



# Introduction to Data Standards

Nelson Rios  
Division of Informatics

YALE  
PEABODY  
MUSEUM



# Standards

Documented agreement containing specifications, guidelines, or characteristics that ensure products, systems, or services work reliably and can operate together.

Ethernet, TCP/IP, HTTP etc.

Developed by recognized organizations (such as the through a consensus-based process, which involves input from industry experts, governments, and other stakeholders

IEEE, ISO, or W3C etc.

# Data Standards

**Semantic Structure:** Data standards provide agreed-upon definitions and formats for data elements. This ensures that, regardless of where the data originates, it follows the same rules for structure and meaning.

**Interoperability:** By using consistent data definitions, different systems can exchange, compare, and analyze data seamlessly.

**Data Quality:** Data standards help maintain data quality by setting rules for what valid data looks like, including definitions, units of measurement, or code sets.

# Evolution of Biodiversity Standards

1993

Fish Gopher

1998

ZBIG (Z39.50)  
Darwin Core 1.0

2001

DiGIR  
Darwin Core 1.2

2009

IPT  
Darwin Core Standard

**FISHGOPHER**

Want to search using FishGopher? Click [here](#).

FishGopher is a collaboration between free-standing museums and university collections in the development of open-access, biological collection community databases. This warmly welcomed project was funded by a grant from the NSF Program on Research Collections in Systematics and Ecology (DEB 932743) and facilitated by support from the MDCI Project formerly at Cornell University and the University of Kansas. Since 1993, the University of Michigan provided critical information analysis and coding. The FishGopher project is dedicated to the thousands of individuals who collected these taxa and their associated data.

The University of Kansas and four other institutions were initially invited to participate in FishGopher for the initial development of a comprehensive searching of ichthyological collections. Users may search one or multiple databases. All of the available searchable databases available today in a single search are listed below.

A search item finds out queries to multiple Gopher servers in the ichthyological community. Today, it permits access to the fish collection catalogs of these institutions:

- ANSP - Academy of Natural Sciences of Philadelphia
- BPBM - Bishop Museum (Hawaii) (limited data)
- CU - Cornell University
- FMNH - Field Museum of Natural History
- MCZ - Harvard University
- KU - University of Kansas Natural History Museum
- UACG - University of Alabama
- UMMZ - University of Michigan

The catalogs are full-text indexed and searchable on every word. Searches are case-insensitive, either lower or upper case characters may be used. Boolean queries are possible: restrict searches with "and" and "not" (without the quotes), broaden searches with "or". The "or" operator is the default for multiple term query specifications. Any number of Boolean operators may be used in a search specification.

Example:

First note that Fish names are in Latin. Do not use common names.

To search on a binomial, use the "and" operator:

Amia and calva (do not use "Amia calva")

To restrict a search for a species in a certain state, use this form:

Amia and calva and Tennessee

This also applies:

Amia and calva and (Tennessee or Ohio or Illinois or Indiana)

Do not use quoted strings.

A maximum of 1500 records per catalog will be returned, i.e. a total of 15000 records for a broad query on all ten servers. Be aware that this is more than some gopher clients are able to handle.

Searches returning a very large number of records may take a minute or two. Response time is also dependent on network latency between Biodiversity and Biological Collections (BBC) Gopher Server (<http://biodiversity.unm.edu/>) and the other servers as well as the latency between your Gopher client and the BBC Gopher.

An asterisk may be used for wild card (terminating) searches at the end of any specified search term. For example, to find all of the collections of *Myoxocephalus microdon* and avoid carpal tunnel syndrome use:

My\* and order

or your search specification.

You can also restrict your search to a particular institution by including the institutional acronym in your specification with the "and" operator, e.g.:

Exoco and american and CU

or

Percia and far\* and UMMZ

The acronyms for the collections are shown in the above table.

