



iDigBio Data Management

Recordset Data Quality



Kevin J. Love, iDigBio

September 15 – 17, 2015

Arizona State University (ASU)

Biodiversity Knowledge Integration Center (*BioKIC*)

Brought to you by: [Tri-Trophic TCN](#), [DataONE](#), [ASU](#), [iDigBio](#), [Symbiota](#), and [NCEAS](#)



An Overview

- **January 2011**

The National Science Foundation (NSF) required a data management plan (DMP) for all full proposals submitted, or due, to NSF on or after January 18, 2011

- **July 2011**

First ADBC Program awards, including iDigBio and 3 Thematic Collections Networks

- **Yada Yada Yada**

13 TCNs, 15 PENS, over 200 collaborating institutions

- **August 2014**

iDigBio Data Management Interest Group holds first Adobe Connect meeting

- **October 2014**

iDigBio Summit IV

- Breakout group to discuss Data Management challenges
- Data Management Interest Group Poster



iDigBio Summit IV

Data Management Breakout Group

- **Attribution**
Copyright, Licensing, Data use, and Citation
- **Data Transformations**
"dwc:stateProvince": "Florida" -> "Country": "United States"
- **Data Cleaning**
"dwc:countryCode": "US" -> "countrycode": "USA"
- **Extending Data Schemas**
Annotations, Media Records, etc
- **Data Products**
Community driven data portals, Reference datasets, Modular workflows



iDigBio Data Management Team

- Data Management Coordinator
- Application Developers
- System Administrators
- Biologists
- PostDoc's and Graduate Students
- iDigBio Senior Staff



iDigBio Data Management Team

Meet Monthly

- Coordinate Activities with DM Interest Group and ADBC community
- Developing Internal DM Practices
- Describe “User Stories” for iDigBio feature development



iDigBio Data Management Team

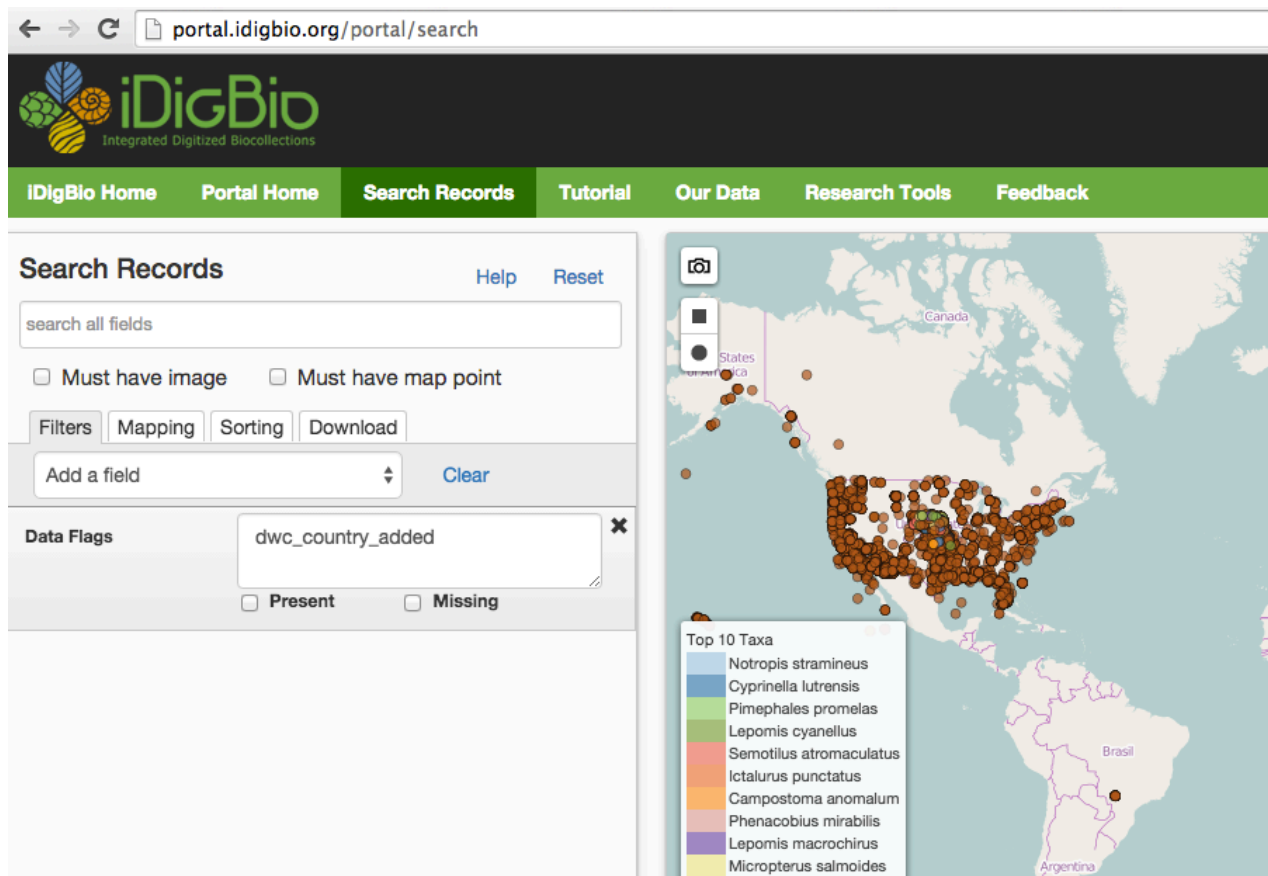
Data Quality Flags added to iDigBio Search API

