Project no. 006463

EuMon

**EU**-wide **mon**itoring methods and systems of surveillance for species and habitats of Community interest

Instrument: SSP

Thematic Priority: Biodiversity conservation

## D16: Framework for integration of different species monitoring schemes

Due date of deliverable: 31 October 2006

Actual submission date: 31 October 2006

Start date of project: 1.11.2004                    Duration: 42 month

National Museum of Natural History (MNHN), Paris, France

Revision : final

| Project co-funded by the European Commission within the Sixth Framework Programme (2002-2006) | | |
|---|---|---|
| **Dissemination Level** | | |
| **PU** | Public | PU |
| **PP** | Restricted to other programme participants (including the Commission | |
| **RE** | Restricted to a group specified by the consortium (including the Commission | |
| **CO** | Confidential, only for members of the consortium (including the Commission Services) | |

# Framework for integration of different species monitoring schemes

Deliverable 16 of EuMon's Work Package 2.1

By Pierre-Yves HENRY (MNHN)
with contributions by Andrej SELIŠKAR (ZRC SAZU), Valerija BABIJ (ZRC SAZU), Tatjana ČELIK (ZRC SAZU), Piotr NOWICKI (UJAG), Bernd GRUBER (UFZ), Klaus HENLE (UFZ), Dirk SCHMELLER (UFZ) and Sandra BELL (UDUR).

Summary. In this Deliverable we describe how to identify schemes that could be combined, either to increase precision or representativity of results from the analysis of monitoring data, or to improve understanding and quantification of changes in biodiversity by using complementarity across schemes. We briefly present the main methods that can be used when combining data or estimates from different schemes.

## Table of Contents

# (1) Objectives

Many biodiversity monitoring schemes exist. While some are part of an international network but most of them are run independently, with a local scope and small geographical coverage. Biodiversity monitoring would thus benefit from integrating monitoring schemes/data into broader monitoring programmes (i.e. unifying existing monitoring schemes) and broader monitoring goals (i.e. combining monitoring schemes with complementary monitoring goals) to monitor biodiversity changes more effectively and more representatively. An output of integration should also be to increase visibility and legitimacy of integrated monitoring schemes compared to 'disconnected' schemes. They would thus get a higher profile in the scientific community and would be of higher value for biodiversity policy.

This deliverable mainly develops the different avenues for integration that could be followed, as well as the methods that can be used to achieve this integration. Avenues for integration were mainly determined from questions in the DaEuMon questionnaire (Deliverable 8). These questions were designed so that they allow a thorough description of monitoring targets, methods, designs, efforts and biological scopes. In the text hereafter, we refer to questions by their number in the questionnaire (legends in Deliverable 8; e.g., S26 for question 26 in Species questionnaire, namely Taxonomic group).

According to spatial scale, we restrict suggestions of integration to two administrative scales: national and Europe. If not mentioned otherwise, integration is focussing at the European level. For all following criteria, if integration is sought at the national level, just add 'same country' criterion. We did not address integration at a smaller scale than the national level because region identity is not available in DaEuMon (but, on the medium-term, that could be changed) and because biodiversity evaluations for environmental policy issues are most frequently undertaken at national scales.

Integration should be thought in two general ways, with complementary benefits:

➢ Combine schemes with similar characteristics. The resulting network of schemes would promote exchanges and integration of data or estimates. Main benefits would be:
- o increase geographical coverage per taxonomic group. If geographical scales differ among schemes to be integrated (e.g. mix of local and national schemes), then one may use post-stratification with weights in statistical analysis to compensate for these differences;
- o increased precision of estimates of states and trends, e.g. by increasing sample sizes (cf. number of sites);
- o increased representativity of results.

➢ Combine schemes with different characteristics. In this case, the main benefits come from complementarity among schemes. Benefits would be:
- o increase biological coverage (complementarity across species and across taxonomic groups), thus allowing characterisation of biodiversity changes with different taxonomic group rather than the one available at hand in each separate schemes;
- o increase geographical coverage;
- o increase temporal coverage;
- o increase coherence of monitoring activities with EU Directives by identifying which schemes need to be combined so that 'state and trends of species and

their habitats' are monitored for all species listed in the Directives in all countries.

From a practical point of view, the goal of work presented in this Deliverable is to design a tool that identifies which schemes could be integrated.

To imagine this 'tool to promote integration across schemes', one may think of an interface that basically works as a filtering tool. You may think of it as a filter in Excel: the more you define constraints of similarity (or dissimilarity) for different questions, the easier it will be to integrate schemes since they will be similar in many points. Then you may relax some constraints, increasing the list of schemes that could be combined, but also increasing difficulties to combine information since schemes will be less and less similar. However, their complementarity will increase.

