Journal of Applied Ecology

DR JAMES W PEARCE-HIGGINS (Orcid ID : 0000-0003-1341-5080)
DR ROB A ROBINSON (Orcid ID : 0000-0003-0504-9906)

PROFESSOR JEREMY D WILSON (Orcid ID : 0000-0001-7485-5878)

Article type : Commentary

Handling Editor: Yolanda Wiersma

Overcoming the challenges of public data archiving for citizen science biodiversity recording and monitoring schemes.

Pearce-Higgins, J.W.[1], Baillie, S.R.[1], Boughey K.[2], Bourn, N. A. D[3], Foppen, R.P.B.[4,5], Gillings, S.[1], Gregory, R.D.[6,7], Hunt, T.[8], Jiguet, F.[9], Lehikoinen, A.[10], Musgrove, A.J.[1], Robinson, R.A.[1], Roy, D.B.[11], Siriwardena, G.M.[1], Walker, K.J.[12], Wilson, J.D.[13]

[1]British Trust for Ornithology, The Nunnery, Thetford, Norfolk, IP24 2PU, UK.

[2]Bat Conservation Trust, London, SE11 5RD, UK

[3]Butterfly Conservation, Wareham BH20 5QP, UK

[4]Sovon Dutch Centre for Field Ornithology, Nijmegen, The Netherlands

[5]Department of Animal Ecology and Ecophysiology, Radboud University, Nijmegen, The Netherlands

[6]RSPB Centre for Conservation Science, The Lodge, Sandy, Beds, SG19 2DL, UK

[7]Centre for Biodiversity & Environment Research, Department of Genetics, Evolution and Environment, University College London, Darwin Building, Gower Street, London, WC1E 6BT, UK

[8]Association of Local Environmental Records Centres.

[9]Centre d'Ecologie et des Sciences de la Conservation UMR7204 MNHN-CNRS-Sorbonne Université, CP 135, 43 Rue Buffon, 75005 Paris, France.

[10]Finnish Museum of Natural History, University of Helsinki, 00014, Finland

[11]Biological Records Centre, Centre for Ecology and Hydrology, Wallingford, Oxfordshire, OX10 8BB, UK

[12]Botanical Society of Britain and Ireland (BSBI), Room 14 Bridge House, 1-2 Station Bridge, Harrogate, North Yorkshire HG1 1SS, UK

[13]RSPB Centre for Conservation Science, RSPB Scotland, 2 Lochside View, Edinburgh Park, Edinburgh, EH12 9DH, UK

Corresponding author: james.pearce-higgins@bto.org

Running title: Open biodiversity data and citizen science

Abstract

1. Public data archiving (PDA) is widely advocated as a means of achieving open data standards, leading to improved data preservation, increased scientific reproducibility and transparency, as well as additional data use.
2. PDA was primarily conceived to archive data from short-term, single-purpose scientific studies. It is now more widely applied, including to large-scale citizen science biodiversity recording and monitoring schemes which combine the efforts of volunteers with professional scientists.
3. This may affect the financial security of such schemes by reducing income from data and analytical services. Communication between scheme organizers and researchers may be

disrupted, reducing scientific quality and impeding scheme development. It may also have an impact on the participation of some volunteers.

4. *Synthesis and applications*. In response to the challenges of PDA for citizen science biodiversity recording and monitoring schemes, the archive function of scheme organisations should be better recognised by those promoting open data principles. Increased financial support from the public sector or from commercial or academic data users may offset financial risk. Those in favour of public data archiving should do more to facilitate communication between non-scheme users and the originating schemes, whilst a more flexible approach to data archiving may be required to address potential impacts on volunteer participation.

Keywords**:** citizen science, open data, biological recording, monitoring scheme, volunteer, data archiving, quality assurance

Introduction

This paper is prompted by the challenges posed by the interaction of three current trends: the growth of 'big data' in the information age, the shift towards an open data culture, and increasing volumes of biodiversity records collected by volunteer observers and collated and analysed within citizen science schemes. As representatives of eleven organisations running and supporting such schemes across a range of taxa and countries in Europe, based on our experience we summarise these challenges for our volunteers and the schemes they contribute to, and suggest some potential solutions.

