Isaac, Nick J.B.; Pocock, Michael J.O. 2015. **Bias and information in biological records** [in special issue: Fifty years of the Biological Records Centre] *Biological Journal of the Linnean Society*, 115 (3). 522-531. 10.1111/bij.12532

Bias and information in biological records

Nick J.B. Isaac & Michael J.O. Pocock

Biological Records Centre, NERC Centre for Ecology & Hydrology, Maclean Building, Benson Lane,
Crowmarsh Gifford, Wallingford OX10 8BB.

Email: njbi@ceh.ac.uk

Abstract

Biological recording is in essence a very simple concept in which a record is the report of a species at a physical location at a certain time. The collation of these records into a dataset is a powerful approach to addressing large-scale questions about biodiversity change.

Records are collected by volunteers at times and places that suit them, leading to a variety of biases: uneven sampling over space and time, uneven sampling effort per visit and uneven detectability. These need to be controlled-for in statistical analyses that use biological records. In particular, the data are ‘presence-only’, and lack information the sampling protocol or intensity. Submitting ‘complete lists’ of all the species seen is one potential solution because the data can be treated as ‘presence-absence’ and detectability of each species can be statistically modelled.

The corollary of bias is that records vary in their ‘information content’. The information content is a measure of how much an individual record, or collection of records, contributes to reducing uncertainty in a parameter of interest. The information content of biological records varies, depending on the question to which the data are being applied.

We consider a set of hypothetical ‘syndromes’ of recording behaviour, each of which is characterised by different information content. We demonstrate how these concepts can be used to support the growth of a particular type of recording behaviour.

Approaches to recording are rapidly changing, especially with the growth of mass participation citizen science. We discuss how these developments present a range of challenges and opportunities for biological recording in the future.

Introduction

A biological record consists of four main pieces of information: the What, Where, When and Who of what was recorded. What refers to the identity of the species; Where is the spatial location; When is the date or, occasionally, the range of dates over which the record was collected; Who is the person who made the record. Individual records can provide evidence about the persistence of rare species, or the spread of invasive species, but most records contain very little information on their own. However, when collated into large databases, the information encoded in biological records is enormous (Hochachka *et al.*, 2011).

Biological records come from a wide range of sources, including systematic population monitoring, professional surveys and mass participation projects [Pocock et al., this volume]. In the UK, the term is most closely associated with activities of volunteers supported by the National Recording Schemes and Societies. These data are characterised by having neither a consistent structure nor a fixed sampling protocol. They constitute a mixture of both opportunistic records and focussed surveys by expert volunteers (e.g. filling gaps in distributions or targeting under-recorded regions). We focus on the properties of these data, and the issues around their use in scientific applications.

The enormous value of biological records for scientific applications is well known [Pocock et al., this volume; Powney & Isaac, this volume]. However, statements about this value are often qualified with references to the biases and limitations inherent in opportunistic volunteer-collected data (Bird *et al.*, 2014). Records are made only for species that were observed (they are often called 'presence-only' data), thus greatly limiting the inferences that may be drawn from them (Tingley & Beissinger, 2009). A deeper problem is that volunteer recorders are highly motivated by encounters with interesting wildlife. This means that the spatial and temporal patterns of recording, and hence records, are very different from the kind of stratified random sampling protocol that a professional ecologist might design (Tulloch, et al., 2012).

This 'recorder effort problem' (Hill, 2012; Prendergast et al., 1993) has limited the scope of scientific applications of biological records, as researchers sought ways to draw robust conclusions in the face of biased data. One common approach was to aggregate the data to some a spatio-temporal scale at which the biases in space and time might be argued to be averaged out, such as counting the number of occupied grid cells across the whole country within atlas periods (e.g. Telfer, Preston, & Rothery, 2002; Thomas et al., 2004). An alternative approach has been to identify and select subsets of records perceived to be free of bias, based on thresholds for data quality (Maes et al., 2012; Roy et al., 2012). Both approaches throw away most of the information contained in the records, and thresholds for data quality are necessarily subjective. Recently, a suite of statistical techniques has emerged that, rather than removing the bias, models the data collection process (MacKenzie, 2006; Szabo *et al.*, 2010; Hill, 2012; van Strien, van Swaay, & Termaat, 2013). A feature of many such methods is the grouping of records into sets that share a common time and place (e.g. the site visit), such that observations of some species can be used to infer a failure to report others. Adoption of these methods has been accompanied by a much greater sophistication in the range of scientific questions being addressed using biological records [Powney & Isaac, this volume].

Against the backdrop of these developments, we take a fresh look at the recorder effort problem. We introduce the concept of the information content of records data, and discuss some of the ways in which an information theoretic approach to biological records data could be useful. We speculate on how small changes in the way biological records are collected and stored would have large benefits in terms of the insights that could be gained from the data.

Biases in biological records

The fact that biological records data are biased is well known (Prendergast et al., 1993), but these biases have rarely been defined and quantified. Isaac et al. (2014) identified four major biases in records data: 1) uneven recording intensity over time, 2) uneven spatial coverage, 3) uneven sampling effort per visit, and 4) uneven detectability across space and time. In this section we explore the nature of these biases, with some examples from recording schemes in Great Britain.

