

Overview

In 2020, Laikre, Hoban and colleagues outlined an approach to assess genetic diversity status and trends *without requiring any genetic data (e.g. DNA sequence)*, using 'genetic indicators'. The indicators were developed in response to a recognized gap in reporting to the U.N. Convention on Biological Diversity (CBD). There was a need for tools to report on 'genetic erosion' and 'safeguarding' or 'maintaining' genetic diversity, within and among populations of species. Two indicators allow a fairly standardized and rapid way to assess *whether a species is likely to be losing or has lost genetic diversity* by quantifying critical aspects of demography that generally correlate to genetic diversity: (1) the size of each population (number of adult individuals), and (2) the geographic range or number of populations relative to historic conditions. A third indicator on genetic knowledge/ genetic studies has also been proposed.

The logic behind these indicators is simple. (1) Small populations lose genetic diversity, and very small populations lose genetic diversity very quickly (Frankham 2021). (2) Loss of populations can result in loss of unique genetic adaptations (Exposito Alonso et al 2022). So, measuring population size and loss of populations is a fairly good proxy or summary of genetic diversity status, without ever measuring the DNA diversity itself.

The indicators should be applied to a *curated, representative set of species* (represent a range of habitats, taxonomic groups, traits, etc.)- 100 to several thousand species per country, depending on a country's biodiversity, data capacity, and resources for CBD reporting. Approximately 100 species is likely a minimum to represent diverse habitats, taxonomic groups, commonness/ rarity, and threat status. However, if 100 cannot be collected, any amount of data is important.

The first two genetic diversity indicators are:

1. The first indicator is based on comparing **the effective population size of each population N_e to a critical threshold** (e.g. $N_e = 500$). For many species, it is sufficient and appropriate to use the **census size N_c (the number of living adults) as a proxy for N_e , and a threshold of $N_c = 5000$ mature individuals**. Below this threshold, a population rapidly loses genetic diversity, can become inbred, and starts to lose ability to adapt to environmental change.
2. The second indicator is based on comparing **the current number of populations that exist to a prior/ historic assessment of the number of populations**. This reflects **loss of populations** to human-induced changes, with 50-200 years ago as a baseline (depending on the country). If counts of populations existing and lost are not available, a proportion (or

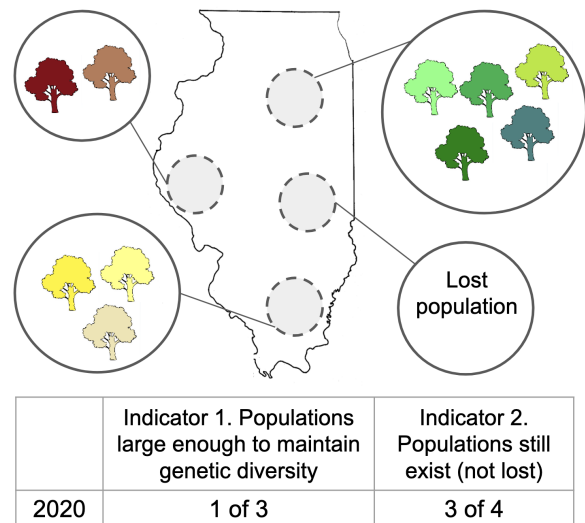


Fig 1: Illustration of genetic diversity indicators, for four populations in Illinois, USA, measured in the year 2020. One tree = 1000 plants. Colors show genetic diversity. In 2020, one population is extinct, and 2 of 3 are too small.

percentage) of the species' range are lost is an acceptable substitute. Lastly, if this is not available, some estimate of overall decline is acceptable (details will be explained below). A third indicator exists, a count of the number of species with genetic studies available. Although this does not directly relate to maintenance of genetic diversity or preventing genetic erosion, performing genetic studies and collecting genetic data correlates to management actions designed to manage genetic diversity- such studies help understand needed management actions and to guide them. So we will also collect information on whether genetic studies exist for species, for indicator 3. However, this project will *not use genetic datasets to calculate the Ne or Nc* because reanalysis of data is time prohibitive- we are not downloading and analyzing genetic data.

The primary undertaking is to gather the data on populations for each species. This is a challenge because there is no global, standard database of population census size (The Living Planet Index for example does not measure full population census, and not all Red List species have census size for each population). But the census size of many populations of many species is available in different reports and databases, in more or less easy ways to extract. **This guidance document will explain how to gather and use the necessary data, from diverse sources, in a standard way.**

The Project Purpose

This project was planned in 2021 and initiated in 2022 to test the indicators in a small number of countries. The project goal is for each country team to evaluate >100 species (per country) to determine (a) how many species have the required data, (b) to extract the data when possible for indicator calculation, and (c) to identify barriers encountered so that this guidance document and indicator calculation can be improved for larger scale use by more countries to inform the CBD framework. The project will also highlight species and regions where data is deficient, or where there is high uncertainty in the estimates.

Getting started- how many species, which species?

Creating species lists

Each country team first will make a list of at least 100 candidate species for which there is likely some data for indicator 1 and/or 2. *'Likely to have some data'* means that the species are *not* recently discovered/ poorly known/ very hard to document population size (e.g. they are countable by observation, camera trap, etc.), do not have taxonomic disputes, etc. This is the list of species to *try* to collect data for.

Following are two ways to make this list of 100 species, though other approaches or a blend of approaches is fine.

1. First, compose a list of *species* at the country level that a national biodiversity expert or panel of experts thinks might have data. Then, 'cross check' this list against relevant sources of data to narrow it down (e.g. removing species for which there are no published reports,

articles, websites, databases, or experts available). This approach could lead to over-representation of well-known, flagship, or economically important species.

