Supplementary Materials for

Global Threats to Biodiversity

L. Joppa[*1], B. O'Connor[2], P. Visconti[2], C. Smith[2], J. Geldmann[3,4], M. Hoffmann[2,5], J. E. M. Watson[6,7], S. H. M. Butchart[8,4], M. Virah-Sawmy[9], B. S. Halpern[10,11,12], S. E. Ahmed[1], A. Balmford[4], W. J. Sutherland[4], M. Harfoot[2], C. Hilton-Taylor[13], W. Foden[14], E. Di Minin[15,16], S. Pagad[17], P. Genovesi[18], J. Hutton[2], N. D. Burgess[2,3,4]Correspondence to: lujoppa@microsoft.com

This PDF file includes:

Author Affiliations
Acknowledgments
Materials and Methods
Fig. S1 to S2
Table S3

Other Supplementary Materials for this manuscript includes the following:

Tables S1 to S2

Author Affiliations

[1]Microsoft Research, One Microsoft Way, Redmond, WA 98052
[2]UNEP World Conservation Monitoring Centre, 219 Huntingdon Road, Cambridge, CB2 ODL, UK
[3]Center for Macroecology, Evolution and Climate, Natural History Museum of Denmark, University of Copenhagen, Universitetsparken 15, DK-2100 Copenhagen E, Denmark
[4] Conservation Science Group, Department of Zoology, University of Cambridge, Downing, Cambridge CB2 3EJ, UK
[5] IUCN Species Survival Commission, IUCN, 28 Rue Mauverney, 1196 Gland, Switzerland
[6] School of Geography, Planning and Environmental Management, University of Queensland, Brisbane, QLD, Australia, 4072
[7] Wildlife Conservation Society, Global Conservation Program, Bronx, NY, USA, 10460
[8]BirdLife International, The David Attenborough Building, Pembroke Street, Cambridge CB2 3QZ, UK
[9]Luc Hoffmann Institute, WWF International, Avenue du Mont-Blanc, 1196 Gland, Switzerland
[10]Bren School of Environmental Science and Management, University of California, Santa Barbara, CA 93106, USA
[11]NCEAS, 735 State St., Santa Barabara, CA 93101, USA
[12]Imperial College London, Silwood Park, Buckhurst Rd., Ascot, UK
[13]IUCN Global Species Programme, The David Attenborough Building, Pembroke Street, Cambridge, CB2 3QZ, UK
[14]Department of Botany and Zoology, University of Stellenbosch, Private Bag X1, Matieland 7602, South Africa and Chair IUCN Species Survival Commission, Climate Change Specialist Group
[15]Finnish Centre of Excellence in Metapopulation Research, Department of Biosciences, University of Helsinki, Helsinki, FI-00014, Finland
[16]School of Life Sciences, University of Kwa-Zulu-Natal, Private Bag X54001, Durban 4000, South Africa
[17]The IUCN Species Survival Commission Invasive Species Specialist Group, University of Auckland, Tamaki Campus, Auckland, New Zealand
[18]ISPRA, Institute for Environmental Protection and Research and Chair IUCN SSC Invasive Species Specialist Group, Rome, Italy

Acknowledgements

This paper was supported by funding from the Cambridge Conservation Initiative (CCI) project “Toward an integrated approach to the mapping and mitigating of global pressures on biodiversity “ involving the following CCI partners: the UNEP World Conservation Monitoring Centre (UNEP WCMC), BirdLife International , International Union for Conservation of Nature (IUCN) and the Department of Zoology, University of Cambridge. The non-CCI partners in the project are: Microsoft Research, University of Copenhagen, University of California (Santa Barbara), the Wildlife Conservation Society, University of Queensland and James Cook University. LNJ conceived the study and wrote the manuscript with contributions from all co-authors. PV, BOC and LNJ designed the analyses, CS, BOC, and PV collected the data. All authors and their affiliated institutions, as well as Ben Phalan (University of Cambridge), Graeme Buchanan (Royal Society for the Protection of Birds), Maria Dias (BirdLife International), Victoria Veach (University of Helsinki), and Steffen Fritz (International Institute for Applied Systems Analysis), were involved in reviewing the datasets included here. BOC, PV and LJ analyzed the data,
All data analyzed in this paper is presented in Tables S1 and S2.

