Getting Data Out There

Now that you have it, what can you do with it?

Janeen Jones
Sharon Grant
Pete Herbst
Kate Webbink
Rob Zschernitz

Axiell North America
Midwest Roadshow
2018

Data Aggregators - where does all the data go?

GBIF Global Biodiversity Information Facility

VerNet

iDigBio Integrated Digitized Biocollections

Atlas of Living Australia ala.org.au

fossilworks

GRBio GLOBAL REGISTRY OF BIODIVERSITY REPOSITORIES

MORPHO SOURCE

EU BON european biodiversity portal

India Biodiversity Portal

BIODIVERSITY INDIA India Biodiversity Portal

IPT - Integrated Publishing Toolkit

GBIF INTEGRATED PUBLISHING TOOLKIT [IPT]
free and open access to biodiversity data

email ••••••••• login ENGLISH

Home About

Hosted resources available through this IPT

Filter:

Logo	Name	Organisation	Type	Subtype	Records	Last modified	Last publication	Next publication
	Field Museum of Natural History (Botany) Seed Plant Collection	Field Museum of Natural History	Occurrence	Specimen	576,367	2018-03-05	2018-03-05	--
	Field Museum of Natural History (Botany) Bryophyte Collection	Field Museum of Natural History	Occurrence	Specimen	115,932	2018-03-05	2016-12-19	--
	Field Museum of Natural History (Botany) Fungi Collection	Field Museum of Natural History	Occurrence	Specimen	62,851	2018-03-05	2018-03-05	--
	Field Museum of Natural History (Botany) Lichen Collection	Field Museum of Natural History	Occurrence	Specimen	55,370	2018-03-05	2018-03-05	--
	Field Museum of Natural History (Botany) Pteridophyte Collection	Field Museum of Natural History	Occurrence	Specimen	70,784	2018-03-05	2018-03-05	--
	Field Museum of Natural History (Geology) Fossil Invertebrates Collection	Field Museum of Natural History	Occurrence	Specimen	62,149	2018-01-23	2017-01-06	--
	Field Museum of Natural History (Geology) Paleobotany Collection	Field Museum of Natural History	Occurrence	Specimen	22,851	2018-01-23	2017-01-27	--
	Field Museum of Natural History (Zoology) Amphibian and Reptile Collection	Field Museum of Natural History	Occurrence	Specimen	285,342	2018-01-23	2017-02-06	--

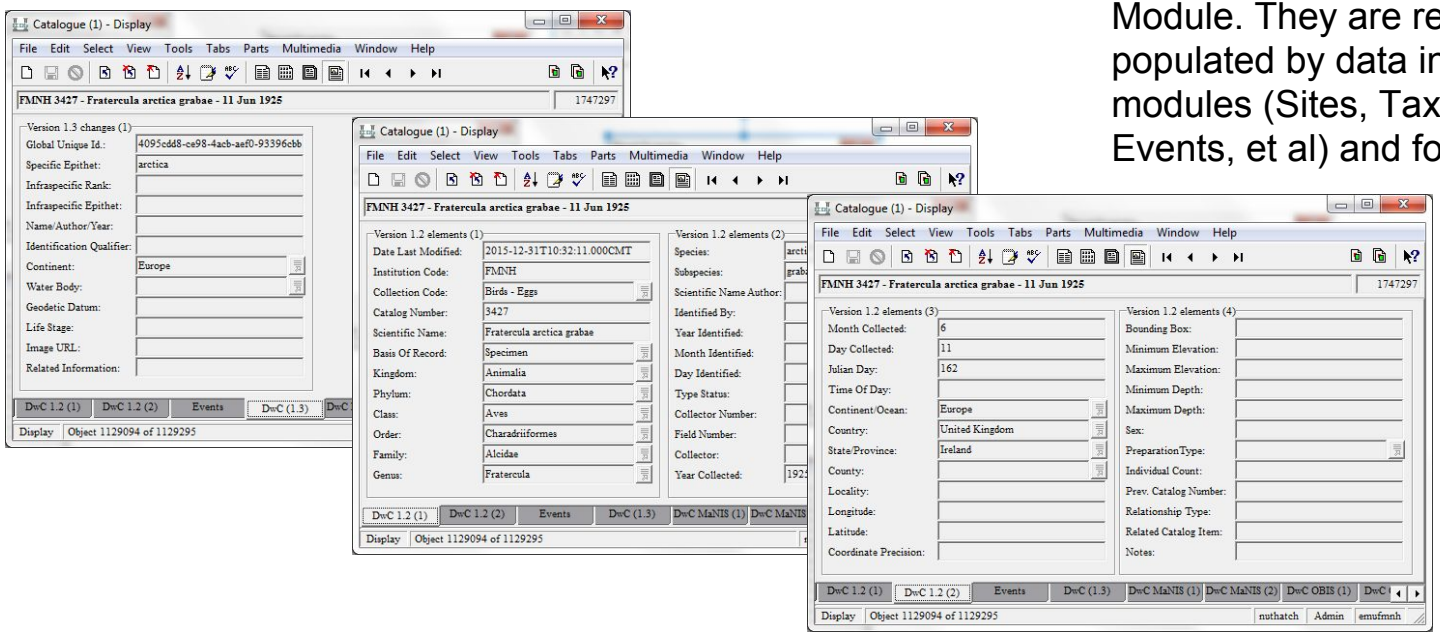
The process is set up to allow **advanced users** to control **what data** is exported and **when** it is exported to refresh to IPT.

This allows users to place their data in one place and make it available for all data aggregators to “pick it up.”

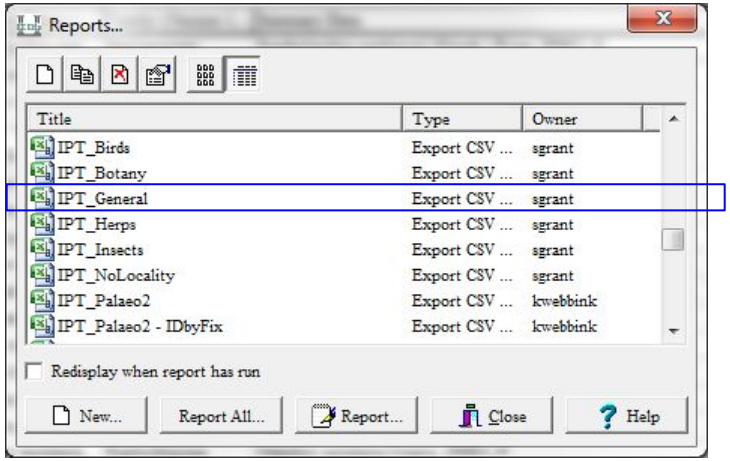
Darwin Core tabs

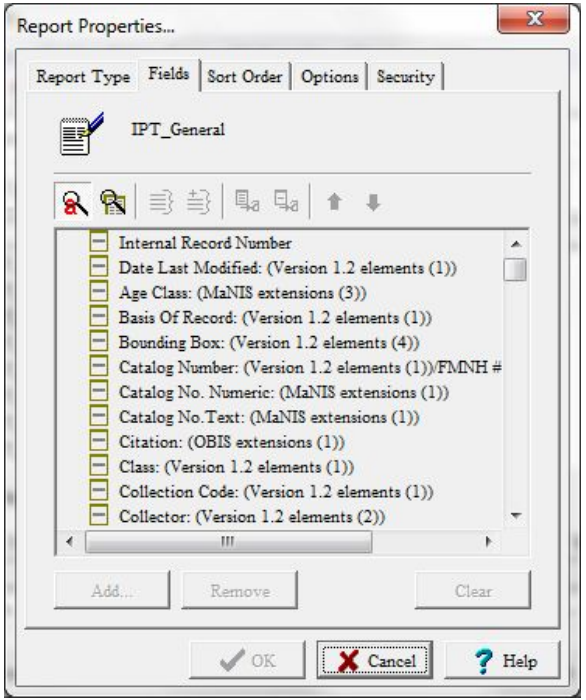
These tabs pull data from other modules and map from various catalogs in a “standard” way. Allows for a single report to pull data to export for all catalogs. Also allows for data standards from each discipline to be used in the same field.