2. Choose one or two prominent data sources (e.g. recovery plans or similar), list all species in that data source, and pick species from this list in a stratified random fashion to cover taxonomy, habitat, etc.. For example, this might involve going through recovery plans for all federally listed Endangered Species, the national Red List, or other lists of conservation concern (e.g. Annex II, IV and V species of the EU Habitats Directive- a defined list of policy importance). This could lead to overrepresentation of species of conservation concern/ underrepresentation of common or "least concern" species. Many countries have Red Lists for various taxonomic groups. These lists could also help ensure each threatened status is represented (Endangered, Least Concern, etc). Note: many LC IUCN species are nevertheless of local or regional conservation concern, and are declining rapidly, etc. so should not be ignored.

It is vital to document how the list is developed in order to identify any biases (e.g. mostly common species). In this project, and in the first use of the indicators by a country for National Reporting, **it is acceptable to have some biases**, but as data quality and collection efforts improve, biases should decrease. Multiple data sources may be needed (for example: scope the Red List to see what species have data available, and consult with experts on other data sources).

It is not necessary for all chosen species to have high quality data across their range. While indicators would be more accurate if all species have data for all populations, complete population data may only be available rarely. *It is ok if data are available for only one of the two indicators or for only some populations of a species* (as explained under Common Issues below). Moreover, upon investigation, species initially deemed likely to have some data, may actually have insufficient data to calculate either indicator. Species should not be removed from the list after the initial list is made. We will calculate the indicators with and without various types and levels of missing data.

There are some species where it will be particularly hard or impossible to quantify Indicators 1 and 2, and they should be excluded from the species list. For example, evaluation of the $N_e > 500$ criterion will be hard in species where natural subpopulations are typically very large and/or hard to measure, such as microcrustaceans, many insects, some fungi, highly clonal organisms, some plants with deep soil seed banks (where all 'individuals' cannot be counted). Populations of such species can also grow in a short amount of time to very large numbers and have high standing genetic variation (Chaturvedi et al. 2021). We advise not attempting to include such species in a country's first evaluation of these indicators due to difficulty in finding and interpreting data.

How to choose a diversity of species

As noted above, candidates for the list of 'selected, representative species' might be species with management or recovery plans, species that have been Red List assessed, species of known ecological or economic importance, species of national concern, or species monitored by a national biodiversity entity or citizen scientists. They should be representative (e.g. able to provide a representative picture of what other species are experiencing for the indicators). This may include species of commercial interest, but should not be limited to commercially important species. If possible, the species should represent a diversity of at least some of:

- Ecosystems / biomes within the country (e.g. temperate forests, rainforest, etc)

- Taxonomic groups (e.g. mammals, invertebrates, herpetofauna, plants, birds, fish)
- Terrestrial and aquatic
- Range size and commonness (see Kobo form for descriptions of terms)
- Threatened status (threatened or not)

If possible, also include a range of:

- Value (economic, ecological, social)
- Traits (body size, lifespan/ generation time, dispersal)

The ability to fully include all of these areas will vary by country, due to native levels of biodiversity and capacity/ data availability. Further discussion of categories of species can be found in [Hollingsworth et al 2020](#), page 22-24. Full taxonomic and ecological diversity will be hard, or impossible, for the first application of an indicator. Bias, including focus on some taxonomic groups e.g. initiated with trees, amphibians, birds, medicinal plants, etc., is ok, as bias will be noted in reporting. Even a first attempt helps a country set up the infrastructure and methods of data analysis. We have created a 'species selection matrix' as a guide to help countries visualize and document the types of species chosen.

For wild species, the species considered should usually be native (e.g. non-introduced, non-invasive). Note however, that this indicator can, and often should, also be applied to some proportion of crops, domesticated animals, and crop wild relatives, and in such cases non-native species would be ok.

Keep the same species list while collecting data

It is important to fill in the Kobo form for every species on the 100+ species list- even if there is little data available, and even if the indicator cannot be calculated. In other words, we expect some proportion of this list to have no useful data and that is ok (see Fig 2). Knowing how much data is missing will provide valuable insight into challenges countries will face in assessing species for these indicators and where data are lacking. However, if missing data is extensive (say, >50% of the species list), country teams may revisit their original species list and continue to add more species.

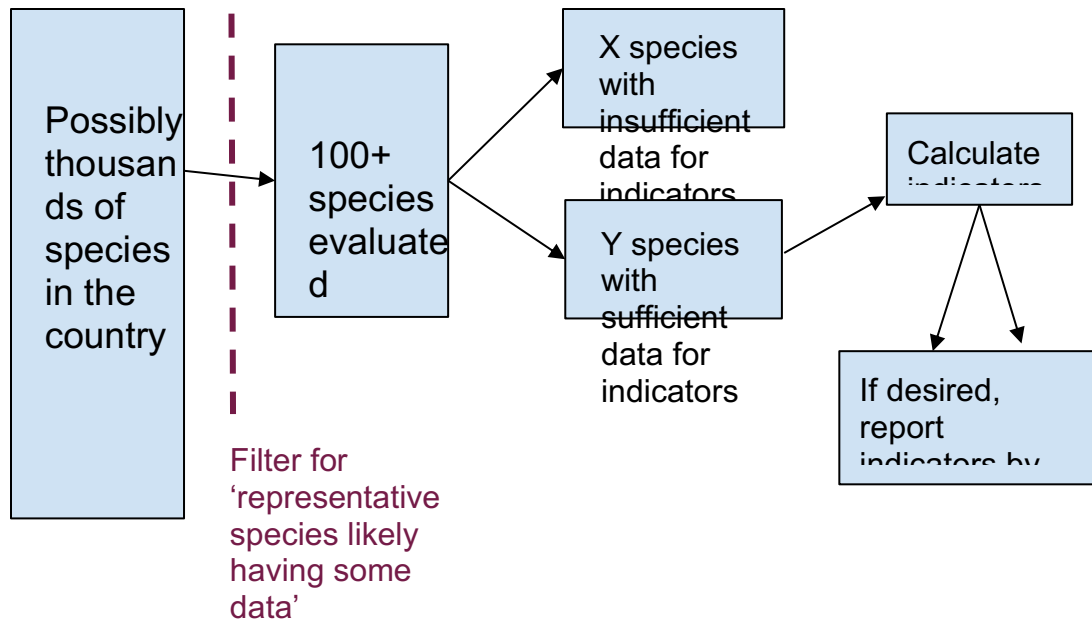


Figure 2: Conceptual illustration of the fact that each country will evaluate >100 species, and that some proportion of species evaluated will have insufficient data for calculation of the indicators. The proportion of $X/(X+Y)$ may (possibly, but not necessarily) suggest some groups or countries that consistently have limited data for calculating indicators.

