



## Review Article

# The ATTAC guiding principles to openly and collaboratively share wildlife ecotoxicology data



Cynthia C. Muñoz<sup>a,\*</sup>, Sandrine Charles<sup>b</sup>, Emily A. McVey<sup>c</sup>, Peter Vermeiren<sup>a,d</sup>

<sup>a</sup> Department of Environmental Science, Radboud University, Nijmegen, GL 6500, the Netherlands

<sup>b</sup> CNRS, UMR 5558, Laboratory of Biometry and Evolutionary Biology, University of Lyon, University Lyon 1, Villeurbanne F-69622, France

<sup>c</sup> College for the Authorization of Pesticides and Biocides, Ede, LL 6717, the Netherlands

<sup>d</sup> Department of Natural Sciences and Environmental Health, University of South-Eastern Norway, 3800 Bø, Norway

## REVIEW HIGHLIGHTS

- The ATTAC workflow sustains Access, Transparency, Transferability, Add-ons, and Conservation sensitivity in wildlife ecotoxicology.
- The ATTAC workflow gathers guidelines supporting both prime movers and re-users in maximising their use of already available data in wildlife ecotoxicology.
- The ATTAC workflow promotes an open and collaborative wildlife ecotoxicology in support of wildlife regulations and management.

## ARTICLE INFO

### Keywords:

Accessibility  
Transparency  
Transferability  
Add-ons  
Conservation management  
Systematic literature search  
Database homogenization and integration  
Heterogenous scattered data analyses  
Biodiversity conservation  
Chemical pollution regulation  
Collaborative workflow  
Data curation  
Reproducibility  
Risk assessment

## ABSTRACT

The inability to quantitatively integrate scattered data regarding potential threats posed by the increasing total amount and diversity of chemical substances in our environment limits our ability to understand whether existing regulations and management actions sufficiently protect wildlife. Systematic literature reviews and meta-analyses are great scientific tools to build upon the current push for accessibility under the Open Science and FAIR movements. Despite the potential of such integrative analyses, the emergence of innovative findings in wildlife ecology and ecotoxicology is still too rare relative to the potential that is hidden within the entirety of the available scattered data. To promote the reuse of wildlife ecotoxicology data, we propose the ATTAC workflow which comprises five key steps (Access, Transparency, Transferability, Add-ons, and Conservation sensitivity) along the chain of collecting, homogenizing, and integrating data for subsequent meta-analyses. The ATTAC workflow brings together guidelines supporting both the data prime movers and re-users. As such, the ATTAC workflow could promote an open and collaborative wildlife ecotoxicology able to reach a major objective in this applied field, namely, providing strong scientific support for regulations and management actions to protect and preserve wildlife species.

\* Corresponding author.

E-mail address: [munozc.cynthia@gmail.com](mailto:munozc.cynthia@gmail.com) (C. C. Muñoz).

<https://doi.org/10.1016/j.mex.2022.101987>

Received 15 November 2022; Accepted 26 December 2022

Available online 28 December 2022

2215-0161/© 2022 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license

(<http://creativecommons.org/licenses/by-nc-nd/4.0/>)

## Specifications Table

Subject area:	Environmental Science
More specific subject area:	<i>Ecotoxicology</i>
Name of the reviewed methodology:	<i>ATTAC workflow (and linked FAIR framework and CRED, CAT and Klimisch methodologies)</i>
Keywords:	Keywords: Accessibility; Transparency; Transferability; Add-ons; Conservation management; Collaborative workflow; Systematic literature search; Database homogenization and integration; Data curation; Heterogenous scattered data analyses; Biodiversity conservation; Reproducibility; Chemical pollution regulation; Risk assessment
Resource availability:	<i>Not applicable</i>
Review question:	<i>What are the critical steps to ensure sustainable reuse of scattered ecotoxicology data? What are concrete guidelines and tools available for both data prime movers and re-users in the context of open, collaborative data sharing in wildlife ecology and ecotoxicology?</i>

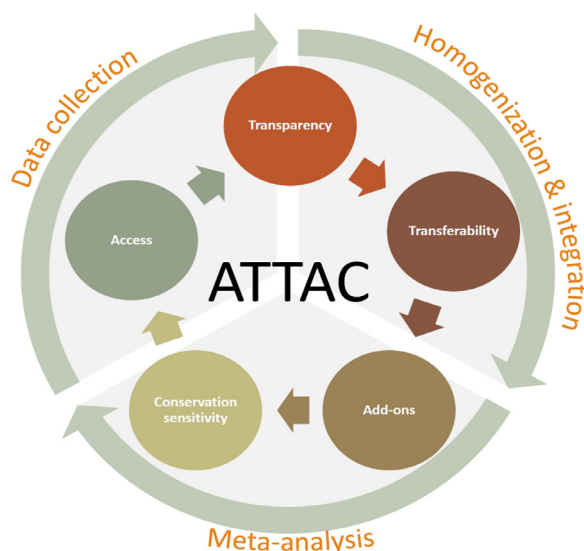
## Method context

Since the early reports on chemical contamination in the 1960s, the total amount and diversity of toxic compounds within the environment have been steadily increasing, gradually pushing our planetary boundaries [9,77]. Despite global efforts over the past few decades to monitor and respond to environmental pollution, the exposure, bioaccumulation and effects of chemical substances on many wildlife species remain poorly quantified [23,49]. Chemical contamination biomonitoring in wildlife is predominantly conducted by isolated research teams [10], often within individual projects with a limited timeframe and budget. Accordingly, data and knowledge regarding threats posed by pollutants on wildlife are commonly fragmentary, and on their own insufficient to inform chemical risk assessment, ecosystem management, or species conservation actions based on comprehensive, quantitative ecotoxicological understanding. Most of our knowledge of wildlife ecotoxicology is currently derived from studies on domestic animals or lab model species [49]. Constrained by ethical, practical, and financial limitations, the inability to quantitatively integrate scattered data and knowledge regarding possible threats posed by chemical pollution to wildlife species, unfortunately, limits our ability to determine whether existing regulations and management actions regarding pollution sufficiently protect their populations.

The increasing attention paid to improved accessibility of research outputs in both the academic and the regulatory and management worlds should, in the short-term, offer new perspectives to transform the capacity of wildlife ecology and ecotoxicology to inform on chemical risk assessment, ecosystem management and species conservation in the context of environmental pollution. Research output accessibility has been heavily promoted under the umbrella of Open Science and the FAIR (Findable, Accessible, Interoperable and Reusable) data principles [81] across the academic community, and in recent policy actions on funding for academic research such as the Memorandum on Ensuring Free, Immediate, and Equitable Access to Federally Funded Research [55] in the USA. In parallel, there is also increased demand for accessibility within regulatory contexts as echoed, for example, in the transparency regulation of the European Commission regarding risk assessment of products in the food chain [31].

Systematic literature review, research synthesis, and meta-analysis provide scientific tools to build upon this renewed push for accessibility of knowledge and data regarding the ecology and ecotoxicology of wildlife species. Such tools allow us to explore new investigation angles from which additional innovative findings can emerge. These findings can go far beyond the original data because they fruitfully gather information from a wide diversity of scattered sources thereby deepening our understanding of the underlying processes beyond what could have been achieved in any individual study. For example, a meta-analysis across 17 available studies regarding sea turtle egg pollution levels [61] established for the first time a global, quantitative overview of a topic that is critical to assess pollution risks to early life stages across a long timeframe and at a large geographic scale. A subset of the integrated database related to paired mother-egg samples, and allowed quantitative differences in maternal transfer for different pollutants to be established and linked to known processes of egg formation. This meta-analysis also highlighted critical research directions which could not have been detected from individual studies alone. In a follow-up, Muñoz et al. [59] succeeded in the integration of 26 studies covering 40 years of biomonitoring data, thereby allowing for a comparative analysis of the internal distribution of persistent organic pollutants (POP) among all extant sea turtle species. Similarly, based on a meta-analysis, Ratier & Charles [75] proposed a new and promising method to assess the bioaccumulation capacity of chemical substances accounting for the uncertainty on the bioaccumulation metric estimates; an innovative approach that could replace the current use of a single median to do so as required in regulatory documents (e.g., [30,65]). Without an extensive integrated database pooled across the literature, development and testing of this new approach would have been limited to prior selected chemical substances. Similar examples of meta-analyses demonstrate the potential value of integrative analyses for wildlife species (e.g., [14,70]). Nevertheless, despite the potential of such integrative analyses, the emergence of innovative findings in ecology and ecotoxicology of wildlife species are still too rare relative to the potential that is hidden across available scattered data.

Multiple facets of our global environment are changing at unprecedented rates, including an ever-expanding universe of chemicals [9]. Simultaneously, there is increasing awareness and clear societal requirement for replacement, refinement, and reduction (3R) in the use of vertebrate toxicity testing [52]. Consequently, there is today an obvious need, if not a duty, for closer collaboration between researchers in academia, industry and regulatory fields collecting field and/or experimental data, and those aiming at reusing such data to investigate new fundamental and applied research questions from sufficiently data-rich meta-analyses. Achieving a successful sharing of data with the perspective of data reuse could at first resemble the ascent of Mount Everest, but quickly an ascent of Mont Blanc if the will and energy of researchers are sufficient, if the means are mutualised, and if the methods of data collection are harmonised.



**Fig. 1.** The ATTAC workflow, including the five key steps corresponding to data Access (1), Transparency (2), Transferability (3), Add-ons (namely, provision of additional metrics) (4), and Conservation sensitivity (i.e., the wise use of conservation-sensitive materials) (5).

Based on our experience in developing integrated databases from extensive systematic literature searches and making the most of these databases in subsequent meta-analyses, we propose in this paper an optimized workflow (ATTAC workflow, Fig. 1) to promote the use and reuse of wildlife ecotoxicology data. The workflow focuses both on opportunities for scientists producing primary data (prime movers) and scientists reusing these data in secondary analyses (re-users). The workflow is suited to wildlife species, and particularly relevant for species of conservation concern where data collection and reuse are critical issues because of their declining numbers and the difficulties in their sampling in the field or experimentally. The apical aim is to provide the needed scientific basis to inform on chemical risk assessment (e.g., retrospective assessment and providing input on protection goals), ecosystem management and conservation of wildlife species with in-depth knowledge about wildlife ecology and ecotoxicology from improved, integrated databases and analyses.