Initially, we want to emphasise our support for data being well-organised and archived for posterity, and widely shared for collaborations that address large-scale questions (Hampton *et al.* 2013). Doing so enables scientific practice to become more transparent as part of a move towards 'open science'

to ensure the reproducibility of scientific findings, and to use data to tackle multiple problems (Vision 2010, Nosek *et al.* 2015). This need for open science is greatest in relation to short-term research grants that address specific questions where, for example, over 50% of ecological data is inaccessible to others (Hampton *et al.* 2013).

Open data and public data archiving

At its broadest, open data can be freely used, re-used and redistributed by anyone (http://opendatahandbook.org/guide/en/what-is-open-data/), although precise definitions of what constitutes 'open' data vary (Groom *et al.* 2016). The open archiving of data has advantages for scientific practice: ensuring data preservation, facilitating independent validation of scientific results, increasing public access to data and information, and providing the opportunity for improved scientific and educational return on research funding through data re-use (e.g. Roche *et al.* 2014).

Public data archiving (PDA), where data are made freely available on demand through recognised data repositories (Reichman *et al.* 2011, Roche *et al.* 2014), is increasingly mandated by funders and journals to promote open data (Hampton *et al.* 2013, Kenall *et al.* 2014). However, more flexible models of open data access including user registration (e.g. Sullivan *et al.* 2014), multi-year embargos (Roche *et al.* 2014) and multiple open data licences (Groom *et al.* 2016) are available (Table 1). As custodians of large European citizen science datasets, we are supportive of the principle of open data. However, as this is a rapidly developing and uncertain area (e.g. Reichman *et al.* 2011, Borgnam 2012, Roche *et al.* 2014, Mills *et al.* 2015, Whitlock *et al.* 2016, Bowser *et al.* 2017), we wish to highlight potential risks to the maintenance of dynamic and robust citizen science biodiversity monitoring schemes if mandatory PDA is the key mechanism for facilitating open data.

Citizen science biodiversity recording and monitoring

In a biodiversity context, citizen science has a long history, stemming from personal interests, enjoyment of, and concern for, nature (Lawrence & Turnhout 2010, Ganzevoort *et al.* 2017). As the insights to be gained from such pursuits have been realised, they have become increasingly coordinated by professionals (Pocock *et al.* 2015). Now, most data describing species' distribution and abundance, especially in developed countries, are collected by volunteers and coordinated within schemes.

The effort contributed is substantial. For example, in the UK, 85 national schemes collect biological records from over 70,000 recorders to cover 11,431 taxa (Roy *et al.* 2014, Pocock *et al.* 2015), whilst some 15,000 recorders are active in the Netherlands (Ganzevoort *et al.* 2017) and 26,000 in France. Over 8,500 volunteers spend 171,000 hours per year monitoring UK abundance trends in 250 bird, 56 butterfly and 20 mammal species (Way & Robinson 2013). In Britain and Ireland, over 2,600 trained individuals annually ring over 1,000,000 birds, often through repeated effort from specific sites that in aggregate provide data for national or continent-scale analyses (Baillie & Schaub 2009).

Although volunteer contributions to data collection are great, and often outweigh scheme costs (JNCC's £1 million annual spend on UK terrestrial biodiversity monitoring levers an estimated £20 million volunteer contribution; Robinson *et al.* 2018), the organisation required for successful schemes can be significant (Wiggins 2013). Schemes divide into the following four types (Pocock *et al.* 2015) which may be affected by PDA in different ways.