Collecting data- where and how

Collection of data will proceed with slight variations for each country and potentially for taxonomic groups and data sources within each country. **Please note that data are only to be collected for populations existing in the reporting country; a country is not obligated or expected to report on populations existing entirely outside the country.** Populations that are transboundary (e.g. part of the population exists in the reporting country) may be reported on.

The goal for indicator 1 is to extract a current census size (and/or, if available, effective size) for each population for each species. It is also important to record associated data such as year, reliability/ uncertainty, population names or geographic information, reference information, and other information of the species. This is why we have created a Kobo form and Kobo guide.

Note: N_c represents the number of adults present in the population. For birds, estimates of numbers of breeding pairs (if available) can be used, which represents $0.5 \cdot N_c$. Thus, multiply the number of breeding pairs by 2 to get the N_c .

The goal for indicator 2 is to extract a count of the number of populations existing today, and which existed at a historic time point. To allow maximum use of each country's data and flexibility, we do not define 'historic time point' specifically but we offer guidance below.

Collection of data may be very straightforward or fairly laborious, depending on the country and data source. Some countries may have a centralized database of many species (across taxonomic groups and levels of rarity) from which population sizes and counts can be extracted

directly and quickly. This may be especially the case for plants, long lived organisms and those of high economic, ecological or cultural value. Extracting data directly from computer files or tables would be very straightforward. For more countries we anticipate that data will need to be extracted manually by humans by reading reports, websites, planning documents, and/or expert consultation. Often it may be necessary to consult more than one resource. Options include

- Endangered or declining species management reports. Some countries have mechanisms to publish recovery plans, status, or other information about selected species. These reports often list current population sizes and current and sometimes historic range or population occurrence.
- IUCN Red List or NatureServe. More than 100,000 species have been assessed by the IUCN Red List at global or national levels. These reports sometimes list current population sizes and current and sometimes historic range or population occurrence.
- Literature or internet searches. Scientific journal articles sometimes present census size and current and historic range information in the Introduction, Methods or Discussion sections. This may be especially useful for species with old, outdated Red List assessments or where population level information is not otherwise available. Wikipedia, NatureServe, WWF, taxon specific NGOs, and other respected websites may also have relevant information.
- Expert consultation. For some taxonomic groups it may be feasible and indeed most efficient to present the list of species to a panel of taxonomic experts (e.g. amphibians, trees, etc.) who would have up to date and possibly unpublished knowledge and can also provide measures of uncertainty. This can be an efficient way to gather data on dozens of species in a short period of time. Convening a workshop (virtually or in person with a goal of discussion and consensus) of experts might be considered as well for gathering information.
- Occurrence points from GBIF or citizen science. In some cases this can help define populations, define habitat area, and possibly even define lost populations if these points have high reliability. They should be interpreted with caution, because (i) not every occurrence is a population (as mentioned above) and (ii) such occurrence points might mix occurrences from different years, so spatial and temporal occurrences can be conflated (e.g. a GBIF map shows an extant population in a place because of old occurrences, but the population is actually extinct). As with all data interpretation, use caution.

If doing manual data extraction, each data source should be read thoroughly. Many Red List assessments and management plans might be 1 to 20 pages long (sometimes longer). The text might clearly state the size of each population, with text, tables, or maps. However, data may be incomplete, such as listing the size of only the largest or smallest populations. As noted above, it is ok to submit species for which some information is missing (see common Issues below). The time necessary to gather data from individual sources (and/or consult experts) may range from 45 minutes (for clear, short reports), to several hours.

Data may be recorded directly on the Kobo form, or on printed paper, or other means, and then transferred to the Kobo form.

Important! In instances where data sources provide different information for species' populations (e.g., different size estimates for the same population or different names of populations) and the assessor cannot determine which is most reliable and/or the data cannot be amalgamated, assessors are **required to submit separate data entries** (Kobo forms) for the species (refer to Step-by-step guide for further guidance and examples). We trust the expertise of the assessor in deciding how many submissions. For a given species, it may be four submissions for example, to reflect uncertainty in 'what is a population' and differences in population size from different reports.

Calculations and reporting

The assessors will not do any direct calculation of indicators. They will fill in a Kobo form for each species with information on Ne and/or Nc, populations maintained, numerous other characteristics, references, maps, etc. For most species the Nc, census size, will be the main data collected, rather than Ne. If Ne is estimated in the report or publication it should be reported, of course. **However, assessors should not calculate Ne from Nc manually at this point-** it will be calculated as follows after all data collection is complete.

After all species and populations are reported for a country, for all cases in which Nc was recorded, the colleagues involved in data analysis (see project roles spreadsheet) will then apply at least two Ne/Nc ratios to all populations in order to obtain Ne estimates from Nc- the 0.1 as a conservative default for all species, and a taxon specific one (either from that species or from a general taxa such as 0.3 for plants). We will do this in multiple ways to generate confidence bounds for reporting in the National Reports (a low and a high estimate of Ne). This will be done also in the case of multiple estimates of Nc for a population of a given species. This would result in, for example, potentially four values: low and high Nc estimates and low and high Ne/Nc ratio assumptions.

Then all Ne values (directly extracted, and calculated from Nc) will be compared to Ne 500. Every species will receive a ratio of populations above Ne 500. This will be reported as a proportion, but the original ratio (including total number of populations) can also be retained.