## Workflow

The ATTAC workflow we propose comprises five key steps along the chain of collecting and homogenizing data for subsequent meta-analysis (Fig. 1). The five steps of the ATTAC workflow along the above-mentioned chain specifically refer to data Access (1), Transparency (2), Transferability (3), Add-ons (namely, provision of auxiliary metrics) (4), and Conservation sensitivity (i.e., the wise use of conservation-sensitive materials) (5). This paper establishes an innovative workflow to successfully address issues arising at each of the five key steps with best practice guidelines for scientists working at both ends of the spectrum, i.e., prime movers and re-users. The ATTAC workflow progressively emerged from experiences and earlier applications in Muñoz & Vermeiren [61], and Muñoz et al. [59], and has been further consolidated when building an integrated database to investigate the maternal transfer of chemical substances across reptile species. The new, extended, and improved ATTAC workflow presented in this paper includes refinements based on further discussion among the authors in gathering ecotoxicological data to support quantitative meta-analyses and modeling from both academic and regulatory perspectives.

We provide guidelines associated with each key step of the ATTAC workflow within the context of wildlife and particularly conservation-sensitive species. Nevertheless, the ATTAC workflow could also be applied in a broader sense in situations where data regarding ecology or ecotoxicology is produced and reused. In this context, the ATTAC workflow complements FAIR data principles as follows:

- The “Access” step integrates the “findable” and “accessible” FAIR principles. The ATTAC workflow places additional emphasis on the type of data that are ideally made accessible (or can be extracted) in an ecotoxicological context, while the FAIR principles primarily relate to how data are made available.
- The “Transparency” and “Transferability” steps apply the “interoperability” and “reproducibility” FAIR principles toward the specific context of wildlife ecology and ecotoxicology. Moreover, we distinguish between transparency (focussed on communication) and transferability (focussed on methodology and data harmonization for easy reuse). Both steps are closely linked but cover distinct issues and associated guidelines within the ATTAC workflow.
- The “Add-ons” and “Conservation sensitivity” steps are going beyond the FAIR data principles and are of relevance with wildlife species for which data collection raises critical ethical, financial and resource issues. The study of wildlife species requires a highly efficient alignment across disciplines to make optimal use of the scarce data that can be collected on these animals with minimal impact.

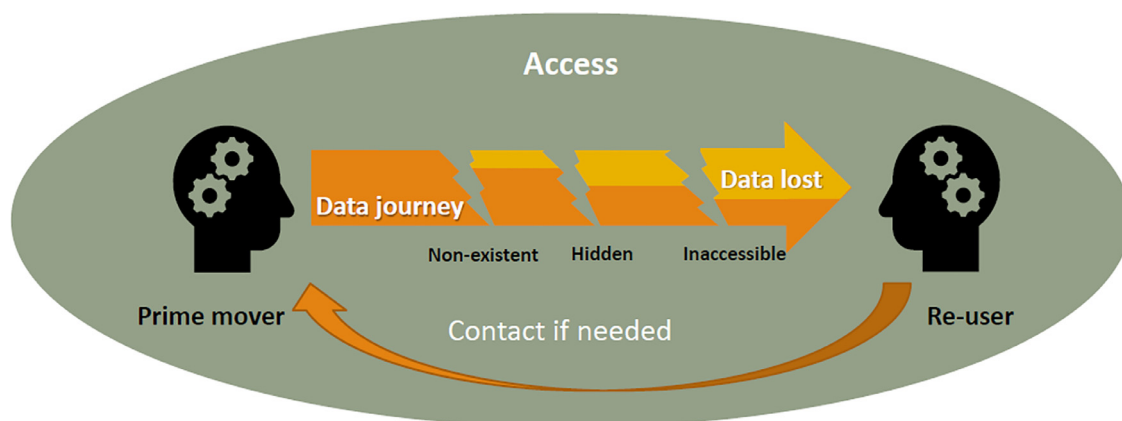


Fig. 2. Types of data loss during the data journey from the prime mover to re-user, including non-existent, hidden, and inaccessible data.

### Access

A critical first step in data reuse is finding and accessing data. Initially, this often involves a systematic search for publications regarding a certain topic. From a regulatory perspective, a systematic literature search is generally required to demonstrate that data supporting an application were selected without bias, rather than cherry-picked to fit the objectives. Guidance regarding systematic search strategies, including steps such as identification, screening, and assessing eligibility for inclusion, can be found in guidelines such as PRISMA [57]. Next, the search for data supporting the publications of interest can begin, as data are usually not reported within the publication itself. Consequently, any appendices, supplementary materials, linked publications or reports, as well as references to repositories should be checked and an inventory made of the available and accessible data. This can then be cross-checked against the required data (e.g., in some cases a database with pollutant concentrations or toxic effects alone is insufficient for a certain meta-analysis aim). Contact with the authors might be required to access any missing or additional data, as well as to clarify unclear or missing meta-data and methodological issues.

In some cases, data are lost to the broader community (Fig. 2). Such lost data include “non-existent” data that were never collected to begin with (e.g., a study might choose to express pollutant concentrations on a wet weight rather than lipid basis and therefore never measure lipid contents). While these data are not technically lost, they can nevertheless present a loss as they limit the future usability of the data. Data can also be lost because they remain “hidden” in grey literature which is more difficult to find and may not show up in typical scientific literature databases. Examples of this include students’ theses, research project reports, as well as dossiers submitted to regulators. Finally, data can be lost because they are “inaccessible”, either fully or partially (e.g., because the data regarding the detection limits are lacking, or because the data are presented in a condensed form such as averages for groups of animals). This includes cases where details of contact authors are no longer up to date and no successful contact can be established with the broader research team (or the broader team also cannot contact the responsible author). Inaccessible data also include those data that have been lost over time, often because they were archived in obsolete technologies or poorly managed filing systems. Last but not least, data can remain inaccessible due to restrictions or embargoes, for instance, concerning privacy issues or legal protection of sensitive data. Likewise, in most regulatory dossiers, the entity that funded the research may “own” the data for a certain amount of time after it is developed. It can be difficult even for regulatory agencies to know when data ownership has lapsed. Consequently, while data summaries may be available in openly accessible reports and assessments, the underlying data can only be purchased from the data owner. A special case of restriction, primarily related to academia, relates to the phenomenon of “authorship bargaining”. Here, authors will only release their data under the condition that they gain authorship of the new work. When publications are accompanied by statements that data will be available upon request, the subsequent inaccessibility due to “authorship bargaining” highlights a broader concern regarding the breaching of FAIR data commitments [81]. Such breaches are of particular concern in the context of helicopter science where research conducted within developing countries by foreign scientists becomes unavailable to local scientists, management and regulatory agencies and the broader public within those developing countries [38,74]. In all these cases of data loss (non-existent, hidden, and inaccessible data), the resulting lack of data presents a serious limitation for wildlife ecotoxicology to build upon valuable, scarce data.

### Data accessibility guidelines for the prime mover

As a prime mover, having a data management plan (DMP) according to FAIR data principles can overcome issues regarding the loss of data or loss of contact [15,43] and is in fact one of the main tenets of Good Laboratory Practice (GLP, [69]). In such a DMP, attention should be given to the long-term sustainability of the database under the FAIR umbrella. Academia today is characterized by short-term contracts, particularly for early career scientists [2,73], making it often hard to keep contacts up to date. Even in the case of an apparent stable contact, some thought should be given to the afterlife of the database. We experienced many cases where the author retired or passed away inclusive of their life’s work, limiting transgenerational use of their data. Likewise, the most

up-to-date technology at the time of the original research is likely to be replaced by newer, state-of-the-art technologies at the time of intended reuse. The increased use of data repositories is a suitable solution to both issues as it provides a long-term, identifiable and findable location for research data, and should also ensure compatibility with future technologies [39]. Journal supplementary materials provide an alternative which is closely connected to the publication, yet have two major drawbacks. Firstly, the data are not immediately identifiable on their own (e.g., supplementary materials do not have their own DOI, while repository entries do), making the data less visible and findable to other researchers. Secondly, when the publication is not open access, the accompanying supplementary materials are also frequently not openly accessible, thereby limiting the potential for reuse. Depositing data into repositories does not necessarily mean that the data are immediately publicly and openly accessible, as access restrictions can still be applied. The choice, however, to not make research data open should carefully consider future sustainability, as loss of contact with the data owner is a real possibility in the future. Meanwhile, the application of methodologies such as blockchain technology within repositories, in principle, makes them less dependent on unique people or infrastructures.

Authorship bargaining presents a conditional and selective release of data that is highly questionable. There is no doubt that the authors who initially collected the data have done so with a large effort and intellectual input. Any subsequent reuse of these data should thus give proper citation of the original study. If the data are published, authors receive credit for the work in the form of their publication and subsequent citations. A further analysis of existing data, using systematic literature search and meta-analysis techniques, goes conceptually beyond the initial data and involves further manipulation and intellectual development. It is therefore considered unreasonable to demand authorship simply because data are provided. Guidelines regarding authorship can be found in many journals, e.g., in Contributor Roles Taxonomy (CRediT) statements [3,17] and guide what constitutes intellectual input into new work. Nevertheless, it is worth noting that a repository database with DOI is also a unique, citable research output in addition to a publication [37] and that providing open access to data increases the confidence in the research and thereby citation rate of studies [8,25]. Furthermore, sharing data presents a networking activity connecting researchers with similar interests and can therefore provide fruitful ground for additional, collaborative research.

A final best practice guideline in the context of data access relates to the type of data (i.e., “*what*”) that is ideally made available. Data are often combined during analyses, for example, because the focus might be on differences between groups of individuals or broad categories of pollutants. When presenting data in such a condensed form, however, some of the information which might turn out to be valuable in the future is lost. It is difficult to know beforehand what types of future analyses might be possible. Therefore, we recommend that data are made available in as raw as possible form and that any complementary data (e.g., detection limits, lipid contents) are also released.