For instance, imagine that an EU-wide programme that monitors population sizes of bats is to be set up. Highly similar, easily combined schemes would be identified by filtering schemes monitoring 'at least bats' (field S26 - Taxonomic group contains 'bat'), at 'national or international or EU' levels (field 'geographical scope'), collecting 'counts of individuals' (field S2 - Data type), with a representative sampling design (field S6 - Choice of sites equals 'random choice' or 'systematic choice' or 'exhaustive'). This filtering would yield a limited list of schemes, and that would be the basis for building an EU-network, with schemes rather easy to combine. Actually that yields a list of six schemes from three countries (Estonia, France, Hungary; DaEuMon database, 16/10/2006).

Then, some constraints maybe relaxed to find more schemes, less straightforward to be integrated, but that still would benefit from each other. For instance, the constraint on data type could be relaxed to increase the biological coverage of the monitoring, although data type will then be a mixture of presence/absence, counts of individuals, capture-mark-recapture, age-structure and phenology data. In the case of European bats monitoring, that adds two more schemes from two countries (Slovakia, Switzerland). These schemes document presence/absence only. On the whole, state and trends in distribution for bats could be integrated across five countries, with trends of population size being documented with good (counts) precision in three countries, and lower precision (presence/absence) in two countries.

Another example would be the identification of schemes running in other European countries than the one of interest. That would correspond to filtering on question 'Country', constraining that schemes 'do not contain' the name of the country of interest.

These suggestions for the design of the 'integration tool' are rather simple, and will be improved by WP5-WP6 to design and produce a user-friendly browser of DaEuMon, while allowing advanced filtering criteria.

# (2) Levels of integration

## I.  <u>Highly similar schemes</u>

Highly similar schemes are defined as having the same entries for taxonomic group (S26), field data type (S2), sampling design (S5, S6), habitats (S30), and countries. If these conditions are verified, integration is straightforward and it has the benefit of combining 'similar' schemes (cf. part 1).

## II.     Species schemes differing in habitats monitored

If different schemes monitor different habitats (S30), integration across schemes will broaden the monitoring scope and representativity of results since more habitats will be included in the monitoring.

## III.     Schemes differing across countries

A major avenue for integration is identifying which monitoring schemes are implemented in several European countries and are still lacking in others. For instance, butterflies are monitored by counts of individuals in 14 European countries (DaEuMon 16/10/2006). All these schemes could be combined in a single international network within which data and estimates are shared (and indeed, that it what is occurring; cf. http://eumon.ckff.si/monitoring/monitor_show_wp23-1.php?mid=239, van Swaay & van Strien 2005). Countries such as Spain, Greece and Poland (among others) are still lacking a national monitoring of butterflies based on individual counts. Design of national butterfly monitoring schemes could thus be recommended in these countries, and based on collaboration with experienced countries.

This kind of gap analysis can be a useful tool for national officers designing environmental policies, in particular those responsible for the reporting on state and trend of national biodiversity, as well as for any institution to justify priorities for a given monitoring that is still lacking in their country whereas it is implemented in several other countries, and that it is worth developing it.

In practice, that can be achieved by searching DaEuMon for similar schemes (e.g. same responses for Data type S2 and Taxonomic groups S26, and national Geographical scope) running in two countries or more in Europe but that are not yet existing for the country of interest. In the search, EU-wide and international schemes (cf. Geographical scope) may be highlighted as schemes that are by definition integrated across countries.

Obviously, reliability and usefulness of such a tool is strongly dependent on exhaustiveness of DaEuMon for the searched country. In any case, if decisions are taken to promote and maintain updated DaEuMon, this is a promising avenue for integrating monitoring schemes throughout Europe, promoting transfer of knowledge/methods and increasing biodiversity trend comparisons throughout Europe.

In addition, once quality criteria have been selected (with the help of Deliverable 17) it will be possible to identify schemes running in other countries than the one of interest and to sort these schemes according to the quality criteria, putting forward schemes of 'good' quality.

## IV.     Schemes differing in sampling design

An intuitive recommendation is to unify existing sampling designs. That is, all monitoring schemes should use the same sampling design, and same methods, so that all monitoring data are strictly similar and compatible. However, in practice, this is far from realistic. Separate schemes often have different goals (translating into different sampling designs), or may have different amounts of manpower (translating in sampling designs with more or less optimized sampling effort). When unifying existing sampling designs is not warranted, then methods for integrating data or results are necessary.

Major differences in characteristics of sampling design are:
- Ability to account for **detection probability** (S7). 69% of species schemes are declared as not accounting for detection probability (N = 297; 11/09/2006). Indeed, 28% of these schemes could partly account for detection probability since they collect some kind of replicates: 16 declare to have capture-mark-recapture data, 46 take two

samples or more per sampling sites (~ spatial replicates), and 26 visit sampling sites at least twice per year (~ temporal replicates). Although accounting for detection probability is considered as a critical issue when designing a monitoring design and analysing data (cf. Deliverable 2; Yoccoz et al. 2001), it is still insufficiently considered in current monitoring practices.