Darwin Core fields are in the Catalog Module. They are read only fields populated by data in Catalog and other modules (Sites, Taxonomy, Collection Events, et al) and formatted in Catalog.

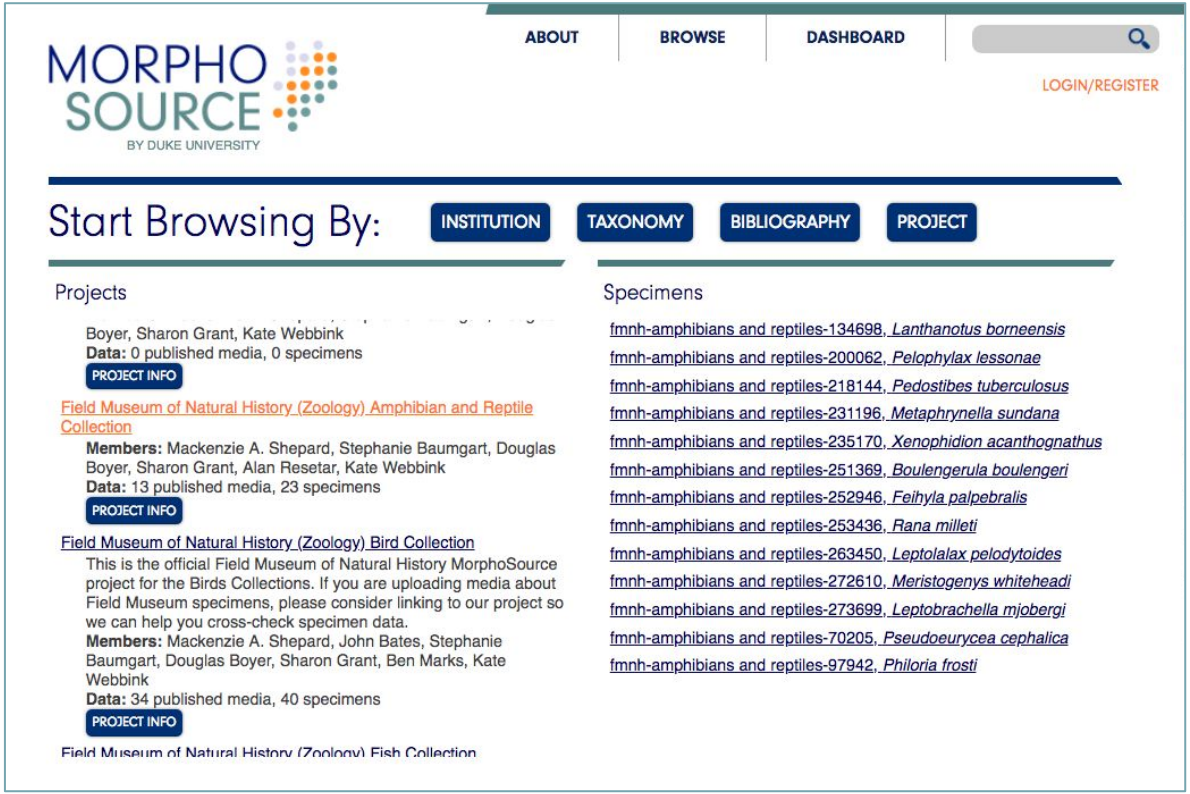


CSV reports to get data from EMu to IPT





Getting data+media back from outside repositories

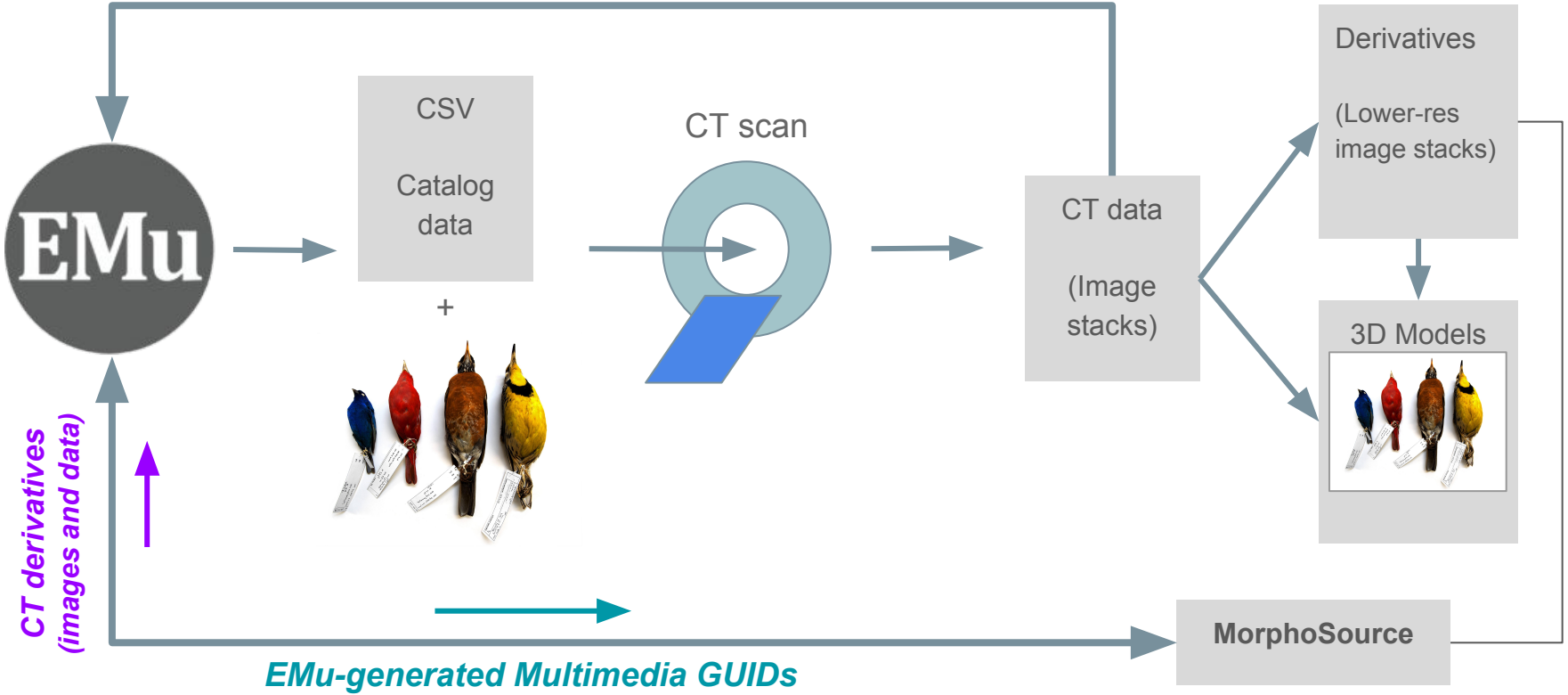


MorphoSource is an online archive for 3D data:

- **CT scan** image stacks
- **Photogrammetry** image sets
- **3D models**…

...Large media files that would otherwise travel between institutions on delicate hard drives...