MASS PARTICIPATION RECORDING

Biological recording of species can engage large numbers of people with minimal professional support, particularly if the species are highly detectable and identifiable (e.g. iNaturalist). Such schemes often have an open data ethos where submitted records are available for others to view and verify, likely to maximise classification accuracy and participation (Lukyanenko *et al.* 2015), and

are most similar to other, non-ecological citizen science schemes that successfully operate an open data policy (e.g. www.cocorahs.org, www.patientslikeme.com). These data may be useable by conservationists and scientists (e.g. Roy *et al.* 2015), providing added motivation for contributors. Although generally associated with significant biases and analytical challenges (Isaac *et al.* 2014), data from mass participation recording can complement other more structured approaches (Dennis *et al.* 2017).

UNSTRUCTURED BIOLOGICAL RECORDING

Taxa that require a high degree of volunteer expertise to gather or identify records are unsuitable for mass participation recording. Schemes that collect data on these taxa are often reliant upon a limited number of contributors and verifiers (Pocock *et al.* 2015), but yield records that are appropriately checked and which generally provide the only large-scale and long-term data for the species concerned. Given their uniqueness, the commercial provision of data and information can be an important source of funding.

FOCUSSED RECORDING

The introduction of structure to the collection of biological data, for example through specific sampling protocols, spatial targeting of recording (e.g. biological atlases) and the collection of null records or complete lists, can significantly enhance the information to be gained (Sullivan *et al.* 2014). This requires communication of instructions to volunteers, bespoke software and web development to guide and enable data capture, scientific analyses of scheme data for reporting, research and development, and communication of the results to volunteers, funders and wider society. The resourcing required to achieve this may be partly derived from charges for data or analytical services.

MONITORING SCHEMES

Monitoring schemes similarly require volunteers to follow standard protocols to track changes in the abundance, demography and behaviour of individual species, usually based on repeat visits to specific locations, which can be self-selected or allocated according to a sampling design. These designs help to maximise the scientific value of the data collected whilst ensuring appropriate levels of participation. The resulting high quality data archived by organisations running monitoring schemes, supported by their rigorous analysis to maximise inference, provide an important resource for research, and to inform policy and management decisions. The necessary investment by scheme organisations and their scientists may be partly offset by data charges for commercial use and bespoke data access for research.

Potential implications of PDA

A model of open data that is based upon mandatory PDA, where archived data are openly available to anyone on demand, may have negative implications for 1) scheme funding, 2) scientific quality and 3) volunteer participation.

SCHEME FUNDING

Depending on the form of data licence adopted, mandatory PDA may threaten the viability of schemes which rely on charges for data or analytical services. Although there are no generalizable statistics available on the value of such charges, which vary between schemes, it is clear that where schemes make significant use of commercial income, the favouring of a Creative Commons – Attribution licence (CC BY) over a non-commercial licence (e.g. CC BY-NC), will reduce funding. Even a CC BY-NC licence may increase competition for analytical funding with researchers not involved in scheme organisation and development, particularly impacting schemes that have a business model

of also securing research grants to analyse the data collected. This could threaten the long-term viability of schemes where costs to the host institution are partly offset by the benefits of enhanced data access for research by their own scientists, potentially weakening the link between scheme development and research. Research scientists at scheme organisations have a successful track record of designing and developing long-term recording and monitoring schemes (e.g. Baillie & Schaub 2009, Balmer *et al.* 2013, Sullivan *et al.* 2014, Pocock *et al.* 2015), a role that cannot be readily undertaken by external scientists alone. Given the current difficulty of funding long-term monitoring (Birkhead 2014), the risk that PDA may pose to the business model of any scheme should be assessed, and where significant, alternative models for data sharing adopted or further sources of income secured.

RESEARCH QUALITY

Mandatory PDA with data freely available to download on demand will reduce communication with non-scheme users, making scheme organisers less aware of independent analyses being undertaken and reducing volunteer feedback. This has a number of consequences.

