

*In this doc, we will find some questions I received from colleagues that aimed to fill some lines in the INVACOST database. I have now organized the doc around different topics (headlines in green).*

*I identified in red the questions concerning columns that have now been changed in or removed from the current version of the database. Nonetheless, I did not suppress anything (i.e. questions were asked regarding the old version of the database) as the answers might always help. Also, I sometimes added an update for the answer to provide most current and relevant information.*

*My answers are provided in this colour ;) Please note that I have left the text unaltered from original exchanges. So, sorry for familiarities and typos :)*

**Content**

Benefits..... 2

Cost estimates..... 6

Cost standardisation ..... 10

Further analyses ..... 2

General..... 10

Improvement of database content and structure ..... 12

Methodology..... 12

Nature of costs ..... 13

References..... 14

Spatial coverage ..... 16

Taxonomy..... 17

Temporal coverage..... 18

Updates ..... 18

## Analyses

1. In order to select the data to work with, I suppose we would have also to pay attention to at least these two variables (Christophe?):

- IMPLEMENTATION: observed vs potential. Maybe use only observed items?

Totally agree with you! We strongly think that working with the more 'realistic' data should be the most optimal strategy (except for a study that would like to compare both categories of costs). And for this purpose, focusing on 'observed' costs seem the first stage to select the data (we advocate for this in our first paper on INVACOST)

- METHOD RELIABILITY - high or low. Maybe we could do separate analyses including or excluding low reliability references in order to assess if low reliable data follow the same pattern?

Here again, I totally agree. Maybe we can show results with 'All' vs 'High'-data based estimates? BUT, it is important to consider that deeming with the reliability of a methodology used to estimate a cost data mainly based on the nature of the publication (e.g. peer-reviewed or not) may be too simplistic. Indeed, irrespective of the type of document analyzed (peer-reviewed or not), cost estimates are sometimes provided without any citation, or derived from untraceable sources (such as unreachable citations, 'unpublished data' or 'personal communication'). Besides, some reachable works cited to support the cost estimates provided seem to not contain the figures exhibited. One way to improve the assessment of this descriptor should be to use a similar approach than those previously developed by Bradshaw et al. (2016, Nat. Commun. 7, 12986 doi: 10.1038/ncomms12986; see section 'Determining cost estimate reproducibility' of this article for details). Their previous study evidenced that assuming the reproducibility of published methods should not rely only on the nature of the materials, but also with the reproducibility of the methodology. However, directly applying this approach to the diverse methods reviewed in the frame of INVACOST (Bradshaw and colleagues' study focused on a single taxonomic class) was complicated and came with a series of constraints when dealing with too large data involving several collaborators. Indeed, the authors recognized that the attribution of reproducibility was a qualitative procedure specific to each monetary value, although such attribution procedure was clear in most cases. For the purpose of applying such a procedure in the most convenient way for the large array of methodologies found in studies recorded in INVACOST, consistent and objective criteria should be clearly defined. With the colleagues from the INVACOST team, we are currently working on defining such a clarification mainly based on the estimation method(s) provided by the authors. Here is another discussion we can have during the workshop. But, for the moment, our categorization 'high' vs 'low' seems to convince most of you. Thus, we can still work using that for now ;)

2. Would be possible to have a set of explanatory data already prepared? So, everyone would use the same data for e.g., country area, population, human population density, GDP, human footprint index, geographical coordinates, etc. The data I used was taken from Google, therefore, they are not very trustful (they were just for exercising).

It would be great. And I think that it is one purpose of the first afternoon when we'll discuss on the potential 'blocks'.

Update: This task is part of several ongoing projects.

3. Thank you for the update. I have been testing the functions on some subsets from various descriptive projects and they are working well and seem easy to adjust. The 'Usage notes' suggest that raw costs or annualised estimates should not be simply summed. I realise that the expanded version has much higher total annualised estimates because it also accounts for the

estimation period, however this causes a disparity between the main database and the expanded one. For example, if one element of analysis presents the cumulative costs (using expanded data), and another element the annualised cost totals across e.g. species groups (using original data), there will potentially be a large difference in what is presented in a single paper.

In other words - should we now always work with the 'expanded' version of the data when quantifying annualised costs, even for basic totals which do not consider temporal elements?

Much of the current analyses has simply used the annualised totals from the original database.

We have now re-worked the way to collect and analyse cost estimates within INVACOST. The aim was to facilitate the process and allow the investigation of different questions regarding the targeted objective(s). Actually, you can work on both (i) quantifying annualised costs using the 'expanded' version of the data and deriving an annual average of cost and (ii) calculating basic totals not considering temporal trends.

For the objective (i), you can use one of the two strategies implemented in our package, mainly depending on the amount of data available.

For the objective (ii), you'll have probably to create a column representing the total cost over a given or specified period. Total costs may be obtained by considering the 'Raw cost estimates' (i.e. exactly how cost estimates appear in the materials) and the 'Time range' columns; for periodic costs, you can consider them as such as total costs, and for year costs, you'll need to consider a duration time (= the number of years between the starting and ending ones that can be changed depending on the 'scenario' tested - for instance, you can test different ending years for costs with 'unspecified' ending year provided if these costs are tagged as potentially ongoing in the 'Occurrence' column) with which you multiply the costs provided for obtaining the total costs.

Therefore, you can choose one or the other strategy depending on which are the targeted aims, focusing on temporal trend of costs, presenting the total cumulative costs, and/or having an average annual value. I hope that it's clear enough, tell me if some parts are unclear.

4. This Invacost package is nice and all for the accumulation of costs, but for the "descriptive papers" I am working on right now, I would like to know which columns we should now refer to. I get that the new columns added are important for the accumulation cost, but I do not see how the description of pure e.g. "taxa/country specific costs" was affected by the update?

We have re-worked the database for both analysis of cumulative costs and descriptive papers.

Considering the 'annualised cost estimate' column from the 1st version of INVACOST does not seem relevant any longer for descriptive studies that simply sum the costs.

Providing realistic total costs for a species/an area etc. rather requires to consider the raw cost estimates (i.e. exactly how cost estimates appear in the papers; this column is now provided in addition to 'Annualised cost estimates' that we renamed 'Cost per year' because while some cost estimates can be expressed yearly, other costs occur only once and should thus not be labelled as annual to avoid a misleading idea of repetition). Considering mean cost per species, per area, etc. also requires to use the raw cost estimates provided as a starting point. The point is that sometimes costs are given for a period of one year while these costs can occur over a period different than a single year. In this case, total costs can be obtained by repeating the cost/year over the given period. One solution we propose is to create a column 'total cost' for each study by combining information from different columns.