Materials and Methods

Introduction

All of the project partners as well as external reviewers engaged in a collaborative process of identifying and cataloguing spatial datasets which could potentially be used for systematically mapping global threatening processes (in the terrestrial, freshwater and marine biomes). Datasets were categorised according to the IUCN/CMP Threat Classification Scheme (version 3.2) and grouped down to level 2 of the scheme.

We used a threat classification scheme (*15*; Main Figure 1), that, while not without shortcomings (*16, 17*), has been widely deployed for tens of thousands of conservation assessments for species, sites, and projects. We followed a structured data collection procedure and associated each dataset with one or more classes of threat (see below). We omit three threat classes from our analysis: two (Geological Events; Other Options [i.e. other threats]) because they are not exclusively anthropogenic drivers of threat (see SOM for details) and the other (Climate Change and Severe Weather) due to the comprehensive treatment of this topic by the Fifth Assessment Report for the Intergovernmental Panel on Climate Change (http://www.ipcc.ch/report/ar5/). For the remaining threat classes restricted our search to spatial datasets (or to datasets that could be spatially determined, e.g. by geopolitical boundaries) with a global extent.

The initial search for datasets was done online using the threat categories listed in the Threat Classification Scheme as search terms. The results of this search were from many different organisations and communities: governmental, non- governmental, research centres and universities, commercial and scientific entities and the broader public. All relevant datasets were classified under the IUCN scheme and recorded in an excel spreadsheet (hereafter referred to as the metadatabase). Sometimes datasets fitted under multiple IUCN categories, in which case they were entered multiple times.

Next a systematic literature review of published datasets was carried out by searching the Web of Science database. The searches required that a paper's abstract, title or key words contained a word or phrase from each of several categories. A 'scale' search term category to specify the scale as global, a 'method' category to specify that the paper discuss a spatial model, remotely sensed data etc., and a 'threat source' or 'threat mechanism' category that related to one of the threat categories. A category of terms relating to conservation targets (e.g. species, ecosystems etc.) was not included as many relevant datasets might not be presented in a conservation context. This document contains a full list of the search terms, how they were combined, and terms initially considered, but later taken out after test searches, either because they were too precise and returned irrelevant material, or because they related to IUCN threat categories which were not analysed further. These were categories 10 (Geological Events) and 12 (Other options) which do not directly address anthropogenic drivers of threat, and category 11 (Climate Change and Severe Weather), which has been given comprehensive treatment by the Fifth Assessment Report for the Intergovernmental Panel on Climate Change and therefore was considered beyond the scope of this current metadatabase.

The Web of Science literature search yielded 1964 papers whose titles and abstracts were imported into referencing software. The initial decision whether to reject or accept a paper was based on the following series of rejection criteria:

- Does not contain spatial information
- Does not address an IUCN direct threat to biodiversity
- About climate change as a direct threat
- Considers a non-anthropogenic threat
- Scale less than global
- Coarser spatial resolution than country level statistics
- Methodology for collecting the data unclear

The remaining papers were then reviewed a second time and kept or rejected on the basis of the abstracts (or the full paper where necessary). Finally the full text of all remaining papers was examined to collect metadata, and the datasets recorded in the metadatabase, while stating that they had been found during the literature review. While examining these papers, references to other relevant datasets were found and added to the metadatabase.