Getting data+media back from outside repositories



EMu-Drupal Interface

Collection Databases served on-line

Collection Databases served on-line

Running Drupal 7 and data served via Apache Solr

- Website search data -- Apache Solr data import via Kiwi* (custom PHP library - https://github.com/palantirnet/kiwi)
  - 1 Solr core = 1 EMu module (e.g. catalogue) for that collection
- Website page and informational content -- stored in the Drupal DB

*Kiwi is a PHP command line application that maps data from KE Software's EMu collection management system into the Apache Solr search server.

Design is to allow advanced users to be able to edited the content of the site. Including searchable and viewable fields, help and information content. Includes list, page, detail views and multiple csv downloadable files.

Home » Administration

Content

CONTENT | COMMENTS | WEBFORMS

+ Add content

SHOW ONLY ITEMS WHERE

status: any

type: any

Filter

UPDATE OPTIONS

Publish selected content

Update

TITLE	TYPE	AUTHOR	STATUS	UPDATED	OPERATIONS
Home Page	Home Page	admin	published	04/11/2017 - 13:11	edit delete
Help	Basic page	admin	published	12/15/2016 - 15:20	edit delete
Collections Data - Birds	Basic page	sgrant	published	06/30/2016 - 12:35	edit delete
Collections Data - Mammals	Basic page	sgrant	published	06/30/2016 - 12:31	edit delete
Collections Data - Amphibians and Reptiles	Basic page	sgrant	published	06/30/2016 - 12:28	edit delete
Collections Data - Insect, Arachnids and Myriapods	Basic page	sgrant	published	06/30/2016 - 12:26	edit delete
Collections Data - Fishes	Basic page	sgrant	published	06/30/2016 - 12:24	edit delete
Collections Data - Invertebrates	Basic page	sgrant	published	06/30/2016 - 12:22	edit delete
Help - How to Parse Measurements	Basic page	kwebbink	published	02/09/2016 - 13:47	edit delete
Page not found	Basic page	admin	published	08/17/2015 - 11:09	edit delete
Contact	Webform	admin	published	08/17/2015 - 11:09	edit delete
Collections Data Survey - Fishes	Webform	sgrant	published	08/17/2015 - 11:08	edit delete

Edit Basic page Collections Data - Invertebrates

VIEW | EDIT

Title *

Collections Data - Invertebrates

Body (Edit summary)

Source | Format | Styles

Welcome to the Invertebrate Zoology Collections Data

The Collection of Invertebrates contains all invertebrate groups except insects and other primarily terrestrial arthropods. The collections currently exceed 353,000 catalogued lots (= specimen series) with continuing growth. Research and collecting traditionally focused on the phylum Mollusca with more than 330,000 cataloged lots. Non-mollusk invertebrates are represented by ca. 13,000 cataloged lots with the Arthropoda (ca. 50%), Annelida (20 %), Echinodermata (8 %), Cnidaria (7%) and Porifera (5 %) best represented.

Other FM invertebrate Searches

- Malacological Journals and Newsletters

For copyright, data usage and citation information please refer to the Field Museum Data Norms and Considerations.

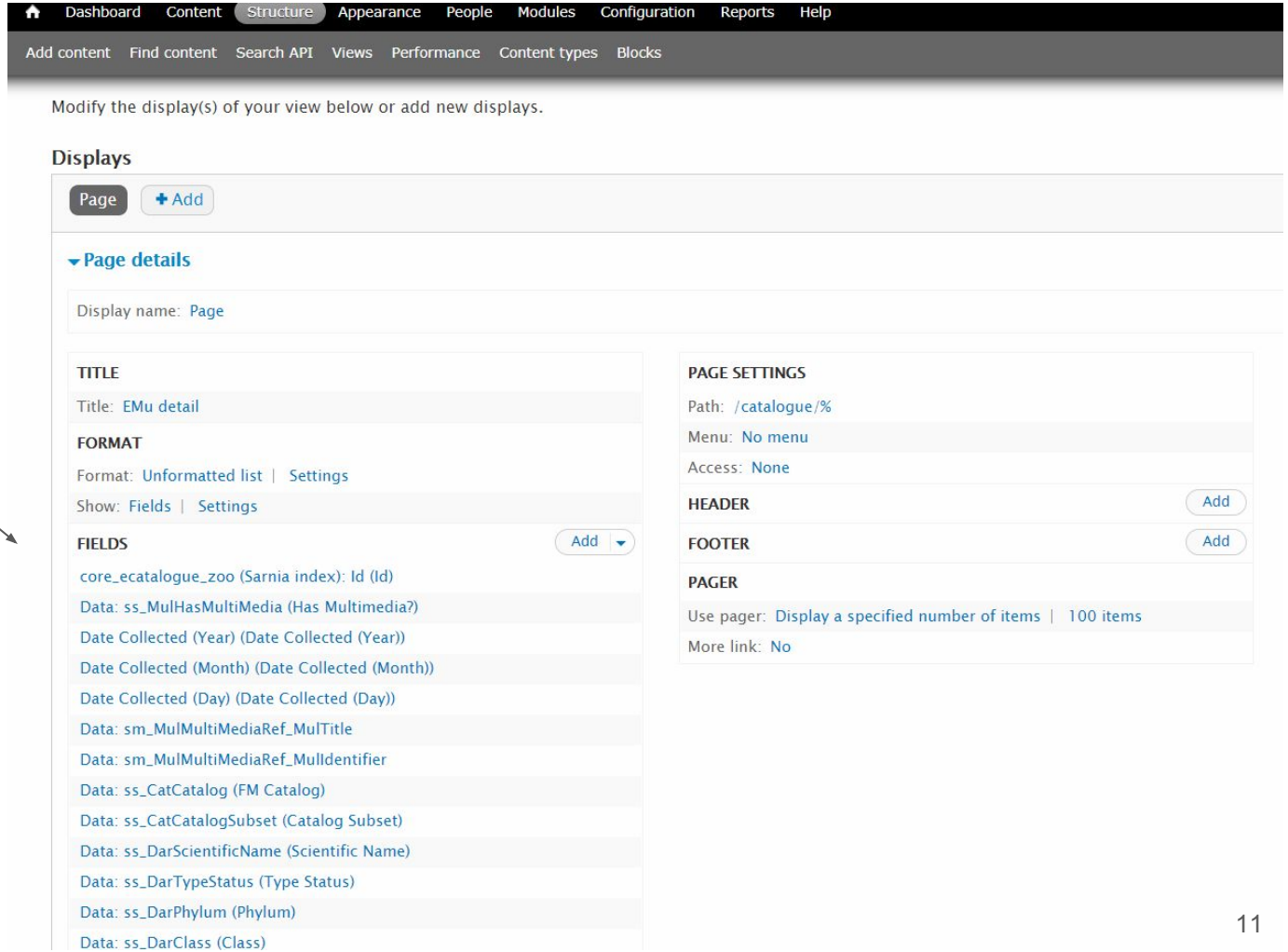
Contact the collection manager for direct assistance related to Invertebrate Zoology data.

Disable rich-text

Text format Filtered HTML

More information about text formats

EMu fields pulled and mapped via Kiwi to separate server.



Zoological Collections | The Field Museum

List | Table | Media | Need help?

Search the Zoological Collections of 1,808,700 Records.