#### *Data accessibility guidelines for the re-user*

When confronted with some of the issues identified under the ATTAC step of “Access”, several best practice guidelines can be employed by the re-user. A first guideline is simply being prepared to face issues and making them part of the systematic literature search strategy. This particularly relates to cases where contact with the authors is required. Data are valuable and definitely worth the effort of unearthing. Nevertheless, seeking and establishing successful contact with authors can become an elusive and highly time-consuming activity. Having a strategy including deadlines regarding the timeframe in which you like to achieve the collection of databases, and a plan of attack regarding how and how often to contact authors can prevent mission creep in the data collection stage of the project. For example, as part of our strategy, we targeted key researchers such as the first, last, and corresponding author from the publication during first contact, and followed this up with a maximum of three additional attempts to all authors. When contact details were no longer up to date, approaches included a web search for the current affiliation, a search on (professional) social media, and contacting closely related colleagues or the secretariat of the institute.

When data are lost, technical solutions are available to (partially) rescue data. Data can, for example, be digitized from graphs or extracted from tables using an increasing variety of image analysis software (e.g., ImageJ, WebPlotDigitizer, GetData Graph Digitizer, Adobe Acrobat). Additionally, data that are presented in condensed form (e.g., pooled over a given number of replicates) can often be analysed by back calculation, applying weighting to data points based on the number of replicates within a study [58]. Likewise, summarized data (e.g., means and standard deviations) can sometimes be rescued by simulating from a distribution with the summary statistics as parameters (e.g., a normal distribution with mean and error) although assumptions might need to be made and checked *a posteriori* (e.g., regarding the underlying distribution, [41]), and results will depend on simulation choices. The application of such rescue solutions implies that in the end, synthetic data will be available, but not the original raw data. Consequently, this is not an ideal situation but provides an acceptable compromise when anything else can be done.

#### *Transparency*

Once the available and accessible data are collected, the next critical step is to understand what these data represent and how they were created. Such “data reproducibility”, above all, requires transparency regarding the detailed study design (while “results reproducibility” also requires details regarding the statistical methods) and would allow one to collect an independent dataset upon which similar scientific insights could be gained. The required information can often be retained from meta-data, the scientific publication, or its supplementary materials, although contact with the authors might be required using similar approaches as for obtaining data access as detailed in section “Access”.

The ATTAC step of “Transparency” involves, firstly, understanding the nature of the different entries (typically rows within a database) and attributes (typically columns within a database) collected for each entry (Fig. 3). For example, it should be clear if data in rows represent individual organisms, subsamples within individuals (e.g., different tissues), or an aggregated grouping of

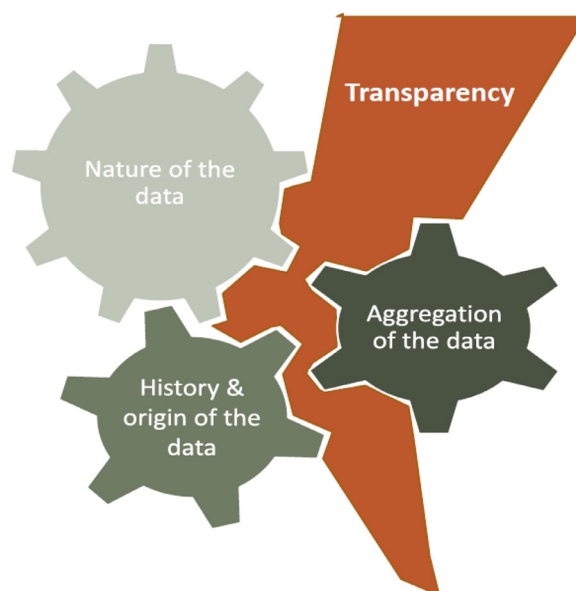


Fig. 3. Issues affecting the transparency of data for reuse can reduce the resulting information content of integrated databases.

individuals (e.g., when multiple individuals are pooled to have sufficient sample material, or when mandatory for studying toxic effects on reproduction of hermaphrodites such as in [66]). For each attribute, the measurement units are of paramount importance for correct interpretation and further analysis. Additionally, for pollutant concentrations, the reference basis (e.g., lipid basis, wet mass, dry mass) is a crucial attribute for ecotoxicological analysis as it directly relates to how the pollutants behave in organisms, allowing for comparisons among organisms or tissues with different physiological profiles. Likewise, the laboratories' reporting limits accompanying pollutant measurements are critical information to be further accounted for in ecotoxicological analyses. Unfortunately, different laboratories (and consequently different studies and databases) provide and calculate different reporting limits. For example, the procedure for calculating detection limits and the subsequent derivation of quantification limits differs between US EPA and IUPAC guidelines ([40] Chapter 3). Moreover, reporting limits even for the same machine and sample matrix will vary within and between laboratories and analysis runs. Consequently, reporting limits from one study cannot be applied to another. Tragically, in our experience, reporting limits are frequently not reported, and among the data most often lost in wildlife ecotoxicology. Further complicating this situation is the fact that some studies substitute data below reporting limits, for example, with 0 or with an artificial number between zero and the reporting limit itself. As a result, homogenized, integrated wildlife ecotoxicology databases will almost certainly contain pollutant concentrations measured with different reporting limits, with some of these data lacking information on reporting limits. Finally, for attributes recording toxicity, the duration, (controlled) conditions, replication and number of individuals per replicate are essential information to accurately understand and interpret the reported values.

A specific issue regarding the nature of the data in the context of wildlife ecology and ecotoxicology is the traceable and identifiable documentation of species and pollutant names. Species names and their taxonomic classifications are known to change [21]. Additionally, for some taxonomically diverse, but less studied wildlife groups, expertise to identify specimens might be limited. The latter can translate itself into a limited taxonomic resolution in the dataset, and a taxonomic mismatch and heterogeneity in homogenized databases [85]. In such cases, the assistance of taxonomic experts as well as connecting the homogenized, integrated database with available taxonomic databases might be required. In parallel, one chemical compound can be known under different names. Using an international naming convention such as CAS or IUPAC names, available in databases such as PubChem [46] or the ECHA REACH [27], provides transparency for integration with other databases.

A second issue regarding the ATTAC step of "Transparency" relates to the aggregation of data. Aggregation often results from the need to present data in a condensed way, highlighting, for example, the main patterns in the data in a publication. Nevertheless, data presented in aggregated form (e.g., when the summary data in a publication are not accompanied by a raw database) require specific attention because they can constrain later analyses. Aggregation can occur in both the entries (e.g., individuals pooled into classes or subpopulations) and in the attributes (e.g., pollutant concentrations reported for broader chemical classes rather than as compound-specific concentrations; conditions at sampling locations classified into broader classes). Aggregation of entries limits the resolution of subsequent analysis, as important biological characteristics of individuals (e.g., sex, age, size, or differences in ecological preferences such as diet and habitat) may be lost when several individuals are pooled. However, individual variability can have important consequences on population dynamics, species conservation, and ecotoxicology [4,16,60]. Likewise, aggregation of attributes limits some analyses to be performed when information is lost, especially if data is presented in coarser classes. For example, it might not be possible to attribute toxic effects to specific pollutants, but rather to a coarse grouping. This might limit the ability



to combine these data with data that are presented at a finer resolution. Particularly, in the context of ecotoxicological analyses, it is important to clarify in the “Transparency” step which compounds and what number of species/individuals were investigated.

Finally, the “Transparency” step in ATTAC also relates to the details on the field and laboratory designs which are critical to trace the origin and processing history of data, as a primary condition to properly conduct further meta-analyses (e.g., accounting for the correct sample sizes when comparing and weighting among studies). For instance, transparency regarding field design is critical to determine whether different entries were collected at the same location and time, and to which attributes (e.g., sex, size) they relate. When working with wildlife, fieldwork can often be unpredictable and involve opportunistic sampling that can deviate from a strict sampling design. This is in response to the logistic and practical difficulties that one might encounter when sampling wildlife species in their natural environment, including stochastic influences such as weather and animal behavior. Additionally, transparency regarding laboratory design is important to understand whether all entries underwent the same analysis or if there were improvements and subsequent deviations in the protocol for some of the samples. In this context, applying standardized methods and protocols greatly improves transparency. Nevertheless, even standardized protocols can change over time and regulatory guidelines can be updated (e.g., [67]), so that deviations might need to be made for specific samples. Hence, an accurate description of the protocols used remains a crucial need even if a general description of the protocol has been published. Further, a specific methodological transparency relates to the batch structure of the analysis (including quality control and standards). Sometimes, samples are analysed for only some compounds as a cost-reduction method, and some samples are lost during laboratory processing e.g., due to failure or accidental loss. As a result, the final number of analysed samples might deviate from the initial study design. In other cases, data are the outputs of a statistical analysis or a simulated model. Here too, transparency regarding the statistical methods or the modeling techniques is needed to ensure the reproducibility of such derived, estimated or simulated data (a detailed description of such a situation is beyond the scope of the current paper).

#### *Transparency guidelines for the prime mover*

Guidelines regarding “Transparency” from the perspective of the prime mover mainly relate to a clear and complete description of all aspects of the data; specifically, the nature of all entries and attributes, their origin and processing history (as related to field and laboratory study design), and clear identification of aggregated data (Fig. 3). In most cases, exhaustive and comprehensive data descriptions (meta-data) with the view of full transparency and data reproducibility are too detailed for a publication and likely exceed manuscript word limits. Documenting meta-data in supplementary materials, or a file directly linked to the dataset (e.g., in the same repository as a “readme” file) are suitable options. For clarity, a tabular representation might present a structured format which ensures that similar information is given for each entry and attribute (e.g., using table headings: entry/attribute name; entry/attribute description; origin; processing history). Developing clear, exhaustive meta-data takes additional research time. Nevertheless, a transparent data description increases trust which eventually contributes to greater citations of one’s work, and also makes it possible to later understand and reuse one’s own data [78]. Additionally, transparent reporting of methods and results is a key tenant for the harmonized and consistent evaluation of reliability and relevance of toxicology and ecotoxicology data in regulatory contexts, hence, using methodologies such as Klimisch [47], Criteria for Reporting and Evaluating ecotoxicity Data (CRED, [56]), and Critical Appraisal Tools (CATs, [28]) can also improve the potential for data reuse.