SEARCH the FISHGopher Databases

Select the databases you wish to search:

- ☐ Academy of Natural Sciences (Philadelphia, PA)
- ☐ Bishop Museum (Hawaii)
- ☐ Cornell University (Ithaca, NY)
- ☐ Field Museum (Chicago, IL)
- ☐ Harvard University (Cambridge, MA)
- ☐ University of Alabama (Tuscaloosa, AL)
- ☐ University of Kansas Natural History Museum (Lawrence, KS)
- ☐ University of Michigan Museum of Zoology (Ann Arbor, MI)

Search Options

Maximum hits desired (Max=1000) 250

Additional fields (currently implemented): Title C No

Home Species Analyst

KU NMH & BRC  
Informatics  
LifeMapper  
Specify  
 SEEK

Species Analyst  
Search  
Collections  
Documentation  
Darwin Core  
DiGIR  
Downloads

ISA Affiliated Projects  
FishNet  
MaNIS  
HerpNet  
DiGIR  
GBIF  
GBIF

© 1998-2004

**Species Analyst**

**Author:** Dave Vieglais  
**Revision:** 1.6  
**Date:** 2003-06-30

The Species Analyst is a research project developing standards and software tools for access to the world's natural history collection and observation databases. The Species Analyst project is based at the [University of Kansas Natural History Museum and Biodiversity Research Center](#).

There are currently over 120 natural history collection databases being served through the Species Analyst Network, with collections located all over the world. All of these collections may be searched simultaneously through the Species Analyst, with the combined information returned to you in one of a variety of formats.

The original, and currently dominant form of the Species Analyst network is based on the ANSI/NISO Z39.50 standard for information retrieval. DiGIR, a replacement for Z39.50, is being actively developed by a number of developers including the creator of the Species Analyst network, and will gradually replace the Z39.50 infrastructure over time.

The Species Analyst project has given rise to, or provides the technical infrastructure for a number of other projects including:

- FishNet
- MaNIS (the first full DiGIR implementation)
- HerpNet (following on from MaNIS)
- Canadian Biodiversity Information Facility
- Global Biodiversity Information Facility (using DiGIR as the primary information retrieval protocol)

View document source: Generated on: 2004-02-04 21:02 UTC. Generated by Docutils from restructuredText source.

**FishNet2 Home : Portal Home**

Powered By **DiGIR**

Select type of network query

- ☒ View specimen records
- ☐ View taxonomic index
- ☐ View geographic index
- ☐ View taxonomy by geography index
- ☐ View gazetteer

Build query

**GBIF** INTEGRATED PUBLISHING TOOLKIT™  
The first open access to biodiversity data

Home About

Hosted resources available through this IPT

Logo	Name	Organisation	Type	Subtype	Records	Last modified	Last publication	Next publication
	Botany Division, Yale Peabody Museum	Not registered	Occurrence	—	151,132	2019-07-24	2019-07-24	2019-07-25 22:02:51
	Entomology Division, Yale Peabody Museum	Not registered	Occurrence	—	337,935	2019-07-24	2019-07-24	2019-07-25 23:45:32
	Invertebrate Paleontology Division, Yale Peabody Museum	Not registered	Occurrence	—	516,125	2019-07-25	2019-07-25	2019-07-26 02:54:52
	Invertebrate Zoology Division, Yale Peabody Museum	Not registered	Occurrence	—	142,283	2019-07-24	2019-07-24	2019-07-25 22:11:41
	Paleobotany Division, Yale Peabody Museum	Not registered	Occurrence	—	95,572	2019-07-24	2019-07-24	2019-07-25 21:38:31
	Vertebrate Paleontology Division, Yale Peabody Museum	Not registered	Occurrence	—	79,725	2019-07-24	2019-07-24	2019-07-25 21:39:31
	Vertebrate Zoology Division - Ichthyology, Yale Peabody Museum	Not registered	Occurrence	—	46,968	2019-07-24	2019-07-24	2019-07-25 21:14:31
	Vertebrate Zoology Division - Mammalogy, Yale Peabody Museum	Not registered	Occurrence	—	33,237	2019-07-24	2019-07-24	2019-07-25 21:07:01
	Vertebrate Zoology Division - Ornithology, Yale Peabody Museum	Not registered	Occurrence	—	15,777	2019-07-24	2019-07-24	2019-07-25 21:06:01
	Vertebrate Zoology Division - Reptiles and Amphibians, Yale Peabody Museum	Not registered	Occurrence	—	150,808	2019-07-24	2019-07-24	2019-07-25 22:14:01

Showing 1 to 10 of 10

The most recently updated resources are also available as an [RSS feed](#)

IPT Version 2.3.5-4b0544 [About the IPT](#) [User manual](#) [Report a bug](#) [Request new features](#)

©2017 Global Biodiversity Information Facility.



