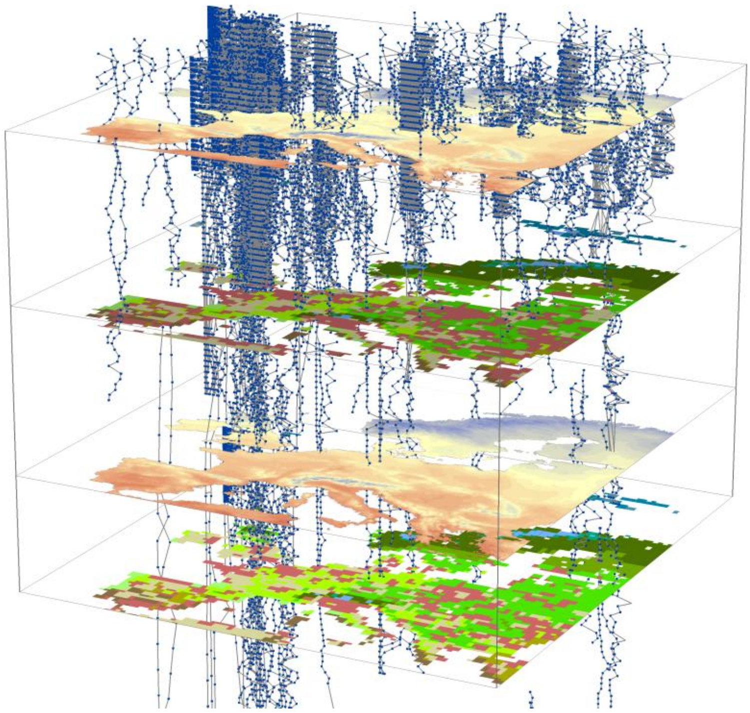
The Global Population Dynamics Database (GPDD)



User Guide

Version 2

July 2010

NERC Centre for Population Biology
Division of Biology, Imperial College London, Silwood Park Campus
Ascot, Berkshire, UK SL5 7PY

Contents

Credits

The GPDD was initially compiled by John Prendargast, Ellen Bazeley-White?, Owen Smith, John Lawton and Pablo Inchausti and released in 1999. Version 2.0 was released in 2010 following a substantial restructuring of the database and the addition of 123 new series by David Kidd and Sarah Knight.

Contact

Department of Biology, Imperial College London, Silwood Park campus, Ascot, Berkshire, SL5 7PY, U.K.

Telephone	44 (0)20 7594 2346
FAX	44 (0)1344 873173
E-mail	cpb-gpdd-dl@imperial.ac.uk

Citing the GPDD

NERC Centre for Population Biology, Imperial College (2010) The Global Population Dynamics Database v2.0. http://www.sw.ic.ac.uk/cpb/cpb/gpdd.html.

Front cover: 2900 abundance time-series from the Global Population Dynamics Database[1].

Series abundances are standardized and displayed with increasing abundance to the east (right) and decreasing abundance to the west (left). Time series intersect layers of annual mean temperature in 1921 and 1981[2] and land use in 1900 and 1950[3].

[1]NERC Centre for Population Biology, Imperial College (1999) The Global Population Dynamics Database. http://www.sw.ic.ac.uk/cpb/cpb/gpdd.htm. 2 Mitchell, T. D., Carter, T. R., Jones, P. D., Hulme, M. & New, M. (2004)

[2]A Comprehensive Set of High-Resolution Grids of Monthly Climate for Europe and the Globe: The Observed Record (1901-2000) and 16 Scenarios (2001-2100) (Tyndall Centre for Climate Change Res., Norwich, U.K.), Working Paper 55.

[3]Global Historical Land Cover and Land Use Estimates (1700-1990) , Klein Goldewijk, K., 2001. Estimating global land use change over the past 300 years: The HYDE database, Global Biogeochemical Cycles 15(2): 417-433.

Introduction

Understanding the way in which populations of wild plants and animals behave over long periods of time is crucial to unravelling the way in which communities are assembled and the way in which they respond to disturbance, control or harvesting. The implications for conservation and agriculture are legion. Aside from practicalities, population variation is also intrinsically interesting, and provides a wealth of opportunity for mathematical innovation or exploration, especially when populations have particular cyclic, outbreaking or chaotic properties. For most students of population behaviour, the limiting factor in investigating any of these phenomena, and the development of theory to explain them, is the availability of suitable data. Usually, where analyses are performed and published, authors work on data that they have collected themselves. By definition, the collection of population time series is a lengthy process, and many ecologists have committed themselves to a lifetime of work in order to accumulate detailed information on populations at certain sites over many years.

Studies of population behaviour tend to address a number of themes, each with a typical taxonomic flavour. Thus, students of the chaotic *vs* cyclic question tend to focus on small mammals, those with an interest in the effects of harvesting or culling generally study fisheries or large mammal data respectively and analyses of insect populations tend to dominate the literature on pest control.

Analysis and subsequent publication can only take place once time series of adequate length have been amassed, but frequently authors will continue to collect data after publication and may follow the first paper with an updated or extended version, or a book or book chapter at a later date. There are examples of data sets that have been assiduously collected but from which no publications have resulted, or from which internal, private or unpublished documents have been generated.