We have not created this column as the goal of INVACOST is to collect and provide raw info, not to make assumptions for further analyses. Thus, you can obtain a specific duration time (= the number of years between the 'Probable\_starting\_year' and the 'Probable\_ending\_year' - to which we add "1" to avoid null values for costs occurring in only one year) by which you could multiply the cost per year. This strategy implies to provide a year where info was missing (unspecified). For instance, one

can opt for providing the same year in both starting year and ending year columns when info was provided for only one of them.

5. According to the notes, Pimentel et al 2000 has the animal papers in the database, but not the plants. As I started looking through the database, I realized that it's a bit more complicated than this. There are quite a few Pimental papers in the database (see attached spreadsheet, which is just his papers), and when you sort by species, you see that quite a few species have the same costs associated with different papers. And this doesn't even account for any papers that might have been found in a Pimental paper, traced back to and cited from the original source, and then subsequently cited from a different Pimental paper. It also doesn't account for cases in which numbers are updated in subsequent papers, so any earlier costs are incorporated into more recent estimates. Thus, we risk having double, or even triple+, entries of costs. It seems like this issue might warrant a systemic review and clean up of the data, so I have done nothing further with this paper.

I totally agree with the fact that multiples entries are likely existing in the DB. The most probable and frequent case is probably when recent costs incorporate earlier ones. Nevertheless, we think that the purpose of the DB is to be a repository that records/provides all the existing costs/situations found in the literature, regardless the potential connections that might exist between them. At the stage of collating data in InvaCost, we must just ensure that obvious/true duplicates (i.e., same cost figures from (non-)identical sources) are not recorded, or removed during the final checking steps (we'll do it when putting together the different databases from the *invacosters*<sup>^^</sup>). This large host of data is important to cover a wide array of situations in order to allow different analyses and working objectives (e.g. for instance, a study focusing on a specific - past - period could be impaired if we initially decided to keep only the most recent cost entries in the general DB).

Each study based on the DB generally uses a subset, and needs to review and clean up the data considered before analyses. I think that this step may be done following different strategies/step, which depends on the amount of data in the subset and the type of 'duplicates'. Indeed, duplicates may be

- (i) temporal: recent cost entries that incorporate older ones;
- (ii) geographic: cost entries that cover a large spatial scale, and that may include some costs provided at smaller scales;
- (iii) taxonomic: cost entries covering several species that may include several costs also provided separately for single species (e.g., 'rodentia' order with 'diverse/unspecified' species, while there are other costs for 'Rattus spp.' and 'Mus spp.' in the same country than the 'rodentia' costs);
- (iv) methodological: same cost entries provided as different ones by (at least two) different studies -> this situation is hard to manage, so that we must acknowledge this possibility and assume that they are (normally) not numerous as we filtered out most of them when collecting the data;

Therefore, following the objectives and subset of the study, there are different options to deal with these duplicates, which include for instance 1/ working at specific spatial scales below (and including) the 'country' level to avoid double counting in some 'continental' or 'regional' costs (while this strategy could not be suitable for species with few costs provided for those spatial scales that will be removed), or 2/ identifying for each species (through a table) cost data that are recorded in the same geographic region/country with the same cost values.

Therefore for the paper you are working on, I agree with you. But given my point above, I think that the best way to proceed is to ensure that the papers he cited are already collected, or at least analysed by other colleagues (see the xlsx file with the assigned papers). I assume that these are already collected, and there would be no need to process them. But in this way, we ensure that no info will be missing. Tell me if I was unclear of if you need some help on that.

Happy to discuss that with everyone to find the best way to proceed in each project ;)

6. I just started playing with the data as I thought I'd do the estimation for the Medit and Bivalves paper again. However, various entries have unspecified starting and ending dates, which is screwing up the cost expansion. So I wanted to ask you guys if you thought about a way we should deal with this issue. I imagine that the number of entries with this issue will vary according to the paper project, but for the Medit paper there is a considerable number of these entries (due to the spanish and french data I'd assume). Replacing the unspecified ending year with 2017 is one way, but if it is the correct one... not sure.

Elena: I think Christophe will respond you with a good way for dealing with that, because InvaCost (the English version) has a very high proportion of entries with such information lacking (in comparison with the Non English version).

Chris: I agree with Elena. It was the same situation than for InvaCost in which we have some unspecified in these columns. You have to make some choices (we did it for calculating cost estimate per year, for entries with this missing information), depending on your assumptions on the cost duration (we provided some indication in different columns to help in that way). Below are some information that will help you (you can find also crucial info in the package document provided by Boris, see section 'How do I know when costs occurred?'):

- create two sets of columns for the purpose of our analysis (we called them 'Probable\_starting\_year\_low\_margin' and 'Probable\_ending\_year\_low\_margin' as we considered very conservative cost duration

- based on duration of impacts indicated by the authors (see 'period-of-estimation' column) and 'publication year' of the materials, we assigned a single year to each column: publication year when no information was available for the duration, and publication +/- duration of costs when this info was provided in the 'period\_of\_estimation' column

- if you want to be less conservative, you can consider that some cost occur until a more recent year; in this case, you change the ending year column with this year, but only for costs that have been considered as 'potentially ongoing' in the 'occurrence' column.

### Benefits

7. "Benefit value(s)" column - my understanding is that this should be noted with "Yes" only if the authors also reported that the given IAS also produces some positive monetary values, and they reported such values in the paper? There is also a case where authors report cost of controlling/eradicating some species, and they also provide some positive values that such control efforts produce (e.g. increased employment), but I think this should not be considered as a benefit value of IAS, as it is not a monetary gain IAS produces, right?

Yes, we consider benefit values only when money was obtained from direct exploitation of IAS. We also considered 'avoiding costs' for some apparent benefit-like values, i.e. money that is not lost any more due to efficient management strategy. For instance, if a species A (impacting agriculture) was successfully controlled (control costs = 10,000\$) and that the net benefit is now 13,000\$, we use one line for control costs and another for avoiding costs (= 13,000\$). Alternatively, if this species A is used for trade activities that generates 14,000\$, we simply mentioned 'yes' in the column "Benefit value(s)" (only if the paper is mentioned in the DB as containing at least one cost estimate; if a paper has only benefit values, it is not mentioned in the DB).

8. I was a bit confused by the explanation in the Table in the Info & Description. mention of if any benefit value was found in the analysed material; *"this might cover different types of benefits (e.g. real benefits provided by direct use or exploitation of invasive species or benefits that are actually avoided costs) - the figure was not recorded or described as being out of the scope of INVACOST"*.

Yes, I understand. I think that we should be clearer for this column. Actually, we have put the second part of the sentence as sometimes the identification of clear benefits is not straightforward (as you already noticed). We wanted to highlight that we are aware that what we have considered as actual benefits might be only avoided costs sometimes. But, we did our best to collect all the infos on these avoided costs when finalizing the database. We'll correct this part of our table in the published version. Thanks a lot for your remark.

