standardised data formats used by specialist public data aggregators. In contrast, the archives in Option 2 can accept any kind of data and these data centres provide some editorial scrutiny of the data and metadata. While Option 3 can also accept any kind of data, it is the responsibility of the authors to check its quality before submission. At present, none of these data set publishing mechanisms can be considered peer-reviewed (Costello et al., 2013b). If data have been made available to the referees of a paper, then the referees may look over the data to check that they appear correct. VegBank provides an interesting step in the right direction by allowing commentaries and recommended corrections to species identifications (Peet et al., 2012).

Another emerging option is to publish the data through journals that provide archiving services, such as Ecological Society of America's (ESA) *Data Papers* and *Ecological Monographs*. Data papers in Pensoft journals (e.g., Zookeys, Photokeys) may be generated automatically from the GBIF Integrated Publishing Toolkit (Penev et al., 2011). These 'data papers' have the additional advantage of data being peer-reviewed before publication (Costello et al., 2013b).

6. Sensitive data

There may be exceptional cases where releasing all the data about a threatened species may expose the species' population to illegal collecting or hunting. In these cases, the information may be generalised or withheld to safeguard the species location, yet still be made available in confidence to conservation authorities (Chapman and Grafton, 2008). The Darwin Core standard defines two terms suitable to alert data consumers of the existence of additional data that may not be in the public domain (Wieczorek et al., 2009). If a paper uses commercially sensitive data, such as fishery catch data, then the data used must still be made public at the time of publication or the results of the analyses are not reproducible or independently verifiable. The sensitivity of these data may also decline over time such that they can be released subsequently.

It is not necessarily the expectation that all the data from a study will be published immediately; only that data supporting a published paper will be. Data may be published at any time, whether or not its custodians plan to use it in further publications. Indeed, one could envisage scientists who specialise in data collection and publication (e.g., monitoring data) and leave much of the analysis to others. In these cases, they may prefer to publish 'data papers' in specialist journals that can be referenced by studies using the data.

7. Conclusions

Opportunities for biodiversity data publication are increasing. The key aspect of data publication is that data should be permanently archived in an online, open-access repository (permitting use without conditions) with sufficient metadata for potential users to understand how and why they were collected. Ideally, the repository should conduct independent quality checks on the data and enable them to be integrated with similar data. Linking such data sets with published papers that used the data provides more confidence in the quality of the data and background on its provenance and thereby better informs about potential further uses (Costello et al., 2013b). Though we have provided some examples of repositories that authors should consider, they should be cognisant of new opportunities that emerge, as this is a rapidly growing activity.

Acknowledgements

We thank Ward Appeltans, Zeenatul Basher, Vishwas Chavan, Kendall Clements, Richard Corlett, William K. Michener, Lyubomir Penev, Tim Robertson, Yuri Roskov, Éamonn Ó Tuama, Leen Vandepitte, Charley Waters, Zhi-Qiang Zhang, the editor and anonymous referees for helpful comments that improved this article.

References

Alroy, J., 2013. Paleobiology Database. <http://paleodb.org> (Accessed 26.02.13).

Andelman, S., Arzberger P., Berkley C., Blankman D., Brunt J., Eddins O., Helly J., Higgins D., Jones C., Jones M., Nottrott R., Rajasekar A., Reichman J., Schildhauer M., Sutton D., Tao J., Waide B., Willig M., 2013. The Knowledge Network for Biodiversity. <http://knb.ecoinformatics.org> (Accessed 29.08.13).

Anon, 2013. Joint Data Archiving Policy. datadryad.org/pages/jdap. (Accessed 16.08.13).

Appeltans, W., Bouchet, P., Boxshall, G.A., Fauchald, K., Gordon, D.P., Hoeksema, B.W., Poore, G.C.B., van Soest, R.W.M., Stöhr, S., Walter, T.C., Costello, M.J., (Eds.), 2013. World Register of Marine Species. <http://www.marinespecies.org> (Accessed 25.02.13).

Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N., Bourne, P.E., 2000. The protein data bank. Nucl. Acids Res. 28, 235–242. http://dx.doi.org/10.1093/nar/28.1.235.