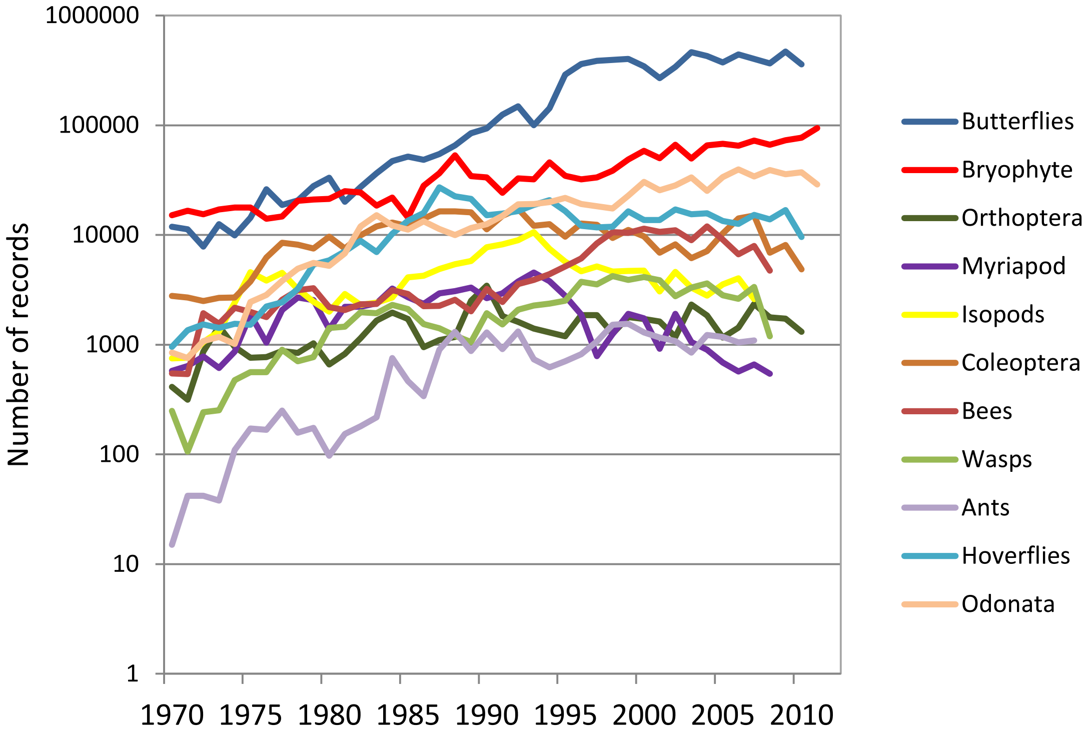

Figure 1: Number of records per year for 11 taxonomic groups in Great Britain, 1970-2010

Uneven sampling over time is the best-known form of bias. The number of records being generated has increased markedly in recent years, and for many groups the growth is approximately exponential (i.e. linear on a logarithmic scale; figure 1). As recording intensity increases, the number of grid cells that appear to be occupied is likely to increase, even for species with stable distributions (Telfer *et al.*, 2002). However, the growth of recording has not been smooth, but punctuated by bursts of activity associated with the production of distribution atlases.

Uneven spatial coverage occurs because most recorders tend to submit records within a well-defined geographic area, e.g. close to where they live or places they enjoy visiting. In addition, recording in the British Isles traditionally operates at a regional level, often within 'Watsonian vice-counties' (regions defined in 1852 to help standardise recording: Dandy, 1969). This often serves to homogenise recording effort within a vice-county (e.g. as associated with the production of local atlases for specific taxa) but can accentuate this apparently arbitrary variation in recording effort between vice-counties (figure 2). When records are aggregated into large datasets, the spatial intensity of recording effort varies markedly, reflecting the spatial distribution of where recorders visit (figure 3). Patterns of variation in species richness may therefore be accentuated because people choose to visit places which are especially diverse for their taxon of interest (Prendergast et al. 1993). Moreover, since recorders are active at different times, the spatial intensity of recording is uneven over time.

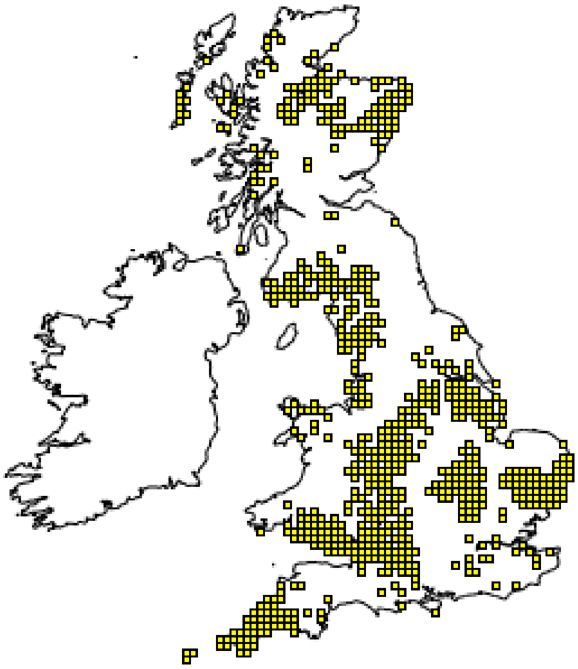

Figure 2: Distribution of moth Chrysoteuchia culmella (Lepidoptera: Crambidae), illustrating strong variation in recording intensity between vice-counties. Data from the National Biodiversity Network.

Sampling effort per visit refers to the degree to which any one set of records is an accurate reflection of the organisms that were actually present. Very few visits (i.e. a set of records from one site collected on the same day) produce a list of all the species that are present and active (i.e. ‘recordable’) on a site, because detection is less than perfect (Chen et al., 2012; Kéry et al., 2009). The sampling effort of a visit has two components: the intensity of the search effort, and the set of species that were surveyed. The first of these reflects the standard concept of a species accumulation curve: the more time you spend searching, the more species you find. The second component is best illustrated by the concept of the ‘complete list’: a complete list of bees means that a record was made for every bee species that was observed on a particular visit. Note that a complete list does not indicate that all species that were present and able to be observed were actually observed, but merely that all the species observed were reported. Incomplete lists occur because many biological records are not visit lists (i.e. there is no ‘survey’), but rather ‘incidental records’ (e.g. just the “interesting” species).