An initial draft of this metadatabase, with various metadata added to each dataset including the scale, resolution, repeatability, availability, relevance, and level of validation efforts, was then sent out for review to the authors who identified additional reviewers. This collective group, further referred as reviewers, assessed the comprehensiveness of the metadatabase and identified any omissions and errors in the catalogue. Each reviewer was sent one review form per threat category which consisted of a table of the catalogued dataset names, links to the paper or website for further information about the dataset and the following fields to be filled in by the reviewer:

- Have you used the dataset in spatial analysis before? (yes/no)
- List one strength of this dataset for global threat mapping
- List one weakness of this dataset for global threat mapping
- List one additional threat category where this dataset could be useful
- List any further comments

At the end of the form reviewers were invited to suggest any datasets that they felt were missing from the threat category they had reviewed. However, while each reviewer was sent all of the forms, the reviewer was asked to review the threat categories in which they felt they had the most expertise. Once the completed forms were returned the reviewer comments were considered and any newly suggested datasets that fit the above criteria were added while suggested datasets that did not fit the criteria were not considered further. The reviewer comments on strengths and weaknesses of the datasets were used to gauge whether the collated datasets were actually in use by the reviewer group and if there were extra metadata that could be extracted and added to the metadatabase. Most comments however were felt to be too subjective to warrant an extra metadata field in the final metadatabase. Generally the question were answered quite selectively and inconsistently across all reviewers which prevented the answers being used more substantially to refine the existing metadata. Their comments did however help to confirm that the existing metadata fields were sufficient to characterise the datasets for the filtering and analysis stage.

<u>Data Type</u>

This metadata field describes the type of dataset in terms of a pre-defined data categorization scheme in order to ascertain if there were any biases in certain threat categories towards data types. The first type is data in its most basic form (primary), i.e. which resulted from the collation of primary observations which are not necessarily spatially referenced but can by bounded by geopolitical boundaries at the sub-national, national or regional level, e.g. national level statistics on agriculture. When these observations are mapped into a regular array (grid) or as objects (vectors), they become primary spatial data, whether *in-situ* or remotely sensed. Therefore primary observations of threat can be both spatially explicit, e.g. through gridded arrays of remotely sensed variables, or non-spatial, e.g. through government held census data.

In the creation of secondary datasets, the primary (spatial or non-spatial) data undergoes various levels of processing or modelling. Basic processing may remove noise or bin data to a coarser spatial or temporal resolution etc. The coverage may then be extended with interpolation between measurements. Through 'inversion' remote sensing observations can be used to infer information about the underlying earth system. The primary dataset may be input to of a process based model, with various output secondary datasets. Sometimes an 'index' is created by overlaying datasets, in an attempt to address combined threats or something more abstract. For example a 'water quality index' might sum together various datasets on hydrological regime and pollutant run off. These are all examples of what we consider 'secondary' data.

The next category, citizen science, captures crowd-sourced, voluntary observations. With the availability of editable 'wiki' type maps and handheld data collection devices, this is an ever increasing mode of data collection. On the other hand, commercial databases are usually extremely detailed and built by a consortium of companies, businesses and other fee-paying organisations. While these are mostly related to the energy and extractive sectors, other databases exist to describe commercial activities related to building and urban development projects. As these are costly to obtain access to they are usually beyond the reach of those conducting threat assessments on behalf of conservation organisations.

Indirect data quantifies threat indirectly, e.g. a threatening state or process, and that does not directly threaten biodiversity but can result in a threat if not managed. For example, by combining commercial drivers with natural commodity values, we can estimate how likely biodiversity is to be threatened such as in areas of concession for timber, hydrocarbon extraction etc. Evaluating the supply and demand of such commodities can provide an indirect indicator for a threat source, e .g. the value of natural commodities which are internationally traded. While these do relate back to a threatening mechanism such as natural resource extraction they cannot be considered directly threatening and are therefore considered to be indirect threat data.

All of these examples serve to illustrate the variety of origins of data sources used in the metadatabase and can highlight where there is over-reliance on one or more data sources as well as where others are under-utilised. Therefore, a metadata field on data source was added and each data set coded according to the following codes: P: Primary, PS: Primary Spatial, S: Secondary, I: Indirect, CS: Citizen Science, C: Commercial. The number of datasets originating from each data source is presented as a proportion of the total number of datasets in each threat category in Figure S1.