Bihn, J.H., Verhaagh, M., Brändle, M., Brandl, R., 2008a. Do secondary forests act as refuges for old growth forest animals? Recovery of ant diversity in the Atlantic forest of Brazil. Biol. Conserv. 141 (3), 733–743. http://dx.doi.org/10.1016/j.biocon.2007.12.028.

Bihn, J.H., Verhaagh, M., Brändle, M., Brandl, R., 2008b. Data from: do secondary forests act as refuges for old growth forest animals? Recovery of ant diversity in the Atlantic forest of Brazil. Dryad Dig. Repos. http://dx.doi.org/10.5061/dryad.3h12j8d3.

Bisby, F., Roskov, Y., Culham, A., Orrell, T., Nicolson, D., Paglinawan, L., Bailly, N., Appeltans, W., Kirk, P., Bourgoin, T., Baillargeon, G., Ouvrard, D. (Eds.), 2012. Species 2000 & ITIS Catalogue of Life, 2012 Annual Checklist. DVD. Species 2000: Reading, UK (Accessed 25.02.13).

Borer, E.T., Seabloom, E.W., Jones, M.B., Schildhauer, M., 2009. Some simple guidelines for data management. Bull. Ecol. Soc. Amer, vol. 90 (2), pp. 205–214. <http://www.nceas.ucsb.edu/files/computing/EffectiveDataMgmt.pdf> (Accessed 16.08.13).

Chapman, A.D., Grafton O., 2008. Guide to Best Practices for Generalising Primary Species-Occurrence Data, version 1.0. Global Biodiversity Information Facility, Copenhagen, 27pp. <http://www.gbif.org/orc/?doc_id=1233> (Accessed 16.08.13).

Chapman, A.D., Wieczorek, J. (Eds.), 2006. Guide to Best Practices for Georeferencing. Copenhagen: Global Biodiversity Information Facility, Copenhagen. 90pp. http://www.gbif.org/orc/?doc_id=1288> (Accessed 16.08.13).

Constable, H., Guralnick, R., Wieczorek, J., Spencer, C., Peterson, A.T., 2010. The VertNet Steering Committee, 2010. VertNet: A New Model for Biodiversity Data Sharing. PLoS Biol. 8 (2), e1000309. http://dx.doi.org/10.1371/journal.pbio.1000309.

Cook, R.B., Olson, R.J., Kanciruk, P., Hook, L.A., 2001. Best practices for preparing ecological data sets to share and archive. Bull. Ecol. Soc. Am. 82 (2), 138–141.

Costello, M.J., 2009. Motivation of online data publication. Bioscience 59 (5), 418–427. http://dx.doi.org/10.1525/bio.2009.59.5.9.

Costello, M.J., Vanden, Berghe E., 2006. "Ocean Biodiversity Informatics" enabling a new era in marine biology research and management. Mar. Ecol. Prog. Ser. 316, 203–214.

Costello, M.J., Stocks, K., Zhang, Y., Grassle, J.F., Fautin, D.G., 2007. About the Ocean Biogeographic Information System. <http://hdl.handle.net/2292/5236> (Accessed 25.02.13).

Costello, M.J., Bouchet, P., Boxshall, G., Fauchald, K., Gordon, D.P., Hoeksema, B.W., Poore, G.C.B., van Soest, R.W.M., Stöhr, S., Walter, T.C., Vanhoorne, B., Decock, W., Appeltans, W., 2013a. Global coordination and standardisation in marine biodiversity through the World Register of Marine Species (WoRMS) and related databases. PLoS ONE 8 (1), e51629. http://dx.doi.org/10.1371/journal.pone.0051629.

Costello, M.J., Michener, W.K., Gahegan, M., Zhang, Z.-Q., Bourne, P., 2013b. Biodiversity data should be published, cited and peer-reviewed. Trends Ecol. Evol. 28 (8), 454–461. http://dx.doi.org/10.1016/j.tree.2013.05.002.

Costello, M.J., Berendsohn, W., Appeltans, W., de Jong, Y., Mees, J., Segers, H., Froese, R., Edwards, M., Bisby, F.A., 2013c. Strategies for the sustainability of online open-access biodiversity databases. Biol. Conserv. http://dx.doi.org/10.1016/j.biocon.2013.07.042.