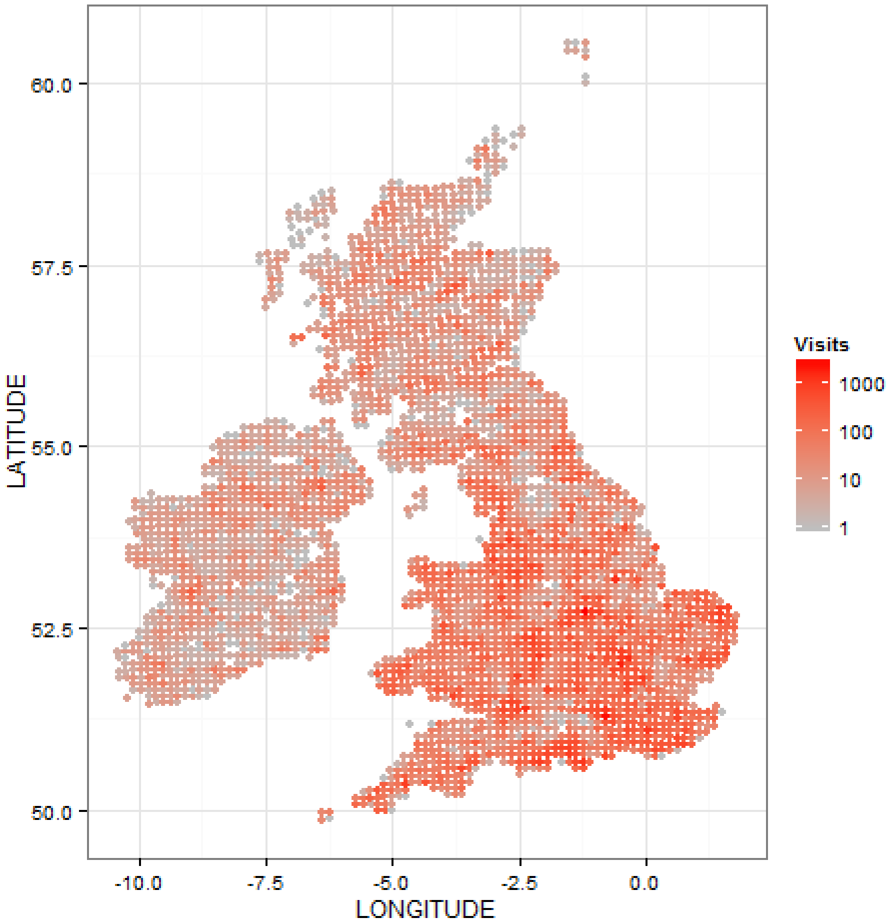

Figure 3: Map of the recording intensity of dragonflies and damselflies since 2000, as determined by the number of visits, i.e. unique combinations of place and date. Data from the British Dragonfly Society.

For complete lists, we would expect the identity of species on the list to be some function of their local population density (Royle & Nichols, 2003) and visual apparency. The length of the list is then an indication of the duration and intensity of the survey (Szabo *et al.*, 2010). Short lists in real datasets reflect both complete lists from brief surveys, complete lists where few species are recordable (e.g. in sites with low diversity, or at the beginning or end of activity periods for invertebrates) and collections of incidental records: real datasets contain an unknown mixture of these three broadly defined data types. There is no direct information on this phenomenon, but the high proportion of short lists among British recording schemes and societies (figure 4) implies that incidental recording is more common than complete listing.

