Who Has Time for Biological Collections Data Quality Feedback? Maybe a Community Can Help

Teresa J. Mayfield-Meyer
Collections Assistant, Museum of Southwestern Biology
Treasurer, Arctos Consortium

SPNHC/TDWG 2018

Website: http://arctosdb.org
Search: http://arctos.database.museum
What do biodiversity collection managers do all day?
What do biodiversity collection managers do all day?

DIGITIZE AND MANAGE DATA
Stuff happens when you digitize…

Arctos
EMu
Specify
Symbiota

COBOL
dBASE
ACCESS
LOTUS123

Cloud Computing

FileMaker
No Excuses (Not)
Who has time for surveys?

iDigBio data quality flags feedback

Survey Design

iDigBio data quality flags feedback
Created on 2/18/2018

10 QUESTIONS

1 PAGES

Responses and Status

TOTAL RESPONSES
3

Collectors

OPEN
I'm not even sure what this is referencing: data in Arctos or data that is being migrated into Arctos! I have never heard this terminology until now.
Wait, there are data flags?

Whether or not you received a data quality email from iDigBio, have you reviewed the data quality flags provided by iDigBio for your collection data?

Answered: 3  Skipped: 0
My initial experience

As my data was recently ingested by iDigBio, I received a huge list of specimens flagged for various corrections (sigh). I wanted to bring this one to the group to see if we should be paying more attention to Darwin Core, or if it is just something to let iDigBio keep "correcting" for.

Some of my specimens on islands in the Pacific, are flagged by iDigBio with "dwc_continent_replaced | Darwin Core Continent Corrected." one example is here: https://www.idigbio.org/portal/records/890158e-d745-430c-bb46-8b250b614cd6

Is Arctos not complying with Darwin Core or is this just an artifact of iDigBio? Do we need to do anything about it or do I just need to know that these flags are not a problem? My main concern is that users of iDigBio will view our data as less reliable with flags attached.

huge list of specimens flagged for various corrections (sigh)

dwc_continent_replaced

users of iDigBio will view our data as less reliable
My initial experience

noticed something similar in GBIF regarding dates

Darwin Core is an exchange standard; Arctos isn't "complying" with any data standards because none exist

I agree with your assessment: User's initial reaction to the flag will be "Arctos is broken," which is absolutely not the case.
My initial experience

“..subjective to how data are stored in Arctos vs. other models. The subjective ones I don't think are worth our time to care about at this point”

“..some data quality flags from iDigBio that are useful because they correct objectively incorrect data”

the DQ tests and methods iDigBio uses are in flux due to work being done in TDWG. ....
My initial experience

I would hope (although I guess hope is the operative word here), that people who are running analyses on or otherwise using aggregator data for something beyond browsing would notice that the flags are doing more standardizing than correcting, and that obviously different collections/databases use different but equally correct ways to say the same thing...
One Collection’s Data Flags
This table shows any data corrections that were performed on this recordset to improve the capabilities of iDigBio Search. The first column represents the correction performed. The last two columns represent the number and percentage of records that were corrected. A complete list of the data quality flags and their descriptions can be found here. Clicking on a data flag name will take you to a search for all records with this flag in this recordset.

<table>
<thead>
<tr>
<th>Flag</th>
<th>Records With This Flag</th>
<th>(%) Percent With This Flag</th>
</tr>
</thead>
<tbody>
<tr>
<td>idigbio_isocountrycode_added</td>
<td>23354</td>
<td>100</td>
</tr>
<tr>
<td>geopoint_datum_missing</td>
<td>22963</td>
<td>98.326</td>
</tr>
<tr>
<td>dwc_datasetid_added</td>
<td>21141</td>
<td>90.524</td>
</tr>
<tr>
<td>dwc_parentnameusageid_added</td>
<td>21141</td>
<td>90.524</td>
</tr>
<tr>
<td>dwc_taxonid_added</td>
<td>21141</td>
<td>90.524</td>
</tr>
<tr>
<td>dwc_taxonomicstatus_added</td>
<td>21141</td>
<td>90.524</td>
</tr>
<tr>
<td>dwc_taxonrank_added</td>
<td>21141</td>
<td>90.524</td>
</tr>
<tr>
<td>gbif_canonicalname_added</td>
<td>21141</td>
<td>90.524</td>
</tr>
<tr>
<td>gbif_genericname_added</td>
<td>21141</td>
<td>90.524</td>
</tr>
<tr>
<td>gbif_taxon_corrected</td>
<td>21141</td>
<td>90.524</td>
</tr>
<tr>
<td>dwc_scientificnameauthorship_added</td>
<td>20615</td>
<td>88.272</td>
</tr>
</tbody>
</table>
Data from source **Arctos** [Classifications] [Top]

