Combining Biodiversity Resurveys across Regions to Advance Global Change Research

KRIS VERHEYEN, PIETER DE FRENNE, LANDER BAETEN, DONALD M. WALLER, RADIM HÉDL, MICHAEL P. PERRING, HABEN BLONDEEL, JÖRG BRUNET, MARKÉTA CHUDOMELOVÁ, GUILLAUME DECOCQ, EMIEL DE LOMBAERDE, LEEN DEPAUW, THOMAS DIRNBÖCK, TOMASZ DURAK, OVE ERIKSSON, FRANK S. GILLIAM, THILO HEINKEN, STEFFI HEINRICHS, MARTIN HERMY, BOGDAN JAROSZEWICZ, MICHAEL A. JENKINS, SARAH E. JOHNSON, KEITH J. KIRBY, MARTIN KOPECKÝ, DRIES LANDUYT, LEEN DEPAUW, THOMAS DIRNBÖCK, TOMASZ DURAK, OVE ERIKSSON, FRIDE HØISTAD SCHEI, WOLFGANG SCHMIDT, TIBOR STANDOVÁR, KRZYSZTOF ŚWIERKOSZ, Karol Ujházy, HANS VAN CALSTER, MARK VELLEND, ONDŘEJ VILD, KERRY WOODS, MONIKA WULF, AND MARKUS BERNHARDT-RÖMERMANN

More and more ecologists have started to resurvey communities sampled in earlier decades to determine long-term shifts in community composition and infer the likely drivers of the ecological changes observed. However, to assess the relative importance of and interactions among multiple drivers, joint analyses of resurvey data from many regions spanning large environmental gradients are needed. In this article, we illustrate how combining resurvey data from multiple regions can increase the likelihood of driver orthogonality within the design and show that repeatedly surveying across multiple regions provides higher representativeness and comprehensiveness, allowing us to answer more completely a broader range of questions. We provide general guidelines to aid the implementation of multiregion resurvey databases. In so doing, we aim to encourage resurvey database development across other community types and biomes to advance global environmental change research.

Keywords: legacy data, (quasi-)permanent plots, community ecology, ground-layer vegetation, temperate forest

Increasing human impacts on the environment have large and pervasive effects on the composition and functioning of ecosystems (MA 2005). This makes it important to document and understand how ecosystems and communities are changing and to determine how the multiple drivers of global change interact. Without such knowledge, we are unable to develop appropriate strategies for the effective conservation and restoration of biodiversity and to maintain desired ecosystem functions.

To improve our understanding of how multiple global-change drivers affect ecosystems, we should combine different methods (Luo et al. 2011). Quantifying how ecosystems and communities vary along environmental gradients is an important source of information in this respect (e.g., Newbold et al. 2015), complementing knowledge gained from experiments and modeling studies (cf. Luo et al. 2011). Environmental gradient studies can give information on ecosystem responses to multiple drivers across space and can also be used to infer how ecosystems may potentially respond to temporally varying drivers. However, such space-for-time approaches rely on many assumptions (e.g., Walker et al. 2010). Repeat observations of the same community over time to quantify how communities are changing are therefore invaluable additional sources of information (e.g., Tingley and Beissinger 2009, Dornelas et al. 2012), particularly when data extend to several decades or longer, because more reliable and informative signals to estimate the nature and rates of change can be obtained (cf. Magnuson 1990, Pauly 1995).

More and more ecologists have started to resurvey communities sampled in earlier decades to determine long-term shifts in community composition and infer the likely drivers of the ecological changes observed. Plant ecologists now use vegetation data from early- to mid-twentieth-century vegetation descriptions to examine long-term changes in these communities (see, e.g., Bakker and colleagues 1996 for an earlier discussion on the topic). Many examples from other communities exist as well (e.g., birds, Tingley and Beissinger 2013; butterflies, Nieto-Sánchez et al. 2015; small-mammal communities, Moritz et al. 2008; and zoobenthos, Olsson et al. 2013).
However, most resurvey studies have worked with data collected in single regions, and their utility is limited if we are to understand the importance of the multiple often interacting global-change drivers that affect plant and animal communities. These drivers vary at multiple spatial and temporal scales and often co-vary in space and time. Proper assessments of the relative importance of multiple drivers and of the interactions among them require us to analyze resurvey data from multiple regions, spanning large environmental gradients and multiple geographic regions.

In this article, we provide arguments as to how pooling resurvey data from multiple regions realizes the potential to make major contributions to the understanding of community dynamics and the response to various interacting environmental changes. We illustrate our arguments with published results from long-term resurveys of temperate forest ground-layer vegetation and share lessons to enable database development and data retention in other community types and biomes. Our approach serves as an example of data sharing and collaboration (Wolkovich et al. 2012, Mills et al. 2015) and furthermore provides an example of how to make best use of legacy data sets, which are often abandoned and at risk of being lost (see also Vellend et al. 2013).