The result of all this fragmentary activity in population dynamics, where data sets are often analysed individually, or in line with certain taxonomic conventions, is that it has been difficult to a) formulate general theory and b) investigate large scale pattern, both spatially and taxonomically. The general unavailability of data has also led to the development of theory through the repeated analysis of the same data sets. The celebrated Canadian lynx/snowshoe hare cycle has been the subject of analyses and publications almost too numerous to count. There is an obvious danger that if individual data sets such as this happen not to be representative of the way in which most populations behave then theoretical understanding may suffer.

History of the GPDD

Initiated in 1994, as a collaborative venture between the NERC Centre for Population Biology at Imperial College and The Department of Ecology and Evolution, University of Tennessee, and subsequently in collaboration with the National Center for Ecological Analysis and Synthesis, Santa Barbara, the GPDD is one of the largest collections of animal and plant population data in the world, bringing together over 5000 time series in one database. The type of data contained in the GPDD varies enormously, from annual counts of mammals or birds at individual sampling sites, to weekly counts of zooplankton and other marine fauna. Time-series' run for a minimum of 10 years for over 1800 animal species across the globe and have been recorded from a variety of sources, such as scientific journals and unpublished data.

The intention is that the GPDD will become a widely available resource to be used to investigate general questions in population biology. As of 2010 the database has been cited in over 20 papers(Brook et al. 2006; Collen et al. 2008; Doncaster 2006; Eberhardt et al. 2008; Fagan 2001; Getz and Lloyd-Smith 2006; Halley and Stergiou 2005; Heering and Reed 2005; Holmes et al. 2007; Inchausti and Halley 2001; John Halley 2002; Kendall et al. 1998; Lotts et al. 2004; Peacock and Garshelis 2006; Reed and Hobbs 2004; Reed et al. 2003; Ross 2006; Sibly et al. 2005; Sibly et al. 2006a; Sibly et al. 2006b; Valone and Barber 2008) as well as providing a resource of data for teaching purposes. For a thorough and penetrative analysis on the behaviour of individual populations to be performed, more information than can be provided in the GPDD may be required. Clearly there is a limit to the type and quantity of information that can be supplied via general resource such as the GPDD. We have therefore not included any extensive life history or demographic data, even where this information is presented in the original publication.

Version 1.0 (released in 1999) has now been superseded by v2.0 which includes the following enhancements,

- A consistent definition of a time-series.
- Consistent metadata.
    - Units.
    - Sampling protocol.
- Consistent temporal coding.
- Addition of missing location information, the spatial bounds of study areas and a spatial accuracy index.
- Abundance data are supplied ‘retro-transformed’ as well as in the published source units.
- Improved documentation.
- 123 additional time-series are included, courtesy of Barry Brook (University of Adelaide).
- Removal of un-cited associated data including body size and biotope information.

See the Appendix for full details of change between versions.

Data Sources

The time series in the database come from both published and unpublished sources. They have been located in a variety of ways:

1. By systematically searching back issues of the predominant ecological and science journals.
2. By following citation trails - when a suitable time series is located in a publication, there are usually one or more references to similar or comparable studies in the citations list - every paper tends to lead to another paper.
3. By searching the World Wide Web, where an increasing number of ecological datasets are being made available.
4. By searching promising book titles. The vintage literature often contains a wealth of tabulated data of varying types, and we have drawn extensively upon the resources of the Imperial College library and the British Library to locate many long-out-of-print volumes.
5. By negotiating access to unpublished data. Through our network of professional contacts we have endeavoured to locate unpublished data which collectors are prepared to contribute to the project. Quite often, and quite understandably, collectors prefer to retain data for their own use, at least until they have published. We have, however been delighted at the selfless response of several collectors, who have donated unpublished data with no, or minimal, restrictions.
6. *Data from Barry Brook* (MainID 20527 – 20663). As part of the upgrade from v1.0 to the current v2.0 123 time series collated from a variety of sources and published as a supplement to an Ecology Letters paper (Brook et al. 2006) were added to the GPDD. The GPDD supplies these data as published by Brook et al. as such they may be derivations of the original source material. Users should refer to the original sources listed in supplementary table S3 of (Brook et al. 2006) sm004.doc on http://www3.interscience.wiley.com/journal/118634071/suppinfo. Barry Brook's ID can be found in the Main table in the Notes column for cross-reference.

As the search protocol followed for v1.0 was not undertaken for v2.0 and the Brook data was collated under a different protocol other series that meet the GPDD criteria and have been published after 1999 undoubtedly exist. *Users should always refer to the original source material to confirm data are fit for purpose.*

Relationship to other time series databases

The GPDD contains some data that also resides in other time series collations. For example, the Living Planet Index (Collen et al. 2008), held at the Institute of Zoology, Zoological Society of London, that provides a measure of global biodiversity contains some series extracted from the GPDD as well as other online-resources and scientific literature. It differs from the GPDD by holding vertebrate-only data and with a reduced minimum time-series length of 2 years. Other series may overlap with data held by the British Trust for Ornithology or the US Bird Census. *When amalgamating data from different repositories users should check source metadata to prevent duplication.*

Data capture

Population data are presented in published material in many ways. Obviously, it is easiest to extract data from a table of numbers. However, population trends and variations are best depicted in a line plot or histogram and it is in these graphical forms that most data are published. In some cases it has been possible to retrieve the raw data from the author(s) in question, but in others it has been necessary to reconstruct the data from the graphics. This has its limitations, and the accuracy of the derived numbers may be compromised where logarithmic scales are used or where printed copy is of poor quality or is very small. Commercial scanning and data retrieval software was used for this purpose and, generally speaking, we believe that errors due to data extraction are within acceptable limits. Users will need to draw their own conclusions about data accuracy.