- Last update: 2018-07-29 14:44:14.0
- Globalnames score not available
- Match type not available

**author_text**: A. Dietr.
**display_name**: Picea A. Dietr.
**nomenclatural_code**: ICBN
**remark**: Imported from ITIS 6 Feb 2007, added author text 7 Jul 2018 from ITIS
**scientific_name**: Picea
**source_authority**: ITIS
**valid_catalog_term_fg**: 1

**Classification**:

- **Plantae (kingdom)** [more like this term] [including rank] [from this source]
- **Tracheophyta (phylum)** [more like this term] [including rank] [from this source]
  - **Pinopsida (class)** [more like this term] [including rank] [from this source]
  - **Pinales (order)** [more like this term] [including rank] [from this source]
  - **Pinaceae (family)** [more like this term] [including rank] [from this source]
  - **Picea (genus)** [more like this term] [including rank] [from this source]

---

**dwc_kingdom_suspect**: 2
**dwc_phylum_added**: 1
<table>
<thead>
<tr>
<th>taxon_match_failed</th>
<th>2274</th>
<th>9.737</th>
</tr>
</thead>
<tbody>
<tr>
<td>dwc_specificEpithet_added</td>
<td>2191</td>
<td>9.382</td>
</tr>
<tr>
<td>dwc_specificEpithet_replaced</td>
<td>1666</td>
<td>7.134</td>
</tr>
<tr>
<td>dwc_family_replaced</td>
<td>1101</td>
<td>4.714</td>
</tr>
<tr>
<td>dwc_infracpecificEpithet_added</td>
<td>1069</td>
<td>4.577</td>
</tr>
<tr>
<td>dwc_order_replaced</td>
<td>739</td>
<td>3.164</td>
</tr>
<tr>
<td>dwc_genus_replaced</td>
<td>645</td>
<td>2.762</td>
</tr>
<tr>
<td>dwc_family_added</td>
<td>524</td>
<td>2.244</td>
</tr>
<tr>
<td>dwc_genus_added</td>
<td>524</td>
<td>2.244</td>
</tr>
<tr>
<td>dwc_phyllum_replaced</td>
<td>521</td>
<td>2.231</td>
</tr>
<tr>
<td>dwc_class_replaced</td>
<td>517</td>
<td>2.214</td>
</tr>
<tr>
<td>dwc_order_added</td>
<td>436</td>
<td>1.867</td>
</tr>
<tr>
<td>geopoint_low_precision</td>
<td>121</td>
<td>0.518</td>
</tr>
<tr>
<td>geopoint_datum_error</td>
<td>100</td>
<td>0.428</td>
</tr>
<tr>
<td>dwc_class_added</td>
<td>9</td>
<td>0.039</td>
</tr>
<tr>
<td>dwc_taxonremarks_added</td>
<td>4</td>
<td>0.017</td>
</tr>
<tr>
<td>dwc_kingdom_suspect</td>
<td>2</td>
<td>0.009</td>
</tr>
<tr>
<td>dwc_phyllum_added</td>
<td>1</td>
<td>0.004</td>
</tr>
</tbody>
</table>
Is Aves not a class?

Family=palinuridae
What the what?

Animalia > Arthropoda > Malacostraca > Decapoda > Palinuridae

Animalia > Chordata > Aves > Aves
The mystery deepens...

Genus = Avus
An answer, I think

Animalia > Chordata > Aves

Identification=scientific name=dwc:scientificName “Aves”

Taxon rank=dwc:taxonrank “class”
What do biodiversity collection managers do all day?
<table>
<thead>
<tr>
<th>We need more help at home!</th>
</tr>
</thead>
<tbody>
<tr>
<td>• under-staffed and under-funded</td>
</tr>
<tr>
<td>• lack of experience/knowledge on the tech end</td>
</tr>
<tr>
<td>• assistance expressing the importance of the endeavor to administrators</td>
</tr>
<tr>
<td>• sustainability</td>
</tr>
</tbody>
</table>
Thank You!

Photos courtesy of: Arctos
Arthur H. Harris
Museum of Obsolete Media
Wikimedia Commons

All opinions expressed are my own. I would like to thank

Everyone in the Arctos Working Group for putting up with my multiple requests and issues on the Arctos GitHub. You all make my life so much richer!

Dr. Art Harris for being my iDigBio data quality flag guinea pig and partner in biological data crime.

Deb Paul for taking my questions seriously and making an honest attempt to answer them even while traveling.