**The added value of multiregion community resurvey data: Representativeness, comprehensiveness, and orthogonality**

Well-reasoned criteria for data-set inclusion are needed to turn a collection of data sets into a powerful ecological research platform. In this section we therefore start by defining the main features of resurvey data sets suitable for inclusion in a multiregion analysis and then compare how a collection of resurvey data sets performs compared with multiregion experiments and *a priori* designed community monitoring networks.

We define a *resurvey data set* as a collection of community surveys sampled at multiple locations within a defined region and across at least two points in time. The two time points typically span a period of at least several decades in order to obtain a true long-term perspective on environmental and community change—that is, the unique, invaluable feature offered by legacy data sets. A region is defined here as a geographic entity with more or less similar site conditions, including climate, major soil types, and levels of atmospheric nitrogen (N) deposition. Regions are defined this way because the main objective of multiregion resurvey data analyses is to quantify the (interactive) effects of multiple drivers, which often vary at different scales. For instance, climate change generally plays out at larger spatial scales, whereas management changes can vary among locations within a single region. However, the combined outcome of both drivers will ultimately determine changes in the local microclimate and the resulting changes in community composition (see figure 1 for an example). A combination of multiple regions with multiple resurveyed locations within each region is therefore a key design feature of a research platform that aims at understanding long-term community changes. Besides these general criteria, specific criteria for the inclusion of data sets in the research platform also need to be defined so that the platform resembles *a priori* community monitoring networks with a standardized design (table 1).

Such a multiregion network of community resurvey data scores well for all three fundamental design criteria for ecological research platforms, notably comprehensiveness, representativeness, and orthogonality (figure 2; Nadowski et al. 2010, Baeten et al. 2013). **Comprehensiveness** in this article relates to the spectrum of ecological questions that can be addressed with a particular research platform. **Representativeness** refers to the relevance of analyzed results for sites that were not included in the investigation. Finally, the **orthogonality** of the platform refers to its ability to disentangle the separate effects of each environmental driver on the response variable(s) under study. Most obviously, the representativeness generally increases when an increasing number of regions are incorporated in the research platform, because sites not initially investigated will more likely fit within the environmental envelope spanned by the platform. This should lead to more reliable inference. The spatio-temporal replication of community data (i.e., resurveys in multiple locations in multiple regions) strongly increases the likelihood of orthogonality within the design. It should be noted that orthogonality and representativeness are not entirely independent in this case: The inclusion of multiple regions is a necessary condition to increase orthogonality for drivers varying at large spatial scales and this will simultaneously increase the representativeness. Finally, repeatedly surveying broadly across multiple landscapes or regions also results in high comprehensiveness, allowing us to more completely answer a broader range of questions, as well as potentially unanticipated ones.

In addition, long-term, multiregion resurveys have the ability to complement the outcomes of globally distributed experiments with environmental manipulations, such as nitrogen addition (figure 2; cf. Fraser et al. 2013, Borer et al. 2014). Although experiments typically score higher on the orthogonality axis, they reduce representativeness and often comprehensiveness by using simplified communities and often extreme (“shock”) treatments (e.g., a sudden shift from low to high temperature regimes) with a limited number of treatment levels. Furthermore, treatment responses are rarely monitored for more than a few years. These elements constrain the spectrum of questions that can be addressed with experiments and therefore their comprehensiveness. Making best use of long-term resurveys from multiple sites as a complement to experimental approaches therefore responds to calls for more integrated approaches to better understand the effects of global changes on complex ecological communities and ecosystem functions (Luo et al. 2011, De Frenne et al. 2013; for a good example, see Frerker et al. 2014).

Parallel to the rise of globally distributed experiments, more and more *a priori* designed community monitoring...
Putting long-term, multiregion resurveys into practice: The forestREplot network as an example

Resurveys of long-term, (quasi-)permanent plots are particularly appropriate for communities that exhibit slow dynamics, such as ground-layer communities in forests. These plant communities often show delayed responses to environmental changes: The long life span of many ground-layer species (Ehrén and Lehtilä 2002) promotes remnant populations and extinction debts (Eriksson 1996, Vellend et al. 2006), whereas slow immigration rates can lead to colonization credits (Verheyen et al. 2003). Because the ground layer in temperate forests constitutes the majority of plant diversity in these systems and has an important impact on their functioning (Gilliam 2007), it is important to document the long-term changes in the ground-layer composition and diversity and to understand the drivers that underlie these changes. Changes documented in forest understories may also serve as early warnings of impacts to even slower canopy dynamics.