First, by enabling researchers to extract data about which they have no personal knowledge, PDA increases the risk of inappropriate manipulation, analysis and interpretation which could affect the reputation of the scheme and the researchers themselves (e.g. Telleria *et al.* 2015) and require effort from scheme organisers to correct (e.g. Siriwardena *et al.* 2014). Although not commonly regarded as a major issue in other disciplines where PDA is the norm, the complexities of ecological data can make archiving for independent use particularly challenging (Kenall *et al.* 2014, Mills *et al.* 2015).

Secondly, PDA without communication can lead to competition between research groups. While this may raise standards and validate findings, in a sector where resources are scarce, unnecessary duplication is inefficient.

Thirdly, any disconnection between scheme organisers and academic researchers reduces the opportunities for schemes being supported by collaborative proposals. This will also impact the potential for future scheme development co-designed by researchers and scheme organisers.

Fourthly, PDA reduces the potential for schemes to promote the work undertaken through such collaboration. This provides important feedback to volunteers and other scheme supporters, and increases the societal impact of the research undertaken.

Fifthly, although the PDA of data associated with specific analyses can provide important validation of scientific reproducibility, this can result in multiple archives of effectively the same data, each associated with different time-points, levels of processing, funding agreements or publications (Kenall *et al.* 2014). Such duplication makes it difficult for external researchers to identify and use the most appropriate dataset for their purposes.

By ensuring data preservation, facilitating independent validation of scientific results, enabling access to data and information, and providing those data for collaboration, the current data archiving function of scheme organisations already fulfils many open data aims (Table 1). They ensure data are 'preserved and usable for decades in the future', a key motivation for organisations mandating PDA (e.g. http://besjournals.onlinelibrary.wiley.com/hub//data_archiving_policy.html). By producing a constantly expanding data-set, these schemes do not easily fit into a PDA model designed primarily to archive scientific data collected initially for a single purpose, that otherwise would be lost.

VOLUNTEER PARTICIPATION

Successful schemes depend upon the coordination, support, and often training of volunteers, whose attitudes towards PDA will affect their motivation. Concerns that open access to data about rare or threatened species can put conservation objectives at risk may lead volunteers to withhold data from schemes that contribute to public archives (Yang & Chan 2015, Eaton *et al.* 2015). PDA of

citizen science data may also conflict with aspects of volunteer identity and privacy (Bowser *et al.* 2017). Whilst knowing the identity of submitting individuals is essential for verification and allows observer effects to be modelled, if not carefully managed, it may lead to individual movements or residential addresses being inferred.

Although standard practice can reduce these risks, volunteer participation may be more influenced by perception than actual risk. Information about the attitudes and practice of biological recorders towards data sharing suggest they are variable. For example, only 12% of Dutch biological recorders support unconditional use of their data (Ganzevoort *et al.* 2017), although 83% of biological recorders in Scotland would be happy for biological records to be openly available (Wilson *et al.* in prep). Based upon figures of actual data availability, 64% of the Botanical Society of Britain and Ireland (BSBI) vice-county (regional) recorders provide access to full resolution (100m) plant records, but 94% to 2km resolution records, whilst 8% of roost data collected by the UK National Bat Monitoring Programme are not openly available at any resolution. Although only 0.09% of c. 19 million records submitted to the Britain & Ireland Bird Atlas 2007–11 (Balmer *et al.* 2013) were submitted confidentially at high spatial resolution, confidential records comprised 10% of the breeding records for 30 species, and more than 50% for six.

Many volunteers retain a sense of ownership of data they have submitted, particularly if significant investment was required, and will withhold them if they perceive those data may be used inappropriately (Ellis & Waterton 2005, Ganzevoort *et al.* 2017). The provision of data by citizen scientists to organisations undertaking biological recording or monitoring is underpinned by trust (Martin *et al.* 2016), which requires those organisations to take account of volunteer perspectives when making decisions about data sharing, and to be open about potential data uses (Groom *et al.* 2016, Ganzevoort *et al.* 2017). Given the variable attitudes and practices towards open data outlined above, mandatory PDA of high resolution biological records has the potential to disrupt that

relationship for some, and could result in a variable proportion of volunteers ceasing to submit data, creating a smaller, biased, dataset. These risks are greatest for schemes capturing data on the occurrence of rare species, or where a large proportion of data can be attributed to a small number of individuals.