### Cost estimates

9. In the column with "Annualized cost estimates" (and in columns with minimum and maximum values), I presented standardized costs, expressed per year, by dividing costs from defined longer periods by the period duration (e.g. if some cost was 5 million \$ over 5 years, I noted in the column "Period of time" that the period duration was 5 years, but as a cost I noted 1 million \$). I hope that this is the correct way to do it. For values reported as periods, with min and max range endpoints, I also entered the median value of the range as the "Annualized cost estimate". Is it correct?

Simply perfect!

10. Authors are often not making it clear when they report periods what are the exact start and end of the reported period - e.g. period that was reported as 1992-1996 can be understood as lasting for either 4 or 5 years, depending on whether those two years are also included or not. This is however very important when one calculates "Annualized cost estimates". If this was not clear in the text, I usually went for conservative approach and included both start and end years in the period - this resulted in longer periods and thus in smaller cost estimates when "Annualized costs" are estimated (e.g. if a cost of 20 million \$ was reported for period 1992-1996, annualized cost will be either 4 or 5 million \$, depending on how you count the duration of the period).

Exactly what we did, so perfect again!

11. For the "Annualized cost value" column, I did what you said, I converted all reported values to annual values when it was possible, so for example if someone reported a cost that covers several years, I divided the cost by the number of years considered. For values with unspecified time range (which is a rare case), I left the cost values unmodified.

Ok!

12. In cases when there were both pooled data and data per species provided, I reported data per species, but in cases when all data were reported per same species or group or species, but they provided both pooled and very detailed data for different cost categories (e.g. price per gas, manpower, overheads, etc.), I went for pooled data. In some cases, they reported different values for costs for different years, even though all the other descriptors (species, locality) were the same, in that case I reported each of the values separately for each year.

Really wonderful!

13. I have a question regarding the annualized cost estimate - do we have to convert it to 2017 US\$ or leave it in its original currency? I know that you eventually convert it in the database, but I am not sure if you would like the raw data at this stage or the conversions.

Indeed, we first need the raw (but annualized) cost data in the dedicated column in the Subset\_INVACOST file. In a second step, each cost will have to be converted using indications and formula provided in the file 'INVACOST-Rationale&Methods.pdf' (see section Standardisation of cost data). So, you can attempt to convert the raw cost data you entered and provide the obtained estimates in new columns I have now created. Please note that there are two types of conversion: one based on the classic exchange-rate and the other one based on the Purchase Power Parity (everything is explained in the dedicated section). This will help us to see if procedures for converting cost data are sufficiently well explained in the doc.

14. There are some studies which have modelled the cost of managing (or containing) invasive species in a landscape. Like other modelling studies, these studies often use known parameters based on a small spatial scale to extrapolate the estimated cost for a larger spatio-temporal scale. While dealing with these studies, should we extract both information or only the extrapolated one?

Yes, we should extract both information (that will allow to further opt for one or the other info depending on the analyses targeted and the research questions).

15. The one other question I have is that I have a list of direct costs and a single "total cost" for each species, where the total cost includes both direct and indirect costs. Right now I've put all of them in as lines in the database (with direct costs separated by sector where possible), but technically that's repeating some of the costs - what would you suggest? I'm sure we'll end up sorting some of this out in person, too.

Actually, I think that we'll have to deal with potential double-counting similar to those you highlight in your message. Maybe you can add this info within the column 'details'? The other solution may be to consider only the total costs, then indicating in the column 'details' that these costs are detailed in a specific part of the paper (and mentioning in the column 'type of costs' all the direct and indirect costs that are concerned?). If you have collected all data, please proceed simply by considering a mention of 'double counting' in the column details before or after the other info you planned to add in this column.

16. How do we deal with costs where we do not have the information to generate an annualised estimate, only giving a single dollar value – e.g. "In 2004/05 dollar terms this equates to a total cost of approximately \$12.7 million

We also got these cases when collecting costs ☹ We mentioned the cost as provided in the dedicated column, and then 'unknown' in 'Period of estimation'. Probably, we should exclude these types of costs when carrying out statistics...

Update: Now, you must simply collect the raw cost provided and we'll be able to make different assumptions on the temporal extent of the costs.



17. I have costs that are X US\$ in total inferred from the last 10 years prior to the publishing year of that study, I can enter the probable starting year and ending year, the raw cost, but Occurrence is then "one time", and Time\_range="Year", right? This way I ensure that this cost X is not repeated 10 times when the database is expanded, right?

Ok for entering 'one time', but if you enter 'year' as a time range, this means that the cost will be repeated ten times when we expand the database. You must enter 'year' if the cost is provided as an annual value, which is expected to occur each year of the last 10 years. If the cost is provided as a total (i.e. when summing all the annual costs within these 10 years), you must enter 'period'. In this way, you ensure that the costs will be automatically divided per year, and that this is this annual cost that will be repeated over the ten years. I hope that it is clear enough :)

18. The paper on economic impact of Johnsongrass (*Sorghum halepense* [L.] Pers.) in cotton production in Turkey (Gunes2008) actually states that this is a native species in Turkey. So I guess we should not include it in Invacost?

Yes, we should not. Please indicate in the 'References\_to\_analyze' file that this paper was 'done' and leave a note on this point (the paper focuses on a native species).

19. I could not find the 2001 version of this pest risk assessment of *Ambrosia* spp (Karnkowski2001). I have a 1999 version that states that "In Poland, *Ambrosia* spp. occur on non-agricultural land, and rarely in field crops. They do not cause economic losses at present." So I guess we should not include it in Invacost?

Yes, you can proceed as proposed for the last point. Thanks!

20. The paper on economic impact of *Acacia* in Israel actually reports the annual mean willingness to pay (WTP) for containment or eradication of *A. saligna* was US\$8.41 and US\$8.83, respectively. This is a non-monetary method. So I guess we should not include it in Invacost?

Actually, we report all economic estimates irrespective of the method used. Especially, the WTP is a method that was used for several cost estimates already recorded in our database. You can therefore enter these cost information as 'containment/eradication' cost (as the type of costs) and mention that the method used is the mean willingness to pay in the 'Details' column.

21. I've found a couple new reference that have costs in them. They're yearly reports from the Florida Department of Environmental Protection, Aquatic plant division. I've found the report for 2 fiscal years, 2005-6 and 2006-7, and in each there's costs for the same several plants, and are presumably recurrent costs, though not particularly even between the years. Given how we're extrapolating out recurrent costs in our analysis, it would obviously be double counting to include both papers and extrapolate them forward. However, they're different enough from year to year, that I'm not sure how representative it is to just put in one. Have you encountered this before? What is the standard for dealing with these?

