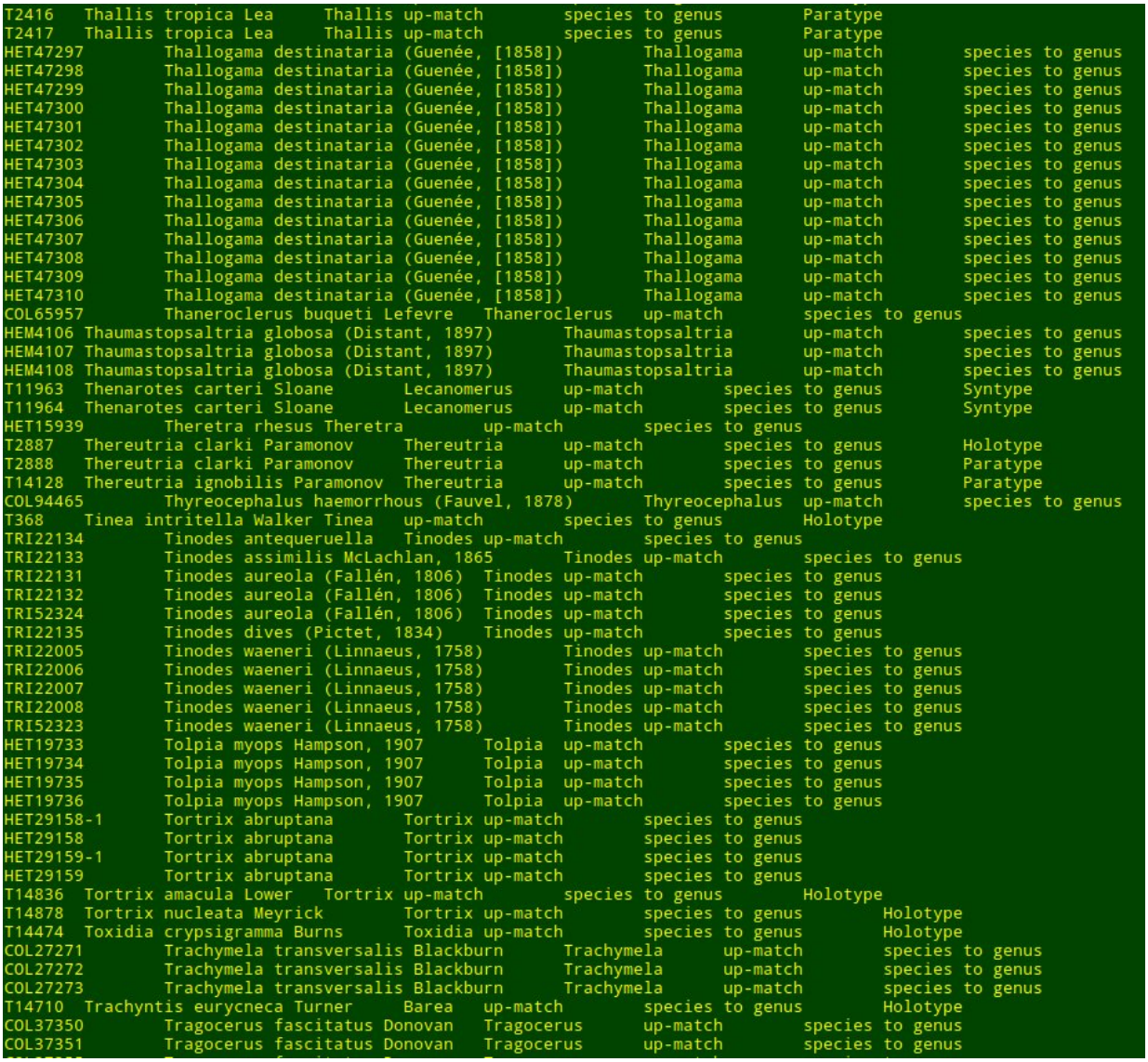
Audit finds biodiversity data aggregators 'lose and confuse' data