<https://github.com/iDigBio/idigbio-search-api/wiki/Data-Quality-Flags>

```
← → C search.idigbio.org/v2/search/records?rq={"uuid":"91efb04-3086-404b-b5e8-1da5f85ef7fe"}
{
  "uuid": "91efb04-3086-404b-b5e8-1da5f85ef7fe",
  "countrycode": "usa",
  "basisofrecord": "preservedspecimen",
  "collector": "c. s. moreau",
  "institutioncode": "cas",
  "datemodified": "2014-03-21T15:28:38.233000+00:00",
  "flags": [
    "rev_geocode_mismatch",
    "geopoint_datum_missing",
    "geopoint_low_precision",
    "dwc_continent_added",
    "idigbio_isocountrycode_added",
    "dwc_country_added"
  ],
  "minlevation": 2,
  "kingdom": "animalia",
  "highertaxon": "animalia;arthropoda;insecta;hymenoptera;hymenoptera;formicidae;myrmicinae",
  "scientificname": "tetramorium simillimum",
  "class": "insecta",
  "occurrenceid": "cas:antweb:fmhins0000088656",
  "country": "united states",
  "locality": "key west, beachfront",
  "collectioncode": "antweb",
  "eventdate": "2009-08-25/2009-08-25",
  "fieldnumber": "csm1255",
  "verbatimeventdate": "8/25/2009/8/25/2009",
  "recordids": [
    "ded380b5-1ba2-4089-8e0c-0a1b4140785\\cas:antweb:fmhins0000088656"
  ],
  "genus": "tetramorium",
  "order": "hymenoptera"
}
```

iDigBio Data Management Team

Search query facets added to the iDigBio Specimen Portal



The screenshot shows the iDigBio Specimen Portal search interface. The browser address bar displays `portal.idigbio.org/portal/search`. The page header includes the iDigBio logo and navigation links: [iDigBio Home](#), [Portal Home](#), [Search Records](#) (active), [Tutorial](#), [Our Data](#), [Research Tools](#), and [Feedback](#).

The **Search Records** section features a search input field labeled "search all fields" and two checkboxes: Must have image and Must have map point. Below these are buttons for [Filters](#), [Mapping](#), [Sorting](#), and [Download](#). A dropdown menu labeled "Add a field" is set to `dwc_country_added`, with a [Clear](#) button. Underneath, the **Data Flags** section shows Present and Missing.