Quality control

The GPDD only contains time series with ten or more records. Usually, this means ten years. Occasionally, where data sets are particularly interesting—they may be of a very poorly studied species, for example—we have included time series which are at least ten years from the beginning to the end of sampling, i.e. they just fulfil our minimum series length criterion, but which also have one or more missing data points.

Most of the datasets are of natural, i.e. unmanaged populations, or of the unmanipulated controls from experimental studies. Even apparently unmanaged populations may be subject to human intervention, for example some of the primate populations contained in the database have been supported by supplemental feeding in some years. In spite of this we have included them because primate data are comparatively rare. Notes in the database record this fact, and the user is referred to the original source to determine whether or to what extent, this is likely to prejudice any analysis. Population data from some laboratory experiments are also included, and are marked as such.

Including, as it does, population counts from a very wide range of sources, there is considerable variation in the quality of the data. Although it does not guarantee accuracy, the peer review process which is applied to much published work may filter out some of the more unreliable data. It is usually difficult to obtain an accurate, objective measure of data quality, and it is often necessary to fall back on a subjective assessment. This we have done, based on criteria such as the type of environment or habitat sampled, the species in question, the area of the sampling site, and the method of sampling. Each dataset has been ranked, on a scale of 1 (low) to 5 (high), for apparent data quality. For example, the database contains numerous very long datasets of fur trapping and export records from North America. As animal population data they are highly unreliable, because the numbers of skins exported depend heavily on factors other than the numbers of animals available for trapping. Nevertheless they are unique, and have been included to provide a context for other contemporaneous datasets, rather than as hard ecological data themselves. These we have given a rank of 1. At the other end of the scale, the database contains a number of UK estuarine datasets collected by automatic sieve sampling, which has been completely consistent over the entire sampling period (17 years). They exemplify the highest possible quality of sampled population data, and these we have ranked as 5. In all cases the ranking is provided as a guide only, and it will be for the user of each dataset to determine whether or not it meets his/her specific requirements.

The collecting of ecological data, especially over long periods of time, may be subject to all sorts of difficulties and variation as circumstances change over the years. Usually, where such changes are relevant they are referred to in the published material, and we have endeavoured to mirror any warnings, caveats or similar points in the Notes field for each dataset.

Temporal Referencing

The GPDD contains time series of abundance records however there is considerable in how data are temporal referenced. Studies rarely adhere to ridged sampling regimes or publish complete information on sampling as even if a systematic sampling method is aimed for circumstance often intervenes resulting in slight differences between the desired regime and reality.

The majority of series are referenced to absolute time by year or year + sub-annual time period. A minority of series, mostly from lab-based experiments, are only referenced to relative time. For example, generation 1, generation 2, etc. or day 1, day 2, day 3, etc.

Series referenced to absolute time differ in the length of the sampling period and the precision with which the sampling period is recorded. The GPDD does not store exact sampling dates, even when provided in the source. Instead observations are temporally coded to whatever temporal unit the data was presented in on the source graph or tables from which it was extracted. Thus, for example data in the GPDD labelled as being from 'May 1965' may be a composite of observations throughout the month or data from any single day in that month; similarly data for '1965' may encompass sampling across any subset of that year. The majority of data in the GPDD is temporally referenced to a year or month and year. Other data is referenced to 3-month, 4-week period, season or other period. GPDD metadata and the original sources should be examined if for further information on temporal sampling.

The GPDD provides three modes of temporal referencing (fig 1).

1. *Timeperiod text description.* All data points have a Timeperiod which provides a text description of the time period of the sample such as a month, season or sequential day number. Timeperiods are grouped into groups such as quarters, months, weeks and days and, where appropriate, ordered within groups. To obtain a text description of the time of a data record referenced in absolute time concatenate the integer year in the data table to the Timeperiod. Data referenced to relative time only have -9999 in year.
2. *Series step* provides a relative integer time stamp for data within a series. The first data value in the series has seriesstep 0. Subsequent values have an increasing seriesstep proportional to the temporal period between the samples.
3. *Decimal Year* provides an estimate of absolute dating across all time series . DecimalYearBegin is the beginning of the period within which sampling occurred, it is simply calculated as year + fraction of the year to the start of the sampling period.

Similarly DecimalYearEnd is year + fraction of the year to the end of the sampling period. Thus an annual value has DecimalYearBegin = Year and DecimalYearEnd = Year + 1, while a monthly value for June has DecimalYearBegin = Year + (5 * 1/12) and DecimalYearEnd = Year + (6 * 1/12).