<u>Data Format and Access</u>

The accessibility of a dataset is influenced by the data type and source. Although all datasets in the metadatabase have a spatial element to them, not all have been put into a spatial format which can be ingested by a geographical information system. Some are simply in a tabular database format, with location information recorded. Some datasets are open access, spatially formatted and can be downloaded from the internet, whilst others are retained by their authors or the industries/ organisations concerned and require permissions or payment for access.

Expert review response

In total 12 reviewers returned review forms but given the scope of the exercise and varying expertise among the group, not all reviewers returned a form for every category, but focused on the categories which best described their expertise. However, even within a review form not all fields were always filled in completely or consistently. The completeness of the returned review forms reflected the reviewer's motivation as well as the level of time available for the review given that it was a voluntary exercise.

Rationale for choice of desirable attributes for a dataset

To enable assessment of the utility of the datasets in the metadatabase we established a set of five criteria to constrain the selection of datasets that could be used for standardised and comparable global assessment of threats. Naturally, to some extent, different threat assessments will require different characteristics of a dataset, making it challenging to set thresholds for some of the criteria when applying them as filters to the full compendium of datasets. Considering specific contexts such as that of an IUCN species Red List assessments, or of reporting progress toward CBD targets enabled us to determine sensible thresholds, however we recognise that these would not be applicable in the context of every possible conservation assessment. Therefore in Table S1 we present the attributes in independent columns for readers to perform their own set of queries.

The more conservative threshold values for each filter and the reasoning for their choice, as described in the main manuscript, is outlined below:

- Accessible: So that a dataset be freely available for scientists and practitioners to access. By 'freely' we mean without payment, and with open access on the internet. If a dataset cannot be accessed it is almost equivalent to it not existing for potential end users.
- Appropriate spatial resolution: We considered all datasets in vector format or raster datasets with a cell size of maximum 100 km^2 at the equator. Clearly, the spatial resolution required of a dataset will vary depending on the application, however when assessing threats to species, as in an IUCN Red List Assessment, anything coarser than 10 by 10 km is of limited use given that 23% of species on the IUCN Red List have ranges ≤1000 km square. For these narrow-range species a spatial layer of threat at a resolution of 10 by 10 km would cover their range with no more than 10 grid cells, insufficient to detect spatial variability in threat level. It could be argued that even finer spatial resolutions are necessary but 10km is taken as a sensible upper limit for work on most species. We tested the sensitivity to this assumption by applying a criterion of by 100 by 100 km which might be appropriate for other applications.
- Up to date: We considered all dataset with a snapshot since 2006. For any information to inform policy it needs to be timely and relevant. For the year of production, **2006** was chosen as a cut-off point, given the fact that IUCN species assessments expire after 10 years. However, again, this criterion will certainly not apply to any conservation assessment. Indeed, some analyses will aim at to study past or future trends and require a different time frame. We tested the sensitivity to this assumption by applying year 2001 as a cut-off year.

- Repeated: Repeated measurements are required to determine how threats are changing over time. To set conservation targets, monitor the impacts of policies or interventions and make future predictions, it is necessary to examine past changes in threat. Regularly updated datasets also allow future projections of threat. The temporal resolution desired in a dataset will vary so greatly with the application that we simply required, as our filter threshold, that the dataset include at least two or more measurements in time.
- Assessed for Accuracy: An understanding of certainty in the information portrayed in a dataset is essential for it to be used with confidence in threat assessments. We therefore only considered datasets consisting of direct measurements/observations or derived through models/extrapolations that had been validated against another independent, validation dataset. The accuracy of the datasets was not under scrutiny. Here we were only interested in identifying whether a measure of accuracy of the dataset was available. The type of validation naturally varies between datasets, with some validated globally against empirical data, others only in certain regions, and others compared against alternative but similar models to estimate variance between model outputs. However, given that inter-comparison between models does not reflect the "truth" on the ground, and the end-user would want to know about the level of confidence of the whole dataset, rather than just one geographic region, we only took validation against globally representative empirical data as criteria for further use. We tested the sensitivity of this assumption by including also an attribute for regionally or globally validated dataset.