The country indicator value is the mean across species (a median could be used for skewed distributions). If taxonomic groups are not represented evenly (as is likely), the indicator value is the mean of each taxonomic group's means, which would downweight taxonomic groups that are overly represented, e.g. mammals or birds. Optionally, each species can be weighted by the proportion of its geographic range in the country, from 0 to 1, to reflect national responsibility, with full weight for endemic species (REF).

The indicator is easily disaggregated to different taxonomic groups by only including species in that taxonomic category. The same can be done for different habitats, species commonness, etc.

How to define populations

This is one of the most important concepts to agree on before searching for data. The definition of a population does differ depending on the species (e.g. the spatial extent for a tree population and a salamander may not be the same) but here follows some general guidance.

Briefly, 'population' refers to a genetically distinct group of individuals. **Within populations the individuals are capable of interbreeding. Genetic distinction of populations is usually due to no or very low gene flow (movement of genes e.g. by movement of offspring) and/or significant adaptive differences, such as due to occurring in different environments (high/ low altitude, different soil types, differences in precipitation or temperature).** Not every occurrence is a population (see Appendix)! The word 'subpopulation' may describe clusters of organisms across a landscape (including family units) which are near enough to exchange gene flow. Usually several 'subpopulations' are considered together as a population (also called a 'metapopulation'). When the potential for gene flow is large, e.g., viable tree pollen that can travel tens of kilometers, 'populations' can range across large distances, sometimes hundreds of km.

Important! In IUCN Red List reports the term 'population' and 'subpopulation' are used differently than intended in this study. An 'IUCN population' refers to the entire species (i.e., the total number of individuals of the taxon). 'IUCN subpopulations' are geographically or otherwise distinct groups of the species between which there is little exchange (IUCN 1995); thus an 'IUCN subpopulation' is what we refer to as 'populations'!

Some published reports will clearly define what a population is based on the knowledge of biodiversity and taxonomic experts. Some judgment may be required in interpreting population designation in a report, but in general the designation of a population by experts should be considered strongly. In other cases, the reports or database may not clearly designate population boundaries and will require interpretation. Visual examination of maps may result in 'merging' occurrences that are likely to experience extensive gene flow- 'likely to experience extensive gene flow' may be defined on spatial proximity e.g. close geographic distance and lack of clear barrier (mountain, fence, etc.).

The distance between subpopulations (or discrete habitat patches where the species occurs) can be used as a proxy to evaluate to what extent subpopulations are likely functionally connected. When the edge of a subpopulation is within a reasonable dispersal distance (for that particular species, a distance within which some 75-90% of realized dispersal distances occur- expert discretion is fine here) of another subpopulation edge, and there are no known physical barriers impeding dispersal, the subpopulations can be considered as part of the same metapopulation. Common trees may disperse pollen tens to hundreds of kilometers, while salamanders or frogs may be dispersing kilometers or less, while wolves may disperse hundreds of kilometers. Designation of populations should consider human induced gene flow e.g. genetic rescue, translocations etc. Experts in the species type (e.g. amphibians) and/or text from the data source can help identify populations.

Some examples may help. A population may consist of a cluster of individuals in a discrete location like an island, lake, river catchment or forest preserve, separate from other discrete locations by some tens of kilometers. A population may constitute a metapopulation consisting of subpopulations (ponds, prairies, etc.) that are separate but not very far (hundreds of meters to

several kilometers), thus capable of exchanging at least 1 migrant (one reproductive adult moving between patches) on average per generation with each other. Well connected subpopulations are sufficiently close for the metapopulation to be called a population, and the population size should be considered the sum of the subpopulations, which may cover tens or hundreds of kilometers.

Genetic data can help define populations, but should be used with some caution. Whether genetic data can detect genetic differences can be a consequence of the type of genetic marker- genomic level markers can detect very fine genetic divisions, including between subpopulations. Not every genetic distinction according to DNA markers is a population. The threshold should still be considered roughly less than 1 migrant per generation (on average). Older genetic markers such as chloroplast and mitochondrial DNA sequence are useful in identifying highly genetically distinct populations.

Populations of common, abundant species with large continuous spatial distributions that are much larger than the dispersal capacity of a single individual (e.g. less than 1 migrant per generation across the extent) are also considered as metapopulations. When very large continuous populations exist, e.g. hundreds of kilometers, then population boundaries should be instituted based on expected adaptive or other differences, e.g. ecological changes such as [ecoregions](#).

Metapopulations should represent stable spatial and temporal units. Many species have ephemeral subpopulations in dynamic source-sink metapopulations- a sink is a spatial location receiving high immigration from adjacent areas and which would not persist on its own without immigration. Sinks are not distinct populations. The conglomerate of connected subpopulations that should be evaluated as a population.

For freshwater fish, the riverscape structure can help define populations or units that can be assimilated to populations/metapopulations with a little GIS work with existing databases (which may have Nc data as well, such as useful databases [here](#) and [here](#) for fish). Individuals inhabiting lakes can be considered as populations, especially for lakes that are disconnected from the hydrographical network. Riverscape (meta)populations can also be defined through their level of connection/disconnection e.g. belonging to different river basins, hydrographical systems, river stretches separated by huge dams, etc.

For trees, the pollination mode and commonness is important. Trees which are wind-pollinated can have continuous populations extend over tens of kilometers, and farther. Trees which are insect pollinated generally but not always, have less gene flow. For common trees, a distinct population may not be easily apparent (e.g. trees that extend across much of a continent in a continuous fashion). In such cases, a "population" may be considered at approximately the level of a state, country, or ecoregion (hundreds of kilometers across).

For amphibians, dispersal capability may vary greatly between species. Toads, for example, are known to be long distance dispersers and, provided they have access to resources, such as water bodies to breed, a single population (meta-population) may span 100s to 1000s of kilometers. Salamanders on the other hand tend to have lower dispersal capabilities, with dispersal often negligible at distances greater than 1.5 km. As such, wetlands within 1 km from each other are likely to constitute a metapopulation. Because amphibians must live in or around freshwater, their dispersal can be constrained when water sources are distant from each other or separated by barriers (e.g., mountain ranges or valleys, highways).