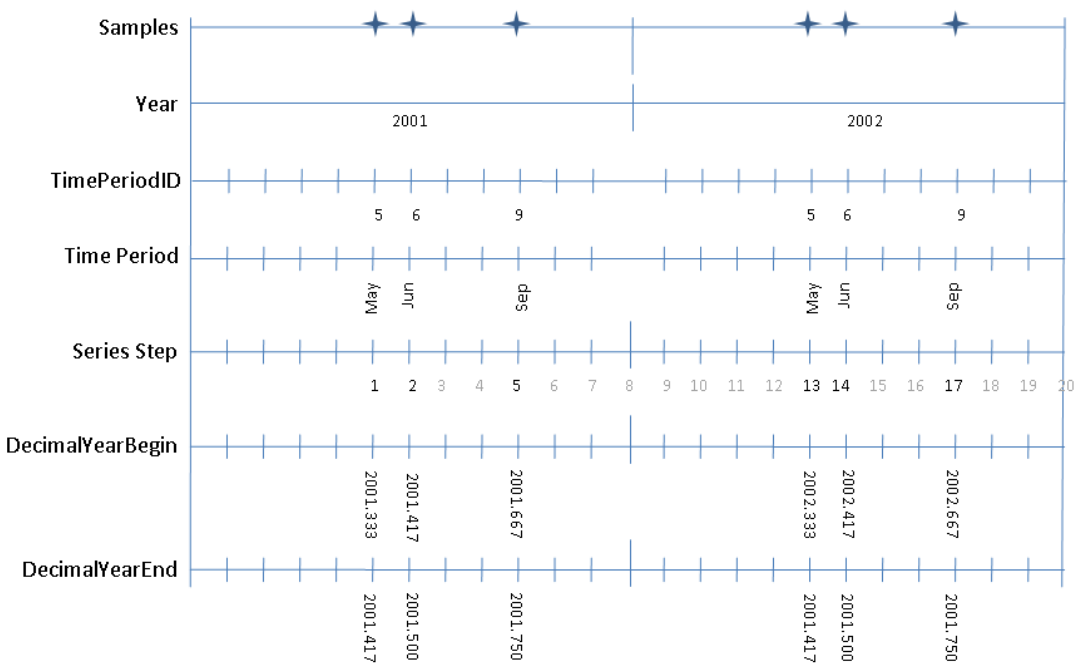


Figure 1. Temporal referencing in the GPDD.

Spatial Referencing

The location contains information defining the spatial location of the populations the time series relate to. If provided coordinates were extracted from the source, otherwise they were estimated from atlases and Google Earth. The SpatialAccuracy column provides a qualitative estimate of accuracy for the coordinates. Where series relate to extensive geographical areas the given coordinates approximate the centroid of the sampled area. Additional information of the geographical extent of a location is encoded in the Area (contains many nulls), LocationExtent and spatial bounding box columns (North, East, South and West

Minimal use of Nulls

The use of null values has been minimised in text fields wherever possible. Instead, 'none' is used where the data source states that there is no information, 'not specified' where the data source fails to state the information, 'unknown' where we were unable access the full data source and 'not applicable' where not relevant. In numerical fields, usually relating to a calendar year, '9999' indicates a null value.

Accessing the GPDD

The GPDD data may be queried and downloaded from the Web portal https://www.imperial.ac.uk/cpb/gpdd2/gpdd.aspx.

Once logged in to v2, the entire database (excluding restricted series) can be downloaded as a Microsoft Access database using the link on the left-hand side of the page (fig 2).