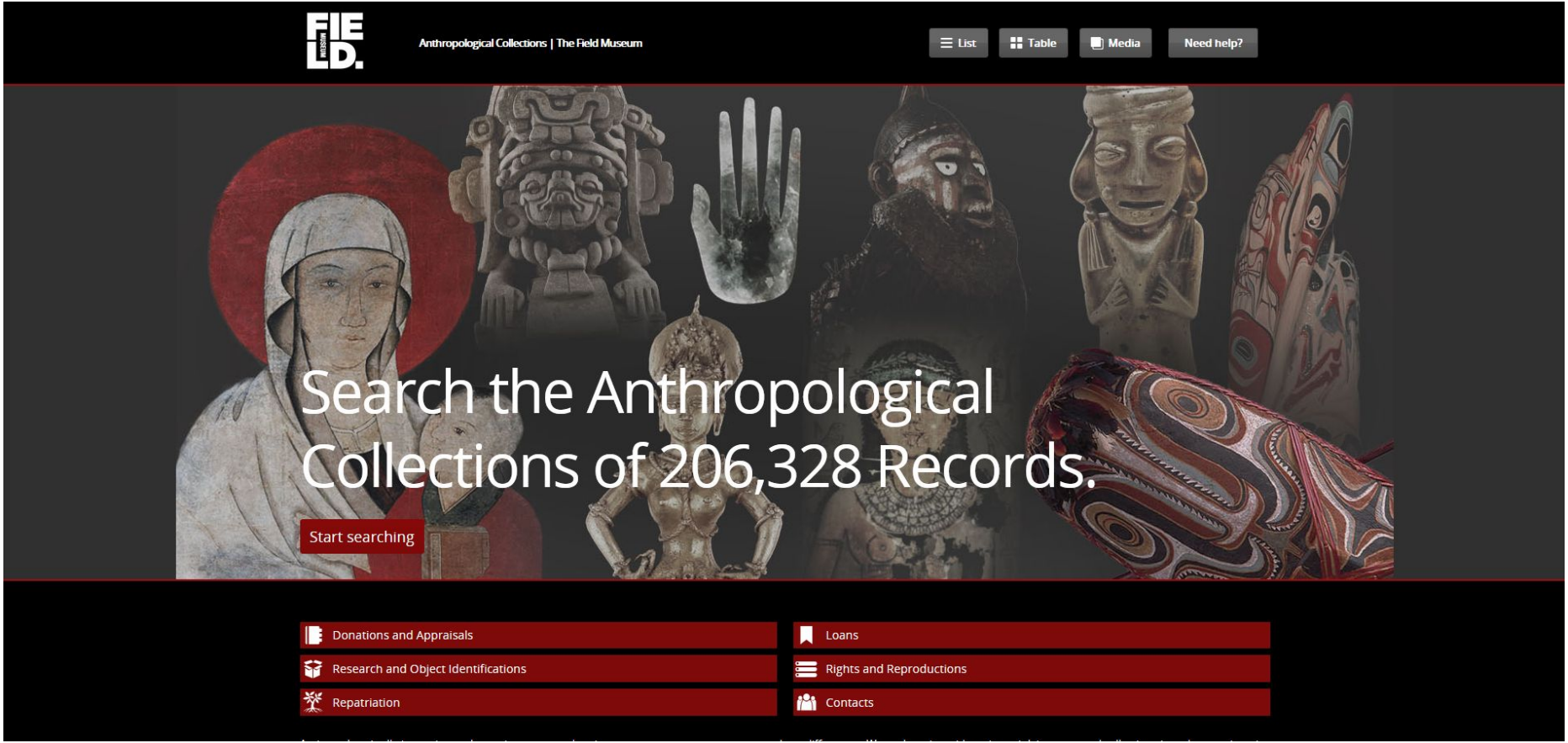
Start searching | Amphibians & Reptiles | Birds | Fishes | Insects | Invertebrates | Mammals

- Policies
- Acknowledgements
- Contact Our Collections Managers
- Database Information

The zoological collections are among the largest and most comprehensive in the world. Holdings include millions of specimens in dry storage (such as bones, feathers, shells, pinned insects), fluid-preserved specimens for anatomical research, frozen tissues for DNA studies, as well as numerous other special collections. In addition to forming the basis for research by Field Museum's zoologists, this enormous resource is utilized by scientists from around the world.

Featured Datasets | Collection Homepage

root.pri
Internet access


Anthropological Collections | The Field Museum
List
Table
Media
Need help?
Search the Anthropological Collections of 206,328 Records.
Start searching
Donations and Appraisals
Loans
Research and Object Identifications
Rights and Reproductions
Repatriation
Contacts


Botanical Collections | The Field Museum
Map
List
Table
Media
Need help?
Search the Botanical Collections of 968,690 Records.
Search Specimens
Seed Plants
Ferns
Bryophytes
Fungi
Lichens
Economic Botany
Algae
Search Taxonomy
The John G. Searle Herbarium is the fifth largest herbarium in the Western Hemisphere and one of the world's preeminent depositories of Central and South American plants. The herbarium contains almost 3 million specimens of flowering plants, ferns and lycopods, bryophytes, lichenized fungi and fungi worldwide. As of January, 2018, we have databased over 954,000 specimens, including 98,000 type specimens.
Featured Projects
Andean Flowering Plants
Costa Rican Fungi
Bryophytes
vTypes
Plowman Erythroxylum
TROPILIT
Berlin Negatives
Policies & Loans
Acknowledgements
History
Field guides and other ID tools
Database information
Staff