The forestREplot network (www.forestreplot.ugent.be) brings together standardized ground-layer vegetation resurvey plots collected in natural or seminatural forests in different regions across Europe and North America (Verheyen et al. 2012, De Frenne et al. 2013, Baeten et al. 2014, Bernhardt-Römermann et al. 2015). Table 1 gives an overview of the criteria used for data-set inclusion in forestREplot. The database currently consists of 55 data sets and nearly 3000 pairs of historically and recently surveyed plots with a mean intercensus interval of 35.7 years (see supplemental appendix 1 for an overview and Depauw and Maes [2015] for more information).

The network aims to (a) collect and archive data sets of resurveyed vegetation plots in temperate forests worldwide and (b) perform analyses across multiple sites to address novel research questions in ecology, with a specific focus on the ground layer and the impacts that various often interacting global-change drivers have on this layer. In many respects, the design and management of the forestREplot network adheres to the guidelines for globally distributed experiments outlined by Fraser and colleagues (2013) and Borer and colleagues (2014).

Here, we illustrate with forestREplot how multiple resurvey data sets can address a broad spectrum of ecological questions (i.e., the comprehensiveness), with results being representative for real-world changes in temperate forest communities (i.e., the increased representativeness). Furthermore, we show how the approach may disentangle networks across large environmental gradients are also being established. They include top-down designed networks, such as the European Level-I and -II monitoring networks of air-pollution effects on forests (http://icp-forests.net) and the UK Countryside Survey (www.countrysidesurvey.org.uk). Multiregion community resurvey networks with a more bottom-up approach, in which regions participate on a voluntary basis, have emerged as well. The Global Observation Research Initiative in Alpine Environments (GLORIA) network (Pauli et al. 2015) can serve as a prime example. The network applies a highly standardized “Multi-Summit Approach” to survey alpine biodiversity and vegetation patterns on four mountain summits per target region. The results of this observation network help us to better understand the response of alpine biota to climate change (see, e.g., Pauli et al. 2012). The first plots were established in 2001 and have been resurveyed at regular intervals since then. Although these multiregion monitoring networks have already produced very valuable results and will certainly continue to do so in the future, they have rarely been established more than one or two decades ago and therefore well after the rise in many anthropogenic pressures. Because insights into longer-term changes are badly needed (cf. Pauly 1995), attempts should be made to make the best use of archived community survey data collected in a more distant past.

In the next section, we illustrate how to put together a network using legacy community resurvey data by introducing forestREplot. In addition, we synthesize already published results from forestREplot to show how new insights can be developed and more general conclusions reached.
the relative importance of multiple drivers of change in the ground layer of forests (i.e., the increase in orthogonality).

**Comprehensiveness.** To quantify the spectrum of ecological questions that can be addressed with multisite resurvey data, here shown using the forestREplot example, we performed a two-step survey among 32 participants of the first forestREplot workshop organized in December 2014 in Ghent, Belgium. All participants in the workshop were data contributors to the forestREplot database. Prior to the meeting, the workshop organizers (KV, LB, LDP, MB-R, PDF, RH, and SLM) quantitatively assessed which of the current 100 fundamental questions in ecology (as listed by Sutherland et al. 2013) could be answered with the forestREplot database by attributing a score between 1 (not suitable) and 3 (very suitable) to all questions. This resulted in a subset of 42 fundamental questions of the original Sutherland and colleagues (2013) list with a score 2 or more. Next, we asked the workshop participants to score the potential of the forestREplot database to answer these 42 questions. The top 10 questions that had the highest probability of being scored very suitable can be found in table 2. The full list with questions and scores can be found in supplemental appendix 2.

**Representativeness.** As we amass resurvey data from more sites, spread over larger regions, we gain a clearer picture of which changes are local or idiosyncratic to a few locations and which reflect more general and widespread changes (figure 2). However, results from any given database are clearly bounded by the variation within the set of species, communities, and environmental conditions present within the database. Resurvey data included in forestREplot, for instance, only come from seminatural and natural forests (see Depauw and Maes 2015). Furthermore, forestREplot is merely a collection of data sets and not a designed monitoring program based on probabilistic sampling, such as the National Forest Inventories (NFI), which reduces its representativeness and makes the statistical analyses more complicated. For instance, many of the first surveys were made for phytosociological purposes, meaning that plot locations are not entirely randomly chosen. These limitations have to be acknowledged when using the data (cf. Holeksa and

---

**Table 1. An overview of the criteria used to decide on the inclusion of data sets in multiregion community resurvey studies, illustrated with the decisions taken to feed the forestREplot network with data sets.**