#### *Transparency guidelines for the re-user*

It is critical to have a clear, defined aim for the intended systematic literature review and subsequent meta-analysis, and consequently a clear view of which data are of priority concern for an intended meta-analysis, which data are potentially valuable add-ons if available, and which data fall outside the scope [22,29]. Given that wildlife ecotoxicological research projects and teams focus on different aspects and present their data in varying, heterogeneous formats, such defined aims and identified data requirements would allow one to quickly categorize individual studies as relevant and reliable for the intended meta-analysis or not. Additionally, methodologies such as Klimisch [47], Criteria for Reporting and Evaluating ecotoxicity Data (CRED, [56]), and Critical Appraisal Tools (CATs, [28]) can be of assistance to the prime mover, as these methodologies include guidance on assessments of relevance, reliability, and adequacy of data, and are frequently employed in regulatory schemes for the evaluation of toxicology and ecotoxicology data. An initial scoping study consisting of a quick literature scan with a limited number of search terms, and a homogenization of a limited set of individual studies, might be needed in order to obtain a first view of the quantity and quality of potentially available data. Systematic literature search and meta-analysis aims can then be re-evaluated and adjusted [22], and the reasons for any potential adjustments transparently justified and documented.

Given the different approaches to reporting data below reporting limits, a specific advice when one aims to reuse data containing pollutant measurements is to carefully check how values below reporting limit were encoded in the dataset. Additionally, careful attention should be paid to the occurrence of zeros (and their effective meaning) as well as any number that would occur with particularly high frequency (the latter might indicate that substitution was used).

When transparency regarding the available databases is opaque, only limited rescue tools are available. The main option is to contact authors (see strategy in section “Access”), as well as check related papers of the author which might use the same dataset or similar methods and protocols. In cases where sample identifiers are given (e.g., wildlife tracking tags), there might also be a possibility to find related data regarding the samples in online data repositories. Alternatively, if transparency cannot be resolved, a decision needs to be made regarding the subsequent use or elimination of the database in the intended meta-analysis. It is almost always a bad idea to make arbitrary assumptions about datasets with the risk to introduce artefactual bias. Nevertheless, it might be worthwhile (rather than completely removing the dataset) to conduct the meta-analysis with and without the opaque dataset, or to

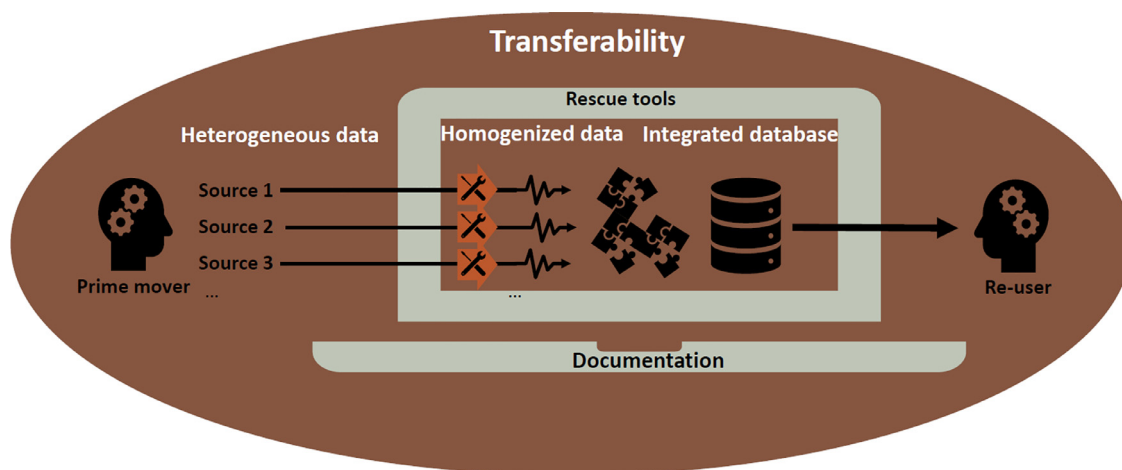


Fig. 4. Pathway of the harmonization and integration of heterogeneous data sources, supported by rescue tools when transparency is opaque.

use the opaque dataset for model validation, as long as the lack of transparency is clearly communicated and its potential influence acknowledged when interpreting the results.

#### Transferability of data

Once data are collected (“access”) and their meaning understood (“transparency”), the combination of data from different sources into a larger database suitable for a further meta-analysis can start. A critical step is then the homogenization and subsequent integration of different types and sources of data. Issues arise when data are highly heterogeneous and not transferable between studies. This is particularly the case in wildlife ecotoxicology where individual databases are frequently created by individual research teams which often use different protocols adapted to their specific target species and laboratory facilities (or simply because there are few standardized protocols available for wildlife species). Consequently, a decisive action is the conversion of heterogeneous data into a homogenized format. An additional issue might arise when such a conversion requires further information (such as the lipid or water content of the sampled tissue) which are not available within a specific database.

#### Transferability guidelines for the prime mover

The harmonization and integration of different databases from individual studies into one database are usually done by the re-user, who is merging several studies to provide a new one (Fig. 4). Nevertheless, the value of an individual database, and its potential for reuse, increase by careful planning ahead for transferability. One strategy to facilitate the potential for data transferability across studies is to apply standardized methodologies which are more likely to result in a database that is presented in a similar format (e.g., units) and with similar information content (e.g., properties) used by other researchers. Such standardized methodologies, however, are often absent in wildlife ecotoxicology. As an alternative strategy, it is worth thinking about how data might be combined with other, similar (existing and future) databases. This can be done by reviewing the type of databases already available or being created in a given research field and thinking of ways how a new database will be compatible with such existing databases. Additionally (and in absence of any comparable databases), one should carefully consider reporting concentrations in a recognized international system of units, and consider measuring additional physiological attributes (e.g., water, lipid contents) that enable future conversions even if not in the immediate focus of the individual study. A third strategy, when planning for data transferability, is to consider existing methodologies (e.g., Klimisch: [47]; CRED: [56]; or the EFSA CATs: [28]) that are specifically tailored to evaluate the reliability (and adequacy) of toxicological data in a harmonized way across available data (studies). Although these methodologies are principally aimed at the data re-user, they provide insight into what criteria (e.g., experimental details) should be recorded and reported when making data available. Following such methodologies can thus also assist the data prime mover in experimental design, even when not utilising standardised methodologies. By considering how the data may be evaluated for potential reuse by those outside the immediate field, the likelihood of providing more standardised information increases. Consequently, planning for transferability might make it more appealing to reuse data in the first place (including by the prime mover their self) and so to increase citations of the corresponding work [78]. The idea is not necessarily to go to complete standardization as different research teams might have different aims and objectives, as well as different facilities and financial opportunities. However, the idea in the “transparency” step of ATTAC is to collect and provide the data in a harmonised way with the view of potential future integration across studies. This specifically means a complete and transparent reporting of protocols and data.

#### Transferability guidelines for the re-user

A necessary step for dealing with heterogeneous data is the homogenization of attributes. In wildlife ecotoxicology studies, this often involves the homogenization of measurement units and the measurement bases of pollutant concentrations. Most conversions



are straightforward (several free, online calculators can assist). Nevertheless, further information might occasionally be needed. For example, given the lipophilic nature of persistent organic pollutants, their concentrations in animal tissues are often reported on a lipid-normalized basis to allow comparison among tissues. However, when an individual study on POP bioaccumulation does not report lipid contents, this might limit the homogenization and integration of the data from this specific study. A potential rescue tool for such databases is the borrowing of information regarding lipid content from studies on the same or phylogenetically closely related species. Such borrowing nevertheless increases uncertainty as the resulting lipid content data will not be directly applicable due to natural, biological variation. In this context, note that databases that present lipid content data in aggregated form (e.g., the average lipid content across all sampled individuals) also increase uncertainty due to inter-individual variability. A second rescue tool, to which both prime movers and re-users can contribute, is the construction of physiological databases, which will eventually also contribute to solving such homogenization issues. For example, the development of a lipid content database by Muñoz et al. [59] allowed comparison among multiple studies and tissues of sea turtles, even when lipid content information for individual studies was missing. In addition to pollutant concentrations, data on toxic effects in wildlife are often heterogeneous and measured against a variety of endpoints for a range of timeframes. It is often exceedingly difficult to quantitatively connect and thus homogenize such different endpoints, or to translate effects measured across different timescales without the need for detailed modeling of underlying mechanisms and dynamics, such as done by using quantitative Adverse Outcome Pathway (qAOP) models for linking effects [33], or Toxicokinetic and Toxicodynamic (TKTD) models for dynamically linking initial exposure concentrations to final individual effects [11,20]. Such models could offer advanced rescue tools in the future but are currently often not focussed on wildlife species, although some databases with wide taxonomic coverage, including wildlife, are already available to support such modeling [1,54].

A special case in the harmonization of ecotoxicological data relates to cases where compounds are reported as groups (e.g., ΣDDTs) in some databases and as individual compounds in others. Here, a re-user needs to apply transparent decision-making regarding the integration of such databases. A first rescue tool is to sum data from studies where individual compounds (o,p'-DDE, p,p'-DDE, etc..) were reported in order to make up the group sum, although this leads to some loss of information. Additionally, often, not all compounds contributing to the group sum are reported in studies (e.g., some report all DDT compounds, while others only report the most common compounds), and thus the sum might not be equivalent across studies [61]. A second rescue tool is to estimate how much each separate compound makes up a group sum. For example, technical grade DDT consists of 77.1 % p,p'-DDT and 19.9 % o,p'-DDE [86]. One can then estimate the total sum based on the percentage of individual compounds that are still missing (e.g., if four DDT compounds were measured, the sum would need to be multiplied by a certain fraction to obtain 100 % across all DDT compounds). Nevertheless, in the environment, different compounds contributing to a group might behave differently and their relative contribution can change over time (e.g., in the environment, p,p'-DDE is more persistent than others, thus the relative ratio of p,p'-DDT/p,p'-DDE changes over time). Hence, this second option should be applied with extreme caution. A final option is to exclude data from the individual studies where compounds are grouped since the resolution is too low for the intended meta-analysis. This, however, reduces the final size of the integrated database. In contrast, the excluded data could further be used for model validation.

During the homogenization and integration process, the origin of each individual database should be clearly labelled so that it can be transparently traced in the final, integrated database (Fig. 4). Likewise, the history of applied rescue tools and manipulations to achieve data transferability should be properly described and documented in supporting documents. In this context, recording such manipulations, including their documentation and decisions, in a modeling notebook, much like a traditional laboratory notebook, allows one to reproducibly trace back the different steps taken during the harmonization and integration of databases [6,32,79]. The origin and history of the new integrated databases will then be fully transparent (as discussed for individual databases in section “Transparency”) and will allow for future reuse of the integrated database, which itself now becomes a prime mover-developed database.