For small mammals habitat specialization and structure are determining factors. In some species, stepping stone dispersal can lead to large contiguous populations, despite restrictions on individual movements that may be limited to at most tens of kilometers. Some species display sex-biased dispersal.

For large mammals, despite greater dispersal ability, many species show habitat specialization. Therefore populations can be connected to different degrees depending on barriers to dispersal. Some habitats are naturally or anthropogenically fragmented (e.g. dense forest separated by agriculture). Fences can also disrupt natural movement patterns. In such instances it is possible to implement metapopulation management, mimicking natural movements between historically connected populations through translocations or dispersal corridors. Some species solely exist in fenced areas; where good records exist, all remnants can together be considered a population.

Defining marine species can be challenging due to environmental barriers found in open or pelagic ocean environments, and due to migration to and from breeding areas. Barriers are often related to temperature, current or depth. Marine species range from very small to large and their dispersal also likewise range from local to planktonic dispersal for many days on currents to traversing the length of a hemisphere. It is important to understand breeding/spawning areas and dispersal patterns/capabilities and environmental covariates to help determine genetically related populations. These distinctions may not be easily defined and some populations may have more or less continuous distributions. In such cases, a “population” may be considered at approximately the level of a state, country, or biogeographic region (hundreds of kilometers across).

On the definition of subspecies, and issues of taxonomy

Generally, we are assessing indicators at the species level. However, there are revisions of taxonomy (lumping and splitting of species) and some species groups are better studied and thus more finely examined than others. Consequently, what may be considered a subspecies by some experts, may be considered a species by others. In such situations, assessors may include subspecies if the entity is an important unit for that country (e.g., it has separate reports or protection or Red List status).

Common Issues and Challenges

Possible issues when recording data in the Kobo form (partial data, missing data, range of values, inference)

The Kobo form should be filled in as completely as possible, ideally all the populations of the species that exist in a country. This includes filling in both Ne and Nc if known, even if from separate data sources. However, complete data may be rarely available, and other issues might arise. Here are numerous examples and how to deal with them

- *Data available as a species census but individual populations listed with percentages.*
Simple math can be used to calculate each population’s census.
 - For example, the IUCN Red List says, “four subpopulations of *Q. engelmannii* were described based on percent of total population: 93% on the Black Mountains of central San Diego County; 6% on the Santa Rosa Plateau of Riverside County; 0.5%

in Orange County, and <0.1% in Los Angeles County.” If an overall census size is given, it is simple to calculate the Nc per population.

- *Data only available for indicator 1 (Ne 500)*, but no information available for historic populations (indicator 2). In other words we do not know the total number of historic populations and thus the number of *extinct* populations
 - You will type in the number of known extant (existing) populations and you will type in -999 for the number of extinct populations. -999 signifies unknown* In addition, an explanation of why we do not know anything about historic populations can be entered in the notes for ‘how populations were defined’.
 - The current (extant) known census or effective sizes should be filled in.
- *Data only available for indicator 2 (populations maintained)*, but no information available for population sizes (neither Ne nor Nc).
 - The “Total number of extinct populations within the country of assessment” and the “Number of extant populations within the country of assessment” should be filled in.
 - The assessor will then click either “No, but Census size data does exist for the species as a whole” and will be given an option to fill in the species’ total size, including if it is likely Nc>5000 or <5000. Otherwise they click “No, the species has insufficient data.”
 - If desired, an explanation of missing population data should go under “Any additional comments or notes with respect to the information provided in this Section.”
- *Partial data available for indicator 1 (Ne 500)*, such that Ne or Nc is known for some but not all populations.
 - The current known census sizes should be filled in for those populations for which it is available; **for those populations with no data, this is indicated by clicking “Point estimate” and entering -999** (no spaces or quotes, just minus sign and 999), as explained in the Kobo manual. **-999 indicates missing data/ data not known.** In addition, an explanation of missing data may be added in the notes (e.g. newly discovered, on private property etc.).
- *Partial data available for indicator 2 (populations maintained)*, such that at least some historic populations are known but there are *likely* to have been more.
 - In this case, click the option “It is possible that other populations exist but these have not been clearly defined.” In addition, an explanation of missing data/ missing population should go in the notes/ comments box.
- *No data on Ne or Nc or historic populations.* Most of the Kobo form therefore cannot be filled out due to lack of estimates of the population Nc or Ne or even lack of knowledge on the number of populations.
 - However, reports or expert consultation may show that it is known that the species is ‘highly fragmented’ or ‘moderately fragmented’ or exists in ‘very small populations’ or severely declined populations etc. Because there is a relationship between fragmentation and loss of genetic diversity there may be some information that can still be reported on. In such cases you should enter a narrative text into “Section 8: Additional information on species range.” We can later decide whether to make an assumption (as was done in the Sweden trial) for example assuming that highly fragmented means 90% of populations are < Ne 500 (a strong assumption).
- *Data from multiple sources, which conflict.* For example, one source indicates that the Nc should be 1000 while another source indicates 100, or in which multiple surveys were

reported from different agencies or at different times of the year. Alternatively one interpretation of the data presented (e.g. a table or map of populations) may indicate there are four populations while another interpretation is ten populations.