<table>
<thead>
<tr>
<th>Data-set inclusion criteria</th>
<th>Criteria</th>
<th>forestREplot</th>
<th>Rationale</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>General criteria</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Suitable for the scientific goals and questions at hand?</td>
<td>Forest ground layer resurveyed at multiple locations within a region with more or less similar site conditions, including climate, major soil types, and levels of atmospheric deposition</td>
<td>This type of data structure is needed to isolate the effects of drivers acting at larger scales, such as changing climate or levels of atmospheric pollutant deposition from the effects of drivers acting at a more local scale, such as management changes (see also figure 1)</td>
<td></td>
</tr>
<tr>
<td><strong>Specific criteria</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Relevant geographic region?</td>
<td>Temperate forest, as defined by Olsen and colleagues (2003)</td>
<td>The ground layer in temperate forest constitutes the majority of plant diversity and has an important impact on ecosystem functioning</td>
<td></td>
</tr>
<tr>
<td>Relevant system characteristics?</td>
<td>Natural and seminatural forests, according to Peterken (1996). Both are composed of locally native trees and shrubs that often derive from natural regeneration or coppicing rather than from planting (in the case of seminatural forests) or have not been managed at all (in the case of natural forest)</td>
<td>Management actions such as soil working and fertilization may completely override the effects of other global-change drivers</td>
<td></td>
</tr>
<tr>
<td>Relevant study design?</td>
<td>(Quasi-)permanent plots</td>
<td>Minimizes so-called pseudo-turnover</td>
<td></td>
</tr>
<tr>
<td>At least 20 plots that can be treated as independent observations (i.e., distributed over a sufficiently large area) per data set</td>
<td>Sufficient replicates within single regions are needed</td>
<td></td>
<td></td>
</tr>
<tr>
<td>At least 20 years between the oldest and most recent survey</td>
<td>Forest ground-layer vegetation often shows delayed responses to environmental changes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Plot size varies between 1 square meter and 1000 square meters</td>
<td>Plots falling within this size range are expected to present a representative picture of the ground-layer vegetation community</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Relevant response variables?</td>
<td>Presence, absence, or cover data of all vascular plants in the ground-layer community</td>
<td>Needed to get a complete view on community change</td>
<td></td>
</tr>
</tbody>
</table>
Figure 2. The comprehensiveness, representativeness, and orthogonality of single-region versus multiple-region resurveys and experiments. Experiments are more orthogonal than observatories. The combination of multiple regions typically creates higher orthogonality and generates more comprehensive and representative results than from a single region. The red arrow indicates that orthogonality and representativeness are not entirely independent in multiregion resurvey observatories: The inclusion of multiple regions is generally a necessary condition to increase orthogonality between drivers of change, and this will simultaneously increase the representativeness.

Woźniak 2005, Michalcová et al. 2011). However, most monitoring programs designed to be representative do not (yet) span long time periods (but see Hedwall and Brunet 2016). Furthermore, the spatial sampling resolution in these monitoring programs is often rather low so that smaller scale changes risk going undetected.

Orthogonality. Single-region studies have shown that ground-layer vegetation in temperate forests responds sensitively to global-change drivers, including forest management, atmospheric N deposition, and climate change (table 3). However, these studies often do not show consistent responses, as is exemplified for species richness in table 3. Furthermore, community responses may not be monotonic over longer environmental gradients. To analyze the orthogonal and interacting effects of these drivers on biodiversity, it is necessary to either include many sites and studied factors within a single large-scale study, or to combine results from several single studies in joint analyses.

For instance, Verheyen and colleagues (2012) presented a meta-analysis of 23 local-scale resurveys from across Europe that focused on the contribution of atmospheric N deposition versus changes in forest management to explain changes in herb layer composition. Shifts in vegetation composition seemed mainly related to management-related alterations in the canopy structure and composition, independent of the N deposition.

An additional study exploring the mechanisms driving temporal changes in biodiversity was performed by Bernhardt-Römermann and colleagues (2015). Using 39 data sets of resurvey data on forest understory communities across Europe, temporal changes in species richness were related to environmental data at multiple spatial scales (continental, regional, and local). These joint analyses were designed to relate temporal changes in species richness with (a) across-site variation in environmental conditions at the time of the initial vegetation survey (i.e., baselines) and (b) temporal changes in environmental conditions between vegetation surveys. No significant and directional changes in local diversity were found, although there was considerable across-site variation, corroborating earlier findings (Verheyen et al. 2012, Vellend et al. 2013). This across-site variation was determined by both local and regional scale drivers (temporal changes in local stand structure and game density). Most excitingly, strong evidence was found that presurvey levels of N deposition determined subsequent changes in biodiversity. Recently, Simkin and colleagues (2016) confirmed the existence of context-dependent effects of N deposition on plant diversity using a large data set from the United States.