### Add-ons

Add-ons refer to any information in addition to the recorded pollutant levels or toxic effects (and critical auxiliary data required to enable transferability of these ecotoxicological data) that is nevertheless relevant to analyse the observed ecotoxicological patterns. Some of these add-ons are frequently collected during sampling because it is *a priori* clear that these are add-ons that provide a basic characterization of the biological variability among samples which should be accounted for as a covariable in an ecotoxicological analysis. Such standard add-ons can often be collected with minimal extra effort. For example, data regarding biological characteristics of sampled individuals such as sex, size and body condition (e.g., approximated as a size-weight quantitative relationship or a qualitative description of health), as well as basic data regarding time, date, and location (and potentially experimental or field conditions) of capture are of high importance as they can influence the exposure, uptake and effects of pollutants within individuals. For example, females can offload some of their pollution burden via maternal transfer, while males do not have this elimination pathway, hence resulting in differences in bioaccumulation patterns between sexes [59]. Additionally, physiological rates such as metabolic, breathing, and heart rates are known to differ between sexes, which can influence toxicokinetics [12,53]. Such life history data are often fundamental to ecologically and ecotoxicologically sensible data analyses and are therefore frequently collected in field and lab datasheets.

During the process of homogenizing and integrating data from various sources into an integrated database for further meta-analysis, careful attention should be paid to recording data regarding the basic characteristics of the different studies being gathered (meta-add-ons, Fig. 5). Particularly, meta-add-ons such as an identifier to trace the original bibliographic reference, the sampling date (which can often be quite distant from the publication date), and the sample size are likely to be included as confounding factors and to enable weighting of individual datasets in meta-analyses.

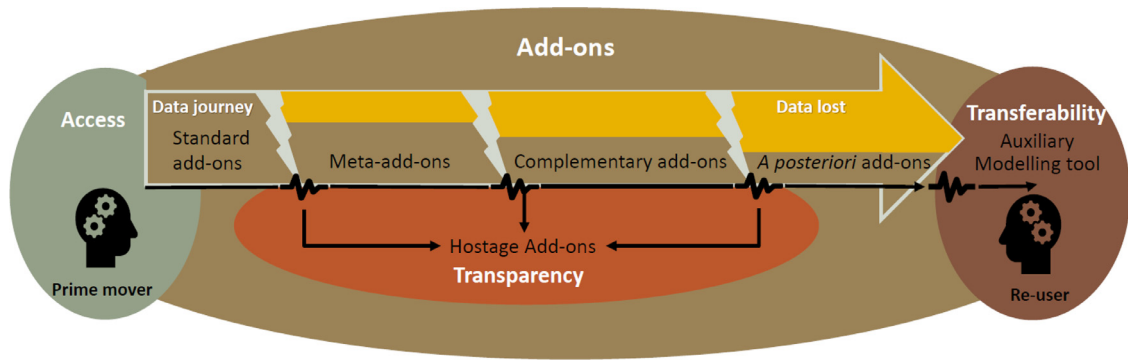


Fig. 5. Relations between different types of add-ons, relevant to support ecotoxicological meta-analyses.

Other add-ons are not typically collected but can nevertheless be of importance in ecotoxicological analyses. Such complementary add-ons are generally beyond the basic characterization of the samples and the study aims, and frequently require a substantial extra effort to collect them (which can include logistic, technical, and financial resources) or data treatment to derive them. Examples of such complementary add-ons include age, reproductive status, and in some cases sex of wildlife species. For instance, sexual maturity and reproductive status can usually only be collected with more invasive techniques (e.g., in reptiles via palpation, endoscopy, imaging analysis, or analysis of blood chemistry, [48]). Similarly, age is often a parameter that is difficult to estimate non-invasively in wildlife during fieldwork [84], and thus is a parameter that might require a model-based derivation from other measured quantities [26]. Another category of add-ons gathers those whose ecotoxicological relevance only becomes apparent in later research (*a posteriori* add-ons). For example, given their lipophilicity, the lipid content of analysed tissues is often reported in studies regarding the bioaccumulation of organic pollutants (and is critical information to ensure transferability, see section “Transferability”). Nevertheless, there is now also evidence that some organic pollutants can associate with specific protein fractions of tissues [72,82]. A relevant *a posteriori* add-on would be to analyse the protein content of any remaining stored tissue to supplement the analysis. Similarly, detailed descriptions of health indicators such as injuries, visible signs of illness, or the presence of parasites can be relevant in ecotoxicological analyses and taken into account as a complementary add-on if existing evidence suggests their interaction with toxicological processes, or as a *a posteriori* add-on when such evidence only emerges later on.

*Add-ons guidelines for the prime mover*

Despite the high relevance of add-ons, their associated data does not always filter through until the stage at which research outputs are published and made available. For example, some results are presented in pooled formats, some add-ons fall outside the scope of a study or analysis, or the raw data on add-ons is not FAIRly stored. Approaches and guidelines discussed above concerning “Access”, “Transparency” and “Transferability” also apply to making the most of valuable add-on data. A particular issue regarding “Transparency” relates to cases where add-ons, although jointly collected with ecotoxicological samples from the same individuals, are subsequently analysed and published in isolation and, in some cases, held hostage (hostage add-ons) in hopes of future analysis and publication. For example, only a subset of individuals might be analysed for diet composition (an important pollution exposure route), and results published on their own for their ecological relevance. Meanwhile, a different, partially overlapping subset might be analysed for pollution levels, with data aggregated by location (while diet data were e.g., aggregated by sex). When such a series of publications using the same samples do not transparently trace back to the unique, individual samples, valuable add-on information is lost. Consequently, including unique identifiers and presenting the data at the individual level is critical, also for add-ons. Additionally, transparent communication of collected add-ons (including add-ons not made available yet) should complement meta-data regarding ecotoxicological datasets. For hostage add-ons, a release plan, including clear and transparent timeframes and conditions should be given.

The selection of add-ons to collect in conjunction with ecotoxicological sampling should be carefully considered because gathering additional add-on data could cost extra effort and also impact the studied species. For example, individuals could experience additional stress and impacts if they need to be kept longer sedated or restrained and if multiple (low) invasive procedures are combined for the collection of add-ons. Standardized protocols for ecotoxicology focus predominantly on the toxic endpoint itself and are typically targeted towards domestic animals or those used in standard toxicity testing rather than wildlife (e.g., [63,64,68]). Nevertheless, for several wildlife species, standardised protocols regarding field monitoring for conservation purposes are available which list standard add-ons that are relevant and feasible to collect for the species of interest (e.g., [35]). Additionally, based on current developments in the broader research field any “low-hanging fruit”-type complementary add-ons can likely be identified during a general literature review regarding the ecology and life history of the species of interest. It is difficult to forecast which further parameters might become relevant as *a posteriori* add-ons. Documenting long-term identifiers such as tag numbers or microchip implants, as well as collecting photographic records [51,80], might enable data to be linked to previous and future research on the same individuals. Likewise, retaining any remaining sample materials not used during the analysis in tissue databanks opens up the possibility for *a posteriori* add-ons to be collected [44,45].

### Add-ons guidelines for the re-user

During the process of data collection, the data re-user is faced with the same issue as the prime data mover, namely: which add-ons to include? Standard add-ons are ideally all collected given that they provide a basic characterization of the biological variability among samples that is likely to interact with ecotoxicological patterns. Additionally, meta-add-ons should be collected to characterize the variation and bias due to the included studies. Just as standard add-ons characterize the biological variability, meta-add-ons capture the study variation (and potentially bias e.g., due to different methodologies, or temporal or spatial heterogeneity in the homogenized, integrated database). Both types of add-ons are likely to be covariates in any subsequent meta-analysis. The inclusion of complementary and *a posteriori* add-ons, meanwhile, is largely dependent on the aims and objectives of the meta-analysis.

While data regarding add-ons should ideally be collected and provided together with the ecotoxicological data, one can try to obtain missing add-ons by searching for identifiers or tag numbers of individuals in biological databases and related publications of the same author team. Alternatively, one can try to make the best out of the available data. For instance, to overcome the absence of clear information on sex and life stage, it is possible to categorise data into classes including one or multiple specific classes for unknowns (e.g., males of unknown life stage or juveniles of unknown sex, [59,83]), although this can render subsequent analyses fragmentary and less information rich. More advanced modeling techniques, e.g., hierarchical modeling, can also provide a solution when data are scattered [5].

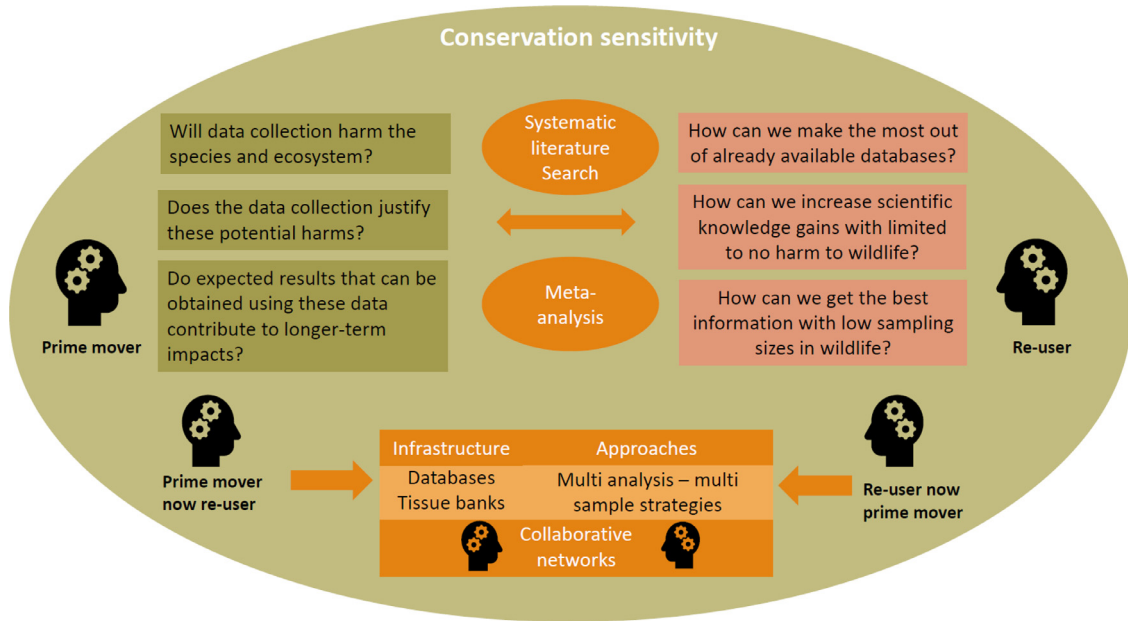
Just as with the prime ecotoxicological data, add-ons can be presented in different formats and units which might require homogenization. Organism size, for instance, is often recorded using different metrics (e.g., snout-vent length vs. total length of crocodiles, [36]) with sometimes different measures pertaining to different life stages (e.g., crown-rump length as a length measurement for embryos, [7]). As a data re-user, this might require the development of auxiliary statistical models. Additionally, the collection of relevant data to support such auxiliary models can go beyond the list of publications that were retained during the primary systematic literature search for ecotoxicological data. Nevertheless, the development of auxiliary models and their supporting databases can be highly useful for the broader community (e.g., tools to enable different morphometric measurements to be converted are applicable across several biology-related fields). Hence, auxiliary models present a valuable research output on their own, whose development should not be underappreciated if we are to achieve integrative data analyses to advance wildlife ecology and ecotoxicology.