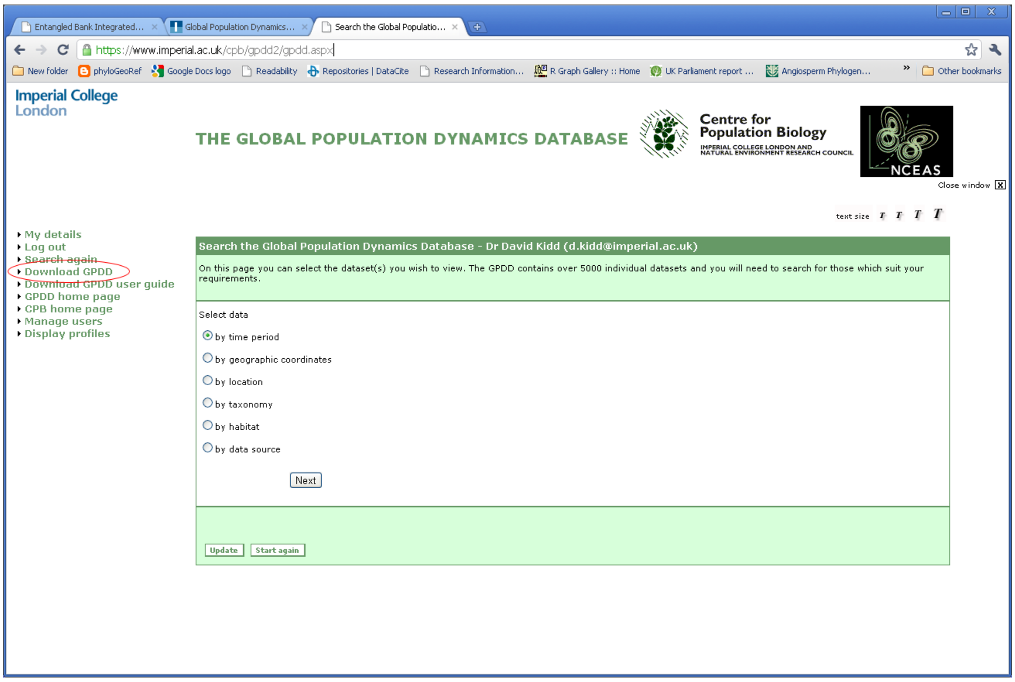


Figure 1. Downloading the GPDD as an MSAccess database.

Restricted access series

Due to licensing restrictions 686 series from 6 sources cannot be distributed without the permission of the owner. These are data from the British Trust for Ornithology's Common Bird Census (97) and Constant Effort Recording Scheme (32), Rothamstead Experimental Station, UK (9), the National monitoring programme for wintering wildfowl in Norway 1980 – 93 (T. Nygard, 23), *Phalacrocorax carbo* (Great cormorant) and *Somateria mollissima* (Common eider) series supplied by N. Rov (2) and data from insect light trapping supplied by H. Wolda (523).

GPDD Data Structure

The GPDD is a relational database comprised of six tables (Fig. 3).

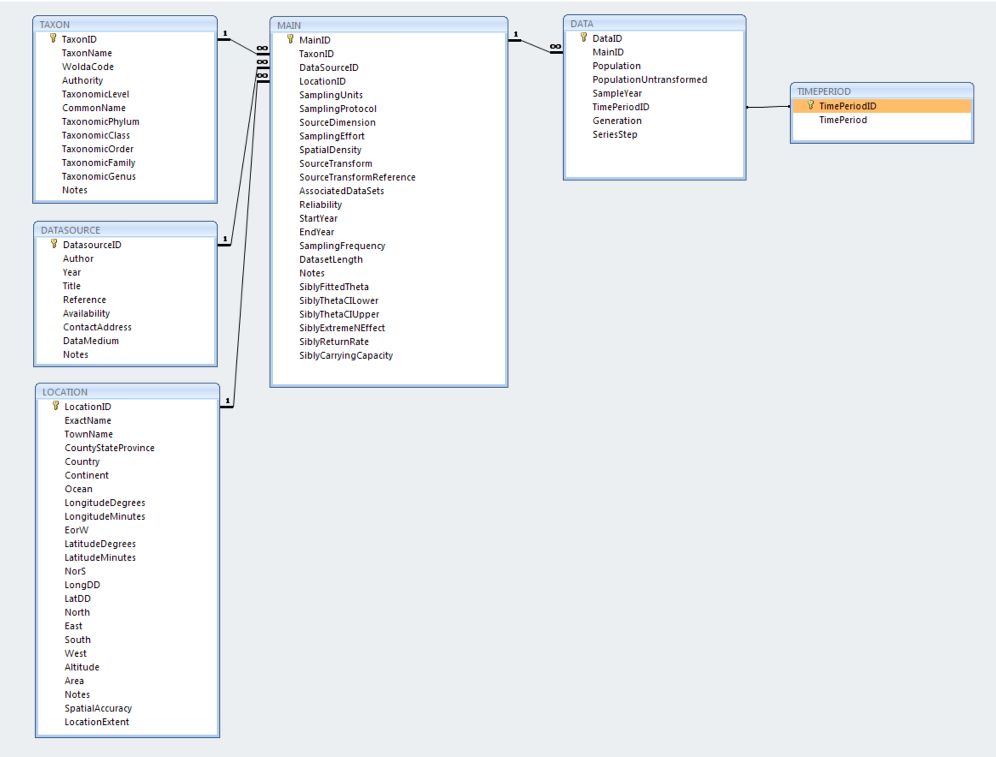


Figure 3. GPDD tables and relations.

Database table fields and data types

Size = size of field, p = precision.

MS Access data types

adInteger	four-byte signed integer (DBTYPE_I4).
adVarWChar	Null-terminated Unicode character string.
adDouble	Double-precision floating-point value (DBTYPE_R8).
adSmallInt	Two-byte signed integer (DBTYPE_I2).
adBoolean	Boolean value (DBTYPE_BOOL).
adNumeric	Exact numeric value with a fixed precision and scale (DBTYPE_NUMERIC).
adLongVarWChar	Long null-terminated Unicode string value.

MAIN table

A MAIN record is a ‘time series’ which is unique Taxon/Location/LifeCycle combination. Sequential data for multiple life stages (e.g. eggs, larve and adults) are split into different Main records and must be amalgamated to create a single time series. Where more than one adult generation occurs per year generation is identified in the generation column of the data table.