For integrating data or estimates across schemes with or without detection probability design, two approaches can be proposed. First, the issue of detection probability can be ignored when analysing data. In this case, it is straightforward to integrate data among schemes (cf. part V.1 of Deliverable 12). Second, estimates of detection probability can be obtained from schemes with a detection probability design. Then, these estimates of detection probability are used to alleviate the bias in estimates from schemes with inappropriate sampling designs. In other words, independent estimates of detection probability are included in the statistical model to account for uncertainty in the measure. For instance, from one scheme, detection probability is known to vary among sites, being twice higher in habitat 1 than in habitat 2. Other schemes do not document detection probability. The most reliable estimates of, e.g., abundance index per habitat over all schemes will be obtained by adjusting all data/estimates for differences in detection probability among habitats. Such statistical methods, integrating independent error measurements in statistical models, are to be developed.

- **Experimental designs** (S5). Only schemes with an (appropriate) experimental design can demonstrate that a given cause of change actually explains observed temporal or spatial trends. All other schemes can only establish correlative relationships between observed temporal and/or spatial trends of biodiversity and independent temporal or spatial trends in covariates (e.g., habitat use, temperature). Results from schemes with an experimental design may thus strengthen conclusions about causal relationships in schemes with no experimental design. For instance, if experimental schemes document the impacts of some agricultural practices on terrestrial insects, their results may help identifying the mechanisms responsible for changes in abundance and community composition of terrestrial insects documented in larger scale monitoring schemes that have no experimental design. Integration of schemes that test the same effect but with or without experimental designs would be implemented by integrating results across schemes with meta-analysis techniques (cf. 'Combine effect sizes', part 3.IV).

- **Stratification** (S4). If the sampling design was defined so that some habitats, regions, or species are more sampled than others, then the sampling design involves stratification. Often, stratification is used to optimize sampling effort according to monitoring goals. Integration of schemes with different stratification designs should be straightforward, i.e. consisting in giving the corresponding weights for each stratum in each monitoring schemes when analysing combined data/estimates (cf. part 3.IV).

- **Methods for choice of sampling sites** (S6):
  - 'random' or 'exhaustive' or 'systematic choice': combination is straightforward since all data/estimates are representative of the area (or at least the habitat, cf. S30) monitored;
  - 'chosen according to personal/expert knowledge' or 'other' or 'not answered': combination requires post-stratification within each scheme, and among all schemes, to correct for poorly represented habitats/regions prior to analysis.

**6**

## V.  Schemes differing in temporal scales

Within-year differences in timing (S16) of monitoring may occur for two main reasons: because different parts of the biological cycle are monitored (*e.g.*, monitoring birds in migration *versus* breeding birds), or because different temporal designs are used (*e.g.*, one *versus* several visits per year)

Within-year timing of monitoring needs to differ when different parts of the biological cycle are monitored. For instance, some schemes may monitor migrating birds at migrating sites, whereas others monitor migrating birds at breeding sites. In this case, the biological sense of combining estimates from both parts of the biological cycle is not straightforward and needs to be thought carefully. Rather than integrating both time series from different parts of the biological cycle, one may want to compare trends from each time series to extract information on population functioning from similarities and dissimilarities in temporal changes. For instance, if long-distance migrants decrease in the local breeding bird monitoring, but they are stable or increase in the migrant monitoring, it may suggest that local populations are decreasing but elsewhere some populations are stable or increasing.

Within-year timing of visits may also be designed to account for phenology. Two approaches may be considered to integrate schemes with different within-year timing. First approach: integrate data, and in this case, resample data so that the sampling design is the same in both datasets. For instance, let consider that most schemes count butterflies only once per year in July (type 1). Some other schemes count butterflies once per week (type 2). To combine data across schemes, one would choose randomly one week of July for each site from type 2 schemes, and take only these data for the global analysis. Second approach: produce estimates for each type of scheme. Theoretically, estimates adjusted for phenology should be more precise that those not adjusted. Thus precision (i.e. standard error) should be used for weighing estimates when computing the average trend over all schemes.

If among-year frequency in monitoring (S15) differs among schemes, then the same two approaches as for different within-year number of sampling could be used. Indeed, years when data are not available for some schemes correspond to missing values in the dataset. If data are exactly the same and can be combined, then use statistical methods that allow accounting for missing values (cf. part 5 of D2).