Frankham, R., Ballou, J.D., Dudash, M.R., Eldridge, M.D.B., Fenster, C.B., Lacy, R.C., Mendelson III., J.R., Porton, I.J., Ralls, K., Ryder, O.A., 2012. Implications of different species concepts for conserving biodiversity. Biol. Conserv. 153, 25–31.

Franz, N.M., Peet, R.K., Weakley, A.S., 2008. On the use of taxonomic concepts in support of biodiversity research and taxonomy. In: Wheeler, Q.D. (Ed.), The New Taxonomy, Systematics Association Special Volume Series 74. Taylor & Francis, Boca Raton, FL, pp. 61–84.

Guiry, M.D., Guiry, G.M., 2013. AlgaeBase. National University of Ireland, Galway. http://www.algaebase.org. (Accessed 25.03.13).

Hagedorn, G., Mietchen, D., Morris, R.A., Agosti, D., Penev, L., Berendsohn, W.G., Hobern, D., 2013. Creative Commons licenses and the non-commercial condition: implications for the re-use of biodiversity information. Zookeys 150, 127–149. http://dx.doi.org/10.3897/zookeys.150.2189.

Hampton, S.E., Strasser, C.A., Tewksbury, J.J., Gram, W.K., Budden, A.E., Batcheller, A.L., Duke, C.S., Porter, J.H., 2013. Big data and the future of ecology. Front. Ecol. Environ. 11, 156–162. http://dx.doi.org/10.1890/120103.

Hook, L.A., Santhana Vannan, S.K., Beaty, T.W., Cook, R.B., Wilson, B.E., 2010. Best practices for preparing environmental data sets to share and archive. Oak Ridge National Laboratory Distributed Active Archive Center, Oak Ridge, Tennessee, U.S.A. doi:10.3334/ORNLDAAC/BestPractices-2010. <http://daac.ornl.gov/PI/BestPractices-2010.pdf> (Accessed 16.08.13).

Lawrence, P.A., 2008. Lost in publication: how measurement harms science. Ethics Sci. Environ. Polit. 8, 9–11. http://dx.doi.org/10.3354/esep00079.

Leleu, K., Remy-Zephir, B., Grace, R., Costello, M.J., 2012. Mapping habitat change after 30 years in a marine reserve shows how fishing can alter ecosystem structure. Biol. Conserv. 155, 193–201. http://dx.doi.org/10.1016/j.biocon.2012.05.009.

Luke, W.J., Siler, C.D., Linkem, C.W., Diesmos, A.C., Diesmos, M.L., Sy, E., Brown, R.M., 2013a. Dragons in our midst: phyloforensics of illegally traded Southeast Asian monitor lizards. Biol. Conserv. 159, 7–15. http://dx.doi.org/10.1016/j.biocon.2012.10.013.

Luke, W.J., Siler, C.D., Linkem, C.W., Diesmos, A.C., Diesmos, M.L., Sy, E., Brown, R.M., 2013b. Data from: Dragons in our midst: phyloforensics of illegally traded Southeast Asian monitor lizards. Dryad Dig. Repos. http://dx.doi.org/10.5061/dryad.d2f79.

Michener, W.K., Brunt, J.W., Helly, J.J., Kirchner, T.B., Stafford, S.G., 1997. Nongeospatial metadata for the ecological sciences. Ecol. Appl. 7 (1), 330–342.

Pankhurst, R.J., 2009. Rosaceae. In: Wilson, K.L., Berendsohn, W.G. (Eds.), 2012. IOPI Global Plant Checklist (version 10.0, Aug 2007). In: Bisby, F., Roskov, Y., Culham, A., Orrell, T., Nicolson, D., Paglinawan, L., Bailly, N., Appeltans, W., Kirk, P., Bourgoin, T., Baillargeon, G., Ouvrard, D., (Eds.), 2012. Species 2000 & ITIS Catalogue of Life, 2012 Annual Checklist. <http://www.catalogueoflife.org/col> (Accessed 25.03.13).

Peet, R.K., Lee, M.T., Jennings, M.D., Faber-Langendoen, D., 2012. VegBank – a permanent, open-access archive for vegetation-plot data. Biodivers. Ecol. 4, 233–241.