- The assessor should first try to judge which source or interpretation is more reliable based on the source reputation, technology/ abilities, recency, etc. If it cannot be determined which is more reliable, *the assessor should submit multiple Kobo submissions, one reflecting each source, with the same species name. You can submit as many as needed.*
- *A range of values is reported for Ne or Nc.* In some species, exact counts or a best estimate is reported, without a range, and in such cases this single number should be reported as a 'point estimate'. In other cases, an exact count might not be possible or may not be reported and therefore other estimates may be presented, such as a range of values (with or without a point estimate e.g. a mean). This may be the case from a mark recapture model, aerial survey, DNA from scat, or expert derived estimate.
 - In cases in which a low and a high estimate are available, this range should be reported using the option "quantitative range" and "estimate is a range". The range (upper and lower values) should be reported, and if any descriptor is available (e.g. that these ranges represent minimum/ maximum, or they represent 90 or 95% confidence) that should be noted under "any additional comments".
- *The Nc or Ne is not known but the area of habitat is known and there is known to be a strong relationship between Ne or Nc and the area of habitat.* This would be the case for example with birds or wolves with a known territory, or that a certain number of ponds supports a certain number of amphibians, or with forest density, etc. There should be good evidence that the species exists and is at the approximate expected density and the habitat is of the appropriate quality.
 - In such cases, the Nc should be reported, **but the assessor should click the box indicating that Nc came from a habitat area.**
- *Data is only available at the species level- no census information is available per population. However it is still useful to record this information.* For example if the whole species Nc is <5000, the proportion of populations greater than this is 0%.
 - Click "No, but Census size data does exist for the species as a whole." and then fill in the point estimate, range, or qualitative information e.g. "Entire species census < 5000."
- Fluctuating population sizes over numerous years. Some species experience substantial fluctuations over time around a stable mean. In such cases the current population size may be either too high or too low to represent the recent fluctuations.
 - We leave the choice to expert biologists, but it may be useful in these situations to take the harmonic mean, which is a mean that gives greater weight to lower values, which is typically done for Ne. Alternatively, separate (e.g. multiple) Kobo form submissions can be made for average, high and low population size values.
- *No values of Nc or Ne or range are given, but a reasonable biologist could conclude that the species size is below a given number.* For example, the population may be listed as "having at most a few thousand individuals". An assessor might interpret this as being <3000 individuals for example based on context, knowledge of the species and country, and other information.

- Similar to the notes above, the “quantitative range or qualitative data” should be chosen, and the option “<5,000 by much” or “<5,000 but not by much” depending on the situation, should be selected. An explanation should be given in the notes under “Additional comments or notes with respect to the information provided in this Section.”
- If an expert can instead give a quantitative reasonable range, e.g. 2000 to 4000, the range option would be better, of course.
- *Only a percentage range loss is known, rather than the number of populations (indicator 2).* This could be the case of a percentage loss or historic and modern estimates of statistics like range extent (km²) or Extent Of Occupancy, from which a percent change is calculable. The report or expert might also give an estimate of ‘overall decline’. Loss from certain provinces/ states or regions might also be known.
 - In all these cases there are no population counts, but a proportion of the range lost is known. In such cases (no obvious way to separate populations using data available), use the box “If available, please record the overall decline in species’ habitat area.” This should be a *percent lost*- entering a value of 60% lost means 40% of the original species remains.
 - Note you can enter information in this box even if populations and their losses are known. This is extra information.

** Note that for all instances of partial/ missing data, and for instances of data with ranges reported, the overall analysis at the country level will be reported in different ways. The indicator will be calculated with all species, with species with only fully complete data, with species assuming mean between range values, with species assuming minimum or maximum values for range, etc.

Other common challenges

- *Populations (say, population location A) being restored or reintroduced.* In the case of reintroductions, an assessor must make the determination as to whether a population represents its own ‘native’ gene pool, e.g. the gene pool that existed prior to reintroduction. Reintroduced individuals may come from another population (and/or may have been through bottlenecks, translocations and mixing with other populations), say from population B. In such a case, the ‘population’ is being created in the location of population A, but the original genetic lineage of that space (population A) no longer exists. We consider that in such cases, the genetic composition now at population A location would not represent the original population A. Thus the individuals now at location A should be considered as an expanded Nc of the lineage that is being translocated from (population B), rather than the original population A. In short, population A should be considered lost, as its original genetic composition is not restored. On the other hand, restoration may be taking place with individuals originally obtained from population A and kept ex situ and then used for restoration. In this case population A may be considered still extant (though if bottlenecks occurred, with a reduced Ne). Additionally, some judgment is required in terms of population restoration- a new restoration (e.g. a planting of just a few years ago) may not yet be considered viable. The assessor should indicate in the notes whatever is known about population viability, and the year (or the range of years) during which the restoration phase (or the releases due to reintroduction) took place.

PLACEHOLDER FOR CONCEPTUAL FIGURE ON THE LAST BULLET POINT ABOUT RESTORATION

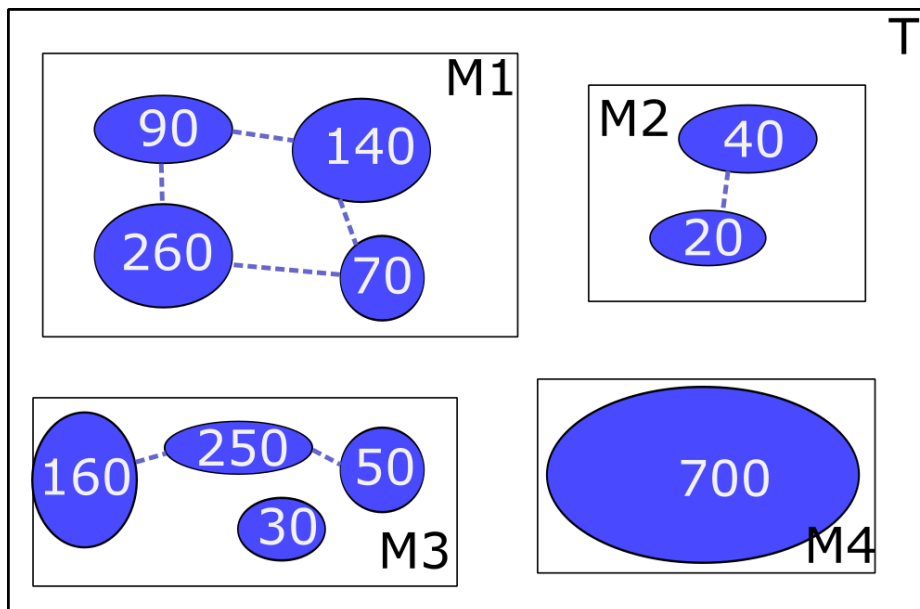
- *A species that is naturally or with human help colonizing new territory, such that it may appear that new populations exist.* This includes populations that might have become 'native invasive' e.g. in a new area they become an 'invasive' pest (koalas). This will also be increasingly common as climate changes and species' move poleward. These may or may not be genetically distinct- in (perhaps) the majority of cases they will not be new genetically distinct populations but will rather be extensions of an existing population and its genetic lineages, and thus should not be considered an increase in the number of populations.
- Another consideration **is of course fragmentation. As habitat is fragmented, the number of distinct groups of individuals may increase, as they are separated. This may appear to be an increase in the number of modern populations relative to historic ones.** Nonetheless, fragmentation is necessarily increasing the number of genetically distinct populations. As with the above point, we should focus on the populations that existed historically and if they have disappeared. If a map of historic and modern occurrences or geographic range exists, this should be easy to handle, because the focus is loss of the population- if a population becomes fragmented, it does not actually increase the number of populations. The number and placement of historic populations should be considered as the baseline, and the consideration is whether these are lost, or not.