<u>Metadata which was considered, but not recorded</u>

In assembling the metadatabase and in considering the reviewer comments described in section 2, more metadata fields were considered to make the metadatabase as comprehensive and useful as possible but were not added to the final version. These are described below and the reasons why they weren't added explained. In all cases it is recognised that these fields would deepen the knowledge base and make a useful resource for threat assessment but given the reasons described could not be added.

Applicability: Some datasets can be used to assess the 'input' distribution of a potential threat to biodiversity, and others look at where that potential threat is actually impacting upon some aspect of biodiversity. A threat impact map for the same potential threat would look quite different depending on the species, population or ecosystem considered. The aim of this exercise was to find datasets assessing threat input (although datasets addressing both input and impact were sometimes included where there was ambiguity). It is for conservation practitioners to take threat input information and decide where and how it is applicable to the conservation target of interest.

Further criteria of 'assessed for accuracy': Besides looking at model validation, we could not provide information about the uncertainty in the datasets, nor their accuracy. The robustness of the datasets could not be addressed by simply looking at which had been published in peer-reviewed journals because many organisations creating these datasets have internal reviewing systems. An example of this would be datasets created by the FAO, which are not published in the scientific literature. Furthermore, simply the documentation of methods was not considered an appropriate indicator of robustness. We could also not guarantee that we had found all published documentation associated with a dataset.

Thematic Resolution: Different datasets provide different levels of thematic detail about a threat. For example, some cropland distribution maps simply group all cropland together and have a low level of discernment between crop types and hence a low thematic resolution, whilst others might provide information about the distributions of different crop types, or the intensity of cropping etc, and have a higher thematic resolution. We do not capture information about the relative thematic resolution of the datasets, not only because it would be very difficult to assess, but also because it is context-dependent. For different conservation targets different thematic resolutions will be relevant. For example, for an organism whose natural habitat is displaced by cropland, the type of cropland involved would be irrelevant. However, for an organism that can inhabit farmland, information regarding cropping methods or crop types would be relevant and necessary for a comprehensive assessment of threat.

Ease of use: Being able to access a dataset does not necessarily mean that it is easy to use. Many comments from the reviewers indicated that datasets were accessible but difficult to work with. If the dataset is overly challenging to manipulate and analyse by end users, it is almost equivalent to not existing. Without having worked with all of the datasets, ease of use cannot be objectively rated, but we recognise that it is important that datasets be not only freely accessible but also intuitive to use and delivered in a readily usable format for spatial mapping. We note that ease of use implicitly includes dataset documentation, detailing the providence, purpose, methods, and other essential details necessary to use the dataset appropriately.

Quantifying the number of species impacted by threats

The percent of threatened terrestrial and inland water and marine species on the IUCN Red List impacted by each threat class was calculated, as shown in figure 1 in the main manuscript. The number of species impacts by each threat was obtained by querying the IUCN Red List online (http://www.iucnredlist.org/). To avoid bias we only included species from taxonomic groups that have been comprehensively assessed, and due to differences in interpretation for several threat classes (e.g. Biological Resource Use) in the marine and terrestrial environments we split our analysis by those two habitat classifications. We retained only species that were listed as 'threatened' on the IUCN Red List, which we considered as listed as Critically Endangered (CE), Endangered (EN), Vulnerable (VU), or Near Threatened (NT). All species included, their Red List ID, formal taxonomy, habitat types, and threat class associations (indicated by 0 for no listed threat, 1 for a listed threat) are available in Table S2.