Name	Type	Size	P	Description
MainID	adInteger		10	Automatic unique ID
TaxonID	adInteger		10	ID Number in Taxon table
DataSourceID	adInteger		10	ID Number in Datasource table
BiotopeID	adInteger		10	ID Number in Biotope table
LocationID	adInteger		10	ID Number in Location table
SamplingUnits	adVarWChar	255		The entity observed. Entries include, ‘adults’, ‘cells’, ‘egg masses’, and ‘pelts’.
SamplingProtocol	adVarWChar	255		How entities were sampled. Values are ‘Count’, ‘Count (millions)’, ‘Harvest’, ‘Index of abundance’, ‘Index of territories’, ‘Leaf area’, ‘Mean Count’, ‘Not Specified’, ‘Percent cover’ and ‘Sample’.
SourceDimension	adVarWChar	255		The dimension of source table or graph. Values are Area coverage, Biomass, Catch, Count, Count (estimated), Density, Density (estimated), Index, Index - CPUE, Mean Area Coverage, Mean Biomass, Mean concentration, Mean Count, Mean Density, Mean Harvest, Minimum Count, Not Specified, Percent Count, Relative Density, Transformed Biomass, Transformed Count, Transformed Density and Unknown.
SamplingEffort	adVarWChar	255		The sampling effort. Usually a measure of temporal effort but also census, density, harvest.
SpatialDensity	adVarWChar	255		The spatial unit of a density.
SourceTransform	adVarWChar	255		Any transformation of the data.
SourceTransformReference	adVarWChar	255		Details of the reference value of any data transformation. Base Year, Log, None, Not Specified, Proportion, Unknown, x 1000 lbs.
AssociatedDataSets	adVarWChar	255		Comma separated list of MainID of other series from the same study or to which the series can be directly compared.
Reliability	adDouble		15	Subjective Rank 1-5, 1 = least reliable, 5 = most reliable (see Quality Control).
StartYear	adInteger		10	The year in which sampling commenced.
EndYear	adInteger		10	Insert the year in which sampling ended.
SamplingFrequency	adVarWChar	50		Approximate number of samples per year.
DatasetLength	adDouble		15	Total number of samples.
Notes	adLongVarWChar			Notes
SiblyFittedTheta	adDouble		15	Calculated as per (Sibly et al. 2005).
SiblyThetaCILower	adDouble		15	Calculated as per (Sibly et al. 2005).
SiblyThetaCIUpper	adDouble		15	Calculated as per (Sibly et al. 2005).
SiblyExtremeNEffect	adBoolean	2		Calculated as per (Sibly et al. 2005).
SiblyReturnRate	adDouble		15	Calculated as per (Sibly et al. 2005).
SiblyCarryingCapacity	adDouble		15	Calculated as per (Sibly et al. 2005).

DATA table

The Data table stores the individual time series abundance records.

Name	Type	Size	P	Description
DataID	adInteger		10	Automatic unique ID
MainID	adInteger		10	Foreign key from MAIN
Population	adDouble		15	Population data as published
PopulationUntransformed	adDouble		15	Retro-transformed population data
SampleYear	adInteger		10	Year of sample (9999 if not applicable)
TimePeriodID	adInteger		10	Foreign key from TimePeriod giving the time of year or sequence of sampling.
Generation	adInteger		10	Generation number if more than 1 generation per year. Values 1, 2 or null.
SeriesStep	adInteger		10	Time-scaled time-series step. Series begin at zero and then increments by the minimum time step required to define all periods between samples.
DecimalYearBegin	adDouble		15	Decimal year of the beginning of the sampling period o
DecimalYearEnd	adDouble		15	Decimal year of the end of the sampling period o

DATASOURCE table

Information on where the data was obtained from, relating to the Main table through a unique DatasourceID. Sources of data include published journals, books and unpublished datasets and the references details are held here. The table also contains information regarding access restrictions, contact details and in what medium the data was obtained.

Name	Type	Size	P	Description
DataSourceID	adInteger		10	Automatic unique ID
Author	adVarWChar	255		Author(s)
Year	adVarWChar	255		Year of publication
Title	adVarWChar	255		Title
Reference	adVarWChar	255		Reference or citation
Availability	adVarWChar	255		Public or restricted access
ContactAddress	adVarWChar	255		Contact details
DataMedium	adVarWChar	255		Format of data acquired
Notes	adLongVarWChar			Notes

LOCATION table

Name	Type	Size	P	Description
LocationID	adInteger		10	Automatic unique ID
ExactName	adVarWChar	255		Name of Location
TownName	adVarWChar	255		Town
CountyStateProvince	adVarWChar	80		County or State
Country	adVarWChar	255		Country
Continent	adVarWChar	50		Continent
Ocean	adVarWChar	50		Ocean
LongitudeDegrees	adDouble		15	Longitude degrees of location/centroid
LongitudeMinutes	adDouble		15	Longitude minutes of location/centroid
EorW	adVarWChar	255		East or West
LatitudeDegrees	adDouble		15	Latitude degrees of location/centroid
LatitudeMinutes	adDouble		15	Latitude degrees of location/centroid
NorS	adVarWChar	255		North or South
LongDD	adDouble		15	Longitude as a decimal degrees
LatDD	adDouble		15	Latititude as a decimal degrees
North	adNumeric		18	Bounding box northern extent
East	adNumeric		18	Bounding box eastern extent
South	adNumeric		18	Bounding box southern extent
West	adNumeric		18	Bounding box western extent
Altitude	adDouble		15	Altitude in metres
Area	adDouble		15	Area in ha
Notes	adVarWChar	255		Notes
SpatialAccuracy	adInteger		10	Value between 0 - 6 indicating the accuracy of the location given. 0 = Unknown, 1 = >100 km radius, 2 = 10 - <100km, 3 = 1 - <9km, 4 = 0.1 - 1km, 5 = 10 - 100m, 6 = accurate survey (incl. GPS) <= 10m.
LocationExtent	adInteger		10	A value between 1 - 4 indicating the size of the study site. Where available absolute size is recorded in the Area field. 1 = Region >10 km radius, 2 = Local Area 1-10 km radius, 3 = Extended Site 0.1-1 km radius, 4 = Precise Site <0.1 km radius.