### Conservation sensitivity

Environmental pollution is ranked among the top five drivers of the current sixth mass extinction by conservation biologists [18]. The contribution of environmental pollution to this ongoing defaunation, the process in which animal populations or species become extinct from ecological communities at global, local, or functional levels [24], has also raised concerns at the international level [42]. Chemical pollution as a contributor to defaunation was already advertised 60 years ago for DDT and related pesticides by the book “Silent Spring” [19] at the origin of the modern environmental awareness. As an applied science that largely emerged from concerns regarding the impacts of environmental pollution [76], a main task of wildlife ecotoxicology is to provide the scientific knowledge needed to address this major societal challenge, including scientific knowledge that can be translated into regulatory decision-making. Nevertheless, given the ongoing mass extinction, it is increasingly critical for researchers to 1) be mindful of any potential harmful impacts that their activities might have on study organisms and ecosystems (i.e., Will data collection harm the species and ecosystem?); 2) balance the scientific gain from their research with these impacts (i.e., Does the data collection justify these potential harms?); and 3) critically evaluate whether research activities contribute to the broader task that ecotoxicology has in addressing ongoing challenges (i.e., Do expected results that can be obtained using these data contribute to longer-term impacts? Fig. 6). The reuse of data aligns with these three critical criteria as it contributes to reducing the need for repeated sampling of wildlife by allowing different studies to build upon each other’s data and results. Data reuse also aligns with the clear societal requirement for replacement, refinement, and reduction (3R) in the use of vertebrate toxicity testing [52]. Against this backdrop, systematic literature searches and meta-analyses are highly valuable tools that can make wise use of ecotoxicological data as conservation-sensitive materials. These tools can ensure long-term impacts. For example, an implicit benefit of these tools is that they contribute to the longevity of data by integrating them and extending results beyond what can be gained from individual studies. Additionally, these tools provide opportunities for large scientific knowledge gains with limited to no harm to wildlife.

### Conservation sensitivity guidelines for the prime mover

Data collection in wildlife ecotoxicology can often contain an element of unpredictability because of variable field conditions, stochastic behavior of individuals, and also the increasing rareness of some species. Nevertheless, a thought-through study design is critical to ensure sufficient replication to produce statistically meaningful results while accounting for conservation sensitivity. Statistical power analyses, as well as looking for examples of sample sizes utilized in previous studies, can help determine more suitable sampling sizes. For example, to characterize levels of organic pollutants in sea turtle nests, previous studies have typically used between one and five eggs per nest [61]. This low number of replicates (including a recommendation for just one replicate) has been justified because, in turtles, all eggs within a clutch undergo vitellogenesis at the same time [13]. Nevertheless, with a sample size of one, it is impossible to estimate the between-egg variability, and thus to compare among levels in different samples. Hence, a higher number of replicates would be needed for comparative statistical analyses, or data regarding variability needs to be obtained from other sources and integrated into new analyses. An additional peculiarity with data regarding pollutant concentrations from environmental samples such as wildlife is that the quantitative information content can be reduced when concentrations in those samples are below the reporting limit of the analytical equipment or laboratory (resulting in censored data). A slightly higher number of replicates would thus be helpful to statistically infer meaningful results when a high number of censored data are expected. A power analysis that accounts for censored data can be applied to estimate sample sizes in such cases [50,71]. An additional strategy for study



**Fig. 6.** Key questions and collaborative tools bringing prime movers and re-users of wildlife ecology and ecotoxicology closer together, towards achieving conservation sensitive research.

design in wildlife ecotoxicology is to include a buffer to account for the unpredictability of wildlife sampling. For example, the study design could include an extra site, which would only be used when sampling at another site is impeded. In this case, the study design should clearly state when sample sizes are achieved to prevent that additional (backup) samples are unnecessarily collected.

*Conservation sensitivity guidelines for both the prime mover and re-user*

A final strategy for wildlife ecotoxicological study design, by data prime movers, is to make the most out of already available databases. Specifically, during analysis, results obtained from the newly collected data can be compared with data extracted from already existing studies (the data prime mover thus *ipso facto* becoming a data re-user). This can, for instance, be achieved via visual comparison of plotted data, or explicitly integrating the prior knowledge with new data via the use of techniques such as Bayesian inference [34]. Data re-users can promote the availability of this “study design” option to data prime movers by making unified, homogenized FAIR databases. The number of sample replication is often quite limited for wildlife animals, hence, gathering different sources of data would both save resources and facilitate work within research partnerships.

To ensure the wise use of conservation-sensitive materials, there is an increasingly pressing ethical duty towards decreasing biodiversity for scientific openness, not only regarding data, but also the underlying samples. For instance, contributing any additional samples to tissue databanks or museum collections helps to build an archive for further (potentially multidisciplinary) use of the samples [62], and avoids valuable materials being thrown away. Likewise, multidisciplinary collaboration can ensure that samples are used to their maximum scientific potential by conducting several analyses on the same sample, rather than “single analysis - single sample” strategies. Infrastructures for data and sample sharing, as well as scientific networks among data prime movers and re-users are critical and should be valued as essential scientific outputs.

**Conclusion**

Data generated by individual research studies strongly deserve to be valued as World Heritage whose long-term preservation needs to be ensured. Indeed, it is crucial to enable the next generation of researchers to tackle urgent societal and environmental challenges using the best possible archive of scientific knowledge and data. This is particularly the case for wildlife ecotoxicology, considering that the number of sample replication within individual studies is often quite limited. Additional data collection is faced with severe ethical, financial, and practical limitations, and needs to be placed within the context of the current global biodiversity extinction crisis. Consequently, the use of already existing databases, by means of their integration and harmonization, can significantly improve the already established results either adding useful complements to the understanding or elucidating innovative findings. This is not a trivial task, which might indeed remain an ascent of Mount Everest if the research community does not take this duty to heart. Among others, this means that the practice of providing well-documented, useful, and preserved data should not be considered as yet another task, conducted as after-thought when publishing research, but should become a standard and valued scientific practice. The ATTAC workflow brings together guidelines supporting both the data prime movers as well as the re-users of these data in five key steps along the chain of collecting, homogenizing, and integrating data for subsequent meta-analyses. As such the ATTAC workflow

promotes an open and collaborative wildlife ecotoxicology capable of guiding the scientific community to the top of Mount Everest in this applied field. In this perspective, the ATTAC workflow could become an essential gateway to provide scientific support for regulations and management actions in protecting and conserving wildlife.

### Ethics statements

There is no ethical statement relating to this work.

### Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

### CRediT authorship contribution statement

**Cynthia C. Muñoz:** Conceptualization, Methodology, Writing – review & editing. **Sandrine Charles:** Conceptualization, Writing – review & editing, Supervision. **Emily A. McVey:** Writing – review & editing. **Peter Vermeiren:** Conceptualization, Methodology, Writing – review & editing, Supervision.

### Data Availability

No data was used for the research described in the article.

### Acknowledgments

This work benefitted from an EU Cost action grant [Action nr. CA18221] for a short-term scientific mission to CCM, and an Erasmus+ staff exchange grant to PV, which allowed for a research visit to the University of Lyon, and subsequent discussions among CCM, PV and SC leading to the conceptualization of this paper.