Third, the increased dominance of warm-adapted plant species (so-called thermophilization) as a result of climate warming has been identified across several ecosystems (Bertrand et al. 2011, Gottfried et al. 2012). However, De Frenne and colleagues (2013) found that this thermophilization was lowest in forests that had become denser over time across Europe and North America, suggesting that reducing management intensity to increase shading can buffer the impacts of global warming (cf. also De Frenne et al. 2015).

These three examples show how multiregion analyses can increase orthogonality compared with single-region studies.

Challenges associated with resurvey data
Despite the great potential that combining long-term resurvey data from multiple regions holds, some important challenges remain, both at the level of the individual resurvey studies and when trying to combine them.

Sources of unwanted variability or bias in resurvey studies have received considerable attention in the scientific literature (e.g., Tingley and Beissinger 2009). Taking the example of vegetation resurveys, studies have been performed to quantify the level of bias introduced because of (a) relocation errors (e.g., Fischer and Stöcklin 1997, Kopecký and Macek 2015); (b) species detectability, observer effects, and sampling exhaustiveness (Archaux et al. 2006, Vittoz and Guisan 2007, Milberg et al. 2008); (c) taxonomic inconsistencies (Jansen and Dengler 2010); and (d) differences in recording dates (Van Calster et al. 2008). Recently, Semboli and colleagues (2014) highlighted a new source of bias, notably a changing vegetation composition after multiple resurvey visits due to, among others, trampling effects. Many of these biases are not easy to solve, particularly when the first surveyors are no longer around. Therefore, there is a need for a robust archiving
Table 2. The top ten most important ecological questions following Sutherland and colleagues (2013) that can be addressed with the multisite ground-layer resurvey data incorporated in the forestREplot database.

<table>
<thead>
<tr>
<th>Rank</th>
<th>Question</th>
<th>Category</th>
<th>Prob[rank = very suitable]</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Can we predict the responses of ecosystems to environmental change on the basis of the traits of species?</td>
<td>Ecosystems and functioning</td>
<td>0.67</td>
</tr>
<tr>
<td>2</td>
<td>How do spatial and temporal environmental heterogeneities influence diversity at different scales?</td>
<td>Communities and diversity</td>
<td>0.64</td>
</tr>
<tr>
<td>3</td>
<td>What is the magnitude of the extinction debt following the loss and fragmentation of natural habitats, and when will it be paid?</td>
<td>Human impacts and global change</td>
<td>0.58</td>
</tr>
<tr>
<td>4</td>
<td>Which ecosystems and what properties are most sensitive to changes in community composition?</td>
<td>Ecosystems and functioning</td>
<td>0.51</td>
</tr>
<tr>
<td>5</td>
<td>To what extent are local species composition and diversity controlled by dispersal limitation and the regional species pool?</td>
<td>Communities and diversity</td>
<td>0.50</td>
</tr>
<tr>
<td>6</td>
<td>How well can community properties and responses to environmental change be predicted from the distribution of simple synoptic traits (e.g., body size and leaf area)?</td>
<td>Communities and diversity</td>
<td>0.48</td>
</tr>
<tr>
<td>7</td>
<td>What are the indirect effects of harvesting on ecosystem structure and dynamics?</td>
<td>Human impacts and global change</td>
<td>0.48</td>
</tr>
<tr>
<td>8</td>
<td>How do natural communities respond to increased frequencies of extreme weather events predicted under global climate change?</td>
<td>Human impacts and global change</td>
<td>0.40</td>
</tr>
<tr>
<td>9</td>
<td>What are the most appropriate baselines for determining the magnitude and direction of ecological changes?</td>
<td>Methods</td>
<td>0.39</td>
</tr>
<tr>
<td>10</td>
<td>In the face of rapid environmental change, what determines whether species adapt, shift their ranges, or go extinct?</td>
<td>Human impacts and global change</td>
<td>0.37</td>
</tr>
</tbody>
</table>

a Taken from the list of Sutherland and colleagues (2013).

b We fitted cumulative link models, which are regression models for ordinal data (cmlm in the R package ordinal; Christensen 2015, R Development Core Team 2015). The results show the estimated probability that a question was rated as very suitable across the 32 respondents.

Table 3. The impact of selected environmental drivers on changes in ground-layer species richness in temperate forests. Shown are exemplarily single-region studies and the estimated general importance of each environmental driver based on multiregion resurvey studies.