Literature Review Search Terms

* Indicates the rest of the word can be filled out with other letters (e.g. threat* could include 'threatening' for example)

$ can be replaced by any character or no character

Use of 'OR' to separate search terms means that the paper's title, abstract or key words had to contain either of the terms

Use of 'AND' to separate search terms means that the paper's title, abstract or key words had to contain both terms

Two separate searches were carried out (using OR to separate the terms within each category in italics):

SEARCH 1: *Approach/ Method* AND *Scale* AND (*Sourc*e OR *Mechanism*)
SEARCH 2: *Database* AND (*Source* or *Mechanism*)

Approach/ Method

Spatial

Remote-sensing

Remote sensing

Map*

Spatial Model*

GIS
Satellite$

Scale
Global scale
Global-scale
Global level
Global-level
Global map
Global analysis
Global study
Global distribution

Database
Global database
Global dataset

Threatening Sources
Urban
Housing
Industr*
Touris$
Recreation*
Agricultur*
Farming
Crop*
Livestock
Pasture$
Grazing
Ranching
Cultivation
Aquaculture
Forestry
Plantation$
Timber
Logging
Charcoal
Firewood
Oil
Gas
Coal
Fossil fuel$
Power plant$
Mining
Quarrying
Aggregate extraction
Renewable energy
Solar power
Wind power
Wind farm$
Wind turbine$
Geothermal power
Tidal power
Hydropower
Transport*
Highway$
Road$
Railroad$
Railway$
Canal$
Shipping
Port$
Airport$
Air-travel
Communication line$
Utility line$
Pipeline$
Powerline$
Power line$
Power cable$
Dam$
Hunting
Poaching
Fishing
Fisher*
Military
Civil unrest
War

Threatening Mechanisms
Land-use change
Land use change
Deforestation
Desertification
Habitat loss
Habitat destruction
Habitat fragmentation
Habitat reduction
Habitat degradation
Ecosystem loss
Ecosystem destruction
Ecosystem fragmentation
Ecosystem degradation
Forest loss
Forest destruction
Forest fragmentation
Forest degradation
Erosion
Sedimentation
Acidification
Deoxygenation
Invasive Species
Alien species
Invasive disease$
Introduced genetic material
Fire$
Water abstraction
Irrigation
Dredging
Pollut*
Effluent$
Harvesting
Killing
Bycatch
Fisheries discard$
Collision$
Disturb*
Litter*

Words initially considered but not included in final search:

From transport
Flight
From secondary Industry/ Commercial areas
Commercial
From scale:
Global
Regional
Regional-scale
Region-scale
Region scale
Region level
Continental
Continent- Scale
Continent scale
Continent level
National
National-scale
National-level
Country
Country-scale
Country-level
Ocean$
Marine
Freshwater
Terrestrial

Entire term categories initially considered but later discarded:

Geological
Volcano*
Earthquake$
Landslide$
Climate
Climate
Climate-change
Drought$
Storm$
Flooding

Generic Words

Threat*
Pressur*
Stress*
Impact*
Effect*

Conservation Target

Biodiversity
Ecosystem
Community
Species
Habitat

Supporting References

15. N. Salafsky *et al*. A standard lexicon for biodiversity conservation: Unified classifications of threats and actions. *Conservation Biology* **22**, 897-911 (2008).
16. A. Balmford, *et al*. Capturing the many dimensions of threat: comment on Salafsky et al. *Conservation Biology* **23**, 482-487 (2009).
17. N. Salafsky, *et al*. Pragmatism and practice in classifying threats: a reply to Balmford et al. *Conservation Biology* **23**, 488-493 (2009).