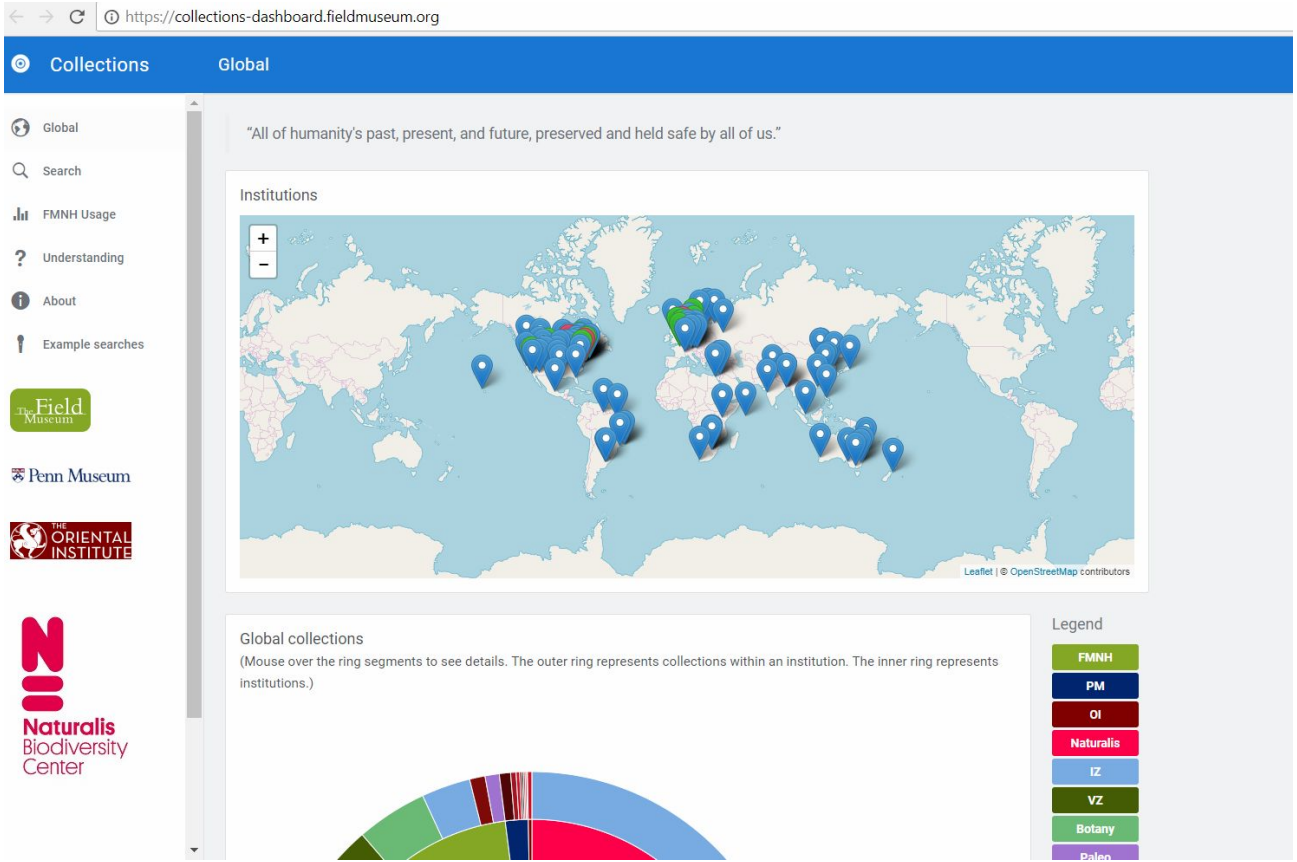
Collection Databases served on-line

Running Drupal 7 and data served via Apache Solr

- Website search data -- Apache Solr data import via Kiwi* (custom PHP library - https://github.com/palantirnet/kiwi)
  - 1 Solr core = 1 EMu module (e.g. catalogue) for that collection
- Website page and informational content -- stored in the Drupal DB

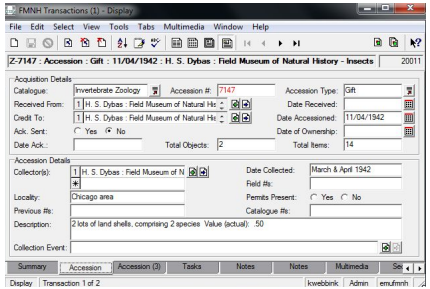
*Kiwi is a PHP command line application that maps data from KE Software's EMu collection management system into the Apache Solr search server.

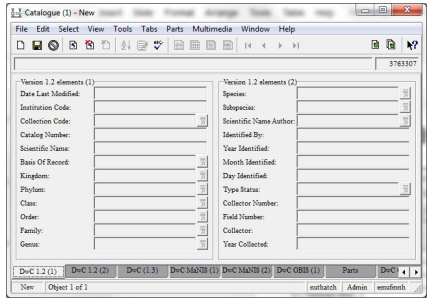
Global Collections Dashboard



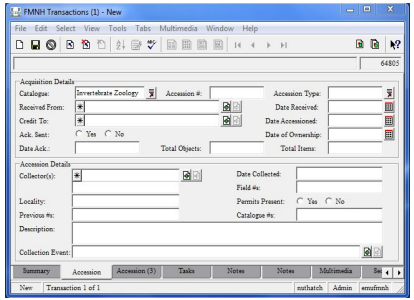
Where is the data in EMu?

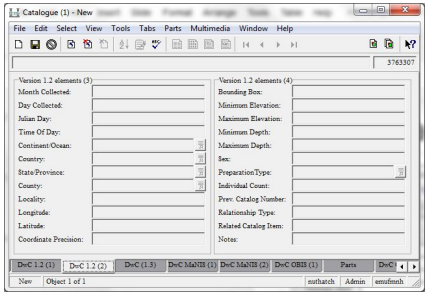
Transactions
(Accession lots)



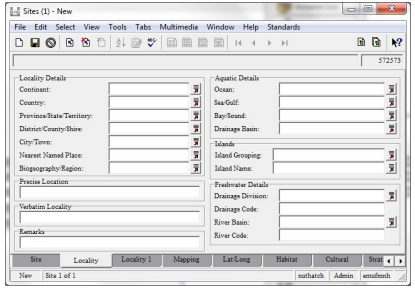


Catalog
(mostly Darwin Core fields)

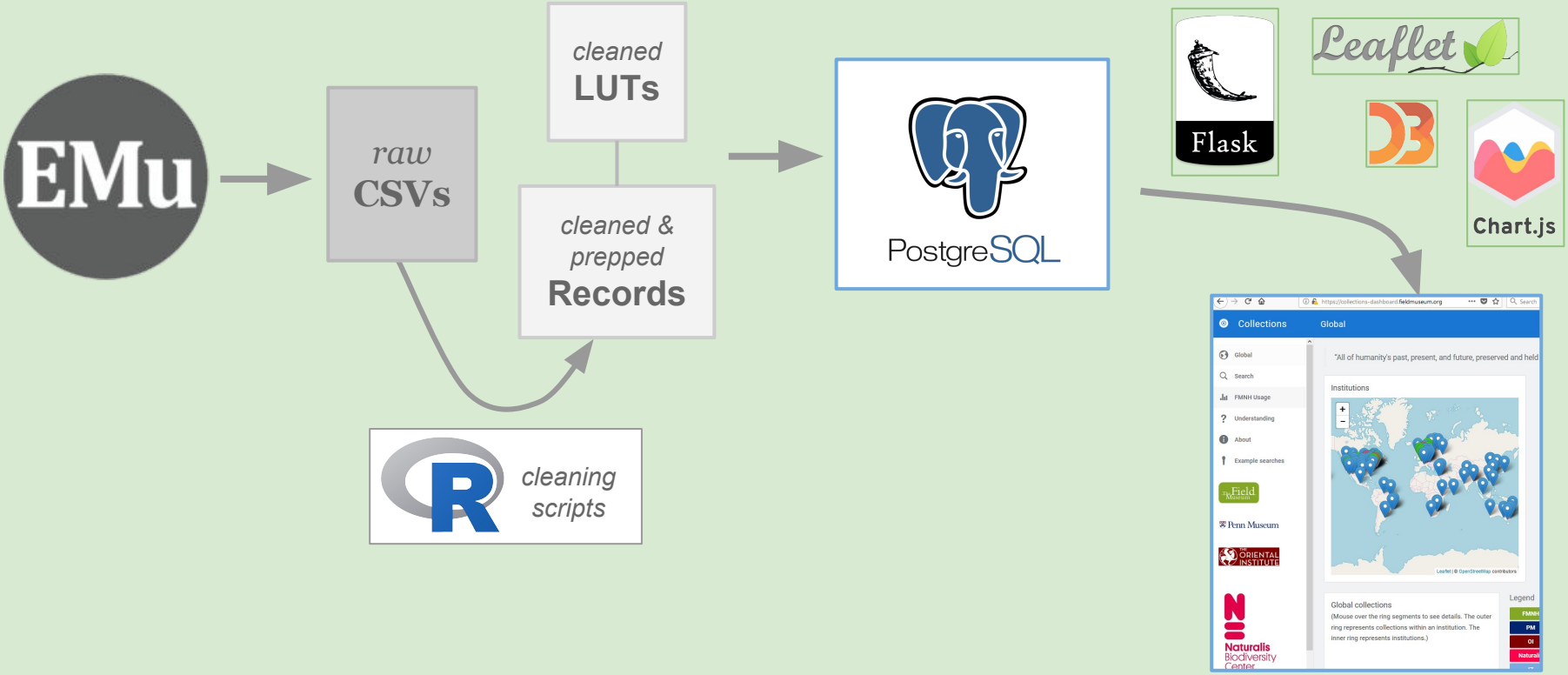




Sites



What happens to the data from EMu?