Solutions

To summarise, the benefits of open data may be offset for citizen science biodiversity recording and monitoring schemes by negative impacts upon scheme funding, scientific quality and volunteer participation. Challenges of scheme funding will most impact organisations with the greatest costs that run focussed recording and monitoring schemes, whilst impacts on scientific quality potentially apply to all schemes. Challenges around volunteer participation are most likely to affect unstructured and focussed recording and monitoring schemes that require a high degree of volunteer expertise or effort, and schemes where a small number of individuals contributed most data. These challenges may be addressed by 1) recognising the archive function of scheme organisations, 2) increasing financial security for schemes, 3) facilitating communication between scheme organisers and non-scheme users and 4) implementing open data flexibly (Fig. 1).

RECOGNISING THE ARCHIVE FUNCTION OF SCHEME ORGANISATIONS

Many challenges of PDA result from a failure of those promoting open data principles to recognise that scheme organisations already provide a long-term archive of citizen science biodiversity data. Our concerns about PDA would be largely eliminated under an alternative open data model in which schemes operate a data request procedure compatible with open data principles, but where the identity and purpose of data download is known. PDA of data associated with specific publications could still occur, but with access restricted to the purpose of scientific validation only.

This combination of PDA of bespoke datasets for scientific transparency and a request procedure to access data for wider use, would deliver the requirements of open science whilst addressing two of our key challenges. The correspondence associated with data requests should improve both scientific quality and feedback to volunteers and scheme funders. By remaining central to the process, schemes can address legitimate volunteer concerns, whilst where appropriate, encouraging participants to adopt a more open approach to data. Through ongoing development, schemes will continue to provide resilient access to data for the future, for example in line with FAIR principles (Wilkinson *et al.* 2016). The main outstanding issue not addressed by this model is funding.

INCREASING THE SECURITY OF LONG-TERM FUNDING

Funding the costs of data archiving is a recognised responsibility of stakeholders that support scholarly publication (Vision 2010). As the maintenance of long-term monitoring schemes is dependent upon sustainable funding models (Groom *et al.* 2016), the risks that open data poses to scheme viability should be assessed by key stakeholders and funders of citizen science. In the case of publically-funded schemes, the provision of sufficient resources to fully support the infrastructure, research and development costs of a scheme would eliminate risk, enabling access through Creative Commons – Attribution (CC BY) or non-commercial (e.g. CC BY-NC) licenses, subject to concerns about communication and volunteer participation being addressed.

In the absence of full-funding, a more market-driven approach may be required where professional non-scheme data users contribute financially to schemes which make their databases openly available. Archives could facilitate more commercial income to schemes that contribute data through CC-BY licenses, by enabling access through commercial subscription and distributing the fees to contributing schemes. Academic researchers could support schemes whose data they require through research grants, which grant awarding bodies could incentivise when scoring proposals.

ENABLING COMMUNICATION BETWEEN SCHEME ORGANISATIONS AND NON-SCHEME USERS

Researchers using citizen science data should welcome communication with scheme organisers (Kenall *et al.* 2014). This would increase their research impact through feedback to volunteers, and more widely to members and supporters of scheme organisations, improving societal understanding of science. Communication would also encourage collaboration in overlapping areas of interest, and improve scientific rigour.

Those using open data should routinely offer due acknowledgement. Where these data are integral to the analysis undertaken, and a meaningful contribution has been made, this may include co-authorship of research papers for the scheme organiser, or where datasets originate from a small number of key individuals, the volunteers themselves. Journals could encourage communication between scheme organisers and non-scheme users by requiring authors to state upon submission that this has occurred. Data archives could fulfil the same function by routinely providing scheme organisers with the identity of those downloading data.