<table>
<thead>
<tr>
<th>Driver</th>
<th>Single-region vegetation resurveys (examples)</th>
<th>Direction of effect on species richness</th>
<th>Multiregion analyses</th>
</tr>
</thead>
<tbody>
<tr>
<td>Increased N deposition</td>
<td>Hédl (2004) Skrindo and Økland (2002) Bernhardt-Römermann et al. (2007)</td>
<td>Negative No effect Positive Positive</td>
<td>Presurvey levels of N deposition determine subsequent changes in biodiversity (Bernhardt-Römermann et al. 2015); actual N deposition is less important than forest management (Verheyen et al. 2012); the exceedance of critical loads favors N-demanding species (Dirnböck et al. 2014)</td>
</tr>
</tbody>
</table>

of survey data so that at least future generations of researchers are not confronted with these issues (see box 1).

When multiple data sets are combined, additional challenges arise that relate to differences in baselines (e.g., due to historical land-use or air-pollution legacies), variation in the time interval between the surveys and variation in the sampling protocols used. For instance, if there is covariation of plot sizes or the time interval between the surveys with environmental changes of interest, then the observed community changes might be principally caused by species-area or temporal effects. These issues require serious attention from the start of any analysis, for instance, by setting
Box 1. Maintaining the resource: Toward a publicly accessible data and metadata archive for resurveys.

Addressing important ecological questions on ecosystem responses to environmental change through the use of long-term data requires the data to exist in the first place, necessitating support for long-term ecological research infrastructure and its integration, such as the European Platform for Biodiversity Research Strategy and the International Long-Term Ecological Research Network. It then requires that these high-quality data survive indefinitely; that, equally important, their accompanying metadata survive indefinitely; and finally, that these data be accessible for analysis. Here, we discuss the need for metadata, the requirement for scientists to know what data are available where, and analysis implications, referring to our experience with vegetation resurveys and forestREplot in particular; the lessons, however, are applicable to all ecological resurveys and long-term data in general.

In the past, ecological researchers tended to maintain their own records, passing on data and their contexts to a relay of successors. However, relay batons have been dropped, successors have not emerged, records have consequently been lost or destroyed, and “information entropy” has ensued (Michener et al. 1997). To avoid unnecessary data loss, well-documented procedures to preserve data with accompanying metadata are required (figure 3). Of fundamental importance is the preservation of the metadata—defined as representing the higher-level information or instructions that describe the content, context, quality (e.g., data anomalies or missing data), structure, and accessibility of a specific data set (Michener et al. 1997). In the context of vegetation resurveys, for example, this includes the detailed descriptions of cover estimation to enable spatial and temporal comparisons and the clear identification of taxonomic authorities and its context of use (see Wiser 2016 for an interesting discussion of the issues associated with nomenclatural or taxonomic changes across space and time). In forestREplot, metadata information is gathered systematically by asking contributors to fill in site and plot information sheets, which characterize the location, land-use history, soil type, and management disturbance between surveys, whereas taxonomic harmonization uses the Plant List (www.theplantlist.org) and, if unresolved there, the Euro+Med PlantBase (ww2.hgbm.org/EuroPlusMed). Without such metadata and the careful integration of primary data, understanding and analyses would be impossible (see also Borer et al. 2014).

The fundamental ecological research questions that broadly distributed vegetation resurveys can answer also require knowing what data are available where. Vegetation databases are rapidly developing at regional and global levels and can be identified through the Global Index of Vegetation Plot Databases (www.givd.info). Automated retrieval and checking systems are increasingly being used to speed up the acquisition, checking, and “wrangling” of data (i.e., their integration) to allow analyses within the broad field of ecoinformatics (Madin et al. 2007, 2008, Michener and Jones 2012, Wiser 2016). Such efforts complement network initiatives such as forestREplot, which have grown informally and identified separate data sets that have been manually integrated to allow synthetic analyses (e.g., Verheyen et al. 2012, De Frenne et al. 2013). All these approaches will be in vain, however, without the required archiving of resurvey data and metadata in the first place.

Ultimately, archiving may be best incentivized for scientists through publication of the data (in “data papers” rather than in typical research articles) using established channels of automated and semi-automated data checking culminating in peer review (Costello et al. 2013). Organizations such as the Global Biodiversity Information Facility can aid this publication and archiving endeavor. Otherwise, the contemporary situation (in which 80% of scientists want to access data created by others but only 20% have actually shared their data) may continue to persist and valuable opportunities to answer fundamental ecological questions may be lost as time-poor scientists prioritize publication over making data available (Costello et al. 2013).