To the right, a map displays specimen locations as brown dots, primarily concentrated in the United States. A legend titled "Top 10 Taxa" lists the following species with corresponding colored markers:

- Notropis stramineus
- Cyprinella lutrensis
- Pimephales promelas
- Lepomis cyanellus
- Semotilus atromaculatus
- Ictalurus punctatus
- Campostoma anomalum
- Phenacobius mirabilis
- Lepomis macrochirus
- Micropterus salmoides



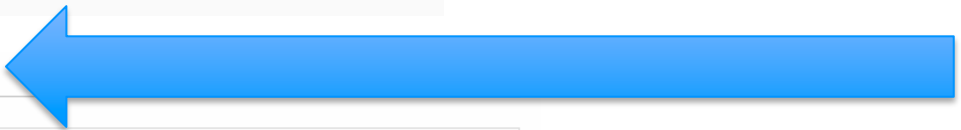
Specimen Record

[Animalia](#) > [Arthropoda](#) > [Insecta](#) > [Hymenoptera](#) > [Formicidae](#)

Tetramorium simillimum

From AntWeb

Continent	North America	Institution Code	CAS
Country	United States	Collection Code	ANTWEB
State/Province	Florida	Catalog Number	Fmnhins 0000088656
Locality	Key West Beachfront	Collected By	C S Moreau
Latitude	24.55	Date Collected	2009-08-25/2009-08-25
Longitude	-8.77		



Data **Flags** Raw

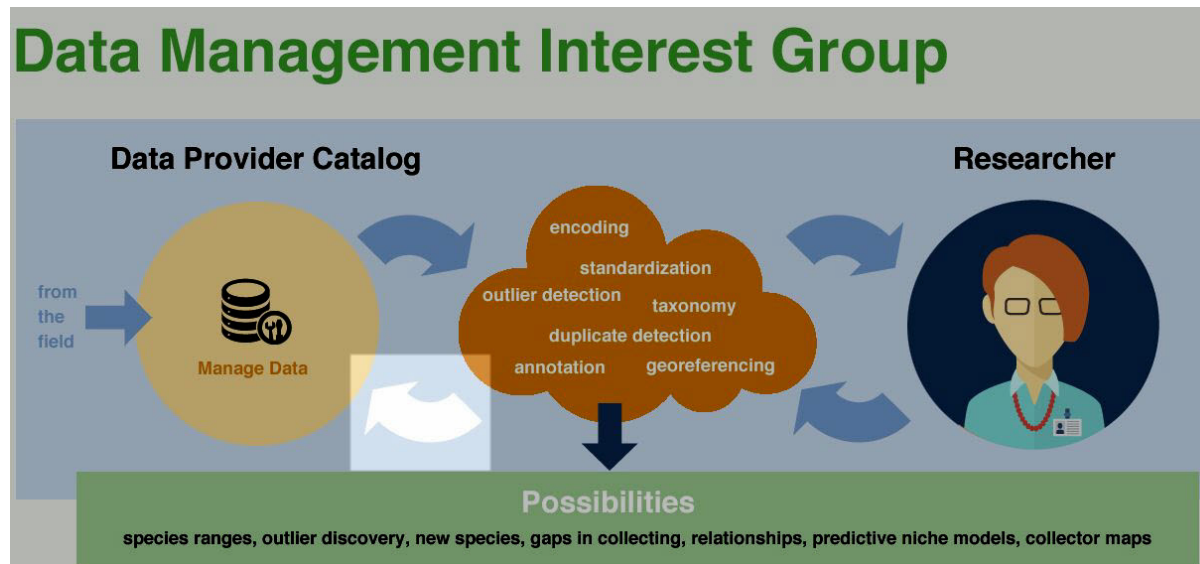
Type	Description
rev_geocode_mismatch	Geographic Coordinate did not reverse geocode to correct country.
geopoint_datum_missing	Geographic Coordinate Missing Geodetic Datum (Assumed to be WGS84).
geopoint_low_precision	Geographic Coordinate has Low Precision.
dwc_continent_added	Darwin Core Continent Added.
idigbio_isocountrycode_added	iDigBio ISO 3166-1 alpha-3 Country Code Added.
dwc_country_added	Darwin Core Country Added.



Join us for a Data Management webinar

Improving Data Quality: iDigBio Recordset data cleaning method, tools, and data flags

<https://www.idigbio.org/content/improving-data-quality-idigbio-recordset-data-cleaning-method-tools-and-data-flags>





Thanks!



www.idigbio.org



facebook.com/iDigBio



twitter.com/iDigBio



vimeo.com/idigbio



idigbio.org/rss-feed.xml



<webcal://www.idigbio.org/events-calendar/export.ics>