Appendix

Background and terminology Indicator 1

The $N_e > 500$ criterion is used because above this N_e population size the importance of genetic drift on genetic diversity becomes very small relative to other evolutionary processes (e.g., selection). What does this mean? It means that evolutionary or adaptive potential in terms of unique alleles or heterozygosity is 'sufficiently' maintained.

Technical note/ footnote: The one migrant per generation rule is a rule of thumb that indicates that above 1 migrant per generation allele frequencies and changes therein are strongly correlated across subpopulations. Above 1 MPG the local gene diversity H_s is mainly driven by genetic drift acting upon the entire metapopulation. Below that threshold gene diversity is mainly driven by genetic drift at the local subpopulation scale, and allele frequencies across subpopulations are only weakly correlated. Gene flow or effective dispersal reflects the movement and exchange of individuals, propagules (resting stages, seeds, spores, ...) and gametes (pollen).



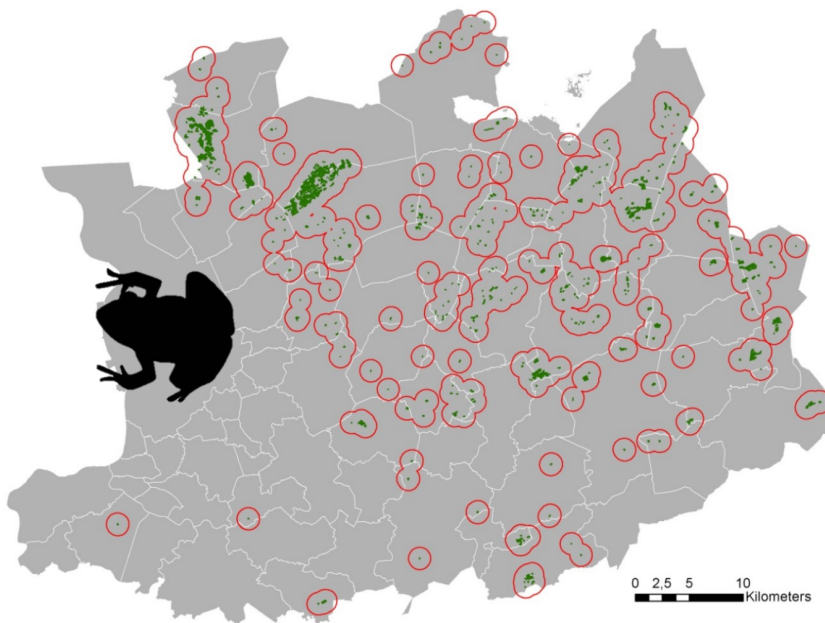
Example: theoretical distributions of subpopulations in metapopulations. Lines connecting subpopulations indicate frequent gene flow (>1 MPG) and are indicative of functional connectivity. T: total population. M: metapopulation. Metapopulations 1 and 4 are large enough (Sum of N_e of connected subpopulations >500). M2 is very low N_e , M3 is close. Note the presence of one isolated subpopulation in the vicinity of the M3 metapopulation. Its N_e is not considered when evaluating M3's N_e size. It should be considered a separate population. Metapopulations or populations are isolated from each other ($\ll 1$ migrant per generation, MPG).

Example

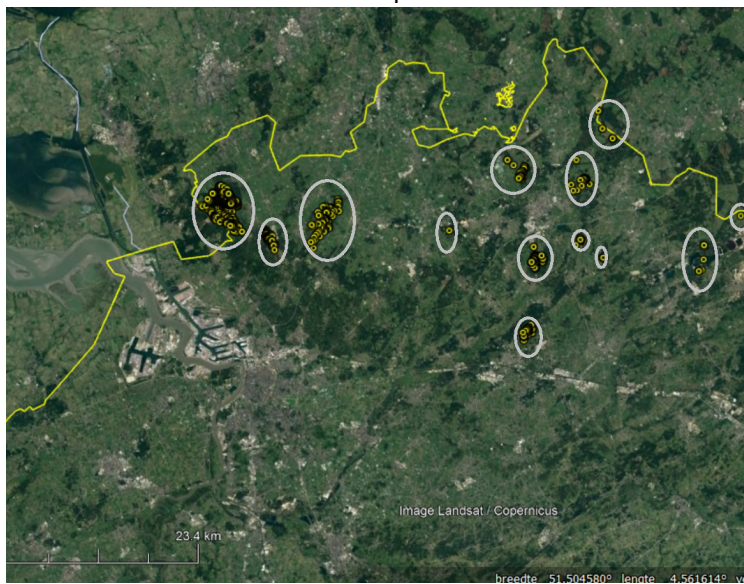
Moor frog:

Suitable habitat + dispersal distance buffers - Actual presence on a map

Both combine into units to be considered as $N_e > 500$



suitable habitat + dispersal distance buffers



Explanation of how to record uncertainty/ multiple values of N_e or N_c .