Archiving may also be encouraged by recent policies to mandate publicly accessible data with journal publications such as in Dryad (http://datadryad.org), sometimes with embargo periods. However, this can be complicated when article authors are not the “owners” of the data, different legislation applies across countries and states, and the databases themselves continue to evolve; efforts to resolve these and other issues are ongoing (Mills et al. 2015, Whitlock et al. 2016). For vegetation resurveys, Wiser (2016) suggested archiving plot data in an established vegetation-plot repository as a first step and then providing data on request. This latter approach is similar to forestREplot, in which the data are archived but not publically accessible. Requests for new analyses are considered by a management committee to avoid overlap with existing projects, and data-set contributors are then contacted to give permission for data use.

Care also needs to be taken with the public accessibility of vegetation data, such as to avoid the explicit location of species of conservation concern. However, arguments exist that we will only get solutions to environmental issues if data are made easily accessible to—and understood by—a broad audience (Peters 2010). Ultimately, records of data existence would be invaluable for researchers—as would instructions for how interested parties can access them with associated rights of use—through, for example, the distributed system of nationally and internationally funded data platforms, as has been proposed by the World Data System of the International Council of Science (Bendix et al. 2012). In addition to electronic data, records that need to be kept according to rigorous procedures include field notes, samples, photographs, and maps.
strict inclusion criteria for resurvey data sets with deviating baseline conditions, resurvey time intervals, sampling unit properties, or internal heterogeneity.

Even if studies are carefully selected on the basis of the methods used to gather the community data, the nature of the temporal data involves several challenges from an analytical point of view. For time-series data, similar difficulties, such as measurement errors and temporal autocorrelation, were identified (Dornelas et al. 2012), but the clever analytical strategies to deal with these do not always easily translate into solutions for typical resurvey studies that provide data for only two time points. For instance, the nature of a temporal trend (e.g., accelerating decrease in diversity) can be quantified with statistical models that account for temporal autocorrelation, but only if sufficient time points are available. Previous studies have used (log) response ratios of old and recent plot values to compare between data sets in a meta-analytical framework (e.g., Verheyen et al. 2012, Bernhardt-Römermann et al. 2015). But although this allows standardizing for particular sampling differences between data sets (e.g., plot size), it does not account for variation in the time interval between surveys unless assumptions are made about the nature of the temporal change (e.g., a [log]-linear response over time; Verheyen et al. 2012). Finally, analyses usually include predictors of change at different scales (plot, study, cross-study) and typically require multilevel models (Qian et al. 2010).

Conclusions
The challenges described above should not discourage researchers from seeking to recover historical legacy data, from working to properly document and archive the data (box 1), or from doing the matched resurveys necessary to document long-term ecological change. Many valuable historical community descriptions exist that can be used to generate and test novel insights into ecological change. Furthermore, insights will be deeper and more general when we can combine data from multiple regions and analyze the results in a comparative context. In this article, we used the forestREplot network as an example of the power that long-term resurvey data have for addressing how communities are responding to a broad range of environmental factors. However, we should bear in mind that forestREplot focuses only on forest ground-layer communities in natural and seminatural temperate forests. We therefore encourage the development of more multiregion resurvey databases for other community types and biomes, as well as new modes of (trait-based) analysis. These will increase the number and nature of the comparisons we can

---

**Box 1. Continued.**

**PROCEDURES FOR ROBUST ARCHIVING OF RE-SURVEY RECORDS**

- **Depository**
  - Records to be reliably stored with explicit archiving function

- **Index and Catalogue**
  - Class mark all items e.g., maps, photograph, digital records
  - Description of contents preferably using Ecological Metadata Language

- **Institutional Memory**
  - Ensure knowledge of data existence within institution and more broadly.
  - Best achieved through online repositories e.g., DataONE.

- **Access and Ownership Arrangements**
  - Clarify rights, costs and legitimate purposes for record use
  - Detail ownership of records

- **Additional Material Deposition**
  - Clear controls and guidelines for additional data deposition

---

**Figure 3. Procedures for the robust archiving of resurvey records.** Data may not need to be made available to all, but it is crucial that its existence not be forgotten given the opportunity they provide to answer fundamental ecological questions. Given the pressure for scientists to publish, archiving may ultimately be best incentivized through credit for data publication. In the meantime and although barriers to this outcome are still present, it is imperative that metadata and the records themselves are robustly archived, with researchers able to find out about their existence through online search tools such as DataONE.
make, allowing us, in turn, to test a wider range of hypotheses and reach more general conclusions. Over time, such tests, performed on replicated sets of regions across many distinct biomes, will allow to more fully assess the several often interacting effects of forces driving ecological change.

Acknowledgments
The ideas for this article were developed at a workshop held in Ghent (BE, December 2014). Financial support was provided by the European Research Council through the PASTFORWARD project (ERC Consolidator Grant no. 614839) to KV. The article also benefited from a collaboration with the LONGWOOD project (ERC Starting Grant no. 278065), coordinated by Péter Szabó, and from the insightful comments of Dr. Scott L. Collins and two anonymous reviewers. None of the authors has a conflict of interest.