Thanks for this interesting question. Actually, this problem is no longer a problem (I hope) as our temporal approach allows to mark the starting/ending years of each cost record. Therefore, if we expand the database with these info known on the cost duration, I guess there will not be a problem. Indeed, both costs re for a single fiscal year (we can consider 2006 in both starting/ending years columns for the first cost, and 2007 in the same way for the second cost). Hence, when expanding these costs, each of them will account only once. The single situation in which it becomes a problem is when we want to consider different scenarios on the duration of costs; for instance, if we considered that each annual costs of the dataset could continue until now. In this case, we would mention 2020 as the ending year; and this cost would be accounted twice. However, we cannot exclude any data for the reasons that some analysis strategies may lead to double counting. It's why we suggest that the general database records all relevant costs, and that each project ensures that any duplicate occurs in its final dataset before proceeding analyses.



22. I'm working on gathering all my papers so I can start slowly working my way through them and adding to the database. One of the ones assigned to me is Pimentel et al. 2000. Environmental and Economic Costs of Nonindigenous Species in the United States. In the notes it says it's in database, but only animals, not plants, so it needs to be checked for plants. I can add the plant details, but in the database I can see, none of this paper is in it. While looking to double check this, I noticed that there are 2 entries from Pimentel's 2005 update, but nothing else from it. While I certainly don't think both papers should be in there, it seems like if we're going to do one paper, it should be the 2005 one. However, since both these papers are well-cited, it seems like you would have made some efforts to get all these costs from the original source, correct? It seems that since these papers are both reviews, there would be a big risk of double-counting unless you were really careful. Maybe I'm just confused and/or missing something, but please let me know how I should proceed with this.

You perfectly understand the situation. Most of the costs cited as a source by Pimentel were searched and costs were retrieved in these refs or the ones that were cited. We also thought that this kind of (biased) review should be avoided if we can have access to the original refs that provide the costs. Concerning plant data, I propose to proceed in the simplest way: you can analyse the Pimentel's paper for plant data (I agree that we should consider the 2005 one) and try to trace back the info in the original sources. Nonetheless, if Pimentel provided some 'original' costs, you can add them, although I think that we will not consider these data when carrying out robust data analyses.

23. The last paper I am adding has avoided costs for two scenarios (depending on climate change and management effort): *“Under moderate climate change (+1.5 8C by 2050) discounted net benefits (4% discount rate) for the whole period range from 19 to 582 million euros depending on the effectiveness of the policy impact. With stronger climate change (+2.4 8C by 2050) the total discounted savings over the full period are expected to range from 42 to 1063 million euros assuming a highly effective policy.”* Which values should I add? (1) pick one scenario (2) add both scenarios as a separate entry (3) average the costs somehow?

We opted to combine the cost figures from the different scenarios to represent the whole situation, as these costs are 'potential'. And we indicate all the details on the different scenarios in the 'Details' column (this will allow other users to proceed otherwise).

Therefore, I'd suggest taking the smallest cost (here, 19 million euros) as the minimum cost estimate and the higher cost (here, 1063 million euros) as the max. From there, you can use the median cost (of all the known values) as the main cost. You can just indicate the strategy used in the 'Details' columns too. I'm aware that there may be another 'good' strategy, but this is how we proceeded to avoid double counting in this situation, while ensuring to cover the global cost figures exhibited in the paper.

24. on this Pimentel 2000 paper. I'm trying to get it entered and finished so I can remove it from my to-do list. First, a heads up that I edited a couple numbers in the original database that appeared to be typos or errors in entering data from the original paper. Also, I'm not entirely done yet, but having gone through it, there are definitely some costs from it that are not in the database at all, so I'm going to work on adding those to my personal Invacost database. There are also some costs in the database that are cited as coming from an on-line press release, but are actually from this paper (but the numbers might have changed a bit from the press release to the published paper?). I'm wondering if I should go ahead and change those in the original database, or if you would prefer I handle this another way? Please let me know how to proceed, and I'll get this done as soon as possible!

Regarding your specific questions:

- when you find errors in original entries, please don't enter them as 'new' costs in your personal database; the best way to proceed in this case is to send an email to "Updates Invacost" <updates@invacost.fr> (and put me in copy) with the required modifications; you can just send an xls (or csv) file with the entries corrected (please just put the Cost ID to allow us to identify those entries

that need further correction); at this stage, we have already planned a first update of the original database that will include both new cost entries from the invacosters and corrections of mistakes that have been detected after the publication of InvaCost v0; these corrections (and thus the way to proceed abovementioned) are also applicable for costs that might have changed between the press release and the published paper

- for costs that have not been yet extracted from this paper, you can add them in your personal database.

### Cost standardisation

25. All the fields in the database have been quite clear to me, except one - 'Type of applicable year'.

It will be nice to have some clarification on it.

The type of applicable year refers to the type of information about the year of the original currency in which the cost estimate was provided. Take as an example a paper published in 2017, which stated 'The invasive plants cost US\$ 272 billions (2015 dollars) annually in this country during the period 2000-2010'. You must provide "2015" as 'applicable year' and "effective" as 'type of applicable year' because the authors explicitly stated (in red) which year was considered regarding the currency. If the case where such info ion (in red) was missing, you would have to mention "publication year" as 'type of applicable year' associated with "2017" in the column 'applicable year' (as you have no info on the year considered for the currency). This info is important as any currency fluctuates in response to inflation and changes in exchange rates.

26. The costs detailed in this paper are expressed in 2004/5 terms (at least some of the time – I have to double check to see if this is true all of the time), no matter what the time period they are from. In the 'Type of applicable year' column should I still label this as 'Effective'?

Yes, exactly! As the authors precise the currency terms without any doubt.

27. If there is no way to infer the type of applicable year, what do I put in that column?

Actually, the applicable year is the year used for standardizing costs to US\$ 2017. You must fill the cell with either 'effective' if the authors specified the info in their paper, or 'publication year' if you can't find any info within the text (in this case, you must use the publication year of the doc). Do you mean that in some cases you don't have the publication year of the doc analyzed?

28. Chris, will you later on do the translation to US\$ 2017 value?

Yes, just put the raw cost (in original currency) and the appropriate information on timing/duration of the costs, and I'll do the rest ;)

### General

29. Sometimes, it is difficult – even impossible – to fill info within a column. What can we do in these situations ?

Sometimes, you may have some columns for which the info to collect from the paper is simply not available or not clearly stated by the authors; in this case, do not hesitate to put 'unspecified' (or 'Diverse/unspecified' in taxonomy-related columns) following what we have mentioned in the table 2 of the doc 'INVACOST-Rationale&Methods.pdf'; just ensure that you did every effort to capture the (missing) if no ;)

30. Regarding the topics, I'm very interested to lead the urbanization paper (or contribute with). But I could not find any data about urban areas in the files. For instance, there are no data if the species occur in an urban area. Or specification of the city or urban area as an "impacted sector" or "location".