How is exported EMu data processed?

A set of R-scripts processes CSVs exported from EMu

***raw* CSVs**

irn, DarGlobalUniqueIdentifier,...

DarEarliestAge, DarEarliestEon,...

DarLatitude, DarLongitude, DarCountry, ...

DarScientificName, ...

DesEthnicGroupSubgroup_tab

cleaning scripts

cleaned **LUTs**

When	*Where*	*What*	*Who*

cleaned **Record Data**

GUID	*When*	*Where*	*What*	*Who*

Collections | Results | This site is a work in progress

- Global
- Search
- FMNH Usage
- Understanding
- About
- Example searches

Summary

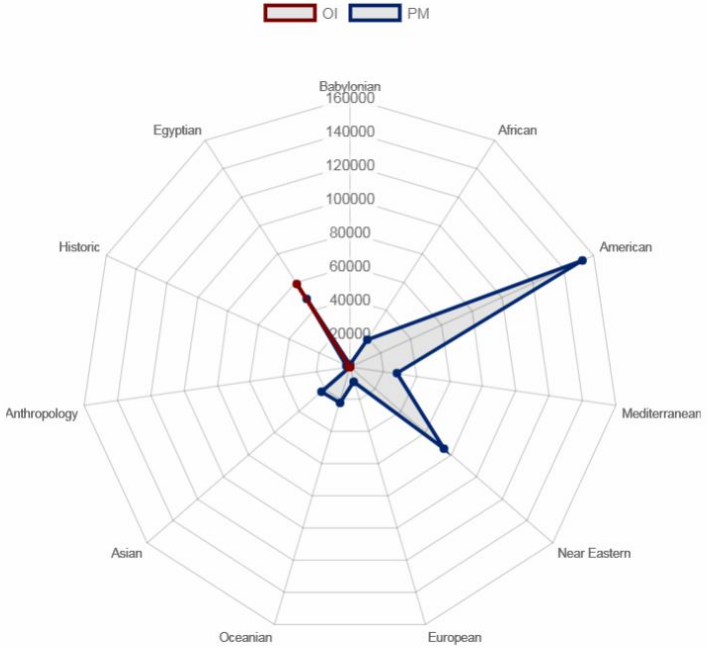
Searched by: **PM**, **OI**
438,590 records
(**379,316** PM records, **59,274** OI records)
At least **0** backlogged items

Select institution(s):
- [x] **PM catalog records**
- [x] **OI catalog records**

View records from all collections

Collection	# of records	# of backlog items
American	152,625	0
Egyptian	106,218	0
Near Eastern	73,990	0
Mediterranean	28,179	0
Asian	22,402	0
Oceanian	21,997	0

Breakdown of collections, by institution



Collections

Catalog Results

This site is a work in progress

- Global
- Search
- FMNH Usage
- Understanding
- About
- Example searches

The Field Museum

Penn Museum

THE ORIENTAL INSTITUTE

Searched by these terms: **OI**, **OI**
We found 59,274 records.

Download CSV of all results

GUID	Collection code	Institution
00034192-5fbd-4857-9d66-23c856f16705	Egyptian	OI
0003e8cc-9e4f-4ebd-a6ac-65305a9ef753	Egyptian	OI
00040f77-1e38-4d31-aeb8-e7b183e10942	Egyptian	OI
00043fac-aad1-499d-9230-237d82654366	Egyptian	OI
0005e011-9229-40d3-90bc-58606207a964	Egyptian	OI
0006c585-3b70-4507-b8eb-5a5b39e85b57	Egyptian	OI
000717ef-bfcb-4f61-ae97-36b8e3c574ec	Egyptian	OI
00098f78-1afe-46e7-a9a6-39c01dce550f	Egyptian	OI
000a78e9-6b2d-4cb3-a8c2-e1148b4b280b	Egyptian	OI
000aee86-6362-4860-9fcd-aecb47287c09	Egyptian	OI
000b63ba-295e-433f-9445-1cfa66b1d06a	Egyptian	OI

Collections Dashboard website

- Exported EMu data is processed with R and exported into CSV format (https://github.com/fieldmuseum/Collections-Dashboard-data-prep)
- Data is imported from CSV into a PostgreSQL database
- Website is a simple Python Flask application that queries the database and serves data via Javascript libraries
  - Chart.js
  - D3
  - DataTables
  - Leaflet (mapping)

Narratives based Web sites

Users can control Web content using the Narratives Module and linking to other modules and Narratives records to create content to be served on-line.