### References

- [1] Add-my-Pet (2021) Add-my-pet database: online database of DEB parameters, implied properties and referenced underlying data. [bio.vu.nl/thb/deb/deblab/add\\_my\\_pet/](https://bio.vu.nl/thb/deb/deblab/add_my_pet/) (accessed 8 November 2022)
- [2] D. Alderson, L. Clarke, D. Schillereff, E. Shuttleworth, Navigating the academic ladder as an early career researcher in earth and environmental sciences, *Earth Surf. Process Landforms* (2022) 1–12.
- [3] L. Allen, A. O'Connell, V. Kiermer, How can we ensure visibility and diversity in research contributions? How the Contributor Role Taxonomy (CRediT) is helping the shift from authorship to contributorship, *Learn Publ.* 32 (2019) 71–74.
- [4] S. Augustine, B. Gagnaire, C. Adam-Guillermín, S.A.L.M. Kooijman, Effects of uranium on the metabolism of zebrafish, *Danio rerio*, *Aquat. Toxicol.* 118–119 (2012) 9–26.
- [5] M. Authier, E. Rouby, K. Macleod, Estimating cetacean bycatch from non-representative samples (I): A simulation study with regularized multilevel regression and post-stratification, *Front. Mar. Sci.* 8 (2021) 1–19.
- [6] D. Ayllón, S.F. Railsback, C. Gallagher, J. Augusiak, H. Baveco, U. Berger, S. Charles, R. Martin, A. Focks, N. Galic, C. Liu, E.E. van Loon, J. Nabe-Nielsen, C. Piuu, J.G. Polhill, T.G. Preuss, V. Radchuk, A. Schmolke, J. Stadnicka-Michalak, P. Thorbek, V. Grimm, Keeping modelling notebooks with TRACE: good for you and good for environmental research and management support, *Environ. Model. Softw.* 136 (2021).
- [7] W.G. Bardsley, R.A. Ackerman, N.A. Bukhari, D.C. Deeming, M.W. Ferguson, Mathematical models for growth in alligator (*Alligator mississippiensis*) embryos developing at different incubation temperatures, *J. Anat.* 187 (1995) 181–190.
- [8] T.M. Beardsley, The biologist's burden, *Bioscience* 60 (2010) 483.
- [9] E.S. Bernhardt, E.J. Rosi, M.O. Gessner, Synthetic chemicals as agents of global change, *Front. Ecol. Environ.* 15 (2017) 84–90.
- [10] P. Berny, Pesticides and the intoxication of wild animals, *J. Vet. Pharmacol. Ther.* 30 (2007) 93–100.
- [11] E. Billoir, H. Delhay, B. Clément, M.L. Delignette-Muller, S. Charles, Bayesian modelling of daphnid responses to time-varying cadmium exposure in laboratory aquatic microcosms, *Ecotoxicol. Environ. Saf.* 74 (2011) 693–702.
- [12] M.J. Binnington, F. Wania, Clarifying relationships between persistent organic pollutant concentrations and age in wildlife biomonitoring: individuals, cross-sections, and the roles of lifespan and sex, *Environ. Toxicol. Chem.* 33 (2014) 1415–1426.
- [13] C.A. Bishop, Water quality and toxicology, in: *Reptile Ecology and Conservation: A Handbook of Techniques*, 1st ed., Oxford University Press, Oxford, 2016, p. 462.
- [14] T. Blacquière, G. Smagghe, C.A.M. Van Gestel, V. Mommaerts, Neonicotinoids in bees: a review on concentrations, side-effects and risk assessment, *Ecotoxicology* 21 (2012) 973–992.
- [15] M. Bloemers, A. Montesanti, The fair funding model: providing a framework for research funders to drive the transition toward fair data management and stewardship practices, *Data Intell.* 2 (2020) 171–180.
- [16] D.I. Bolnick, P. Amarasekare, M.S. Araújo, R. Bürger, J.M. Levine, M. Novak, V.H.W. Rudolf, S.J. Schreiber, M.C. Urban, D.A. Vasseur, Why intraspecific trait variation matters in community ecology, *Trends Ecol. Evol.* 26 (2011) 183–192.
- [17] A. Brand, L. Allen, M. Altman, M. Hlava, J. Scott, Beyond authorship: Attribution, contribution, collaboration, and credit, *Learn Publ.* 28 (2015) 151–155.
- [18] P. Cafaro, Three ways to think about the sixth mass extinction, *Biol. Conserv.* 192 (2015) 387–393.
- [19] R. Carson, *Silent Spring*, 3rd (ed), Hamish Hamilton, London, 1964.
- [20] S. Charles, C. Lopes, S.D. Wu, A. Ratier, V. Baudrot, G. Multari, Taking full advantage of modelling to better assess environmental risk due to xenobiotics — the all-in-one facility MOSAIC, *Environ. Sci. Pollut. Res.* 29 (2022) 29244–29257.
- [21] K.A. Connors, A. Beasley, M.G. Barron, S.E. Belanger, M. Bonnell, J.L. Brill, D. de Zwart, A. Kienzler, J. Krailler, R. Otter, J.L. Phillips, M.R. Embry, Creation of a curated aquatic toxicology database: EnviroTox, *Environ. Toxicol. Chem.* 38 (2019) 1062–1073.
- [22] C. Cooper, A. Booth, J. Varley-Campbell, N. Britten, R. Garside, Defining the process to literature searching in systematic reviews: a literature review of guidance and supporting studies, *BMC Med. Res. Methodol.* 18 (2018) 1–14.



- [23] R. Dietz, R.J. Letcher, J.P. Desforges, I. Eulaers, C. Sonne, S. Wilson, E. Andersen-Ranberg, N. Basu, B.D. Barst, J.O. Bustnes, J. Bytingsvik, T.M. Ciesielski, P.E. Drevnick, G.W. Gabrielsen, A. Haarr, K. Hylland, B.M. Jenssen, M. Levin, M.A. McKinney, R.D. Nørregaard, K.E. Pedersen, J. Provencher, B. Styrisshave, S. Tartu, J. Aars, J.T. Ackerman, A. Rosing-Asvid, R. Barrett, A. Bignert, E.W. Born, M. Branigan, B. Braune, C.E. Bryan, M. Dam, C.A. Eagles-Smith, M. Evans, T.J. Evans, A.T. Fisk, M. Gamber, K. Gustavson, C.A. Hartman, B. Helander, M.P. Herzog, P.F. Hoekstra, M. Houde, K. Hoydal, A.K. Jackson, J. Kucklick, E. Lie, L. Loseto, M.L. Mallory, C. Miljeteig, A. Mosbech, D.C.G. Muir, S.T. Nielsen, E. Peacock, S. Pedro, S.H. Peterson, A. Polder, F.F. Rigét, P. Roach, H. Saunes, M.H.S. Sinding, J.U. Skaare, J. Søndergaard, G. Stenson, G. Stern, G. Treu, S.S. Schuur, G. Víkingsson, Current state of knowledge on biological effects from contaminants on arctic wildlife and fish, *Sci. Total Environ.* 696 (2019) 133792.
- [24] R. Dirzo, H.S. Young, M. Galetti, G. Ceballos, N.J.B. Isaac, B. Collen, Defaunation in the anthropocene, *Science* 345 (2014) 401–406.
- [25] C.S. Duke, J.H. Porter, The ethics of data sharing and reuse in biology, *Bioscience* 63 (2013) 483–489.
- [26] M.J. Eaton, W.A. Link, Estimating age from recapture data: integrating incremental growth measures with ancillary data to infer age-at-length, *Ecol. Appl.* 21 (2011) 2487–2497.
- [27] ECHA REACH - Registration, evaluation, authorization and restriction of chemicals regulation. <https://echa.europa.eu/information-on-chemicals/registered-substances> (accessed 8 November 2022)
- [28] EFSA, Tools for critically appraising different study designs, systematic review and literature searches, EFSA Support Publ. 12 (2017) 1–65.
- [29] EFSA GOF, Food E, Authority S, Submission of scientific peer-reviewed open literature for the approval of pesticide active substances under regulation (EC) No 1107/2009, *EFSA J.* 9 (2011) 1–49.
- [30] EPA (1996) Ecological effects test guidelines OPPTS 850.1730: Fish Bioconcentration Factor (BCF). 25.
- [31] European Commission (2019) Regulation (EU) 2019/1381 of the European Parliament and of the Council of 20 June 2019 on the transparency and sustainability of the EU risk assessment in the food chain and amending Regulations (EC) No 178/2002, (EC) No 1829/2003, (EC) No 1831/2003, (EC) No 2065/2003, (EC) No 1935/2004, (EC) No 1331/2008, (EC) No 1107/2009, (EU) 2015/2283 and Directive 2001/18/EC. 28.
- [32] L. Figueiredo, C. Scherer, J.S. Cabral, A simple kit to use computational notebooks for more openness, reproducibility, and productivity in research, *PLoS Comput. Biol.* 18 (2022) 1–12.
- [33] V.E. Forbes, N. Galic, Next-generation ecological risk assessment: predicting risk from molecular initiation to ecosystem service delivery, *Environ. Int.* 91 (2016) 215–219.
- [34] A. Gelman, J.B. Carlin, H.S. Stern, D.B. Dunson, A. Vehtari, D.B. Rubin, in: *Bayesian Data Analysis*, 3rd (ed), Taylor & Francis, 2014, p. 675.
- [35] Government of Western Australia Dept. Biodiversity C and A (2013) Standard operating procedures (SOPs). <https://www.dpaw.wa.gov.au/plants-and-animals/monitoring/standards-and-protocols/99-standard-operating-procedures> (accessed 8 November 2022)
- [36] G. Grigg, D. Kirschner, in: *Biology and Evolution of Crocodylians*, 1st (ed), CSIRO Publishing, 2015, p. 649.
- [37] P. Groth, H. Cousijn, T. Clark, C. Goble, Fair data reuse – the path through data citation, *Data Intell.* 2 (2020) 78–86.
- [38] D. Haelewaters, T.A. Hofmann, A.L. Romero-Olivares, Ten simple rules for Global North researchers to stop perpetuating helicopter research in the Global South, *PLoS Comput. Biol.* 17 (2021) 1–8.
- [39] M. Hahnel, D. Valen, How to (Easily) extend the fairness of existing repositories, *Data Intell.* 2 (2020) 192–198.
- [40] D.R. Helsel, in: *Statistics for Censored Environmental Data Using Minitab and R*, Wiley Publishing, 2012, p. 325.
- [41] S.P. Hozo, B. Djulbegovic, I. Hozo, Estimating the mean and variance from the median, range, and the size of a sample, *BMC Med. Res. Methodol.* 5 (2005) 1–10.
- [42] IPBES, BrondizioES, Settele J., Diaz S., Ngo H.T. (2019) Global assessment report on biodiversity and ecosystem services of the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services. 39.
- [43] N. Jeliakova, M.D. Apostolova, C. Andreoli, F. Barone, A. Barrick, C. Battistelli, C. Bossa, A. Botea-Petcu, A. Châtel, I. De Angelis, M. Dusinska, N. El Yamani, D. Gheorghe, A. Giusti, P. Gómez-Fernández, R. Grafström, M. Gromelski, N.R. Jacobsen, V. Jeliakov, K.A. Jensen, N. Kochev, P. Kohonen, N. Manier, E. Marussens, A. Mech, J.M. Navas, V. Paskaleva, A. Precupas, T. Puzyn, K. Rasmussen, P. Ritchie, I.R. Llopis, E. Rundén-Pran, R. Sandu, N. Shandilya, S. Tanasescu, A. Haase, P. Nymark, Towards FAIR nanosafety data, *Nat. Nanotechnol.* 16 (2021) 644–654.
- [44] Jürgens M.D. (2015) Biomonitoring of wild fish to assess chemical pollution in English rivers – an application of a fish tissue archive. 338.
- [45] J.M. Keller, R.S. Pugh, P.R. Becker, in: *Biological and Environmental Monitoring and Archival of Sea Turtle Tissues (BEMAST): Rationale, Protocols, and Initial Collections of Banked Sea Turtle Tissues*, NIST Pubs, 2014, p. 76.
- [46] S. Kim, J. Chen, T. Cheng, A. Gindulyte, J. He, S. He, Q. Li, B.A. Shoemaker, P.A. Thiessen, B. Yu, L. Zaslavsky, J. Zhang, E.E. Bolton, PubChem in 2021: new data content and improved web interfaces, *Nucleic. Acids. Res.* 49 (2019) D1388–D1395.
- [47] H.J. Klimisch, M. Andrae, U. Tillmann, A systematic approach for evaluating the quality of experimental toxicological and ecotoxicological data, *Regul. Toxicol. Pharmacol.* 25 (1997) 1–5.
- [48] G. Köhler, Reproduction, in: C.K. Dodd (Ed.), *Reptile Ecology and Conservation: A Handbook of Techniques*, 1st ed., Oxford University Press, Oxford, 2016, pp. 87–96.
- [49] H.R. Köhler, R. Triebkorn, Wildlife ecotoxicology of pesticides: can we track effects to the population level and beyond? *Science* 341 (2013) 759–765.
- [50] G. Kon Kam King, P. Veber, S. Charles, M.L. Delignette-Muller, MOSAIC\_SSD: a new web tool for species sensitivity distribution to include censored data by maximum likelihood, *Environ. Toxicol. Chem.* 33 (2014) 2133–2139.
- [51] H.S. Kühl, T. Burghardt, Animal biometrics: quantifying and detecting phenotypic appearance, *Trends Ecol. Evol.* 28 (2013) 432–441.
- [52] A. Lillcrap, S. Belanger, N. Burden, PasquierD Du, M.R. Embry, M. Halder, M.A. Lampi, L. Lee, T. Norberg-King, B.A. Rattner, K. Schirmer, P. Thomas, Alternative approaches to vertebrate ecotoxicity tests in the 21st century: a review of developments over the last 2 decades and current status, *Environ. Toxicol. Chem.* 35 (2016) 2637–2646.
- [53] C.P. Madenjian, Sex effect on polychlorinated biphenyl concentrations in fish: a synthesis, *Fish Fish* 12 (2011) 451–460.
- [54] G.M. Marques, S. Augustine, K. Lika, L. Pecquerie, T. Domingos, S.A.L.M. Kooijman, The AmP project: comparing species on the basis of dynamic energy budget parameters, *PLoS Comput. Biol.* 14 (2018) 1–23.
- [55] Memorandum (2022) Memorandum on ensuring free, immediate, and equitable access to federally funded research. 27.
- [56] C.T.A. Moermond, R. Kase, M. Korkaric, M. Ågerstrand, CRED: criteria for reporting and evaluating ecotoxicity data, *Environ. Toxicol. Chem.* 35 (2016) 1297–1309.
- [57] D. Moher, A. Liberati, J. Tetzlaff, D.G. Altman, D. Altman, G. Antes, D. Atkins, V. Barbour, N. Barrowman, J.A. Berlin, J. Clark, M. Clarke, D. Cook, R. D'Amico, J.J. Deeks, P.J. Devereaux, K. Dickersin, M. Egger, E. Ernst, P.C. Göttsche, J. Grimshaw, G. Guyatt, J. Higgins, J.P.A. Ioannidis, J. Kleijnen, T. Lang, N. Magrini, D. McNamee, L. Moja, C. Mulrow, M. Napoli, A. Oxman, B. Pham, D. Rennie, M. Sampson, K.F. Schulz, P.G. Shekelle, D. Tovey, P. Tugwell, Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement, *BMJ* 399 (2009) b2535.
- [58] S.G. Moreno, A.J. Sutton, J.R. Thompson, A.E. ADES, K.R. Abrams, N.J. Cooper, A generalized weighting regression-derived meta-analysis estimator robust to small-study effects and heterogeneity, *Stat. Med.* 31 (2012) 1407–1417.
- [59] C. Muñoz, A.J. Hendriks, A.M.J. Ragas, P. Vermeiren, Internal and maternal distribution of persistent organic pollutants in sea turtle tissues: A meta-analysis, *Environ. Sci. Technol.* 55 (2021) 10012–10024.
- [60] C.C. Muñoz, T. Saito, P. Vermeiren, Cohort structure and individual resource specialization in loggerhead turtles, long-lived marine species with ontogenetic migrations, *Mar. Ecol. Prog. Ser.* 671 (2021) 175–190.
- [61] C.C. Muñoz, P. Vermeiren, Maternal transfer of persistent organic pollutants to sea turtle eggs: a meta-analysis addressing knowledge and data gaps toward an improved synthesis of research outputs, *Environ. Toxicol. Chem.* 39 (2020) 9–29.
- [62] T. Odsjö, J. Sondell, Eggshell thinning of osprey (*Pandion haliaetus*) breeding in Sweden and its significance for egg breakage and breeding outcome, *Sci. Total Environ.* 470–471 (2014) 1023–1029.
- [63] OECD (1984a) Test No. 205: Avian dietary toxicity test. OECD Guidel Test Chem:1–10.
- [64] OECD (1984b) Test No. 206: Avian reproduction test. OECD Guidel Test Chem:1–12.
- [65] OECD (2012) Test no. 305: bioaccumulation in fish: aqueous and dietary exposure. 10.1787/9789264185296-en (accessed 8 November 2022)