IMPLEMENTING OPEN DATA FLEXIBLY

There are legitimate cases where providing open data through PDA would be inappropriate, for example to protect vulnerable species or personal information. Although many existing open data policies recognise this, by scheme organisations remaining central to the facilitation of data access, rather than a public archive, all data remain available to that organisation for monitoring and conservation purposes, even if some records remain confidential to others.

When PDA is not possible, alternative approaches may achieve the societal benefits of open data without relying upon immediate open access. Firstly, as many archives allow, data could be deposited under embargo, enabling volunteers (some of whom may be contributing their

professional studies) and scheme organisers, to analyse and use these data for their own research without fear of competition, and potentially reducing the risk to vulnerable species. Up to a five-year embargo period would seem appropriate for long-term data, providing it does not conflict with species conservation needs (Mills *et al.* 2015, Whitlock *et al.* 2016). More recent data could, naturally, be requested.

Secondly, summarised, rather than raw information, could be archived (e.g. converting counts into biological records of presence), at coarser spatial or temporal resolution, or as derived data-products (e.g. modelled surfaces of occurrence rather than raw records). This would enable rapid access to sufficient data and information to satisfy many open science requirements, such as providing public access to easily interpretable scientific information, whilst limiting uses most likely to conflict with data ownership, species conservation and privacy (e.g. https://www.ndff.nl/english/). Although this may not enable the open validation of research from the analysis of capture-resolution data, data could be separately made available for that purpose.

Providing volunteer perspectives are protected, schemes are sufficiently resourced and there are mechanisms in place to facilitate communication and collaboration over data use, then schemes could better encourage an open data culture amongst data contributors without risk to their long-term viability, for example by engaging those volunteers in the design of data-use policies (Luyakenko *et al.* 2015), and offering appropriate data sharing licenses (Groom *et al.* 2016), such as CC BY or CC BY-NC, to participating volunteers.

Conclusions

Open data makes an important contribution towards open science but if driven by PDA, poses significant risks for many citizen science schemes, particularly those with substantial professional co-ordination or that are supported by a small number of key volunteers. These risks are reduced if the existing data archive function provided by schemes is recognised, and they remain the primary access route for scheme data, in line with open data principles. In some cases, this may require scheme development (Groom *et al.* 2016). PDA would be less challenging for schemes if they were financially secure, there were mechanisms to ensure communication between non-scheme data users and scheme organisers, and if volunteer interests were protected. If these suggestions are followed, access to most citizen science biodiversity data should become more open, without threatening long-term data supply.

Acknowledgements

We thank Chris Cheffings, Teresa Frost, Katherine Hawkins, Theo Verstrael and Quentin Groom for comments on previous versions of this manuscript, as well as from two anonymous referees, Yolanda Wiersma (Associate Editor) and Phil Stephens (Editor).

Data accessibility

Data have not been archived because this article does not contain data.

References

Baillie, S.R. & Schaub, M. (2009) Understanding changes in bird populations – the role of bird marking. *Ringing and Migration*, 24, 189-198.

Balmer, D.E., Gillings, S., Caffrey, B.J., Swann, R.L., Downie, I.S. & Fuller, R.J. (2013) *Bird Atlas 2007-11: the breeding and wintering birds of Britain and Ireland.* BTO Books, Thetford.

Birkhead, T. (2014) Stormy outlook for long-term ecology studies. *Nature*, 514, 405.

Bowser, A., Shilton, K., Preece, J. & Warrick, E. (2017) Accounting for Privacy in Citizen Science: Ethical Research in a Context of Openness. In Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing (CSCW '17). ACM, New York, NY, USA, 2124–2136. DOI: http://dx.doi.org/10.1145/2998181.2998305

Dennis, E.B., Morgan, B.J.T., Brereton, T.M., Roy, D.B. & Fox, R. (2017) Using citizen science butterfly counts to predict species population trends. *Conservation Biology*, 31, 1350-1361.