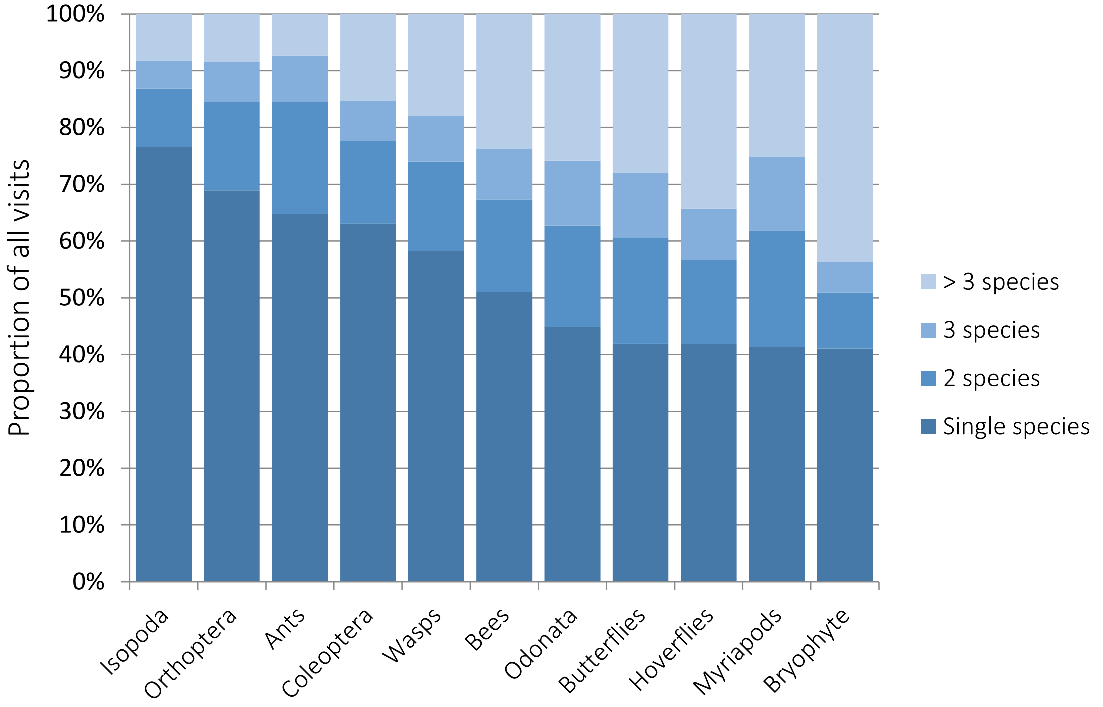

Figure 4: The distribution of list lengths among 11 datasets in Great Britain, 1970-2010. A visit is defined as a unique combination of date and grid reference. Taxa are ordered from highest to lowest proportions of single species visits. Note that grid references are used at the same precision as noted by the recorder, so records with high precision (e.g. 1 m^2) are treated as separate visits even if they were part of the same original survey. Converting the grid references to a common precision (e.g. 1 km^2) would reduce the proportion of single species visits among all groups, but we expect that the relative position of each group would be broadly similar.

As noted above, detection is less than perfect in most situations, and some species are easier to spot than others. The statistical basis for handling imperfect detection is well developed (MacKenzie, 2006), but problems can arise when detection varies substantially in space or time. For example, detectability may be strongly influenced by vegetation structure or successional change (Isaac *et al.*, 2011). Like the biases described above, detectability also reflects the behaviour of individual recorders and the tools available to them. The publication of a new field guide can facilitate the identification of species that were previously hard to separate. New survey methods can also contribute to changing recording effort. For example, moths such as the small ranunculus *Hecatera dysodea* (Lepidoptera: Noctuidae) and butterflies such as the brown hairstreak *Thecla betulae* (Lepidoptera: Lycaenidae), are much more easily recorded as larvae than as adults. As this kind of knowledge spreads, so does the number of records for these species. Similarly, the use of bat detectors by entomologists provides a substantial increase in the detectability of orthopterans that stridulate at ultrasonic frequencies (Benton, 2012).

The four biases of Isaac et al. (2014) are clearly not exhaustive. It is well known that appearances of rare and unusual species (especially migrants) attract 'twitchers' and producing large numbers of records in a short space of time, often producing multiple records of the same individual. Another potential bias comes from 'annual listing', in which only one (usually the first) record of a species is recorded for a site in any year. It is unknown how prevalent these kinds of recording practices are, or what their effects on statistical inference might be.

Information content

The adoption of sophisticated methods for analysing biological records has focused attention on the process by which records are collected. The word 'bias' has negative connotations, so this section explores the opposite concept: the amount of 'information' contained in the records (Munson *et al.*, 2010). Here, we are referring to the mathematical definition of 'information' (as in the discipline of information theory). One way of describing the information content of a record is its contribution to the accurate and precise estimation of any underlying parameter of interest, e.g. distribution size, trends in distribution, position of range margins etc. Since the data are heterogeneous, and because the biases reflect the way in which individual recorders behave when making observations about wildlife, it makes sense to think about the information content of an individual recorder, as well as that of the collated dataset.

For most scientific applications, the information content of a dataset is measured by the sample size (i.e. the number of records). Biological records data are different because information content can be expressed in at least three dimensions, each of which has several sub-dimensions or components. The key components of information content are related to the biases discussed above: the temporal footprint, the spatial footprint and the field sampling method. This multi-dimensionality means that the information content of any one set of records is not a fixed quantity, but is dependent on the question to which the data are being applied. For example, incidental records are sufficient to characterize unusual phenomena, such as first sightings or the spread of potentially harmful invasive species, but are much less useful to assess trends in species' status over time (van Strien et al., 2010).

The temporal footprint has several components, the simplest of which is the time span covered by the records. Whole datasets span many years, but with low recording intensity prior to 1980 (figure 1). For an individual recorder, the time span is simply the years over which a recorder is active. The second component is the rate at which records are produced each year. The time span and the recording rate determine the total output (number of visits and/or number of records) of a recorder. For nearly all recording schemes, output is highly skewed, and the information in the dataset is dominated by contributions of just a few individuals. This inequitable distribution of contribution (known as the Pareto distribution) is characterised as the 90:10 principle, and is exemplified by recording birds in the USA in which 90% of recorded visits are by 10% of observers (Wood *et al.*, 2011). The situation for British biological recording is often more extreme than this. For example, 14% of the total visits between 1970-2010 were made by the four most prolific Orthoptera recorders (of which there are over 2000), and half the visits were from just 38 recorders. Among ants, bees, wasps, centipedes and millipedes, the number of recorders contributing half the visits since 1970 is ten or fewer. The third temporal component is the phenological footprint (c.f. Bishop et al., 2013): i.e. is the recorder active all year, or only during specific times (e.g. summer holidays)?

The spatial footprint simply measures the geographic spread of records: some recorders travel widely but most record in places close to where they live (figure 5). Some public-facing citizen science projects, such as the *RSPB Big Garden BirdWatch,* encourage contributors to sample wildlife from a very restricted area (i.e. their own gardens), so the spatial footprint of each participant is very small.