Supplemental material
Supplementary data are available at BIOSCI online.

References cited


Kris Verheyen (Kris.Verheyen@UGent.be), Pieter De Frenne, Lander Baeten, Michael P. Perring, Haben Blondeel, Emiel De Lombaerde, Leen Depauw, Dries Landuyt, and Sybryn L. Maes are affiliated with the Forest & Nature Lab in the Department of Forest and Water Management at Ghent University, in Melle-Gentrode, Belgium. MPP is also affiliated with the Ecosystem Restoration and Intervention Ecology Research Group at the School of Plant Biology, The University of Western Australia, Crawley, Australia. Donald M. Woller and Daijiang Li are with the Department of Botany at the University of Wisconsin–Madison. Markéta Chudomelová, Martin Kopecký, and Ondřej Vild are affiliated with the Department of Vegetation Ecology at the Institute of Botany of the Czech Academy of Sciences, in Brno. RH is also with the Department of Botany at Palacky University in Olomouc, in the Czech Republic; MK is also with the Department of Forest Ecology in the Faculty of Forestry and Wood Sciences at the Czech University of Life Sciences, in Prague, Czech Republic; and OV is also with the Department of Botany and Zoology in the Faculty of Sciences at Masaryk University, in Brno, Czech Republic. Jörg Brunet is with the Southern Swedish Forest Research Centre at the Swedish University of Agricultural Sciences, in Alnarp, Sweden. Guillaume Decoq and Jonathan Lenoir are affiliated with Écologie et Dynamique des Systèmes Anthropisés FRE 3498 (EDYSAN) at the Jules Verne University of Picardy, in Amiens, France. Thomas Dirnböck is affiliated with the Department for Ecosystem Research at Environment Agency Austria, in Vienna. Tomasz Durak is with the Department of Botany at the University of Rzeszów, in Poland. Ove Eriksson is affiliated with the Department of Ecology, Environment, and Plant Sciences at Stockholm University, in Sweden. Frank S. Gilliam is affiliated with the Department of Biological Sciences at Marshall University, in Huntington, West Virginia. Steffi Heinrichs and Wolfgang Schmidt are with the Department of Silviculture and Forest Ecology of the Temperate Zones at the Burchhardt Institute of Georg August University, in Göttingen, Germany. Martin Hermy is affiliated with the Department of Earth and Environmental Sciences at the University of Leuven, in Belgium. Bogdan Jaroszewicz is with the Białowieża Geobotanical Station, Faculty of Biology, at University of Warsaw, in Poland. Michael A. Jenkins is affiliated with the Department of Forestry and Natural Resources at Purdue University, in West Lafayette, Indiana. Sarah E. Johnson is with the Department of Natural Resources and Biology at Northland College, in Ashland, Wisconsin. Keith J. Kirby is with the Department of Plant Sciences at the University of Oxford, in the United Kingdom. Martin Macek and Petr Petrík are affiliated with the Department of Geographical Information Systems (GIS) and Remote Sensing (RS) at the Institute of Botany at the Czech Academy of Sciences, in Pírůvonic, Czech Republic. František Malíš and Karol Ujházy are with the Department of Phytology in the Faculty of Forestry at the Technical University in Zvolen, in Slovakia. Fraser J. G. Mitchell is with the School of Natural Sciences at Trinity College Dublin, in Ireland. Tobias Naaf and Monika Wulf are affiliated with the Institute of Land Use Systems at the Leibniz Centre for Agricultural Landscape Research (ZALF), in Müncheberg, Germany. George Peterken is an honorary professor in the School of Geography at the University of Nottingham, in the United Kingdom. Kamila Reczyńska and Krzysztof Świerkosz are affiliated with Wrocław University’s Museum of Natural History, in Poland. David A. Rogers is with the Department of Biological Sciences at the University of Wisconsin–Parkside. Frade Haistad Schei is affiliated with the Division of Forest and Forest Resources at the Norwegian Institute of Bioeconomy Research, in Fana, Norway. Tibor Standovár is affiliated with the Department of Plant Systematics, Ecology, and Theoretical Biology at Eötvös Loránd University, in Budapest, Hungary. Hans Van Calster is affiliated with the Department of Biometry and Quality Assurance at the Research Institute for Nature and Forest, in Brussels, Belgium. Mark Vellend is with the Département de Biologie at the Université de Sherbrooke, in Québec, Canada. Kerry Woods is affiliated with the Département de Sciences Naturelles at Bennington College, in Vermont. Markus Bernhardt-Römermann is with the Institute of Ecology at Friedrich Schiller University Jena, in Germany.