Eaton, J.A., Shepherd, C.R., Rheindt, F.E., Harris, J.B.C., van Balen, S.B., Wilcove, D.S. & Collar, N.J. (2015) Trade-driven extinctions and near-extinctions of avian taxa in Sundaic Indonesia. *Forktail,* 31, 1–12.

Ellis, R. & Waterton, C. (2005) Caught between the cartographic and the ethnographic imagination: the whereabouts of amateurs, professionals, and nature in knowing biodiversity. *Environment and Planning D: Society and Space*, 23, 673–693.

Ganzevoort, W., van den Born, R.J.G., Halffman, W. & Turnhout, S. (2017) Sharing biodiversity data: citizen scientists' concerns and motivations. *Biodiversity & Conservation,* 26, 2821-2837.

Groom, Q., Weatherdon, L. & Geijzendorffer, I.R. (2016) Is citizen science an open science in the case of biodiversity observations? *Journal of Applied Ecology*, 54, 612-617.

Hampton, S., Strasser, C.A., Tewksbury, J.J., Gram, W.K., Budden, A.E., Batcheller, A.L., … Porter, J.H. (2013) Big data and the future of ecology. *Frontiers in Ecology and the Environment,* 11, 156–162

Isaac, N.J.B., van Strien, A.J., August, T.A., de Zeeuw, M.P. & Roy, D.B. (2014) Statistics for citizen science: extracting signals from noisy ecological data. *Methods in Ecology & Evolution*, 5, 1052-1060.

Kenall, A., Harold, S. & Foote, C. (2014) An open future for ecological and evolutionary data? *BMC Ecology,* 14, 10

Lawrence, A. & Turnhout, E. (2010) Personal meaning in the public sphere: the standardisation and rationalisation of biodiversity data in the UK and the Netherlands. *Journal of Rural Studies,* 26, 353–360.

Lukyanenko, R., Parsons, J. & Wiersma, Y.F. (2015) Emerging problems of data quality in citizen science. *Conservation Biology* 30, 447-449.

.Martin, V.Y., Christidis, L. & Pecl, G.T. (2016) Public interest in marine citizen science: is there potential for growth? *Bioscience,* 66, 683–692.

Mills, J.A., Teplitsky, C., Arroyo, B., Charmantier, A., Becker, P.H., Birkhead, T.R., … Zedrosser, A. (2015) Archiving primary data: solutions for long-term studies. *Trends in Ecology & Evolution,* 30, 581-589

Nosek, B. A., Alter, G., Banks, G.C., Borsboom, D., Bowman, S.D., Breckler, S.J., … Yarkon, T. (2015) Promoting an open research culture. *Science,* 348, 1422-1425.

Pocock, M.J.O., Roy, H.E., Preston, C.D. & Roy, D.B. (2015) The Biological Records Centre: a pioneer of citizen science. *Biological Journal of the Linnean Society,* 115, 475-493

Reichman, O., Jones, M.B. & Schildhauer, M.P. (2011) Challenges and opportunities of open data in ecology. *Science*, 331, 703-705

Robinson, A., Newton, N. & Cheffings, C. (2018) *Terrestrial Evidence Review Part 1: The value of JNCC's Terrestrial Evidence Programme*. JNCC, Peterborough.

Roche, D.G., Lanfear, R., Binning, S.A., Haff, T.M., Schwanz, L.E., Cain, K.E., … Kruuk, L.E.B. (2014) Troubleshooting public data archiving: suggestions to increase participation. *PLoS Biology*, 12, e1001779

Roy, H.E., Rorke, S.L., Beckmann, B., Botham, M.S., Brown, P.M.J., Noble, D.G. …Walker, K.J. (2015). The contribution of volunteer recorders to our understanding of biological invasions. *Biological Journal of the Linnean Society*, 115, 678–689.