## VI.  Schemes differing in field data type / aim of scheme

Aim of scheme (S1: population trend; distribution trend; community/ecosystem trend) is more or less redundant with field data type (S2: presence/absence; counts; mark-recapture; age/size structure; phenology)[1]: if a scheme monitors only presence-absence, it should not declare that it monitors population trend; if a scheme monitors counts of individuals but declares that it monitors distribution only, they could also have declared that they monitor population trend (although the coordinator does not see it this way). Thus, here, we only rely on declared field data type to identify the main possible monitoring goals.

Combining schemes with different field data types may follow three approaches:
- **Homogenize data type across schemes**. That is transforming data to have the same data type in all monitoring schemes, i.e. degrade information to the lowest complexity

---

[1] Aim of scheme was requested to identify what is the aim of scheme. Field data type was requested to identify what is the data type actually collected. Field data type conditions what are the possible aims for the schemes, which can be broader than those identified by the scheme coordinator.

level available in the datasets to be combined. Capture-Mark-Recapture (CMR) data would be transformed into counts of individuals (i.e. number of individuals captured per unit of sampling effort; see e.g. part V.1 of Deliverable 12) and counts of individuals (and CMR data) would be transformed into species presence-absence data.

- **Homogenize parameters to be estimated**. Quantify the same biological parameter, for instance population growth rate, from different field data type (e.g. CMR, count, and presence-absence data), and then combine estimates of population growth rate across schemes. Differences in precision of estimates among field data types should be accounted when combining estimates by the use of appropriate weights (e.g. Gregory et al. 2005).

- **Use heterogeneity in data types to obtain complementary information**:
    - all data types can document distribution;
    - all data types can document population size (Strayer 1999; Pollock 2006);
    - CMR data can document population structure and demographic processes underlying changes in population size and distribution.

Goals of monitoring to fulfil member states reporting requirements using Natura 2000 and 2010 indicators (see also Deliverable 21 of WP4) can be summarized as: 'quantify state and trend in distribution and population size'. Thus, it is important to explicitly address integration by identifying which schemes should be combined so that distribution *and* population size are monitored per taxonomic group per country. This can be implemented by relying on heterogeneity in data type as suggested before. Note that only schemes with national goal should be included since, schemes with local scope cannot be considered as representative at the country level.

## VII. <u>Schemes differing in species identities</u>

**Increase coverage of the taxonomic group by integrating schemes monitoring different species**. If schemes monitor the same taxonomic groups (S26) but different species (S28), they could be integrated so that they complement each other and provide a broader coverage of the taxonomic group. For instance, if a scheme monitors wolves, and another one monitors bears, both schemes could be integrated into a 'large carnivores' monitoring programme, thus increasing biological scope and representativity, and sharing benefiting from results and experience of each schemes.

**Increase coverage of the taxonomic group by promoting the monitoring of more species per scheme**. If some schemes do not record all species they contact at sampling sites, it could be proposed to their coordination to increase the number of monitored species. If the same field method is suitable for more species than actually monitored, then with no or little additional costs, the biological coverage of the monitoring can be much increased. For instance, if a scheme monitors game birds only in agricultural lands with count points, then participants of the scheme could also monitor raptors and/or songbirds. Schemes that use closed lists of species to be monitored instead of monitoring all contacted species could be identified on the basis of the declared number of species monitored (S27): if the number of species monitored is lower than the average number of species monitored for the same taxonomic group (S26) and the same geographical scope, then they supposedly could monitor more species than they do.
Note that opening the list of species monitored may also be proposed across taxonomic groups. For instance, if a scheme monitors vocal night birds only, persons may also be

requested to record vocal night mammals and vocal night insects. For the same monitoring costs, the biological scope would be much increased. This issue is further developed from the sociological (i.e. motivating volunteers to increase the number of species recorded) and organisational (i.e. inter-organisational cooperation) points of view in outputs of WP1. An important point is that participants may need extra training for acquiring necessary identification skills. Organisations that have a system of individual awards and recognition, channels for feedback and good two-way communication with volunteers are most likely to succeed in increasing the number of species and taxonomic groups that participants are expected to monitor.

**Integration of schemes to produce multi-species indicators of state and trend.** Integration across species can be achieved by combining single-species trends to produce multi-species trends, usually used as biodiversity indicator (e.g. Gregory et al. 2005). This across-species integration has the advantage of producing information with a broader biodiversity scope than multiple single-species approaches (and other advantages of biodiversity indicators; Balmford et al. 2005; Mace 2005). However, a limitation of multi-species combinations is that heterogeneity among species may render interpretation of the index poorly reliable. For instance, a combined trend of top predator carnivores and of micro-insectivores would be hard to interpret. At this level of among-species heterogeneity, complementarity, instead of homogeneity, should be used. Biodiversity indicators should be built from the different schemes by grouping species per habitat, functional trait, trophic level, specialization index, or any other grouping level that make sense for the purpose of monitoring changes in biodiversity. As the aim is to monitor the state and trend of biodiversity as robustly as possible the higher the number of species monitored, the better within-taxonomic group diversity will be covered.