In the absence of information on N_c , alternative approaches can be used, by using knowledge on both:

- **A. the distribution of species** (from biodiversity portals such as Gbif, Inaturalist, artportalen.se, waarneming.nl, waarnemingen.be, observations.org, inpn.mnhn.fr, <https://www.ala.org.au/>, or other species observation databases, species atlas projects,...). This is [being used for species of concern in Belgium](#). It presents for circa 60 species criteria for N_{e95} (the effective size of a population maintaining 95% of its gene diversity over 100 years) or the corresponding amount of habitat required, which can easily be translated into N_{e500} criteria as well. This has been evaluated for a dozen species so far (not necessarily the same set as in the list; consists of amphibians, reptiles, small mammals, a few butterflies, a few plant species), but can relatively easily be transferred to other species too.
- **B. the density of a species** in optimal habitat or home range sizes (from literature, e.g. Santini et al. 2018). See also [TetraDensity](#)

This allows one to calculate what constitutes a minimum area network to host a $N_{e>500}$ or $N_{c>5000}$, and then to evaluate if the sites where the species is minimally known to occur are large enough. This doesn't evaluate the N_c directly, but indicates to what extent a present population has the potential to be large enough.

Evaluating the extent of habitat matrix requires expert knowledge, but can be helped by the availability of satellite images. The extent of connectivity depends not only on distance but also on barriers to gene flow.

The spatial scale of evaluation

Not all species and populations should be evaluated at the same spatial scale (a 1 km² may not suite all species): the unit of interest is the metapopulation- think from the perspective of that species. The categories below can help to evaluate the appropriate scale at which information should be sought.

Category 1 are populations of species with clearly distinct subpopulations and metapopulations at a local scale (subpopulations are typically < 1 km²). Often individuals of such species have very low dispersal capacity. The distribution across landscapes is patchy, and (for animals) landscapes are experienced by such species as coarse-grained. Dispersal events are rare across landscape matrices that do not resemble their habitat at all. Very often discrete habitat patches represent subpopulations. *Examples: Succisa pratensis (plant), most newts and salamanders, non-aquatic snakes, small rodents such as Eliomys quercinus, ...*

Category 2 are populations of species for whom the boundaries of individual metapopulations are hard to define. For animals, these are species for which individual movements can frequently occur across an entirely unsuitable landscape matrix (in which they don't breed or forage; fine-grained landscape perception). In plants, these will often be species with wind-dispersed seeds or pollen. Individuals of such species typically have intermediate dispersal capacities, and animal species typically express active habitat selection during dispersal. Distributions may be patchy or nearly continuous. Landscapes are experienced by animals as fine-grained (high dispersal capacity) or coarse-grained (low dispersal capacity but widely distributed). Within a biogeographical region multiple distinct metapopulations may exist. The evaluation of certain metapopulations may require transboundary information exchange. *Examples Non-migratory bird and bat species, many tree species and medium-sized mammals are typically in this category.*

Category 3 is used for species with high dispersal capacity (experiencing landscapes as fine-grained) and often having broad geographic distributions spanning entire continents or biogeographical regions. Sometimes, only a few (clearly genetically distinct) populations exist globally. Evaluation of the $N_e > 500$ often only makes sense at biogeographical scales. Physically clearly isolated subpopulations, or subpopulations not experiencing functional gene flow ($> 1\text{MPG}$) with other populations are evaluated separately. Migratory bird species and some large mammals belong in this category.

GLOSSARY of TERMS

Census Size, N_c

The number of reproductively mature individuals in a population. This number can be obtained via physical counts; however, it should be noted that these are not necessarily exact counts, but rather estimates given that the detection probability is dependent on several factors, including population density, observer effort (time) and experience. With that said, census sizes provide a conservative estimate of population size. There are instances where it is impractical or impossible to estimate the size of a population. In such instances, other methods, such as capture-mark-recapture, are extremely effective.

Effective population size, N_e

Effective population size is the number of reproductively mature individuals actually contributing to the gene pool. For the purposes of this study, N_e refers to contemporary effective population size which measures ongoing and near-term genetic erosion. Estimating contemporary N_e does not require knowing mutation rates (which is required for estimating historic N_e), but does require demographic knowledge (e.g. census size, N_c , [the number of reproductively mature individuals]; average number of, and variance in, offspring), genetic data (to measure linkage disequilibrium, kinship or genetic drift), OR a “rule of thumb” to convert N_c to N_e (Frankham 1995; and see Hoban et al. 2020, 2021).

50/500 rule

The 50/500 rule has been used as a guiding principle in conservation for assessing minimum viable effective population size (N_e) (Soule 1980; Franklin 1980). The overall premise behind this rule is that in order for a species or population to persist, a minimum N_e of 50 is needed in the short term to combat inbreeding, while a minimum N_e of 500 in the long term is needed to reduce genetic drift and enhance/promote evolutionary potential in perpetuity. A recommendation to double these minimum N_e estimates was made, increasing N_e to 100 and 1000 respectively (Frankham et al. 2014). While some have adopted this change, the 50/500 rule remains more widely used.

Species distributions

Wide-ranging (or widespread) species

Wide-ranging species have an extent of occurrence (EOO) greater than 20 000 km² or are present in more than one-third of a country's 20-km grid squares.

Restricted range species

Restricted range species are those that have an EOO <20 000 km² or an area of occurrence (AOO) < 2 000 km². A list of restricted species has been prepared using KBA criteria and can be found [here](#); however, this list should only be used as a guide as not all species have been assessed.

Rare species

Rare species are those that occur in low densities, typically as single individuals or as very small subpopulations with fewer than 50 individuals. They may occur within a restricted range, but are typically scattered over a wide area.