TAXON table

The taxon table stores the taxonomic names relating to Main records. It is links to the MAIN table with a unique TaxonID. Most series are for species. Some extra information regarding breeding habitats etc may be found in the notes column.

Name	Type	Size	P	Description
TAXON				
TaxonID	adInteger		10	Automatic ID number (Range 1-12151, 1896 rows)
TaxonName	adVarWChar	255		Name of taxon. May be binomial, higher taxon rank or user-defined
WoldaCode	adVarWChar	50		Code used by H. Wolda to identify unnamed/indentified species
Authority	adVarWChar	255		Taxon definition authority (many missing).
TaxonomicLevel	adVarWChar	255		Species, Genus, Family etc. plus 'Virus'
CommonName	adVarWChar	255		Single common name
Taxonomic Phylum	adVarWChar	255		Taxonomic Phylum
Taxonomic Class	adVarWChar	255		Taxonomic Class
TaxonomicOrder	adVarWChar	255		Taxonomic Order
Taxonomic Family	adVarWChar	255		Taxonomic Family
Taxonomic Genus	adVarWChar	255		Taxonomic Genus
Notes	adVarWChar	255		Notes

TIMEPERIOD table

TimePeriod is a look-up table that provides text descriptions of the temporal period the sample relates to such as 'January', 'Spring', 'Week 1' and 'Day 1'.

Name	**Type**	**Size**	**P**	**Description**
TIMEPERIOD				
TimePeriodID	adInteger		10	Automatic ID number (range 1-408, 407 rows)
TimePeriod	adVarWChar	255		Time period name
TimePeriodGroup	adVarWChar	255		Group time period belongs to, e.g. month, quarter, season, wet/dry etc.
TimePeriodOrder	adInteger		10	Order of timeperiod within time period group, e.g. January = 1, February = 2, etc. within the month group. Some groups do not have orders e.g wet/dry.
Begin	adDouble		15	Decimal year of the beginning of the time period
End	adDouble		15	Decimal year of the end of the time period

Appendix

Detail of GPDD change between v1.0 to v2.0

Main table

1) ID column renamed MainID.
2) BiotopeID column deleted – Biotope table deleted.
3) DataType, SamplingUnits, SourceUnits – Unit columns updated.
4) Zeros, NumberOfMissingValues, Native, Introduced, Average, Variance, Maximum, Minimum, Mode, Median, Autocorrelation, DensityDependence, Trends, Attractors, CoefficientOfVariation and HurstConstant columns removed – redundant fields with large volumes of missing information.
5) OldDataID column deleted – redundant data.
6) New units columns created - SamplingUnits, SamplingProtocol, SourceDimension, SamplingEffort, SpatialDensity, SourceTransform, SourceTransformReference.
7) AssociatedData column – empty entries filled with "None".
8) StartData and EndDate columns renamed to StartYear and EndYear for clarity. Missing values entered if found, '9999' if unknown.
9) SamplingFrequency – missing values entered if found, 'Generations' used instead of iterative count when calendar dates are not applicable.
10) DatasetLength – missing values entered.
11) Notes – Barry Brook's ID added to new datasets.
12) Main ID 9912-9916 deleted – erroneous data with incorrect and untraceable data source information.
13) Main ID 20670 added – erroneously missed from the previous version.
14) All spelling and grammar errors corrected.

Data table

1) PopulationUntransformed column created – this is a copy of the Population column but with transformations removed where applicable (such as logarithms and multiplication). Several datasets are entered as indexes and have not been untransformed. This column may be useful for statistical analyses.
2) TimeOfSample renamed to SampleYear – where generative datasets apply instead of calendar months, the values mirror the SeriesStep.
3) TimePeriod column created – refers to a look up table that specifies day/week/season of sampling.

4) Generation column created to differentiate between generations within the same Main ID. This is applicable to species such as butterflies (e.g. Main ID 9285-6).
5) SeriesStep column created – sequential values within each MainID to indicate chronology irrespective of sampling time regime.
6) Flag column deleted – empty and redundant column.

Datasource table

1) SourceNo. Column deleted – contained only zeros.
2) Flag column deleted – empty.
3) Author, Year, Title, Reference columns – missing information entered if found.
4) All spelling and grammar errors corrected.