## VIII. <u>Schemes differing in taxonomic group</u>

As the aim is to monitor the state of biodiversity as broadly as possible, the greater the number of taxonomic groups monitored, the more representatively biodiversity will be covered. When combining data or trends from different taxonomic groups, the main limitation comes from lack of scientific/theoretical knowledge (rather than from statistical methods or data availability). The same methods as combining estimates across species (part 2.VII) could be used, but the results will be difficult to interpret (see for instance the Living Planet Index, Loh et al. 2005, that combines all available data into a single index, whatever realms, habitats and life-history traits). We may prefer relying on a theoretical framework that formally links the different taxonomic groups (see for instance the Marine Trophic Index that relies on knowledge of trophic networks; Pauly & Watson 2005). Currently, no general recommendation for such integration approaches is available. They depend on the theoretical understanding of mechanisms linking species and taxonomic group within monitored ecosystems and the goals of the monitoring. For example if we are interested in monitoring the effects of habitat loss and fragmentation on biodiversity, we may combine species from different taxonomic group that share some biological properties making them sensitive to these processes (Freudenberger & Brooker 2004; Henle et al. 2004).

Different taxonomic groups are expected not to have the same response to environmental changes. For instance, butterflies may react more rapidly (short lived) and at a small spatial scale (highly specialized) to habitat changes, whereas bird would react more smoothly (longer-lived) and at a broader spatial scale (less specialized). Thus monitoring birds and butterflies may provide a complementary understanding of biodiversity changes. Butterflies would document changes on the short-term, whereas bird would document changes on the long-term (e.g. Thomas 2005). With such an approach, combining schemes monitoring

different taxonomic groups should increase the biodiversity coverage and representativity of conclusions drawn from monitoring data. However, we should care not to define fixed taxonomic group indicators, which may have the same weaknesses as generally brought forward against the use of indicator species: lack of scientifically robust demonstrations, lack of generality across biogeographic regions, static view of ecological processes ignoring evolutionary mechanisms, etc.

## IX.    Integrating species and habitat monitoring schemes

A tool providing guidance for integrating schemes that monitor species and schemes that monitor the habitats of these species may serve several purposes:
- member states will have to report on state and trend in distribution and population size of species and their habitat to evaluate the conservation status of species in the annexes of the Bird and Habitat Directive. Thus, they will explicitly need to combine information from schemes monitoring species and schemes monitoring habitats;
- coordinators of species monitoring schemes may use this tool to find who monitors habitats of their focal species;
- it would allow obtaining datasets to test whether it is "better" to monitor a set of species to evaluate habitat status, or to monitor habitat quality criteria to evaluate species status, "better" being evaluated according to quality criteria proposed in Deliverable 17. This is particularly relevant when evaluating species conservation status: if habitat distribution decreases by more than 5% per year, then it is not necessary to gather numerical data for distribution and population size: according to Natura 2000 guidelines, the only conservation status possible is 'bad'.


If habitat preferences of species are known, we could link Species schemes and Habitat schemes in DaEuMon to identify which schemes monitor which species, and which schemes monitors their habitats. Indeed, in the Habitat Directives, species are linked to certain habitats by definition, often called typical or characteristic species (cf. outputs of Scientific Working Group). So we may use these fixed species-habitat links to integrate species and habitat schemes. These links are fully relevant for integration guided by environmental policy, although they are questionable according to the ecological processes behind them (e.g., several of the Annex species are widely roaming species that cannot be directly linked to a habitat in the sense of being a typical species characterising the habitat, so how to account for this uncertainty?). This approach is further developed in recommendations for integrating habitat monitoring schemes (Deliverable 19) and will be taken up by WP5 for the development of a general framework for integrating monitoring schemes.

Integration of species and habitat monitoring and to link species to habitats requires a standardization of habitat descriptions throughout Europe. The current process for producing species indicators per habitat is (cf. Headline Species indicators, e.g. Gregory et al. 2005):
- experts attribute supposedly specialist species to broad habitat types;
- compute geometric mean trend for these species among countries per habitat type.

A weakness of this approach is that specialists of broad habitat types differ across biogeographical regions. For instance, woodpeckers are supposedly indicators of forest condition, but woodpecker species vary throughout Europe, and some are more specialized than others. Likewise, species of the butterfly genus *Maculinea* differ in their habitat requirements across their distribution (see results of the EU-project MacMan; Settle et al. 2005). Standardized definitions of habitats throughout Europe and standard indices of species specialisation would allow computing at national or biogeographical region level, at which a species is a specialist of which habitat (cf. e.g. Julliard et al. 2006). That would allow

discarding *a priori* attribution of species to given habitats at the European scale. This issue may be developed further by WP5.