Unfortunately, it will be hard, even impossible, to identify directly costs data in urban areas. INVACOST was designed with in mind to provide the widest possible descriptive variables regarding time and energy available as well as potential needs of future users. Obviously, it was impossible to anticipate all the further aims that will be raised by this database. Nonetheless, we are confident that most of the topics proposed by you and other colleagues can be treated by combining information from different columns, and/or also by re-analysing papers recorded within INVACOST (in this case, our database appears relevant again, as it compiles a large set of existing papers that have worked on this topic). - Moreover, here is another opportunity to remind that INVACOST is updatable and should benefit from new inputs from users working on a specific topic, for which data at hand seem not sufficient to allow deep investigation. –

Regarding more specifically your case (urbanization paper), I think it should be not so difficult to find relevant refs. A possible way to do it could be to search and filter all the costs retrieved in cities by considering the column 'Location' in which you can identify some obvious cases of urban costs (maybe you can also consider the column 'details' that can provide sometimes useful info for this purpose). Performing this step might allow you to categorize the refs into three groups: those that are 'certainly' relevant for your topic, others that are 'certainly not' relevant, and the other refs for which you can't take any decision without going through the documents. Such a strategy may save both time and energy by reducing the host of refs that would need to be re-analysed for ensuring if the cost collected should (or not) be included in your analysis.

31. Presumably, if we incorporate all extracted costs to date, this would also have to follow the checking/validation protocol of the entries by the InvaCost team. Do you know how long this would take ordinarily for the papers currently collected by the invacosters, or when this would be done?

Great idea to include complementary data that can be of interest regarding the project you lead. The double checking/validation step can be relatively quickly, if cost data are well/clearly/fully entered someone is available to do it. However, our team is currently very busy. Therefore, it would be hard to help in cross-checking new cost entries now (we are planning for doing this work when all data will be collected by all invacosters, in order to avoid multiple efforts).

Nevertheless, there is a solution for including these data in your current project paper:

- you can collate the data from the targeted papers/invacosters' database, and proceed to double checking of data (if needed, with one of the co-authors), meaning that - ideally - you need to have 2 co-authors for each references; I can do the cost standardisation if I get a final dataset with all new entries

- in the paper, we should state that 'we used data from the original InvaCost DB + non-english dataset, and then added new cost data from complementary searches that led to xxx new cost entries, which were added to the final dataset of the paper and then sent to the InvaCost team (updates@invacost.fr) for further consideration/integration in the global public database' (in that way, you can also correct things and present the dataset used in your paper as a supplementary material); this means that the team will perform the final validation at a later stage, for instance for

the next planned update of the DB; this also means that you can transparently assess that the data were not yet double-checked, which would allow to save time and energy for the data collation as only a single colleague can process each ref. I can help to write this section obviously.

### Improvement of database content and structure

32. I have been actively extracting data for the InvaCost, and it occurred to me that a structure of the database could be slightly optimized (in my opinion), which would also make the recorded cost estimates a bit more transparent. Maybe I am overthinking all this, but knowing how much work the plant team still has ahead of us, I am trying to find ways how to get things done with the lowest possible amount of errors as fast as possible. Have you thought about that?

Future changes are expected and welcome to the database. However, if we want to efficiently advance current works & projects, maybe we should consider that all these relevant and interesting proposals will be integrated (to the whole database) at a later stage. The issue might be that everyone has to repeat the work several times. I received some proposals to correct and upgrade the database by colleagues working on specific topics. What I proposed them is (i) to do these changes in their specific projects/papers, (ii) dedicate a specific paragraph in the methods section to state that/why/which changes have been made in the original database to get a specific subset, and (iii) that these changes were sent to updates@invacost.fr (as indicated in our accepted descriptor paper) for further consideration in the public global database. This is the case when errors or obvious duplicates are identified and removed, or when novel data are entered. I think that this solution allows to each user to save precious time in his/her work while contributing to the global improvement of the database.

### Methodology

33. "Method reliability" column - I am not sure if the criteria I was applying for this column is a proper one. If the authors did some estimation/calculation/extrapolation themselves, or cited others who did it, but in either case if they did not explain clearly how they came up with the value (in a clear enough manner so others could repeat the whole procedure) I was marking it as "No", as a non-repeatable value. This is I think as it should be.

Yes!

Update: Please refer to the criteria presented in the "Info&Description-Updated version" doc in the dropbox as well as in the doc explaining our package.

34. However, if they reported some investment (e.g. cost of eradication in some area) but didn't provide clear data about it (e.g. who funded it, who and what was exactly funded, etc.), e.g. if they only cited briefly some other paper as a source, I was also marking it as "No", as non-repeatable - I am not sure if this is OK?

It's okay.

Update: Please refer to the criteria presented in the "Info&Description-Updated version" doc in the dropbox as well as in the doc explaining our package.

35. On one hand, nobody can check and confirm the data cited about the eradication or research investments if there is insufficient information provided, but then again, this is not based on some estimation method, it is just investment record, so one can also consider it as "Yes"? My first idea was that such cases should be considered as non-repeatable, but while reading the data paper I started thinking that I might be wrong.

Your thinking is same than mine. Better is to be conservative. I should write it better in the data paper. Nevertheless, we might consider one exception for this case: for peer-reviewed papers as well as official reports that just report cost estimates not based on estimation method, we should consider it as 'reproducible'. In all cases, we can manage it later by simply using 'filters' in the DB.

Update: Please refer to the criteria presented in the "Info&Description-Updated version" doc in the dropbox as well as in the doc explaining our package.

### Nature of costs

36. In some papers, they report that control/eradication investment contributed to certain economic value that was saved (e.g. ecosystem services, agriculture production, etc.) - something that would be destroyed by IAS if such control effort was not implemented (e.g. "It has been estimated that if sea lamprey was not controlled, the loss of fishing opportunities and indirect economic impacts could be greater than \$500 million annually"). I was also including such costs as "Extrapolation"/"Potential" costs - on one hand, that is something that did not happen and won't it happen, because there was control effort and such loss was prevented, but on the other hand that is a nice indication of the potential of IAS to cause economic cost, so it seemed appropriate to enter such economic benefits saved/prevented through control as estimates of potential damage some species can produce. What do you think?

Really nice. You explain better my thinking above about the 'avoiding costs'. For me, what you explain here perfectly correspond to this situation in which we attributed these potential cost estimates to such category 'avoiding costs'.