# Darwin Core

Darwin Core is a standard maintained by the [Darwin Core Maintenance Interest Group](#). It includes a glossary of terms (in other contexts these might be called properties, elements, fields, columns, attributes, or concepts) intended to **facilitate the sharing of information about biological diversity** by providing identifiers, labels, and definitions. Darwin Core is primarily based on taxa, their occurrence in nature as documented by observations, specimens, samples, and related information.



# Occurrence

occurrenceID

catalogNumber

recordNumber

recordedBy

recordedByID

individualCount

organismQuantity

organismQuantityType

sex

lifeStage

reproductiveCondition

caste

behavior

vitality

establishmentMeans

degreeOfEstablishment

pathway

georeferenceVerificationStatus

occurrenceStatus

associatedMedia

associatedOccurrences

associatedReferences

associatedTaxa

otherCatalogNumbers

occurrenceRemarks

Occurrence		Class
Identifier	<a href="http://rs.tdwg.org/dwc/terms/Occurrence">http://rs.tdwg.org/dwc/terms/Occurrence</a>	
Definition	An existence of a dwc:Organism at a particular place at a particular time.	
Comments		
Examples	<div>a wolf pack on the shore of Kluane Lake in 1988</div> <div>a virus in a plant leaf in the New York Botanical Garden at 15:29 on 2014-10-23</div>	

# Event

eventID

parentEventID

eventType

fieldNumber

eventDate

eventTime

startDayOfYear

endDayOfYear

year

month

day

verbatimEventDate

habitat

samplingProtocol

sampleSizeValue

sampleSizeUnit

samplingEffort

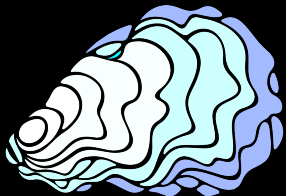
fieldNotes

eventRemarks

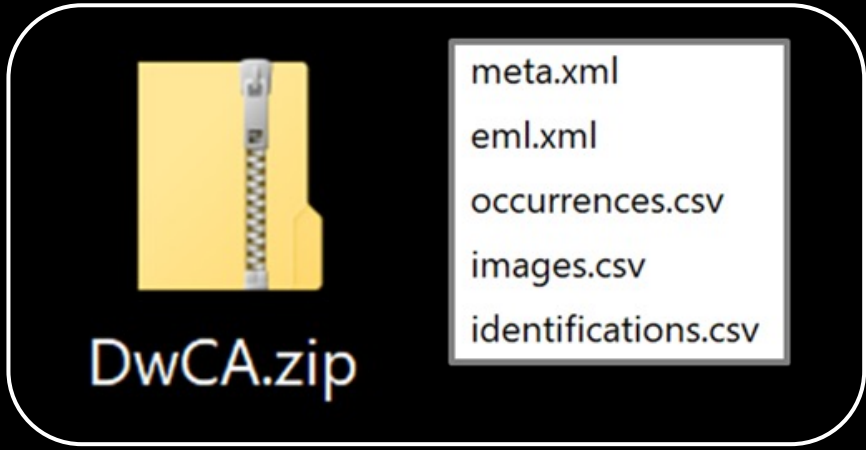
Event		Class
Identifier	<a href="http://rs.tdwg.org/dwc/terms/Event">http://rs.tdwg.org/dwc/terms/Event</a>	
Definition	An action that occurs at some location during some time.	
Comments		
Examples	<div>a specimen collecting event</div> <div>a camera trap image capture</div> <div>a marine trawl</div>	

**vitality**

Identifier	<a href="http://rs.tdwg.org/dwc/terms/vitality">http://rs.tdwg.org/dwc/terms/vitality</a>
Definition	An indication of whether a dwc:Organism was alive or dead at the time of collection or observation.
Comments	Recommended best practice is to use a controlled vocabulary. Intended to be used with records having a dwc:basisOfRecord of <b>PreservedSpecimen</b> , <b>MaterialEntity</b> , <b>MaterialSample</b> , or <b>HumanObservation</b> . This term has an equivalent in the dwciri: namespace that allows only an IRI as a value, whereas this term allows for any string literal value.
Examples	<div><b>alive</b></div> <div><b>dead</b></div> <div><b>mixedLot</b></div> <div><b>uncertain</b></div> <div><b>notAssessed</b></div>