### X.   Integrating schemes according to the causes of change of biodiversity they monitor

If spatial and temporal trends in biodiversity are monitored, it is usually with the hope of being able to identify and to quantify causes of biodiversity change. Two avenues for integration can be proposed:
-   schemes with similar causes of change (S32): potential for meta-analysis and thus more robust tests and quantification of causes of change;
-   schemes with different causes of change (S32), including identification of causes of change not monitored at all: their integration would help in securing that at least one scheme aims at documenting the impact of each major causes of change of biodiversity per country.

For schemes with an experimental design, and if schemes were originally designed to quantify the effect of a specific cause of change, they have good chances theoretically to end up with conclusive results according to the role of the given cause of change. Indeed, only 20% of schemes follow an experimental design (S5) and declare to monitor at least one cause of change (S31; N=372, DaEuMon, 27/10/2006). More commonly, causes of change are identified by *post-hoc* correlative comparisons, though these correlative comparisons can come very close to an experimental design (e.g. when different areas are monitored, and some are affected by a change whereas others are not; compare Henle 2005). Thus, combining correlative information from different, independent schemes strengthens conclusions on the likely causes of change.

If species indicators or taxonomic group indicators of given causes of changes are available (e.g. Henle et al. 2005), combination of information from these different indicator species/taxa would be a way to obtain a synthetic evaluation of these causes of change. While a combination of these indicator values is always possible care must be taken that they are not combined rather arbitrarily which would make them not very helpful as measurements because then each scheme may probably trying to tackle a specific threat and not biodiversity as a whole. If we go for *a priori* determined indicators, such as in the focal species approach (Freudenberger & Brooker 2004), we may face weaknesses that may plague the practical application of bioindicators in general: often their robustness has not been tested sufficiently, they may lack generality across biogeographic regions, and they may take a static view of ecological processes ignoring evolutionary mechanisms (but see e.g. McGeoch 1998, Dziock et al. 2006). The BioScore project[2] will develop a set of biodiversity indicators for the major threats to biodiversity. Recommendations for integration across taxonomic groups for producing biodiversity threat indicators should be based on their work.

# (3) Statistical methods for integration

A preliminary discussion of methods can be found in Chapter (7) of D2 (+ Figure 4)

### I.   Combine data

For combining data across schemes, the measurement unit needs to be the same (e.g. number of individuals per unit of sampling effort). Then, it is rather straightforward to combine

---

[2] http://www.ecnc.nl/StateOfEuropeanNatur/Bioscore_529.html

schemes, although weighing issues have to be considered for compensating differences in representativity among schemes.

## II. <u>Combine estimates</u>

For combining estimates, measurement units do not need to be the same, but the same biological process has to be monitored (e.g. annual change in population size, annual rate of local extinction; cf. part 2.VI, see also part V.1 of Deliverable 12). After producing estimates of the same biological process from all datasets, one needs to decide to how to weigh estimates from different schemes when combining them into a single one (cf. part 3.IV).

## III. <u>Combine effect sizes</u>

Different monitoring schemes may document the same causes of change of biodiversity. However, they may use too different sampling designs, data types, or biological models, making combination of data or estimates difficult. However, if they document the same cause of change, they have one measurement in common: the effect size for the covariate along which the change of biodiversity indicator is tested. This effect size concept comes from statistical methods for meta-analysis (e.g. Hedges & Olkin 1985; Gurevitch et al. 2001; Gates 2002). Effect size quantifies the strength of the effect of an explanatory variable in a statistical analysis. Effect size $z$ is the estimate of the slope for the cause of change included in the model, divided by the standard error of the slope estimate. In practice, in an experimental design testing for the effect of increased nutrient concentration on plant height, the effect size will be the slope linking plant height to nutrient concentration. Another example, more widely transferable to monitoring data, is the test of the effect of average yearly temperature to document climate change. In this case, the effect size will be the slope linking the annual measurements (e.g. date of first reproductive event of the year, total biomass) to average annual temperatures. Effect size can also be computed for categorical explanatory variables, for instance, with or without addition of nutrients.

The principle of meta-analysis is then to combine estimates of effect sizes from all available schemes, and to quantify the average strength of the cause of change under study. Combining effect size estimates from different schemes will increase representativity of conclusions drawn from monitoring data and statistical power to detect effects of investigated causes of change.