April 23 2018



```
T2416   Thallis tropica Lea     Thallis up-match        species to genus        Paratype
T2417   Thallis tropica Lea     Thallis up-match        species to genus        Paratype
HET47297        Thallogama destinataria (Guenée, [1858])        Thallogama      up-match        species to genus
HET47298        Thallogama destinataria (Guenée, [1858])        Thallogama      up-match        species to genus
HET47299        Thallogama destinataria (Guenée, [1858])        Thallogama      up-match        species to genus
HET47300        Thallogama destinataria (Guenée, [1858])        Thallogama      up-match        species to genus
HET47301        Thallogama destinataria (Guenée, [1858])        Thallogama      up-match        species to genus
HET47302        Thallogama destinataria (Guenée, [1858])        Thallogama      up-match        species to genus
HET47303        Thallogama destinataria (Guenée, [1858])        Thallogama      up-match        species to genus
HET47304        Thallogama destinataria (Guenée, [1858])        Thallogama      up-match        species to genus
HET47305        Thallogama destinataria (Guenée, [1858])        Thallogama      up-match        species to genus
HET47306        Thallogama destinataria (Guenée, [1858])        Thallogama      up-match        species to genus
HET47307        Thallogama destinataria (Guenée, [1858])        Thallogama      up-match        species to genus
HET47308        Thallogama destinataria (Guenée, [1858])        Thallogama      up-match        species to genus
HET47309        Thallogama destinataria (Guenée, [1858])        Thallogama      up-match        species to genus
HET47310        Thallogama destinataria (Guenée, [1858])        Thallogama      up-match        species to genus
COL65957        Thaneroclerus buqueti Lefevre   Thaneroclerus   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 Sloane       Lecanomerus     up-match        species to genus        Syntype
T11964  Thenarotes carteri Sloane       Lecanomerus     up-match        species to genus        Syntype
HET15939        Theretra rhesus Theretra        up-match        species to genus
T2887   Thereutria clarki Paramonov     Thereutria      up-match        species to genus        Holotype
T2888   Thereutria clarki Paramonov     Thereutria      up-match        species to genus        Paratype
T14128  Thereutria ignobilis Paramonov  Thereutria      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
TRI22134        Tinodes antequeruella   Tinodes up-match        species to genus
TRI22133        Tinodes assimilis McLachlan, 1865       Tinodes up-match        species to genus
TRI22131        Tinodes aureola (Fallén, 1806)  Tinodes up-match        species to genus
TRI22132        Tinodes aureola (Fallén, 1806)  Tinodes up-match        species to genus
TRI52324        Tinodes aureola (Fallén, 1806)  Tinodes up-match        species to genus
TRI22135        Tinodes dives (Pictet, 1834)    Tinodes up-match        species to genus
TRI22005        Tinodes waeneri (Linnaeus, 1758)        Tinodes up-match        species to genus
TRI22006        Tinodes waeneri (Linnaeus, 1758)        Tinodes up-match        species to genus
TRI22007        Tinodes waeneri (Linnaeus, 1758)        Tinodes up-match        species to genus
TRI22008        Tinodes waeneri (Linnaeus, 1758)        Tinodes up-match        species to genus
TRI52323        Tinodes waeneri (Linnaeus, 1758)        Tinodes up-match        species to genus
HET19733        Tolpia myops Hampson, 1907      Tolpia  up-match        species to genus
HET19734        Tolpia myops Hampson, 1907      Tolpia  up-match        species to genus
HET19735        Tolpia myops Hampson, 1907      Tolpia  up-match        species to genus
HET19736        Tolpia myops Hampson, 1907      Tolpia  up-match        species to genus
HET29158-1      Tortrix abruptana       Tortrix up-match        species to genus
HET29158        Tortrix abruptana       Tortrix up-match        species to genus
HET29159-1      Tortrix abruptana       Tortrix up-match        species to genus
HET29159        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
T14474  Toxidia crypsigramma Burns      Toxidia 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 eurycneca Turner     Barea   up-match        species to genus        Holotype
COL37350        Tragocerus fascitatus Donovan   Tragocerus      up-match        species to genus
COL37351        Tragocerus fascitatus Donovan   Tragocerus      up-match        species to genus
```

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

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

Provided by Pensoft Publishers