Roy, D.B., Harding, P.T., Preston, C.D. & Roy, H.E. (eds.) (2014). Celebrating 50 years of the Biological Records Centre. Centre for Ecology & Hydrology. ISBN: 978-1-906698-52-2

Siriwardena, G.M., Baillie, S.R., Fuller, R.J. & Robinson, R.A. (2014) How can functional space for farmland birds best be studied? A comment on Butler and Norris (2013). *Agriculture, Ecosystem and Environment,* 192, 8-11.

Sullivan, B. L., Aycrigg, J. L., Barry, J. H., Bonney, R. E., Bruns, N., Cooper, C.B., ... & Kelling, S. (2014)The eBird enterprise: an integrated approach to development and application of citizen science. *Biological Conservation,* 169, 31–40.

Tellería, J.L., Ramírez, Á & Aguirre, J.I. (2015) Retraction: 'Are European birds leaving traditional wintering grounds in the Mediterranean?' by José Luis Tellería, Álvaro Ramírez and José Ignacio Aguirre. *Journal of Avian Biology,* 46, 323-330.

Vision, T.J. (2010) Open data and the social contract of scientific publishing. *Bioscience,* 60, 330-331.

Way, L. & Robinson, A. (2013) Refreshing JNCC's surveillance and monitoring evidence role. JNCC13 D07. Available from http://jncc.defra.gov.uk/PDF/comm13%20D07.pdf

Wilson, E., Johnston, C., Edwards, L., Stroud, R., McLeod, C. and Bamforth, L. (in prep.). *The SBIF Review of the Biological Recording Infrastructure in Scotland: Questionnaire Findings.* Scottish Biodiversity Information Forum Commissioned Report No. 1.

Wilkinson, M. D., Dumontier, M., Aalsberg, I.J., Appleton, G., Axton, M., Baak, A… & Mons, B (2016) The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, 3, 160018

Whitlock, M.C., Bronstein, J.L., Bruna, E.M., Ellison, A.M., Fox, C.W., McPeek, M.A., … Shaw, R.G. (2016) A balanced data archiving policy for long-term studies. *Trends in Ecology & Evolution,* 31, 84-85

Wiggins, A. (2013) Free as in puppies: compensating for ICT constraints in citizen science. *CSCW* 13, Ferbruary 23-27.

Yang, J.H. & Chan, B.P. (2015) Two new species of the genus Goniurosaurus (Squamata: Sauria: Eublepharidae) from southern China. *Zootaxa,* 3980, 67-80.

Fig. 1. Suggested solutions to the challenges of PDA for citizen science biodiversity recording and monitoring schemes.

Table 1. A comparison of current access to biodiversity data, mandatory PDA and options for a more flexible model of open data provision.

	Current situation	**Mandatory PDA**	**Flexible open data options**
Level of access	Variable, but often requires data request. Summarised products sometimes available.	Capture-resolution data openly available.	Capture-resolution data openly available if not restricted for conservation / privacy concerns. Otherwise, PDA of simplified / summarised data available. Capture-resolution data may be requested.
Timescale of access	Variable. Summarised data provided within one or more years. Capture-resolution data sometimes available over varying timescale.	Immediate, although one-year embargo possible (rarely longer).	Up to five-year embargo of capture-resolution data. PDA of simplified / summarised data provided within one year.
Re-use restrictions	Data-use normally specific to each request.	CC-BY or CC BY-NC licences the norm.	Data-use specific to request or PDA with CC-BY or CC BY-NC licences. PDA of bespoke data available for scientific validation.
Cost of access	Variable. Can be free for academic or educational purposes. Charges may be required to support data	Free at point of access	Free access if scheme operation and development fully funded. Otherwise, non-scheme users incentivised to support

	extraction. Charges for commercial use usually applied.		schemes providing open data access through licence fees or grant applications.
Feedback to scheme organiser / volunteer	Facilitated through data request.	Limited	Facilitated through data request. Journals encourage communication between non-scheme users and schemes. Data archives provide data downloader identity to schemes.