37. **Observed vs. potential costs:** I sometimes struggle to distinguish between those two. For example, I read the following "We estimate hogweed prevalence at circa 0.1 hectare per district. Given 323 rural districts, yields an expectation of 0.323 km<sup>2</sup> inhabited by hogweed, whose removal would cost over € 5.6 million. However, these control efforts are not being undertaken, it must be further assumed that less than 10 % of this infested area is being dealt with (author's estimate)." Then later in the text these € 5.6 million listed as "annual costs incurred by giant hogweed infestation in Germany." The authors classify it as an incurred --> observed cost, despite previously mentioning that only 10% of this estimated amount is actually being spent. I have classified it as "potential" although I still have doubts because the estimate is based on the species' current distribution. What do you say?

The point you raised here is exactly what we have identified as one of the ways of further improvement of INVACOST. I cced Anne-Charlotte to this message as we started together reflecting on this point since several weeks (potential vs observed costs, considering the distribution of the invasive species, the nature (realized or not) as well as the temporality (past/current or predicted) of the quantified impact in the study area). Re-working info from this column would have required a lot of work and time, substantially delaying the potential papers from the workshop. We think that we can use the current classification for the first papers, and propose this improvement in few months, with other inputs to the database (as INVACOST is updatable). From there, I totally agree with the choice you did: classify this cost as 'potential' because the control efforts were not undertaken.

38. **estimated vs. extrapolated costs:** In the Info& Description, you mention that extrapolation relies on computational modeling. Is it always the case? For example, if it's a simple area-based upscaling without any modeling involved, would it be estimation or extrapolation?

No, it is not always the case, surely. It is even likely that a large part of reported/estimated costs comes from spatial up-scaling (for instance, costs reported at a 'national' scale are probably not systematically the result of estimations from all the regions of the country, but often derive from estimations made for a one or some areas within this country) but in most cases the authors do not explicitly state that. Thus, we have decided - to limit subjective bias - to put 'extrapolation' only when it is clearly indicated in the paper (for instance, via formula or calculations over a long time period). Here again, we work on refining this descriptor for further update ;)

39. Please, could you give some advice? First, the author estimates an annual cost of USD 127–291 mill. in lost ecosystem services. Second, the author provides a cost of USD 284–447 ha<sup>-1</sup> of land currently infested by the invader. Third, the author estimates a total cost of USD 7400 ha<sup>-1</sup> for eradicating the invader and restoring native riparian communities throughout the region. In other words, the first case is the “real” cost, the second is the "real" cost per ha, and the third is the potential cost (if someone decides to eradicate the species and restore the areas). I would say, I must insert the first as “observed” in the column “Implementation” and the third as “potential”. I would skip the second because it's the same as the first, but per ha. Am I right?

You are perfectly right!! I couldn't have been clearer!

40. I have two papers which report avoided damage costs using cost-benefit analyses. I interpret these as benefit values, and if I understand correctly we do not include those in Invacost? I have included excerpts from the paper here below. What is your advice?

These costs should not be considered as actual benefits because they are not the results of the 'exploitation' of the invasive species. We decided to consider them as 'avoided costs' (for the type), and mark them as 'potential' costs (for their implementation). Indeed, they represent what should have been lost if nothing was done. In addition, you can use the range of costs provided in the minimum-maximum costs columns, and take the median as the main costs

41. Where the costs are for eradication, what is the appropriate “implemented sector”? Is it the sector where most of the costs of the invader would be if it weren't eradicated, or should it be authority/stakeholder costs because they are the ones paying for the eradication?

In these cases where we don't have exact information, we put 'Diverse/unspecified' - as usual - when the costs are explicitly related to 'damage'. When this is about management actions - like in your case with the eradication costs - we enter 'authority/stakeholders' for the reason you mention above. We are aware that this column should be further refined, and we hope that the work currently made by Anna Turbelin and Emma Hudgins will allow to get more precise information on these impacted sectors;)

## References

42. In cases when I found reported costs in some paper for which another paper is cited as a source, I would first check whether I can find and access that cited paper directly. If I couldn't, I would still cite its name, authors and the publication year in columns "Reference title", "Authors" and "Publication year" (I would also provide full citation in the "Details" column), and I would include the full citation of the paper where I found the value and the citation in the "Previous materials" column.

Exactly!

43. If I could access the cited paper in full, I would just enter that paper there, without any mention of the paper where I originally found it, since it was not needed anymore. Is that correct?

In fact, we proceed in the same way than in case of no full access to the paper. We systematically put the (citing) reference in which we found the citation in column 'Previous materials'. It is just for full traceability if the paper could not be attached to a unique numerical code and repository (= 2 first columns).

44. I am not completely sure about the "Availability" column, to which publication it is referred (e.g. see the issue I described in the previous point). If a paper A cites paper B regarding some cost, and paper A is available through online access but paper B is not, I was still noting that as "Freely accessible", that the paper I describe in the entry is freely available (even though the cited one is not). This is a bit confusing I guess, since in such situations we have two papers, citing and the cited one, so it is question which one are we defining here?

It refers to the cited one. If paper B is not accessible, we mentioned "Reference title", "Authors" and "Publication year" associated to this paper B, then put 'No' in the "Availability" column (and the full citation of paper A in 'Previous materials'). We also indicate the full sentence (from paper A) in which paper B was mentioned in the column 'Details'. Do you think it is confusing?

45. I couldn't get Mendeley to work on my computer so far, I am using Ubuntu so it makes finding a proper version and installing it a bit more difficult. Nevertheless, I can access online list of papers for which you sent me a link, so I am using that to download papers from Internet, and so far I managed to get nearly all I tried to download like that. This is also good because that way I use the opportunity to check whether they are available and accessible on Internet, which is one of information we have to enter in the database ("Availability").

Great, you perfectly take benefit from this bad situation. Don't hesitate to tell me if you need some help about Mendeley.

46. In some cases I am not sure if the paper is a mismatch or a proper paper for the database. For example, in one paper they mention only in passing, for example in the Introduction or Discussion, some IAS cost, but they are citing another publication which I cannot access, e.g.: "Removal of water hyacinth (*Eichhornia crassipes*) in Wenzhou city cost 128 million US\$ in 1996 (Ding and Xie 1996). An estimate for the economic loss caused by *Mikania micrantha* on Neilingding Island ranges from 0.56 to 1.26 million US\$ (Zhong et al. 2004)." These cited publications seem to be in Chinese, so I cannot access the original material, and the reported information in the accessed is very scarce, so I would not be able to get all the information we need (e.g. time scale, spatial coverage, etc. - for the paper in question I think there will be no data for 90% of types of information we need, simply because they just briefly cited other papers in two short sentences). What should I do in such cases, include this paper and the brief information they report in those sentences, or exclude it from the dataset?