Location table

1) BiogeographicalZone column deleted – unknown source.
2) Flag column deleted – empty.
3) Bounding box coordinates – 4 new columns (North, East, South, West)
4) SpatialAccuracy column created – a value ranging 0-6 indicating the level of accuracy of the bounding box to the site of sampling.
5) LocationExtent column created – a value ranging 1-4 indicating the size of the site of sampling.
6) All missing information entered if found.
7) All spelling and grammar errors corrected.

Taxon table

1) BodyLength, BodyWeight, TrophicLevel, FeedingMethod, SexualDimorphism, Palaearctic, Nearctic, Ethiopian, Madagascan, Oriental, Neotropical, Notogaea, Wallacea, BiogeographicStatus, ConservationInformation, Taxoncheck and our_comments all deleted.
2) Missing information entered if found.
3) All spelling and grammar errors corrected.

Timeperiod table

1) New table created containing TimeperiodID and TimePeriod columns.

Biotope table

1) Deleted as un-cited information. It contained information on the habitat that the series originates, and is dated and without source information. Local habitat descriptions are grouped into 25 ‘biotope type’ classes; Aerial, Agricultural, Arid, Coastal, Coniferous forest, Deciduous forest, Evergreen forest, Fluvial, Grassland, Heath or Moor, Island, Largely unvegetated, Limnetic, Marine, Mixed habitat, Mixed or unspecified forest, Montane, Polar, Savanna, Scrub, Specified plants as habitat, Tidal/Intertidal, Unspecified habitat or no information, Urban/Suburban, Wetland.

References

Brook, B. W., L. W. Traill, and C. J. A. Bradshaw. 2006. Minimum viable population sizes and global extinction risk are unrelated. Ecology Letters 9:375-382.

Collen, B., J. Loh, S. Holbrook, L. McRae, R. Amin, and J. E. M. Baillie. 2008. Monitoring Change in Vertebrate Abundance: the Living Planet Index. Conservation Biology 9999.

Coulson, T., and S. Tuljapurkar. 2008. The Dynamics of a Quantitative Trait in an Age-structured Population Living in a Variable Environment. American Naturalist 172:599-612.

Doncaster, C. P. 2006. Comment on "On the Regulation of Populations of Mammals, Birds, Fish, and Insects" III, Pages 1100c-.

Eberhardt, L. L., J. M. Breiwick, and D. P. Demaster. 2008. Analyzing population growth curves. Oikos 117:1240-1246.

Fagan, M. P. F. K. 2001. Characterizing population vulnerability for 758 species. Ecology Letters 4:132-138.

Getz, W. M., and J. O. Lloyd-Smith. 2006. Comment on "On the Regulation of Populations of Mammals, Birds, Fish, and Insects" I, Pages 1100a-.

Halley, J. M., and K. I. Stergiou. 2005. The implications of increasing variability of fish landings. Fish and Fisheries 6:266-276.

Heering, T. E., and D. H. Reed. 2005. Modeling Extinction: Density-Dependent Changes in the Variance of Population Growth Rates. Journal of the Mississippi Academy of Sciences 50:183-194.

Holmes, E. E., J. L. Sabo, S. V. Viscido, and W. F. Fagan. 2007. A statistical approach to quasi-extinction forecasting. Ecological Letters 10:1182-1198.

Inchausti, P., and J. Halley. 2001. Investigating Long-Term Ecological Variability Using the Global Population Dynamics Database. Science 293:655-657.

John Halley, P. I. 2002. Lognormality in ecological time series. Oikos 99:518-530.

Kendall, B. E., J. Prendergast, and O. N. Bjørnstad. 1998. The macroecology of population dynamics: taxonomic and biogeographic patterns in population cycles. Ecological Letters 1:160-164.

Lotts, K. C., T. A. Waite, and J. A. Vucetich. 2004. Reliability of Absolute and Relative Predictions of Population Persistence Based on Time Series. Conservation Biology 18:1224-1232.

Peacock, E., and D. L. Garshelis. 2006. Comment on "On the Regulation of Populations of Mammals, Birds, Fish, and Insects" IV, Pages 45a-.

Reed, D. H., and G. R. Hobbs. 2004. The relationship between population size and temporal variability in population size. Animal Conservation 7:1-8.

Reed, D. H., J. J. O'Grady, J. D. Ballou, and R. Frankham. 2003. The frequency and severity of catastrophic die-offs in vertebrates. Animal Conservation 6:109-114.

Ross, J. V. 2006. Comment on "On the Regulation of Populations of Mammals, Birds, Fish, and Insects" II, Pages 1100b-.

Sibly, R. M., D. Barker, M. C. Denham, J. Hone, and M. Pagel. 2005. On the Regulation of Populations of Mammals, Birds, Fish, and Insects. Science 309:607-610.

—. 2006a. Response to Comment on "On the Regulation of Populations of Mammals, Birds, Fish, and Insects", Pages 45b-.

—. 2006b. Response to Comments on "On the Regulation of Populations of Mammals, Birds, Fish, and Insects", Pages 1100d-.

Valone, T. J., and N. A. Barber. 2008. An empirical evaluation of the insurance hypothesis in diversity-stability models. Ecology 89:522-531.