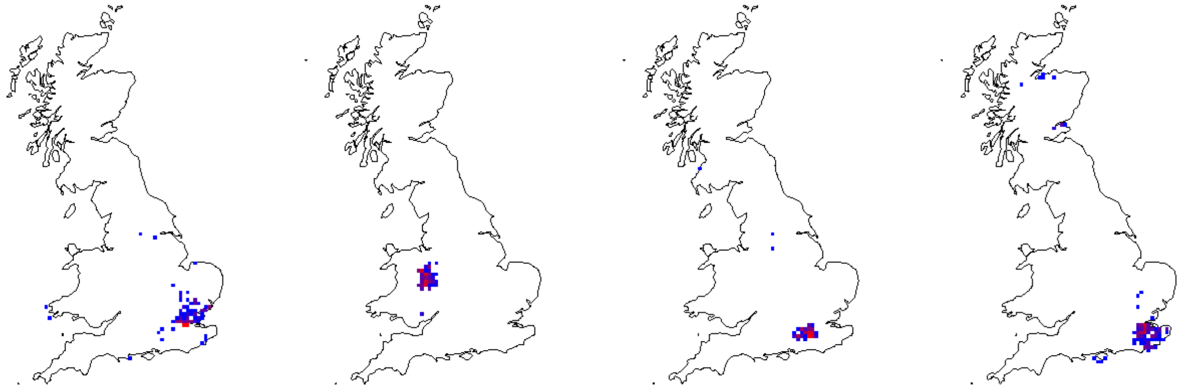

Figure 5: Recording footprints of four moderately prolific wasp recorders in Great Britain, 1970-2010. Red colours indicate grid cells with the highest number of visits; blue colours indicate a single visit.

The field sampling method refers to the process of recording on individual site visits: whether the records come from systematic searches (information-rich) or whether they are mostly incidental records of rare and charismatic species (information-poor). Of particular interest is the notion of a 'complete list', which is used by the BirdTrack database to indicate that a particular set of records constitutes all the species that were observed (Baillie et al., 2006). As noted above, 'complete' refers to the lack of deliberate omissions. These different field sampling methods can be placed within a pyramid of increasing information: the pyramidal shape reflects a suspicion that information-poor strategies contribute a far greater proportion of biological records than do information-rich ones (figure 6). Complete lists are an improvement on incidental records, because they contain information about non-detections. The systematic surveys are complete lists derived from some kind of search protocol: this could be as simple as a minimum time spent searching, or the use of a particular method. The notion of a fixed site means that records gathered at different points in time are directly comparable with one another. Repeated visits within years (on fixed sites) deliver the highest form of information, because they allow detectability to be estimated directly (MacKenzie, 2006).

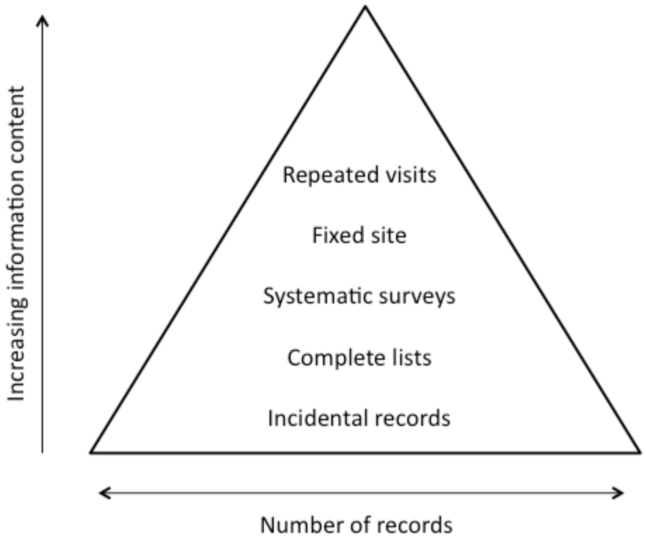

Figure 6: A conceptual representation of the quality (information content) and quantity of records under different sampling strategies.

A further attribute of information content reflects variation in taxonomic expertise and field skills. Common and easy-to-identify species are likely to be over-represented in the records submitted by less-experienced recorders. These factors mean that the less-experienced recorders are likely to have lower list lengths from a thorough search, and so contribute less information. Experienced recorders know what they are looking for and where to look: many rare invertebrates have specific microclimatic requirements, and this kind of natural history knowledge is the difference between easily detectable and impossible to find (S.P.M. Roberts, pers comm.). Search image is especially important for very small organisms, or those with characteristic flight behaviour. Taxonomic identification of some organisms takes years to master. For some invertebrates identification is only possible by microscopy of genitalia, which requires killing voucher specimens, access to a microscope and skill in specimen preparation. For field-based recorders these species will either be recorded as an aggregate group or simply ignored. These factors mean that the recorders differ greatly in the potential list length that might result from a thorough search.

Of course, we have assumed that information in the records (the What, Where and When) is correct. Mis-identifications provide mis-information and so reduce information content. Data verification is therefore an essential step in the collation of these datasets. Volunteer experts and organisers of national recording schemes have a vital, and often under-appreciated, role in verifying records [Pocock et al., this issue], which supports excellent quality science.

Different research applications place different demands on the data. When mapping species richness, we should be primarily concerned about the spatial footprint of the data. For estimating trends in species' status, we should be mindful of the temporal pattern of recording. We might also wish to use biological records to measure turnover in community composition, is which case the sampling strategy is likely to be a key concern [REF Barwell, Isaac & Kunin, in press].

