

From: Neville Barrett  
 To: Graham Edgar, Rick Stuart-Smith, Lizzzi Oh  
 Cc: Antonia Cooper  
 Subject: RE: Database corrections mechanism - potential loss of observation-level metadata  
 Date: Thursday, 19 January 2023 12:14 PM  
 Attachments: [image005.png](#)

Hi Lizzzi,

I'm fine with that. All our original data was entered into those bins anyway in the initial Excel process (but recorded underwater to the nearest mm), and in reality many of the current lobster sizes are estimated rather than measured so a bin is better than an exact size guess anyway and I don't think the new database discerns measured vs estimated sizes? So no issue there. It was only when Peter developed our initial database that we were able to enter the original length measurements (which I think made data entry a bit less error prone when placing sizes into bins) so we are just reverting to the initial way of doing things by the sound of it.

Cheers,  
 Nev

From: Graham Edgar <g.edgar@utas.edu.au>  
 Sent: Wednesday, 18 January 2023 8:19 AM  
 To: Rick Stuart-Smith <rick.stuartsmith@utas.edu.au>; Lizzzi Oh <lizzzibeth.oh@utas.edu.au>; Neville Barrett <neville.barrett@utas.edu.au>  
 Cc: Antonia Cooper <antoniac.cooper@utas.edu.au>  
 Subject: RE: Database corrections mechanism - potential loss of observation-level metadata

Hello Lizzzi

Thanks for progressing.

For me, trying to keep those old data is not worth the effort – a legacy dataset should be fine.

Nev might think otherwise though.

Cheers

gedgar

From: Rick Stuart-Smith <rick.stuartsmith@utas.edu.au>  
 Sent: Saturday, 14 January 2023 4:46 PM  
 To: Lizzzi Oh <lizzzibeth.oh@utas.edu.au>; Graham Edgar <g.edgar@utas.edu.au>; Neville Barrett <neville.barrett@utas.edu.au>  
 Cc: Antonia Cooper <antoniac.cooper@utas.edu.au>  
 Subject: RE: Database corrections mechanism - potential loss of observation-level metadata

Hey Lizzzi,  
 That is excellent news about the corrections tool getting close!  
 I can't see how losing any of the things mentioned below will affect me, so although worth a quick chat this week, all likely fine by me.  
 This really is more a question for Gedgar and Nev though.

Cheers  
 Rick

From: Lizzzi Oh <lizzzibeth.oh@utas.edu.au>  
 Sent: Thursday, 12 January 2023 4:56 PM  
 To: Graham Edgar <g.edgar@utas.edu.au>; Neville Barrett <neville.barrett@utas.edu.au>; Rick Stuart-Smith <rick.stuartsmith@utas.edu.au>  
 Cc: Antonia Cooper <antoniac.cooper@utas.edu.au>  
 Subject: Database corrections mechanism - potential loss of observation-level metadata

Hi All,

We are in our testing phase for the major data correction applications the AODN have created for the NRMN database (yay!).

Before we are able to release it (and finally correct the backlog of documented errors) a decision has to be made about some "metadata" attached to many observations in the database. Some of the of the information I would say is redundant. The most tricky are ATRC-related info, eg. the fine-scale measurements of lobsters and abalone (mm) that we used to record – these have been binned into the new invert-sizing scheme (5mm res to 1cm res for bigger sizes) to match new incoming data, but the original measurements are still attached to the binned observation. These records and certain other info (explained in the file attached and table below) were migrated to the new database, but stored as observation "attributes" rather than an integrated part of the schema, supported in future data ingest. We've now realised that they will potentially become lost or separated from observations if the survey is corrected. So we need to decide if creating a separate legacy dataset of these and moving on is sufficient (and just moving forward with our new methods for handling data), or whether it is imperative that we find a way of keeping this information attached to their observation when corrections are made in case it is of use in the future.

The short story is that the corrections mechanism will load selected survey data into a staging area, then re-ingest it once the required correction are made. This will of course not change the survey\_id, but the database ids for each observation must change (this must happen to give us the best flexibility to actually correct observations). Unfortunately even know we may only change some observations in a survey, all of them will have to be staged and thus get new id when re-ingested. This is where any observation "attributes" will become detached/lost because the processes of staging and re-ingesting corrected data do not cater for the historical observation "attributes" – the process is heavily based off the data ingest mechanism.

If we make a legacy dataset of attributes like lobster mm sizes, it could still be traced to a survey\_id in the database, we just won't know which lobster in the survey was the 113mm one or which was the 114mm one, for example. Potentially this would become more problematic if our corrections involve splitting or merging surveys when their data have been accidentally pooled, or accidentally split by assigning slightly different depths, for instance, but that is probably a rare occurrence.

METADATA/ATTRIBUTES	Description/Examples	count	Comments - Liz
Observation level			
Notes	These are observation-level notes (ie. survey-level notes still preserved). Most are descriptions of bulk corrections (RLS?), but some are not. Eg. "SizeClass corrected from 20.0 to 3.5. See NRMN #107"; "SizeClass corrected from 20.0 to 3.5. See NRMN #107"; "No code, species #N/A; there was also an 'orange eel' on the datasheet", "Fish photographed", "SpeciesID changed 6450 -> 3373 (RLS_085_SpeciesIDChange)". See attached file for all distinct notes found.	4217	
SizeRaw	Original ATRC invert string (from resolution) for lobsters and abalone, and many cases for other M2 animals in certain time periods/projects. See attached file for all examples	33654	
SizeEstimated	Yes or No - whether specimen (eg. lobster) was estimated for size or caught and measured (ATRC)	33664	
SpeciesSex	Male/Female eg. for lobsters and some wrasse (ATRC)	32058	
Biomass	The old database stored biomass values for observation	1137408	redundant - as this is now calculated on data-out based on the stored a's and b's
DescriptiveName	seems to be a big mess of migrated info eg. common names, synonyms or spelling mistakes?	80793	redundant
LegalSize	Yes or No	150	too few to be consistent or used for anything. Assume mostly RLS but unsure why some non commercial species have this populated?
SimulatedAbsence	"No Species Found" records	300	These would be regenerated when corrected observations are re-ingested
SurveyMethodLevel			
NonStandardData	Eg. "Site sampled due to oil spill", "poor visibility", "carried out on seagrass bed"	255	These should be migrated to survey notes?
LegacyMethod	eg. Parallel fish survey Jarvis Bay.	1808	redundant as data has been transformed into new blocks

Cheers,

Lizzzi Oh

Project officer  
 Ecology and Biodiversity  
 Institute for Marine and Antarctic Studies (IMAS)  
 College of Sciences and Engineering  
 University of Tasmania  
 Ph: +61475 465 380  
 Mailing Address: 25-21 Nubenna Crescent, Taroona, Tasmania, Australia, 7054  
[lizzzi.oh@utas.edu.au](mailto:lizzzi.oh@utas.edu.au)

I acknowledge the Traditional Custodians of the Land and Sea Country on which I live and work, and pay respect to their elders, past, present and future.



ORCID: 00090880

University of Tasmania Electronic Communications Policy (December 2014). This email is confidential, and is for the intended recipient only. Access, disclosure, copying, distribution, or reliance on any of it by anyone outside the intended recipient organisation is prohibited and may be a criminal offence. Please delete if obtained in error and email confirmation to the sender. The views expressed in this email are not necessarily the views of the University of Tasmania, unless clearly identified otherwise.