

Audit finds biodiversity data aggregators 'lose and confuse' data

23 April 2018

```

T2416 Thallis tropica Lea Thallis up-match species to genus Paratype
T2417 Thallis tropica Lea Thallis up-match species to genus Paratype
MET47297 Thallopsa destinataria (Guenee, 1858) Thallopsa up-match species to genus
MET47298 Thallopsa destinataria (Guenee, 1858) Thallopsa up-match species to genus
MET47299 Thallopsa destinataria (Guenee, 1858) Thallopsa up-match species to genus
MET47300 Thallopsa destinataria (Guenee, 1858) Thallopsa up-match species to genus
MET47301 Thallopsa destinataria (Guenee, 1858) Thallopsa up-match species to genus
MET47302 Thallopsa destinataria (Guenee, 1858) Thallopsa up-match species to genus
MET47303 Thallopsa destinataria (Guenee, 1858) Thallopsa up-match species to genus
MET47304 Thallopsa destinataria (Guenee, 1858) Thallopsa up-match species to genus
MET47305 Thallopsa destinataria (Guenee, 1858) Thallopsa up-match species to genus
MET47306 Thallopsa destinataria (Guenee, 1858) Thallopsa up-match species to genus
MET47307 Thallopsa destinataria (Guenee, 1858) Thallopsa up-match species to genus
MET47308 Thallopsa destinataria (Guenee, 1858) Thallopsa up-match species to genus
MET47309 Thallopsa destinataria (Guenee, 1858) Thallopsa up-match species to genus
MET47310 Thallopsa destinataria (Guenee, 1858) Thallopsa up-match species to genus
COL65957 Thanaoscleris buqueti Lefevre Thanaoscleris up-match species to genus
HEM4106 Thaumastopsaltria globosa (Distant, 1897) Thaumastopsaltria up-match species to genus
HEM4107 Thaumastopsaltria globosa (Distant, 1897) Thaumastopsaltria up-match species to genus
HEM4108 Thaumastopsaltria globosa (Distant, 1897) Thaumastopsaltria up-match species to genus
T11963 Thenarotes carteri Gloane Lecanomerus up-match species to genus Syntype
T11964 Thenarotes carteri Gloane Lecanomerus up-match species to genus Syntype
MET15999 Theretra rhesus Theretra up-match species to genus
T2887 Theretria clarki Paramonov Theretria up-match species to genus
T2888 Theretria clarki Paramonov Theretria up-match species to genus Paratype
T14128 Theretria signabilis Paramonov Theretria up-match species to genus Paratype
COL94465 Thyreocephalus haemorrhous (Fauvel, 1878) Thyreocephalus up-match species to genus
T368 Tinea intritella Walker Tinea up-match species to genus Holotype
TR122134 Tinodes antequerella Tinodes up-match species to genus
TR122133 Tinodes assimilis McLachlan, 1865 Tinodes up-match species to genus
TR122131 Tinodes aureola (Fallen, 1806) Tinodes up-match species to genus
TR122132 Tinodes aureola (Fallen, 1806) Tinodes up-match species to genus
TR152324 Tinodes aureola (Fallen, 1806) Tinodes up-match species to genus
TR122135 Tinodes dives (Pictet, 1834) Tinodes up-match species to genus
TR122005 Tinodes waeneri (Linnaeus, 1758) Tinodes up-match species to genus
TR122006 Tinodes waeneri (Linnaeus, 1758) Tinodes up-match species to genus
TR122007 Tinodes waeneri (Linnaeus, 1758) Tinodes up-match species to genus
TR122008 Tinodes waeneri (Linnaeus, 1758) Tinodes up-match species to genus
TR152323 Tinodes waeneri (Linnaeus, 1758) Tinodes up-match species to genus
MET19733 Tolpia myops Hampson, 1907 Tolpia up-match species to genus
MET19734 Tolpia myops Hampson, 1907 Tolpia up-match species to genus
MET19735 Tolpia myops Hampson, 1907 Tolpia up-match species to genus
MET19736 Tolpia myops Hampson, 1907 Tolpia up-match species to genus
MET29155-1 Tortrix abruptana Tortrix up-match species to genus
MET29158 Tortrix abruptana Tortrix up-match species to genus
MET29159-1 Tortrix abruptana Tortrix up-match species to genus
MET29159 Tortrix abruptana Tortrix up-match species to genus
T14836 Tortrix amacula Lower Tortrix up-match species to genus Holotype
T14878 Tortrix nucleata Meyrick Tortrix up-match species to genus Holotype
T14879 Tortrix nucleata Meyrick Tortrix up-match species to genus Holotype
COL27271 Trachymela transversalis Blackburn Trachymela up-match species to genus
COL27272 Trachymela transversalis Blackburn Trachymela up-match species to genus
COL27273 Trachymela transversalis Blackburn Trachymela up-match species to genus
T14710 Trachyntis eurychne Turner Barea up-match species to genus Holotype
COL37350 Tragoecerus fasciatus Donovan Tragoecerus up-match species to genus
COL37351 Tragoecerus fasciatus Donovan Tragoecerus up-match species to genus

```

A snippet of the results from a data processing event. Credit: Dr. Robert Mesibov

In an effort to improve the quality of biodiversity records, the Atlas of Living Australia (ALA) and the Global Biodiversity Information Facility (GBIF) use automated data processing to check individual data items. The records are provided to the ALA and GBIF by museums, herbaria and other biodiversity data sources.

However, an independent analysis of such records reports that ALA and GBIF data processing also leads to data loss and unjustified changes in scientific names.

The study was carried out by Dr Robert Mesibov, an Australian millipede specialist who also works as a data auditor. Dr Mesibov checked around 800,000 records retrieved from the Australian Museum, Museums Victoria and the New Zealand Arthropod Collection. His results are published in

the open access journal *ZooKeys*, and also archived in a public data repository.

"I was mainly interested in changes made by the aggregators to the genus and species names in the records," said Dr Mesibov.

"I found that names in up to 1 in 5 records were changed, often because the aggregator couldn't find the name in the look-up table it used."

Another worrying result concerned type specimens - the reference specimens upon which scientific names are based. On a number of occasions, the aggregators were found to have replaced the name of a type specimen with a name tied to an entirely different type specimen.

The biggest surprise, according to Dr Mesibov, was the major disagreement on names between aggregators.

"There was very little agreement," he explained. "One aggregator would change a name and the other wouldn't, or would change it in a different way."

Furthermore, dates, names and locality information were sometimes lost from records, mainly due to programming errors in the software used by aggregators to check data items. In some data fields the loss reached 100%, with no original data items surviving the processing.

"The lesson from this audit is that biodiversity data aggregation isn't harmless," said Dr Mesibov. "It can lose and confuse perfectly good data."

"Users of aggregated data should always download both original and processed data items, and should check for data loss or modification, and for replacement of names," he concluded.

More information: Robert Mesibov, An audit of

some processing effects in aggregated occurrence records, ZooKeys (2018). [DOI: 10.3897/zookeys.751.24791](https://doi.org/10.3897/zookeys.751.24791)

Provided by Pensoft Publishers

APA citation: Audit finds biodiversity data aggregators 'lose and confuse' data (2018, April 23) retrieved 20 May 2019 from <https://phys.org/news/2018-04-biodiversity-aggregators.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.