It is a common situation, unfortunately. In these cases, you're right: there would be many missing info for the values. We opted for the following (certainly disputable) strategy:

- we include the papers Ding and Xie 1996 and Zhong et al. 2004 (each in one line) as main papers (for which we give the dedicated title, publication year, authors, etc.) then indicated 'no' in



Availability section and mentioned the reference from which the unreachable papers came in the column 'Previous materials' (if we screen successive citations to access the reference, we mentioned here all the refs scrutinized separating them by "à")

- the host of associated info (geographical situation, time range, etc.) are captured in the accessible paper analysed with mention of the sentence (e.g., "Removal of water hyacinth (*Eichhornia crassipes*) in Wenzhou city cost 128 million US\$ in 1996 (Ding and Xie 1996)" in the column Details; in this column, you can add any info you deem necessary, for instance in this case, you can mention "this value come from an inaccessible reference")

47. I also wanted to ask you about the ID I should assign to each entry. This explanation was provided in material you sent me: "The identifier of each material collected and screened. This ranges from 1 to the maximum value of the unique materials retrieved from each repository preceded by WOS, GS or Go when materials were retrieved from Web of Science, Google Scholar, Google search engine respectively. For materials gathered opportunistically (i.e., elsewhere than from the three original repositories), the numerical values are preceded by Op. These identifiers correspond to those provided in the Data citation 1. For materials collected as a result of (successive) citations from an Previous materials previously analysed, we mentioned the ID of the original material preceded by 'From!.'" However, I could not find the information about the original material where these papers were discovered as citations, so I used for the time being combination of "FROM", number of the paper as I check them, going from the oldest towards the newest entry in Mendeley, and my initials.

Don't worry about that. I will manage it once the database will be completed. In the new and more simple (I hope) database, there are three first columns that will allow the link with another dataset containing all bibliographic details on the refs mentioned in the database focusing on cost values. Just give a unique number to each line you create.

Also, I will also manage two other types of information that you do not have to fill (you can also leave the columns empty): (i) the taxonomic details (columns K to P), just mention the scientific (with "/" for separating when several species concerned and list them by alphabetical order) and common names (if possible); (ii) the 9 last columns (AO-AW) that I use to convert/actualize values in 2017US\$.

### Spatial coverage

48. Some sources reported the costs as US\$ per day or per area. For instance, the costs to control grass invasion were reported per hectares (lines 274-277). How to deal with this properly? Because I wrote "Per hectare" in the column "Details".

For these cases, we opted to create a dedicated category in the column 'spatial scale', that is 'unit' followed by the indication of the unit in which the cost was expressed (in brackets). In your case, it would be 'unit (day)' or 'unit (area)'.

49. How do we treat costs that are given at a per individual level and/or are estimated as a total savings due to some effective control measure – e.g. "Reduced respiratory ailments costing approximately \$23.85 per sufferer induced by the weed per annum. With a reduction of approximately 350,000 people suffering from hay-fever as a result of biocontrol in 2005, this equates to an estimated health benefit of approximately \$8.4M in 2004/05 dollar terms" and

another example “Biocontrol is estimated to reduce production losses in Victoria by approximately \$4.4M per annum through the following channels...”

For the first case (i.e. *costs that are given at a per individual level*), we opted to create a dedicated category in the column ‘spatial scale’, that is ‘unit’ followed by the indication of the unit in which the cost was expressed (in brackets). In your case, it would be ‘unit (sufferer)’.

For the second case (i.e. *costs that are estimated as a total savings due to some effective control measure*), we considered them as such by assigning them to ‘avoided costs’ in the column ‘Type of cost’ and ‘potential’ in the column ‘Implementation’. Indeed, these costs have not been suffered by people, but should have been if there was not effective control.

50. What should we do with information like “Colombia/Ecuador/Peru”? We can keep that. Or to add the cost in the three countries (I mean replicating the costs). Or we can have a look at the paper again.

I think it is not realistic to replicate the same cost data for each country when such cost was provided at a 'regional' (and not a 'country') scale, as this cost was provided as occurring in an area vocering the three countries. One could suggest to divide the total cost by 3, but we have no precise idea on how the cost is specifically distributed within these countries (probably not equally). Maybe one can proceed as you proposed, but all this must be clearly stated before implementing the study.

51. If an article proclaims a loss in property values per house on a lake side, but does not tell us how many houses there actually are nor tells us if this is for one year or a loss over a certain period of time, how should I deal with it?

You must record the info as they are provided, i.e. take the cost as given by the authors (not necessary to multiply by the number of houses), and precise 'unit (house)' in the "spatial scale" column, and 'unspecified' in the "period of estimation", "starting year" and "ending year" columns (if unknown). We can fill these columns later, for instance with the year of publication, for analysis purposes.

52. In that example above, where is the affected environment, terrestrial as it is a house AT the lake, or is it aquatic as the invader is an aquatic plant that causes the cost?

I would tell 'aquatic' because of the species, but I agree that it's a complex situation in which we can consider 'terrestrial' for the reason you highlighted...

### Taxonomy

53. In the species column, there are many lines with “Diverse/Unspecified”. Again, we should keep like that or we should try to determine each species. Or at least split between diverse and unspecified.

I think there is no 'best' solution. It may depend on the research question as well as the degree of precision specifically needed for each study. We have mentioned that when it was not possible to determine specifically the nature of the information (nonetheless, one can go through papers again to look for it if needed for a specific subset, for instance data related to South America). One study focusing on just depicting costs should consider them, while a study attempting to rank costs within a defined column should remove them. The same situation may occur for cells that do not disentangle specific info: the case you proposed in your question 1 perfectly illustrated that. One study that focuses on relations between cost and GDP should probably remove this info, while a global study that aims to compare costs between continents will consider it.

More broadly, a case-by-case strategy has to be implemented, while accepting some compromises depending on the scientific questions.

### Temporal coverage

54. "Time range" column - it seems that one of the categories, "periodic", is not mentioned in the data paper, I made a comment there regarding that. My understanding was that "punctual" are all those one-time costs and investments, "annual" are those that we can standardize (annualize) to express them per year, and "periodic" are those that we cannot express per year (either the period is not with annual dynamics, or we had no data about that) - for all costs where they do not provide any information about duration I was marking them as periodic.

I think that I was not sufficiently clear in the paper/table, sorry. For this column, it only depends on what has been provided by the authors, not our ability to annualize or not the estimates. For "punctual", it is okay. For "annual", it is when the authors explicitly mentioned that the costs are suffered every year (the authors usually used some terms such as 'per year', 'annually', 'each year', etc.). "Periodic" was used when the authors expressed the estimates along a specified period time, whether or not we can annualize the estimates. I do not know if it sufficiently clear.