Penev, L., Mietchen, D., Chavan, V., Hagedorn, G., Remsen, D., Smith, V., Shotton, D., 2011. Pensoft data publishing policies and guidelines for biodiversity data. Pensoft Publishers, <http://www.pensoft.net/J_FILES/Pensoft_Data_Publishing_Policies_and_Guidelines.pdf> (Accessed 16.08.13).

Piel, W.H., Vos, R.A., Auman, J. Chan, L., Dominus, M.J., Gapeyev, V., Gujral, M., Guo, Y., Lapp, H., Ruan, J., Shyket, H., Tannen, V., 2013. TreeBASE: a database of phylogenetic knowledge. Version 2. <http://treebase.org> (Accessed 29.08.13).

Reichman, J.H., Uhlir, P.F., 2003. A contractually reconstructed Research Commons for scientific data in a highly protectionist intellectual property environment. 66 Law and Contemporary Problems, pp. 315–462. <http://scholarship.law.duke.edu/faculty_scholarship/1515>. (Accessed 16.08.13).

Reichman, O.J., Jones, M.B., Schildhauer, M.P., 2011. Challenges and opportunities of Open Data in ecology. Science 331, 703–705. http://dx.doi.org/10.1126/science.1197962.

Remeš, V., Matysioková, B., Cockburn, A., 2012a. Nest predation in New Zealand songbirds: exotic predators, introduced prey and long-term changes in predation risk. Biol. Conserv. 148 (1), 54–60. http://dx.doi.org/10.1016/j.biocon.2012.01.063.

Remeš, V., Matysioková, B., Cockburn, A., 2012b. Data from: nest predation in New Zealand songbirds: exotic predators, introduced prey and long-term changes in predation risk. Dryad Dig. Repos. http://dx.doi.org/10.5061/dryad.6q81t4m4.

Remy-Zephir, B., Leleu, K., Grace, R., Costello, M.J., 2012. Data from: mapping habitat change after 30 years in a marine reserve shows how fishing can alter ecosystem structure. Dryad Dig. Repos. http://dx.doi.org/10.5061/dryad.6vr28.

Robert, V., Stegehuis, G., Stalpers, J., 2005. The MycoBank engine and related databases. <http://www.mycobank.org>. (Accessed 1.03.13).

Santos, C., Blake, J., States, D.J., 2005. Supplementary data need to be kept in public repositories. Nature 438, 738. http://dx.doi.org/10.1038/438738a.

Schrag, D.P., DePaolo, D.J., Richter, F.M., 1995a. Reconstructing past sea surface temperatures: correcting for diagenesis of bulk marine carbonate. Geochim. Cosmochim. Acta 59 (11), 2265–2278. http://dx.doi.org/10.1016/0016-7037(95)00105-9.

Schrag, D.P., DePaolo D.J., Richter F.M., 1995b. Age and oxygen isotope data for bulk carbonate from DSDP Sites 41–366 and 72–516 and ODP Sites 111–677A and 130–807. doi:10.1594/PANGAEA.707926.

Steinke, D., Hanner R., 2011. The FISH-BOL collaborators' protocol. Mitochondrial DNA 22(S1), pp. 10–14. doi:10.3109/19401736.2010.536538.

Vision, T.J., 2010. Open data and the social contract of scientific publishing. Bioscience 60, 330–331. http://dx.doi.org/10.1525/bio.2010.60.5.2.

Whitlock, M.C., 2011. Data archiving in ecology and evolution: best practices. Trends Ecol. Evol. 26, 61–65. http://dx.doi.org/10.1016/j.tree.2010.11.006.

Wieczorek, J., Döring, M., De Giovanni, R., Robertson, T., Vieglais, D., 2009. Darwin Core Terms: A quick reference guide. <http://rs.tdwg.org/dwc/terms> (Accessed 16.08.13).

Wieczorek, J., Bloom, D., Guralnick, R., Blum, S., Döring, M., Giovanni, R., Robertson, T., Vieglais, D., 2012. Darwin core: an evolving community-developed biodiversity data standard. PLoS One 7 (1), e29715. http://dx.doi.org/10.1371/journal.pone.0029715.

Please cite this article in press as: Costello, M.J., Wieczorek, J. Best practice for biodiversity data management and publication. Biol. Conserv. (2013), http://dx.doi.org/10.1016/j.biocon.2013.10.018