In practice, effect size is computed independently for each scheme. Then the mean effect size is computed by summing effect sizes from all schemes and dividing it by the square-root of the number of degrees of freedom (i.e. number of schemes – 1). If the supposed cause of change has no effect, the mean effect size has a standard normal distribution N(0,1). When there is an effect, the expectation will depart from 0. Whatever the magnitude of the true effect in each scheme, the expectation of the test statistic will be negative, if there is a general negative effect, or positive, if there is a general positive effect. The power of the resulting meta-analysis will depend on the magnitude and precision of the effects in the various schemes. An advantage of this procedure is that the power will be non-negligible in the case of small to moderate effects in all schemes, i.e. the test has a good probability of detecting the effect of the cause of change, which is not the case for separate tests on each scheme.

Note that if you combine effect sizes for linear trends among schemes (e.g., slope for the year effect treated as a continuous variable in a GLM), this is equivalent to combining standardized estimates of linear temporal trends, as suggested in part (3).II, but in addition, you can test that the mean linear temporal trend is significantly different from stability (no trend) overall datasets.

If qualitative information only is available for the cause of change (e.g., significant positive, non-significant, significant negative), non-parametric tests can be used for identifying

whether the proposed cause of change has a significant effect, on average, over all schemes (e.g. Zar 1996).

## IV. <u>Weights</u>

When combining information from different schemes, a general issue is what is the relative importance of each separate scheme: same importance or is one more important than others (because it is more representative or has higher monitoring effort)? This issue is usually solved by attributing different weights to data (or estimates) from different schemes.

Below follows a listing of the major sources of difference among schemes and associated possible weights to be used in analyses:

- If estimates have different precisions (i.e. standard errors), the weight to be used should be $1/(se)^2$.
- If sampling effort varies across schemes to be integrated, this should translate into different precisions. Indeed, standard error is driven by two components: natural variability of the variable monitored and sample size. Thus when standard errors are not available, sampling effort (e.g. number of sites sampled per year, S12) may be used as a surrogate variable to at least partly correct for differences in sample size among schemes. A possible weight could be $N^2$ (or $S12^2$).
- If population sizes differ across monitored geographical countries (or regions) to be combined, a weight to use would be the proportion of the European population size per country.
- If different habitats are monitored, and a single measure is to be obtained for all habitats at a national scale, a weight for combination across habitats would be the percentage of coverage per habitat type at national scale.
- If sampling designs differ across schemes, e.g. when sites are chosen according to expert knowledge (or "other"; cf. S6) or when stratification across habitats is different among schemes (cf. S4), raw data are not representative of state at the country level because some habitat types are more sampled than others. Then, data (or estimates) need to be post-stratified for habitat availability. A possible weight per habitat type is the percentage of total habitat coverage that was actually sampled.
- If different species or different taxonomic groups are to be combined, several weights can be considered. (i) *No* weight is used when biological knowledge of the relationship among species and taxa is insufficient (that is the same weight is given to all species and taxonomic groups; e.g. Living Planet Index, Loh et al. 2005). (ii) Values per scheme can be combined by geometric mean (instead or arithmetic mean) to give a higher weight to low values (Buckland et al. 2005; Gregory et al. 2005). (iii) *Ad hoc* weights can be used to give priority to a given biological property, e.g. degree of specialization, rarity, ecosystem function, functional trait, trophic levels, etc (e.g. Butchart et al. 2005), policy goals or national responsibility (see D21 of WP4). Except of national responsibilities, these specific weights are not developed here since they are to be defined according to analysis goals. (iv) A particular way of weighing can also be used to adjust for phylogenetic non-independence across species. Species have different phylogenetic relationships. The same response to a given environmental change from distant species is more convincing about the general impact of the change on species trends than a same response exhibited by closely related species. Therefore, the analysis should thus account for these differences in phylogenetic relationships by giving different weights to the different species. Comparisons among distant species should be given a higher weight (because addressing highly independent species) than comparisons between closely related species (which may exhibit the same response

because of their common evolutionary history rather than because of a similar ecological response). This can be viewed as a particular way of weighting data across species. See presentation in Harvey & Pagel 1991). Note also that pairs of closely related species, with contrasted ecological requirements, and that exhibit different responses to changes in their environment are also highly informative when linking ecological requirements to environmental changes (Møller & Birkhead 1992).

# (4) References

Balmford, A, Bennun, L, Ten Brink, B, Cooper, D, Côté, IM, Crane, P, Dobson, AP, Dudley, N, Dutton, I, Green, RE, Gregory, RD, Harrison, J, Kennedy, ET, Kremen, C, Leader Williams, N, Lovejoy, TE, Mace, G, May, RM, Mayaux, P, Morling, P, Phillips, J, Redford, K, Ricketts, TH, Rodriguez, JP, Sanjayan, MA, Schei, PJ, Van Jaarsveld, AS, & Walther, BA 2005. The Convention on Biological Diversity's 2010 target. Science 307: 212-213.