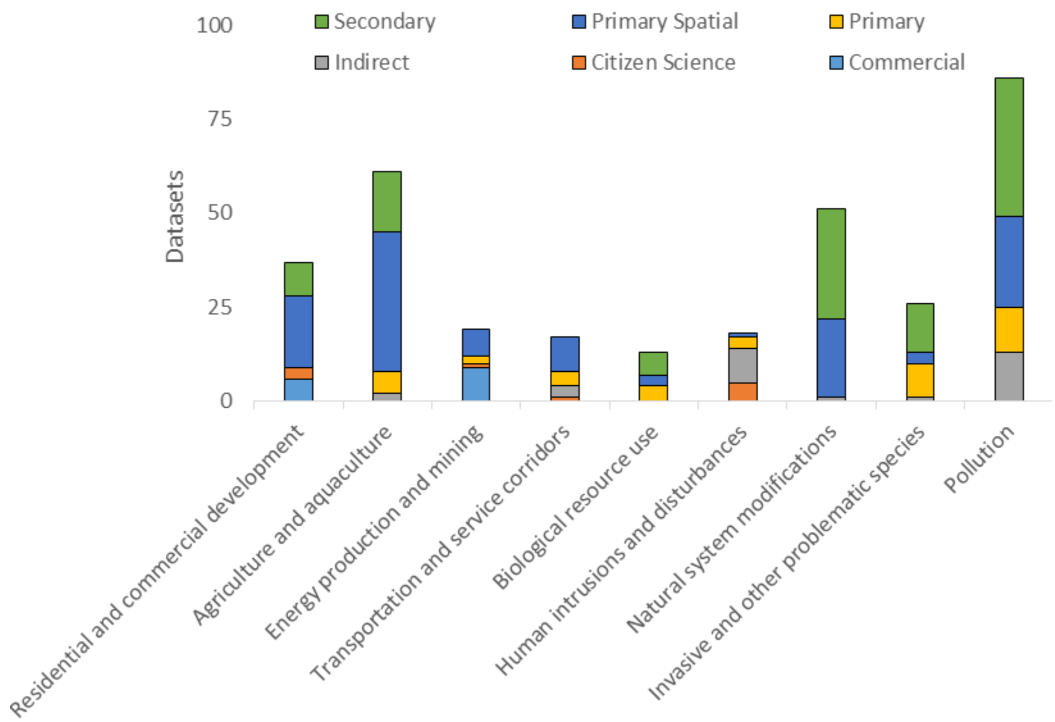

Fig. S1.
The number of datasets per IUCN threat category with the proportion of datasets by source.