**Eastern Seaboard**  
Mobilizing Millions of Marine Mollusks



Home

Hosted resources available through this IPT

10 resource(s) currently available

Filter:

Logo	Name	Organization	Type	Subtype	Records	Last modified	Last publication	Next publication
	Botany Division, Yale Peabody Museum	Yale University Peabody Museum	Occurrence	—	186,529	2025-01-07 13:00:07	2025-01-07 13:00:59	2025-01-08 13:00:00
	Entomology Division, Yale Peabody Museum	Yale University Peabody Museum	Occurrence	—	413,821	2025-01-07 13:05:09	2025-01-07 13:06:50	2025-01-08 13:05:00
	Invertebrate Paleontology Division, Yale Peabody Museum	Yale University Peabody Museum	Occurrence	—	557,942	2025-01-07 13:10:00	2025-01-07 13:12:21	2025-01-08 13:10:00
	Invertebrate Zoology Division, Yale Peabody Museum	Yale University Peabody Museum	Occurrence	—	149,090	2025-01-07 13:15:01	2025-01-07 13:15:42	2025-01-08 13:15:00
	Paleobotany Division, Yale Peabody Museum	Yale University Peabody Museum	Occurrence	—	98,035	2025-01-07 13:20:02	2025-01-07 13:20:33	2025-01-08 13:20:00
	Vertebrate Paleontology Division, Yale Peabody Museum	Yale University Peabody Museum	Occurrence	—	79,770	2025-01-07 13:25:03	2025-01-07 13:25:24	2025-01-08 13:25:00
	Vertebrate Zoology Division - Herpetology, Yale Peabody Museum	Yale University Peabody Museum	Occurrence	—	75,760	2025-01-07 13:30:04	2025-01-07 13:30:31	2025-01-08 13:30:00
	Vertebrate Zoology Division - Ichthyology, Yale Peabody Museum	Yale University Peabody Museum	Occurrence	—	77,592	2025-01-07 13:35:01	2025-01-07 13:35:22	2025-01-08 13:35:00
	Vertebrate Zoology Division - Mammalogy, Yale Peabody Museum	Yale University Peabody Museum	Occurrence	—	18,866	2025-01-07 13:40:02	2025-01-07 13:40:13	2025-01-08 13:40:00
	Vertebrate Zoology Division - Ornithology, Yale Peabody Museum	Yale University Peabody Museum	Occurrence	—	163,560	2025-01-07 13:45:03	2025-01-07 13:45:44	2025-01-08 13:45:00

Showing 1 to 10 of 10

previous1next

The most recently updated resources are also available as an [RSS feed](#).

GBIF Integrated Publishing Toolkit (IPT) Version 3.0.6

[About the IPT](#) | [User manual](#) | [Report a bug](#) | [Request new feature](#)



## Search Records

[Help](#) [Reset](#)

search all fields

☐ Must have media ☐ Must have map point

[Filters](#) [Mapping](#) [Sorting](#) [Download](#)

Add a field ▼ [Clear](#)