Information theory posits that the amount of information contained in a dataset is not merely an arbitrary concept, but rather an estimable quantity. Shannon's entropy is widely-used as a measure of alpha diversity (Jost, 2006), and it could useful for measuring aspects of the information content in biological recording datasets. For example, it could be used quantify the cross-taxon variation seen in our figures 1 and 4. We could further quantify the spatial and temporal footprints of individual recorders using the standard deviations in latitude, longitude and date. Other statistics could be derived to measure other properties of interest, such as the characteristic phenology, list length and predilection for rare or common species. Developing statistics of this nature would make it possible to formally compare the information content of different datasets (c.f. Munson et al., 2010) for a range of different scientific and policy-relevant applications.

Characterising & Motivating Recorders

Increasing the information content in biological records data requires an investment of time and money to motivate and train volunteer recorders. By identifying the motivations of recorders and the traits of information-rich data, schemes and societies can more effectively allocate resources for training to address their goals. For example, would it be more effective to provide taxonomic skills to large numbers of novice recorders, or train mid-level recorders (already committed and proficient) in concepts around sampling effort and repeatability? In other words, is it better to broaden the base of the information pyramid, or to encourage existing participants to move up within the pyramid? If the aim is to produce data that is minimally biased and maximally informative, it may be necessary to supplement the records by paying recorders (or professional surveyors) to visit under-recorded parts of the country.

As we have seen, recorders differ greatly in the information contained within their records. In part this reflects the fact that records come into schemes from a variety of sources including regional surveys, targeted surveys (including ecological consultancies surveying for rare species), incidental observations and, increasingly, citizen science programs. We believe that much of the variation in information content can be explained by the motivations and characteristic behaviours of individual recorders, as well as their taxonomic expertise.

It could be possible to use the characteristic traits of the recording process to categorise different types (or 'syndromes') of recorder based on their pattern of recording. From this it would be possible to assess how much information each type of recorder provides to the overall dataset for any particular potential use of the dataset (Munson et al. 2013). Table 1 defines the traits of seven hypothetical syndromes in terms of their motivations, behaviours and the information profile of their records.

The degree to which these syndromes reflect reality is likely to vary from one recording scheme to another. For example, we suspect that datasets for taxonomic groups with small numbers of recorders (e.g. Auchenorrhyncha) consist of an unusually high proportion of Taxon Specialists. Conversely, we expect that charismatic groups such as butterflies and birds contain large quantities of data from Casual Recorders. This contrast highlights the fact that simply counting the number of records in a dataset provides a poor indication of whether a dataset is suitable for addressing any given research question.

		Hypothetical recorder syndrome				
Trait	Relevance to information content	Taxon specialist	Patch/county specialist	General naturalist	Casual recorder	Pan-lister
Complete lists?	An indication of the typical effort per survey	Yes	Yes	Varies	No	No
Coverage of 'rare' species	Predilection for reporting unusual sightings	Varies	Varies	Low	Low	High
Coverage of difficult species	Taxonomic expertise	High	High	Low	low	Varies
Length of activity of reporting	Temporal footprint	High	High	High	Varies	Varies
Frequency of recording	Productivity & consistency	High	high	varies	Low	High
Spatial variation in recording	Spatial footprint of the data	High	low	varies	varies	High

Variation in recording across taxa	Consistency of recording across taxa (taxonomic specialist versus jack-of-all-trades)	low	low	High	High	High

Table 1. Traits of recorders that could be influential in describing different recorder 'profiles' or 'syndromes'. A range of potential profiles have been identified.

This concept of participant 'syndromes' is receiving increasing interest in citizen science (Furtado et al., 2013; Ponciano et al., 2014), particularly around increasing rates of participation and data flow for online citizen science (in which participation can be accurately assessed). It has some similarities to the recording pyramid (figure 6), but provides a more multi-dimensional view of this data. A key goal in these types of projects is to consider how participants can be motivated to participate in a way that is typical of another syndrome in order to increase the information content of the whole dataset (Furtado *et al.*, 2013). In the context of biological recording, this might be persuading Casual Recorders to start recording complete lists (Baillie *et al.*, 2006; Wood *et al.*, 2011). Patch Watchers are likely to be motivated by a sense of place and local pride, and they could be persuaded to undertake some recording of under-recorded taxa in that location. County or Taxon-specialists may be more likely to be motivated to consistently record in apparently under-recorded grid cells than other syndromes, even if, in some instances, they are simply recording common species in places that are relatively poor in biodiversity. Some recorders might be motivated to collect more and better data by introducing a competitive or game-based element to recording (Greenhill et al., 2014; Nov et al., 2014), although this strategy requires caution because others would be repelled by this idea.

The principle of identifying recorder syndromes is fairly straightforward, but the practice will be challenging, especially if classification is fuzzy (because no one will perfectly fit a single 'syndrome') or because recorders' syndromes change over time. A big limitation is that the way records are currently collected, collated and curated makes impossible to unambiguously identify unique recorders. The use of unique usernames presents an obvious solution for data submitted to online systems (e.g. iRecord, iSpot or eBird), but using this information has a number of issues regarding privacy (Bowser et al., 2013) that have not yet been fully resolved. Further complexities arise when records are attributed to multiple individuals (e.g. when on a society's field trip or during a 'bioblitz'), although the proportion of such records is likely to be small in most datasets.

Modelling assumptions & Metadata

Until recently, the 'presence-only' nature of biological records presented serious problems for scientific applications of biological records: it would be naïve to assume that failure to record a species indicated its absence (Tingley & Beissinger, 2009). Modern analytical techniques treat these pseudoabsences as data, by employing a conceptual or statistical model of the recording process. This feature is most clearly expressed in occupancy-detection modelling (van Strien *et al.*, 2013; Isaac *et al.*, 2014). Failure to detect is inferred from records of other species, which in turn assumes that species are recorded as an assemblage. In effect, the assumption is that recorders are ticking species off a pre-defined list of all potentially recordable species in the assemblage. As discussed, differences in recorder motivation, skill and behaviour mean that the list of potentially recordable species varies markedly between recorders and visits. The extent of this variation in space and time is therefore a critical issue for the robustness of scientific inferences from biological records (Isaac et al. 2014).