55. I have a quick question about the columns "probable starting year" and "probable starting year low margin", respectively the ending year... So, first of all, what is the difference between them? Secondly, if a cost was estimated for a year, but is "potentially ongoing" with 1 million of costs per year, but the probable starting year and probable starting year low margin is 2005, but the ending year and ending year low margin differ, how should we deal with it? I assume that it will affect the expansion, but should these ending years be that different?

The "probable starting year" & "probable ending year" columns (named 'original columns' hereafter) contain the information extracted from the papers processed, with the possibility to have no clear information retrieved; this explains why we have 'unspecified' in some cases. The "probable starting year\_low margin" & "probable ending year\_low margin" columns are those we created for the analyses purposes made in our overarching paper. These columns have the same info than the 'original columns', except in cases where we have 'unspecified' as provided information. We then filled these cells with an actual year based on educated guesses (e.g., corresponding to publication year or calculated from the duration time, when provided). In that way, you don't have 'unspecified' in these 'low margin' columns. This also explains the difference you mentioned in your second question (i.e., ending year vs ending year low margin). In these specific situations where we have 'potentially ongoing' costs, we may assume that they still occur beyond the specified (or not) ending data. Therefore, as you said, these columns impact the expansion of cost data and would allow everyone to make different assumptions on the cost duration.

56. Where the start and ending year aren't clear, is it better to guess based on the publication date, or to put unspecified?

At your level, you can put 'unspecified' as the info is not clearly provided. We'll change things (for instance, by considering the publication year) in another set of columns we'll use for cost standardisation (these are the columns called, that you can see after the 'Contributors' column in the DB).

### Updates

57. Does InvaCost\_3.0 include new columns not integrated in the original version of InvaCost (deposited in figShare) and/or in the previous version 2.1 (deposited in the dropbox)?

Yes, it does. We have added a file 'Descriptors.xlsx' in the figShare (<https://doi.org/10.6084/m9.figshare.12668570.v2>) that provides a full list/definition of these descriptive columns. Note that all columns are now usable and citable in the papers without the

need of fully explaining/describing the procedures used to generate them. You can simply use the right name of the column, refer to the link/doi of the figShare page, and take the definition provided in the 'Descriptors.xlsx' file (for instance, you can add information from the columns you use in your paper(s)). Nonetheless, pay attention to adapt the text according to the purpose/context of your paper(s).

58. Are some existing columns modified in the version 3.0?

We have changed the name of three columns between InvaCost\_2.1 and InvaCost 3.0:

- **Type\_of\_cost\_2** is called now **Type\_of\_cost\_merged**
- **Probable\_starting\_year\_low\_margin** and **Probable\_ending\_year\_low\_margin** are called now, respectively, **Probable\_starting\_year\_adjusted** and **Probable\_ending\_year\_adjusted**.

59. Are some existing columns removed from the version 3.0?

We have removed some columns that are not necessary and/or that can be easily derived from existing

- **Impacted\_sector\_2**: this column was exactly the same than the **Impacted\_sector** one; the single difference is that we grouped under the term 'mixed' all entries that have more than one category for this descriptor; this data processing step can be easily done by everyone prior to the analyses
- **Probable\_starting\_year\_high\_margin** and **Probable\_ending\_year\_high\_margin**: these columns were created based on information provided in the **Probable\_starting\_year\_adjusted** and **Probable\_ending\_year\_adjusted**, but with modifications based on assumptions made on the duration of 'potentially ongoing' costs; these columns can also be added – and then designed according to particular expectations on the duration of costs – by everyone prior the analyses
- **To\_remove**: this column gives information about which cost entries should be removed from the analyses made considering the **Cost\_estimate\_per\_year** column(s); now, these latter columns are now empty when we were not able to calculate or estimate cost duration (we used cost duration – in years – to derive yearly cost values based on the raw costs collated).

60. Why do some cost entries have no information in the **Probable\_starting\_year\_adjusted** and **Probable\_ending\_year\_adjusted** columns?

When we were not able to estimate them based on the ones provided by the authors (e.g. **Period\_of\_estimation** column), we leave them blank, and then we don't calculate the cost estimates per year. This means that the cost entries that have no information for at least one of these columns are not included in the cost analyses implemented in the package. Nonetheless, the corresponding raw costs have been converted in 2017 US\$ (**Raw\_cost\_estimate\_2017\_USD\_exchange\_rate**), and then this information can be used if needed. **BUT** remember that using this raw cost column does not allow getting relevant summations of total costs over time. This is because some raw costs are given as annual values (costs that are 'year' in **Time\_range** column), while they occur for several years (in case a paper explicitly provides the duration time of the costs). It is therefore important to keep it mind that the total costs over a specific period cannot be obtained through simply summing values from this raw cost column.

For the entries that have values in **Probable\_starting\_year\_adjusted** and **Probable\_ending\_year\_adjusted** columns, the right way to get the total cost over time is either to use the dedicated functions from the invacost package or summing values in the **Cost\_estimate\_per\_year\_2017\_USD\_exchange\_rate** column after having expanded the database over time.

61. How can I refer to the citations/doi related to these data?

In addition to the usual references (*original InvaCost*: Diagne et al. 2020 Scientific Data; *Non-English data*: Angulo et al. submitted), you can also refer to:

- **InvaCost\_3.0:** Diagne C., Leroy B., Gozlan E. R., Vaissière A-C., Assailly C., Nuninger L., Roiz D., Jourdain F., Jaric I., Courchamp F. (2020). *InvaCost: References and description of economic cost estimates associated with biological invasions worldwide*. figshare. Dataset. doi: <https://doi.org/10.6084/m9.figshare.12668570>
- **InvaCost\_1.0 (original version):** Diagne C., Leroy B., Gozlan E. R., Vaissière A-C., Assailly C., Nuninger L., Roiz D., Jourdain F., Jaric I., Courchamp F. (2020). *InvaCost: References and description of economic cost estimates associated with biological invasions worldwide*. figshare. Dataset. doi: <https://doi.org/10.6084/m9.figshare.12668570.v1>
- **Non-English database (original version):** Angulo E., Diagne C., Ballesteros-Mejía L., Adamjy T., Ahmed D., Akulov E., Banerjee A.K., Capinha C., Dia C.A.K.M., Dobigny G., Duboscq-Carra V.G., Golivets M., Haubrock P., Heringer G., Kirichenko N., Kourantidou M., Liu C., Nuñez M., Renault R., Roiz D., Taheri A., Verbrugge L., Watari Y., Xiong W., Courchamp F. (2020): *Non-English database version of InvaCost*. figshare. Dataset. doi: <https://doi.org/10.6084/m9.figshare.12928136>
- **Complementary search database (original version):** Ballesteros-Mejía, L. Angulo E.; Diagne C.; Courchamp F. Consortium Invacost (2020): *Complementary search database for Invacost*. figshare. Dataset. doi: <https://doi.org/10.6084/m9.figshare.12928145>