**Scientific Name**  ✕

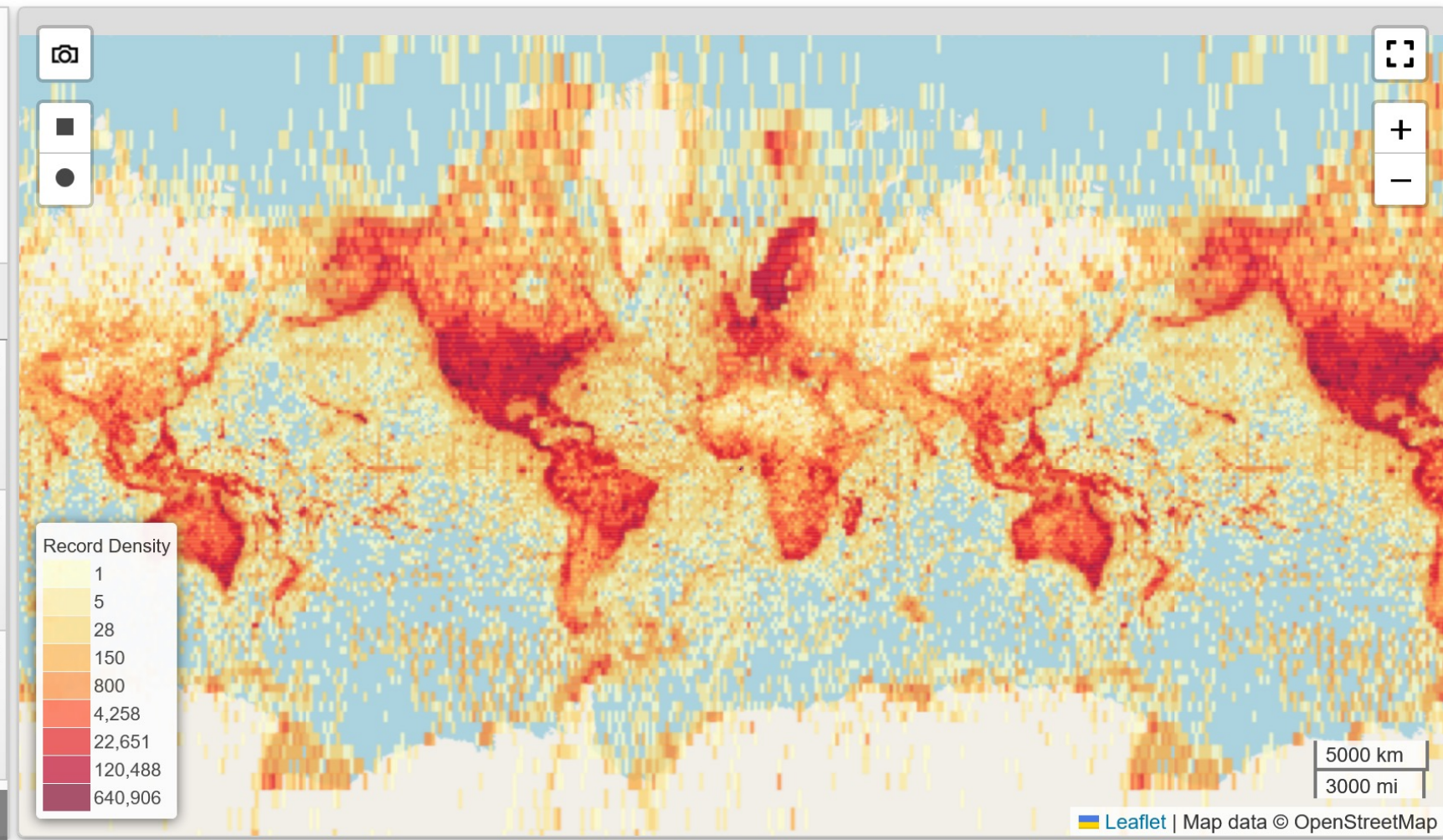
☐ Present ☐ Missing ☐ Fuzzy

**Date Collected** Start:  End:  ✕

☐ Present ☐ Missing

**Country**  ✕

☐ Present ☐ Missing



# Thematic Collections Networks

- InvertNet
- Tri-trophic (TTD)
- Lichens & Bryophytes (LBCC)
- New England Vascular Plants (NEVP)
- PaleoNICHES
- MacroFungi Collection Consortium (MaCC)
- Symbiota Collections of Arthropods (SCAN)
- Fossil Insect Collaborative (FIC)
- Vouchered Animal Communication Signals (VACS)
- Great Lakes Invasives
- InvertEBase
- The Key to the Cabinets
- Microfungi Collections Consortium (MiCC)
- Fossil Marine Invertebrates EPICC
- Cretaceous World
- LepNet
- The Mid-Atlantic Megalopolis MAM
- oVert
- SoRo
- Endless Forms
- The Pteridological Collections Consortium
- Capturing California's Flowers
- Pacific Island Land Snails
- Vascular Flora of the South-Central United States
- Parasite-Host Associations
- DigIn
- ESB
- Bryophytes and Lichens
- Tropical African Plants (TAP)

Thank You