As discussed above, recorders differ in the degree to which they record rare, common and difficult-to-identify species, as well as in their propensity to report complete lists (table 1). Thus, biological records data are extremely heterogeneous, comprised of numerous sub-datasets each with a different but unknown check-lists form which the records are drawn. For example, a list of hoverfly species could be drawn from the whole fauna, the flower-visiting species, or the subset that is easy to identify? What hope, then, that sophisticated statistical analyses will ever deliver substantive and robust insights from biological records data? Whilst the heterogeneity in human behaviour is baffling, it may be possible to model it statistically. The heterogeneous behaviour of human recorders is analogous to a situation commonly encountered in mark-recapture studies, where some animals are much more (or much less) prone to recapture than others, leading to biased population size estimates (Otis *et al.*, 1978). The solution was to assume that assume that animals are drawn from two (or more) populations each with a characteristic capture probability (Pledger, 2000). These 'mixture models' effectively solved the heterogeneity problem at the cost of just a few parameters: in the case of recorders, these parameters would express the probability that an individual recorder (or visit) is a complete or incomplete list.

The assumption that species are recorded as an assemblage is essentially an attempt to reverse-engineer the data collection process. As implied above, that data collection process is often known, and could itself be recorded as metadata. The simplest example is the 'complete list' checkbox employed by Birdtrack (Baillie *et al.*, 2006). For a citizen science project or targeted survey, the species on the checklist are often known, but this information is not retained when records are collated into recording schemes, the NBN and GBIF. Small amounts of metadata about survey methods and scope would add structure to biological records data, and would contribute greatly to the scientific inferences that could be drawn from them, e.g. through the application of hierarchical models which explicitly take account of the recording process (Pagel et al. 2014). Smartphone apps and other technology have enormous potential to harvest metadata at the point of data collection with little or no effort from recorders [August et al. this volume].

Concluding remarks

Biological recording is changing fast. The modelling techniques are becoming more sophisticated, as are the range of scientific and policy-relevant applications [Powney & Isaac this volume]. In parallel, new technologies are changing the way that records are collected and stored [August et al. this volume; Pocock et al., this issue]. The enormous potential of biological records data is increasingly valued by agencies with statutory responsibilities to conserve and report on the status of biodiversity both in the UK (i.e. Natural England, Joint Nature Conservation Committee, Scottish Natural Heritage etc) and internationally (Danielsen *et al.*, 2014). These agencies and funding bodies see enormous potential in engaging the enthusiasm of volunteers; it is a cost-efficient method for gathering data in the face of limited budgets. For this reason new schemes, such as the nascent National Plant Monitoring Scheme [Pescott, this volume], are likely to involve contributions from volunteers in some shape or form. The growth of smartphone apps for recording wildlife and the growth of public-facing citizen science programs mean that data collected in the future are likely to have different biases and different 'information content' than records from the recent past. By understanding the various sources of bias and the characteristics of information-rich data, we will be able to make better use of biological records in policy, conservation and science.

Acknowledgements

The concepts in this paper have emerged from detailed discussions with many users and producers of biological records. In particular, we are grateful to Elizabeth Boakes, Steve Cham, Mike Edwards, Mark Hill, Marc Kéry, Georgina Mace, Roger Morris, Andy Musgrove, Steve Prentice, Chris Preston, Stuart Roberts, Dave Smallshire, Bill Sutherland, Pam Taylor, Chris Thomas, Arco van Strien, one anonymous reviewer and all our colleagues within the Biological Records Centre for stimulating discussions on this topic, and for comments on previous versions of the manuscript.

References

Baillie SR, Balmer DE, Downie IS, Wright KHM. **2006**. Migration Watch: an Internet survey to monitor spring migration in Britain and Ireland. *Journal of Ornithology* **147**: 254–259.

Benton T. **2012**. *Grasshoppers and Crickets (Collins New Naturalist 120)*. HarperCollins.

Bird TJ, Bates AE, Lefcheck JS, Hill N a., Thomson RJ, Edgar GJ, Stuart-Smith RD, Wotherspoon S, Krkosek M, Stuart-Smith JF, *et al.* **2014**. Statistical solutions for error and bias in global citizen science datasets. *Biological Conservation* **173**: 144–154.

Bishop TR, Botham MS, Fox R, Leather SR, Chapman DS, Oliver TH. **2013**. The utility of distribution data in predicting phenology (R Freckleton, Ed.). *Methods in Ecology and Evolution* **4**: 1024–1032.

Bowser A, Wiggins A, Stevenson RD. **2013**. *Data Policies for Public Participation in Scientific Research: A Primer*. Ithaca, NY.

Chen G, Kéry M, Plattner M, Ma K, Gardner B. **2012**. Imperfect detection is the rule rather than the exception in plant distribution studies. *J Ecol*: n/a–n/a.

Dandy JE. **1969**. *Watsonian Vice Counties of Great Britain*. The Ray Society, London.

Danielsen F, Pirhofer-Walzl K, Adrian TP, Kapijimpanga DR, Burgess ND, Jensen PM, Bonney R, Funder M, Landa A, Levermann N, *et al.* **2014**. Linking Public Participation in Scientific Research to the Indicators and Needs of International Environmental Agreements. *Conservation Letters* **7**: 12–24.