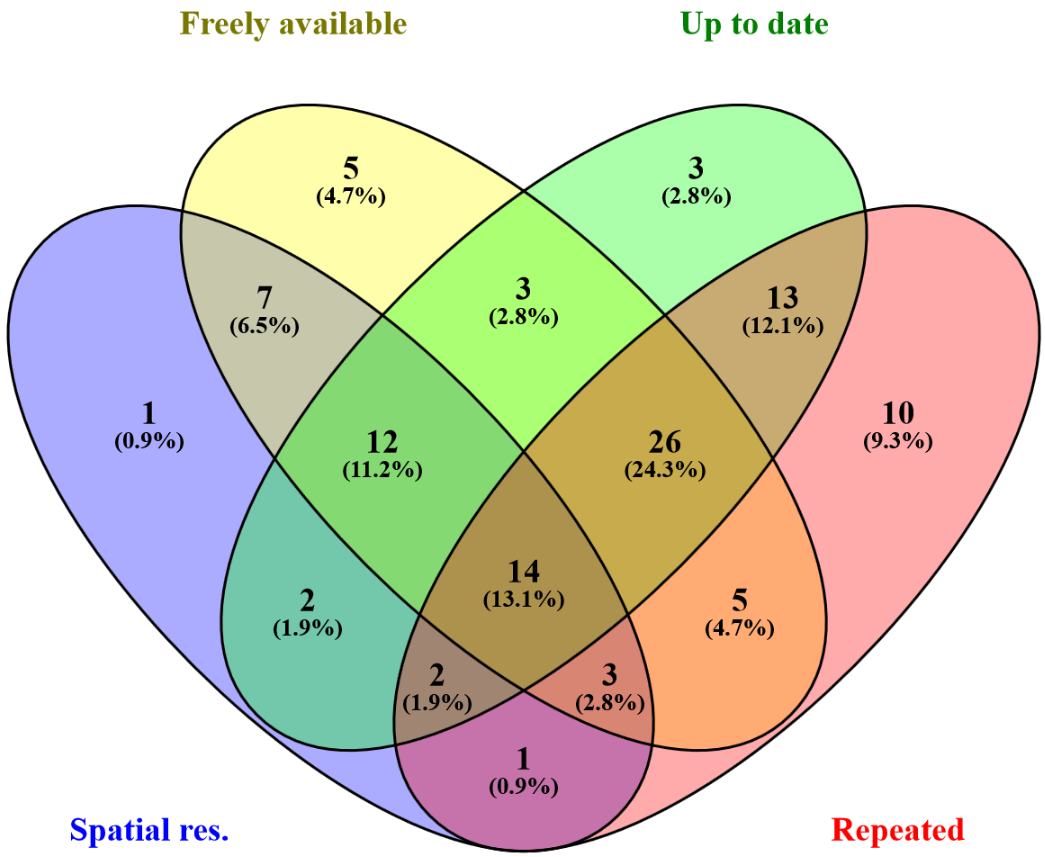

Figure S2. The number of datasets that meet each of four desirable dataset attributes outlined in Table 1 as well as being global in coverage and representing either models assessed for accuracy or empirical observations. Numbers in each intersection represent the number of datasets that meet those constraints. See Table S1 for a full list of datasets and their quality attributes.

Table S1: This table provides the metadatabase analyzed in this study. The classification keys used in the columns are described in Table S3. Filters are provided to identify which datasets did or did not meet each of our five filters.

Table S2: This table provides information for the Red Listed species analyzed in Figure 1, including the taxonomy as well as a binary classification of whether or not a species is threatened by each of the nine classes of threat considered in our manuscript.

Table S3: This table outlines the metadata column headings used in the metadatabase (Table S1).

Column Heading	What The Column Records	Key For Codes Used To Record Information (where not self-explanatory)
IUCN Direct Threat Level 1	The relevant level 1 threat category number from the IUCN's 'Direct Threat' classification scheme	
IUCN Direct Threat Level 2	The relevant level 2 threat category number from the IUCN's 'Direct Threat' classification scheme	x' indicates that a level 2 threat category is unclear
IUCN Direct Threat Level 3	The relevant level 3 threat category number from the IUCN's 'Direct Threat' classification scheme	x' indicates that a level 3 threat category is unclear
Brief Data Description/ Layer Name	A name for the dataset (the 'official' name if the dataset has one or a brief description if not)	
Link	A hyperlink to a website containing the dataset	no link' indicates that no website with the dataset exists
Link To Published Paper	A hyperlink to a free pdf of the peer reviewed paper about the dataset	N: Paper not available for free online n/a: Not applicable
Data Type	General data categories	PS: Primary data (spatial – including remotely sensed datasets) P: Primary data (non-spatial – including statistical databases and national or regional level tabular data) S: Secondary data - modelled, interpolated, or post-processed primary data CS: Citizen science data C: Commercial data I: Indirect data on threats - not a direct estimate of a threat source but describes something that implies or could lead to a threat, e.g. status and trends in natural resource commodity values
Data Format	The format of the dataset	V: Vector GIS layer R: Raster GIS layer T: Tabular U: Uncertain
Spatial Resolution	The spatial resolution of the dataset	A number: The (approx) pixel size of raster in km squared A number x a number: Pixel size of raster in km x km (where not square) C: Data is at country level R: Data is at a sub- country, regional level V: Vector data or associated with point locations (but not necessarily available as shapefile) O: Other (see 'notes' column for details) U: Uncertain
Start Year	The first year of data coverage	F: Future P: Present
End Year	The final year of data coverage	F: Future P: Present

Temporal Resolution	The temporal resolution of the dataset between the start and finish dates above	A: Annual M: Sub-annual to Monthly S: Sub monthly N: No repeat (static) O: Other (multiple snapshot but spaced at interval neither monthly, nor annual, nor sub monthly) U: Uncertain
Data Availability	The availability of the dataset	F: Free access (online or download) P: Pay to access A: Author/ creator has the data- and can be asked to them U: Uncertain
Assessed for Accuracy?	Has the dataset been validated against an independent dataset?	G: Validated against global empirical/observed dataset (including a representative sample of locations from across globe) R: Validated against empirical/ observed dataset in a small area/ for summed regional statistics M: Validated by comparison with a modelled dataset N: No n/a: not applicable as the dataset includes direct observations and does not involve modelling U: Uncertain