Buckland, ST, Magurran, AE, Green, RE, & Fewster, RM 2005. Monitoring change in biodiversity through composite indices. Philosophical Transactions of the Royal Society B-Biological Sciences 360: 243-254.

Butchart, SHM, Stattersfield, JA, Baillie, J, Bennun, LA, Stuart, SN, Akçakaya, HR, Hilton-Taylor, C, & Mace, GM 2005. Using Red List Indices to measure progress towards the 2010 target and beyond. Philosophical Transactions of the Royal Society of London Series B-Biological Sciences 360: 255-268.

Dziock, F, Henle, K, Foeckler, F, Follner K & Scholz, M, 2006. Biological indicator systems in floodplains - a review. International Review of Hydrobiology. 91: 271-291.

Freudenberger, D, & Brooker, L 2004. Development of the focal species approach for biodiversity conservation in the temperate agricultural zones of Australia. Biodiversity and Conservation 13: 253-274.

Gates, S 2002. Review of methodology of quantitative reviews using meta-analysis in ecology. Journal of Animal Ecology 71: 547-557.

Gregory, RD, van Strien, A, Vorisek, P, Gmelig Meyling, AW, Noble, DG, Foppen, RPB, & Gibbons, DW 2005. Developing indicators for European birds. Philosophical Transactions of the Royal Society of London Series B-Biological Sciences 360: 269-288.

Gurevitch, J, Curtis, PS, & Jones, MH 2001. Meta-analysis in ecology. In: Advances in Ecological Research, Vol 32. eds., pp. 199-247.

Harvey, PH, & Pagel, M 1991. The comparative method in evolutionary biology. Oxford University Press, Oxford. pp.

Hedges, LV, & Olkin, I 1985. Statistical methods for meta-analysis. Academic Press, Orlando.

Henle, K 2005. Lessons from Europe. S. 64-74 in Lannoo, M.: Amphibian Declines. The Conservation Status of United States Species. University of California Press, Berkeley, USA.

Henle, K, Davies, KF, Kleyer, M, Margules, C, & Settele J 2004. Predictors of species sensitivity to fragmentation. Biodiversity and Conservation 13: 207-251.

Julliard, R, Clavel, J, Devictor, V, Jiguet, F, & Couvet, D 2006. Spatial segregation of specialists and generalists in bird communities. Ecology Letters 9: 1237-1244.

Loh, J, Green, RE, Ricketts, T, Lamoreux, J, Jenkins, M, Kapos, V, & Randers, J 2005. The Living Planet Index: using species population time series to track trends in biodiversity. Philosophical Transactions of the Royal Society B-Biological Sciences 360: 289-295.

Mace, GM 2005. A user's guide to biodiversity indicators. European Academies Science Advisory Council, The Royal Society, London. 42 p.

McGeoch, MA, 1998. The selection, testing and application of terrestrial insects as bioindicators. Biological Review 73: 181-202.

Møller AP & Birkhead, TR 1992. A pairwise comparative method as illustrated by copulation frequency in birds. The American Naturalist 139: 644-656.

Pauly, D, & Watson, R 2005. Background and interpretation of the 'Marine Trophic Index' as a measure of biodiversity. Philosophical Transactions of the Royal Society B-Biological Sciences 360: 415-423.

Pollock, JF 2006. Detecting population declines over large areas with presence-absence, time-to-encounter, and count survey methods. Conservation Biology 20: 882-892.

Settele, J, Kühn, E & Thomas, JA (eds) 2005. Studies on the ecology and conservation of Butterflies in Europe. Vol 2: Species ecology along a European gradient: *Maculinea* butterflies as a model. Pensoft Publishers, Sofia, Bulgaria.

Strayer, DL 1999. Statistical power of presence-absence data to detect population declines. Conservation Biology 13: 1034-1038.

Thomas, JA 2005. Monitoring change in the abundance and distribution of insects using butterflies and other indicator groups. Philosophical Transactions of the Royal Society of London Series B-Biological Sciences 360: 339-357.

Van Swaay, CAM, & van Strien, A 2005. Using butterfly monitoring data to develop a European grassland butterfly indicator, pp. 106-108. In: Studies on the ecology and conservation of Butterflies in Europe. Vol 1: General concepts and case studies. Kuehn, E, Thomas, JA, Feldmann, R, & Settele, J, eds., pp. 128. Pensoft Publishers, Sofia, Bulgaria.

Yoccoz, NG, Nichols, JD, & Boulinier, T 2001. Monitoring of biological diversity in space and time. Trends in Ecology & Evolution 16: 446-453.

Zar, JH 1996. Biostatistical analysis. Prentice-Hall Inc., Uppler Saddle River, New Jersey. 662 pp.