Furtado A, Andrade N, Oliveira N, Brasileiro F. **2013**. Contributor profiles, their dynamics, and their importance in five q&a sites. Proceedings of the 2013 conference on Computer supported cooperative work - CSCW '13. New York, New York, USA: ACM Press, 1237.

Greenhill A, Holmes K, Lintott C, Simmons B, Masters K, Cox J, Graham G. **2014**. Playing with science : gamised aspects of gamification found on the Online Citizen Science Project - Zooniverse.

Hill MO. **2012**. Local frequency as a key to interpreting species occurrence data when recording effort is not known. *Methods in Ecology and Evolution* **3**: 195–205.

Hochachka WM, Fink D, Hutchinson RA, Sheldon D, Wong WK, Kelling S. **2011**. Data-intensive science applied to broad-scale citizen science. *Trends in ecology & evolution* **27**: 130–137.

Isaac NJB, Cruickshanks KL, Weddle AM, Marcus Rowcliffe J, Brereton TM, Dennis RLH, Shuker DM, Thomas CD. **2011**. Distance sampling and the challenge of monitoring butterfly populations. *Methods in Ecology and Evolution* **2**: 585–594.

Isaac NJB, van Strien AJ, August TA, de Zeeuw MP, Roy DB. **2014**. Statistics for citizen science: extracting signals of change from noisy ecological data. *Methods in Ecology and Evolution* **5**: 1052–1060.

Jost L. **2006**. Entropy and diversity. *Oikos* **113**: 363–375.

Kéry M, Dorazio RM, Soldaat L, van Strien A, Zuiderwijk A, Royle JA. **2009**. Trend estimation in populations with imperfect detection. *Journal of Applied Ecology* **46**: 1163–1172.

MacKenzie DI. **2006**. *Occupancy estimation and modeling: inferring patterns and dynamics of species occurrence*. Academic Press.

Maes D, Vanreusel W, Jacobs I, Berwaerts K, Van Dyck H. **2012**. Applying IUCN Red List criteria at a small regional level: A test case with butterflies in Flanders (north Belgium). *Biological Conservation* **145**: 258–266.

Munson MA, Caruana R, Fink D, Hochachka WM, Iliff M, Rosenberg K V., Sheldon D, Sullivan BL, Wood C, Kelling S. **2010**. A method for measuring the relative information content of data from different monitoring protocols. *Methods in Ecology and Evolution*: no–no.

Nov O, Arazy O, Anderson D. **2014**. Scientists@Home: what drives the quantity and quality of online citizen science participation? (J Bar-Ilan, Ed.). *PloS one* **9**: e90375.

Otis D, Burnham K, White G, Anderson D. **1978**. Statistical inference from capture data on closed animal populations. *Wildlife Monographs* **61**: 1–135.

Pledger S. **2000**. Unified Maximum Likelihood Estimates for Closed Capture-Recapture Models Using Mixtures. *Biometrics* **56**: 434–442.

Ponciano L, Brasileiro F, Simpson R, Smith A. **2014**. Volunteers' Engagement in Human Computation Astronomy Projects. *Computing in Science & Engineering*: 1–1.

Prendergast J, Wood S, Lawton J, Eversham B. **1993**. Correcting for variation in recording effort in analyses of diversity hotspots. *Biodiversity Letters* **1**: 39–53.

Roy HE, Adriaens T, Isaac NJB, Kenis M, Martin GS, Brown PMJ, Hautier L, Frost R, Zindel R, Vlaenderen J Van, *et al.* **2012**. Invasive alien predator causes rapid declines of native European ladybirds. *Diversity and Distributions* **18**: 717–725.

Royle JA, Nichols JD. **2003**. Estimating abundance from repeated presence–absence data or point counts. *Ecology* **84**: 777–790.

Van Strien AJ, Termaat T, Groenendijk D, Mensing V, Kéry M. **2010**. Site-occupancy models may offer new opportunities for dragonfly monitoring based on daily species lists. *Basic and Applied Ecology* **11**: 495–503.

Van Strien AJ, van Swaay CAM, Termaat T. **2013**. Opportunistic citizen science data of animal species produce reliable estimates of distribution trends if analysed with occupancy models. *Journal of Applied Ecology* **50**: 1450–1458.

Szabo JK, Vesk P a, Baxter PWJ, Possingham HP. **2010**. Regional avian species declines estimated from volunteer-collected long-term data using List Length Analysis. *Ecological Applications* **20**: 2157–2169.

Telfer MG, Preston CD, Rothery P. **2002**. A general method for measuring relative change in range size from biological atlas data. *Biological Conservation* **107**: 99–109.

Thomas JA, Telfer MG, Roy DB, Preston CD, Greenwood JJD, Asher J, Fox R, Clarke RT, Lawton JH. **2004**. Comparative losses of British butterflies, birds, and plants and the global extinction crisis. *Science* **303**: 1879–81.

Tingley MW, Beissinger SR. **2009**. Detecting range shifts from historical species occurrences: new perspectives on old data. *Trends in ecology & evolution* **24**: 625–33.

Tulloch AIT, Mustin K, Possingham HP, Szabo JK, Wilson KA. **2012**. To boldly go where no volunteer has gone before: predicting volunteer activity to prioritize surveys at the landscape scale. *Diversity Distrib.*: n/a–n/a.

Wood C, Sullivan B, Iliff M, Fink D, Kelling S. **2011**. eBird: engaging birders in science and conservation. *PLoS biology* **9**: e1001220.