- [66] OECD (2016) Test No. 243: *Lymnaea stagnalis* reproduction test. OECD Guidel Test Chem:1–13.
- [67] OECD (2019) Test No. 203: fish, acute toxicity testing. <https://www.oecd-ilibrary.org/content/publication/9789264069961-en> (accessed 8 November 2022)
- [68] OECD (2022) Test No. 470: mammalian erythrocyte pig-a gene mutation assay. OECD Guidel Test Chem:1–37.
- [69] OECD Good Laboratory Practice. <https://www.oecd.org/chemicalsafety/testing/good-laboratory-practiceglp.htm> (accessed 11 November 2022)
- [70] M.E. Ortiz-Santaliestra, J.P. Maia, A. Egea-Serrano, C.A. Brühl, I. Lopes, Biological relevance of the magnitude of effects (considering mortality, sub-lethal and reproductive effects) observed in studies with amphibians and reptiles in view of population level impacts on amphibians and reptiles, EFSA Support Publ 14 (2017).
- [71] K. Owzar, Z. Li, N. Cox, S.H. Jung, Power and sample size calculations for SNP association studies with censored time-to-event outcomes, Genet. Epidemiol. 36 (2012) 538–548.
- [72] Z. Pedrero Zayas, L. Ouerdane, S. Mounicou, R. Lobinski, M. Monperrus, D. Amouroux, Hemoglobin as a major binding protein for methylmercury in white-sided dolphin liver, Anal. Bioanal. Chem. 406 (2014) 1121–1129.
- [73] A.M. Petersen, M. Riccaboni, H.E. Stanley, F. Pammolli, Persistence and uncertainty in the academic career, Proc. Natl. Acad. Sci. U. S. A. 109 (2012) 5213–5218.
- [74] N. Pettorelli, J. Barlow, M.A. Nuñez, R. Rader, P.A. Stephens, T. Pinfield, E. Newton, How international journals can support ecology from the Global South, J. Appl. Ecol. 58 (2021) 4–8.
- [75] A. Ratier, S. Charles, Accumulation-depuration data collection in support of toxicokinetic modelling, Sci. Data 9 (2022) 1–7.
- [76] B.A. Rattner, History of wildlife toxicology, Ecotoxicology 18 (2009) 773–783.
- [77] J. Rockström, W. Steffen, K. Noone, Å. Persson, F.S. Chapin, E.F. Lambin, T.M. Lenton, M. Scheffer, C. Folke, H.J. Schellnhuber, B. Nykvist, C.A. de Wit, T. Hughes, S. van der Leeuw, H. Rodhe, S. Sörlin, P.K. Snyder, R. Costanza, U. Svedin, M. Falkenmark, L. Karlberg, R.W. Corell, V.J. Fabry, J. Hansen, B. Walker, D. Liverman, K. Richardson, P. Crutzen, J.A. Foley, A safe operating space for humanity Identifying, Nature 461 (2009) 472–475.
- [78] G.K. Sandve, A. Nekrutenko, J. Taylor, E. Hovig, Ten simple rules for reproducible computational research, PLoS Comput. Biol. 9 (2013) 1–4.
- [79] S. Schnell, Ten simple rules for a computational biologist's laboratory notebook, PLoS Comput. Biol. 11 (2015) 5–9.
- [80] C.W. Speed, M.G. Meekan, C.J.A. Bradshaw, Spot the match - wildlife photo-identification using information theory, Front. Zool. 4 (2007) 1–11.
- [81] S. Stall, L. Yarmey, J. Cutcher-Gershenfeld, B. Hanson, K. Lernhart, B. Nosek, M. Parsons, E. Robinson, L. Wyborn, Make all scientific data FAIR, Nature 570 (2019) 27–29.
- [82] J. Sundberg, B. Ersson, B. Lönnerdal, A. Oskarsson, Protein binding of mercury in milk and plasma from mice and man - A comparison between methylmercury and inorganic mercury, Toxicology 137 (1999) 169–184.
- [83] F. Thirion, M. Tellez, R. Van Damme, L. Bervoets, Trace element concentrations in caudal scutes from *Crocodylus moreletii* and *Crocodylus acutus* in Belize in relation to biological variables and land use, Ecotoxicol. Environ. Saf. 231 (2022) 113164.
- [84] A.D. Tucker, Skeletochronology of post-occipital osteoderms for age validation of Australian freshwater crocodiles (*Crocodylus johnstoni*), Mar. Freshw. Res. 48 (1997) 343–351.
- [85] P. Vermeiren, P. Reichert, W. Graf, P. Leitner, A. Schmidt-Kloiber, N. Schuwirth, Confronting existing knowledge on ecological preferences of stream macroinvertebrates with independent biomonitoring data using a Bayesian multi-species distribution model, Freshw. Sci. 40 (2021) 202–220.
- [86] WHO (1979) DDT and its derivatives.