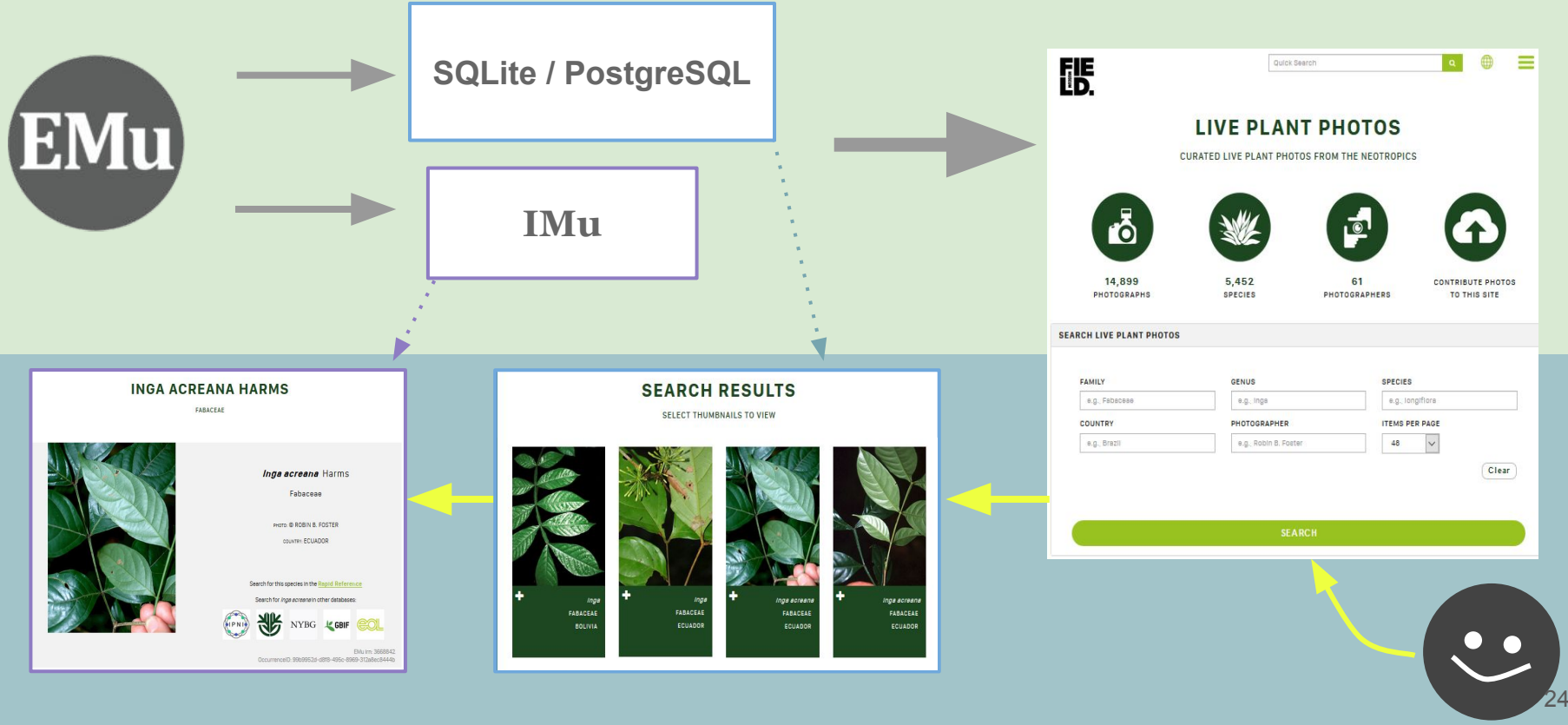
Virtual Silurian Reef Site

LinEpig

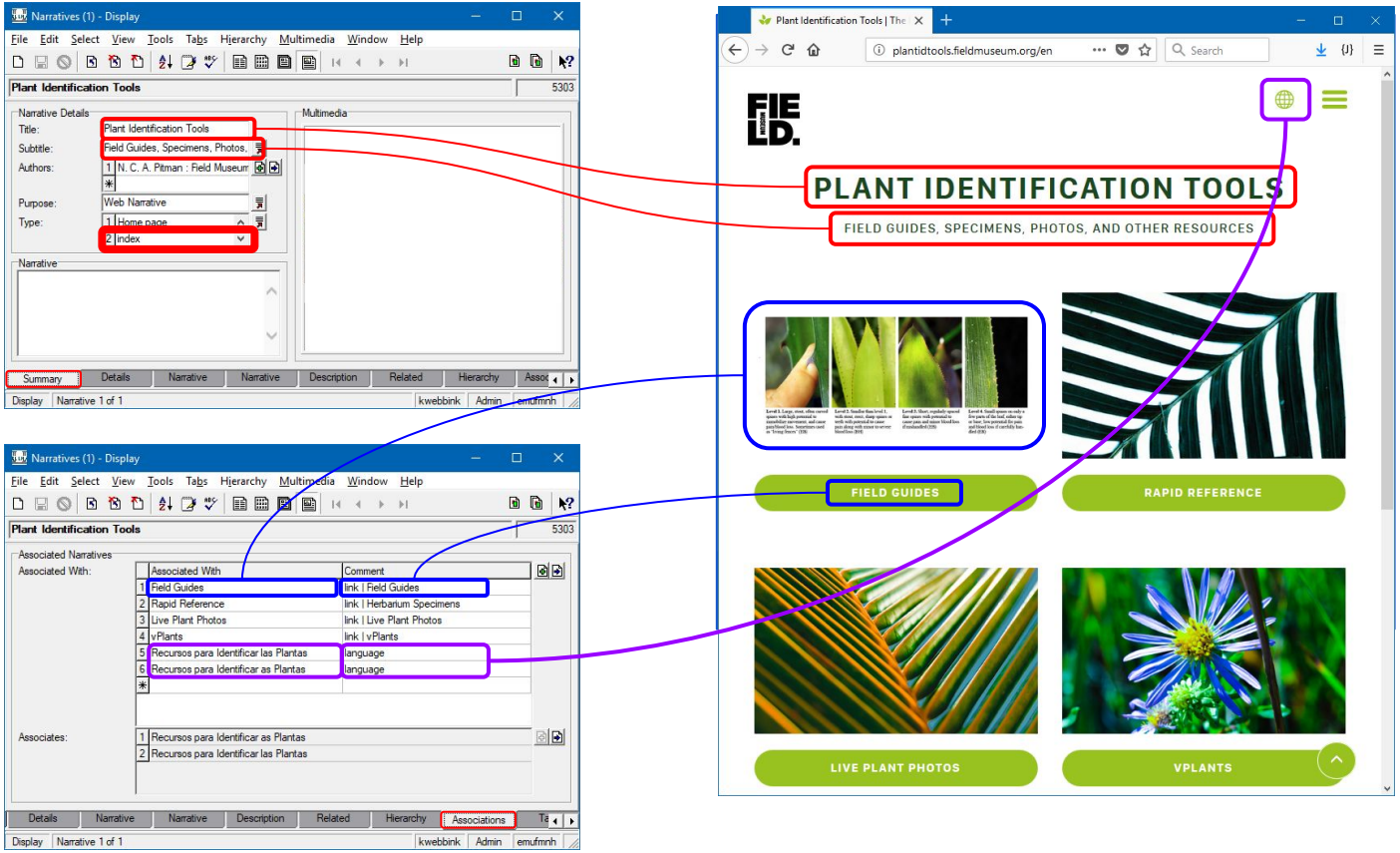
Philippines Co-Curation Portal

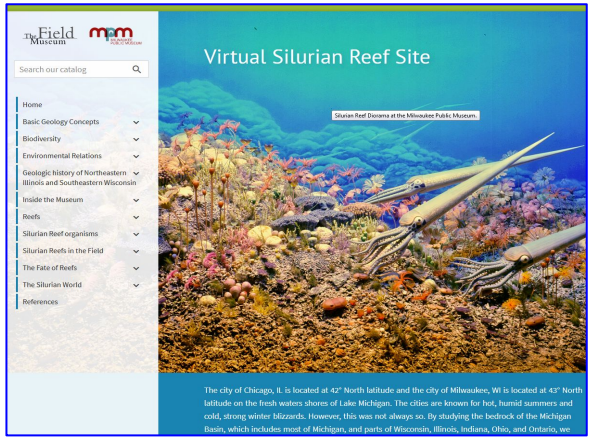
Plant Identification Tools

Getting data into [& out of] our Narrative-based websites

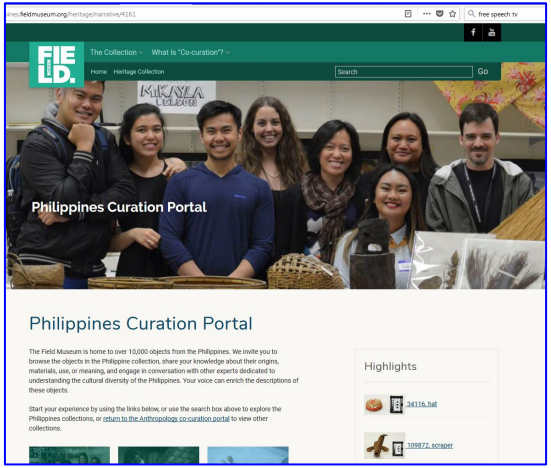


Pulling data to a Narrative-based website:

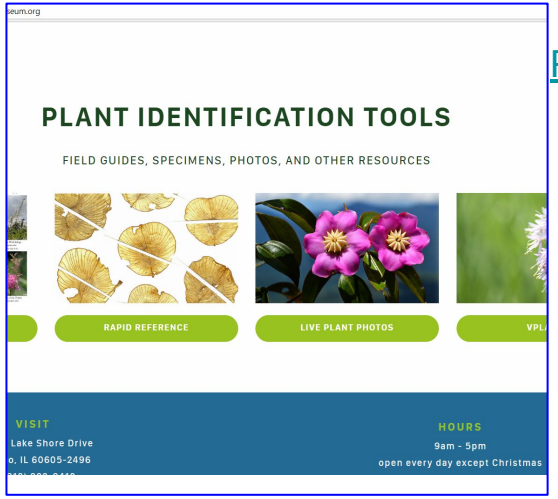




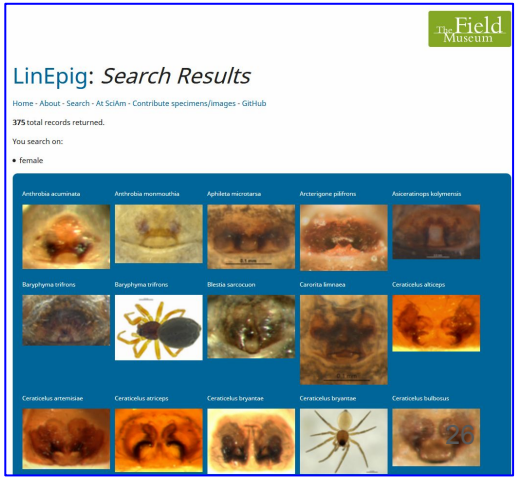
Virtual Silurian Reef Site



Philippines Curation Portal



Plant Identification Tools



LinEpig - spider ID gallery

Narratives websites are designed to handle:

- Different disciplines
- Different institutions
- Different purposes
- Different requirements

But that requires each project to follow standards so that the content can be handled and documented in EMu.

And in the interest of reusability we have 4 sets of documents for each project:

Development:
1) Requirements (wireframes)
2) Comments and feedback
Handover:
3) Wrap-up
4) How-to

Getting data into our Narrative-based websites

Export data from EMu via:

- IMu API (http://imu.mel.kesoftware.com/doc/)
- Exports module (XML export)

We typically use the PHP web framework Laravel (https://laravel.com)

Website data is retrieved in two ways:

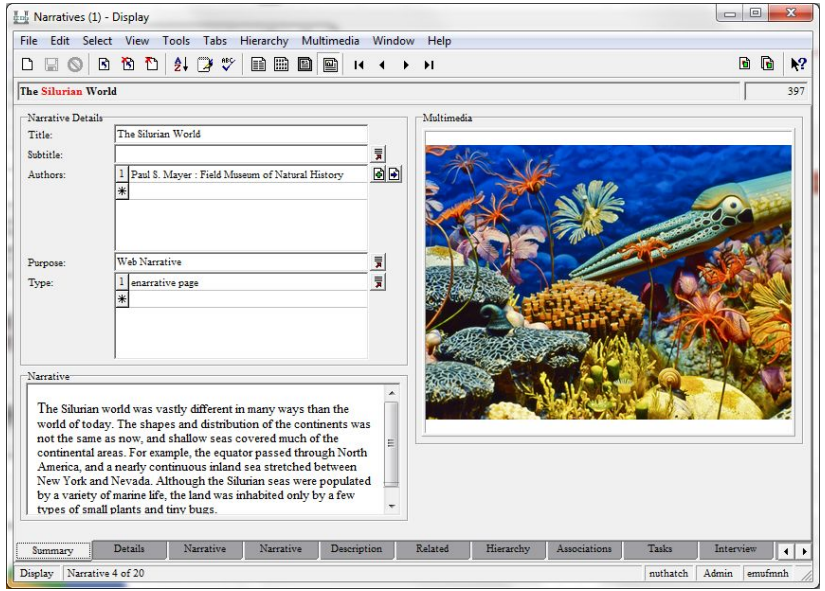
1. Stored in an intermediary database (e.g. SQLite, PostgreSQL)
2. Retrieved on-the-fly using IMu

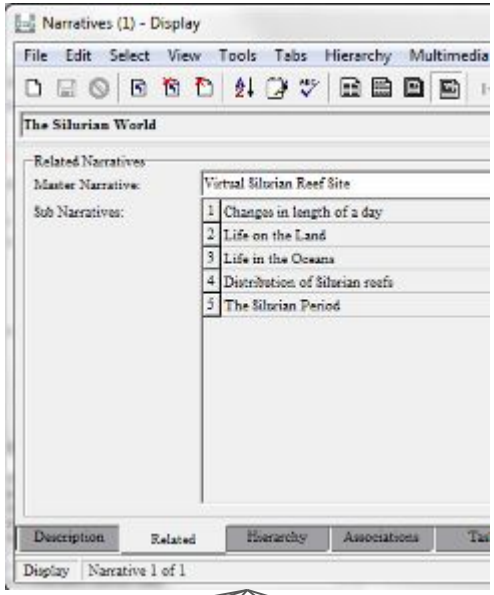
Searching -- store search data in an intermediary database for search speed. Querying via IMu is either too slow or will not work.

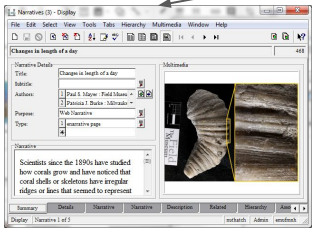
Individual module record pages -- pull data directly from EMu (using IMu), so we are able to serve a more up-to-date version of the data. Data is cached weekly, but can be manually refreshed.

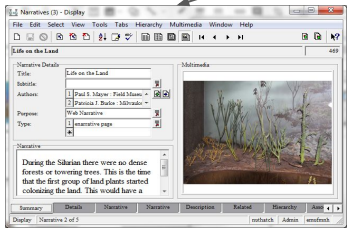
Standards - what we learned vs what we taught

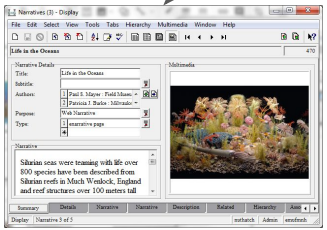
- Data standards
- Workflow standards
- Field standards
- Mapping standards

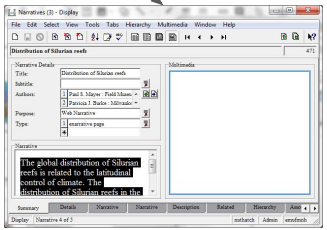

Narratives (1) - Display
File Edit Select View Tools Tabs Hierarchy Multimedia Window Help
The Silurian World
397
Narrative Details
Title: The Silurian World
Subtitle:
Authors: 1 Paul S. Mayer : Field Museum of Natural History
Purpose: Web Narrative
Type: 1 enarrative page
Multimedia
Narrative
The Silurian world was vastly different in many ways than the world of today. The shapes and distribution of the continents was not the same as now, and shallow seas covered much of the continental areas. For example, the equator passed through North America, and a nearly continuous inland sea stretched between New York and Nevada. Although the Silurian seas were populated by a variety of marine life, the land was inhabited only by a few types of small plants and tiny bugs.
Summary Details Narrative Narrative Description Related Hierarchy Associations Tasks Interview
Display Narrative 4 of 20
nuthatch Admin emufmnh

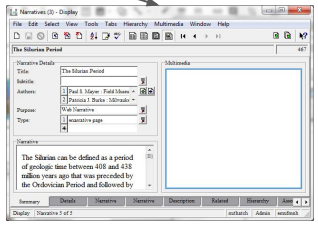

Narratives (1) - Display
File Edit Select View Tools Tabs Hierarchy Multimedia
The Silurian World
Related Narratives
Master Narrative: Virtual Silurian Reef Site
Sub Narratives:
1 Changes in length of a day
2 Life on the Land
3 Life in the Oceans
4 Distribution of Silurian reefs
5 The Silurian Period
Description Related Hierarchy Associations
Display Narrative 1 of 1

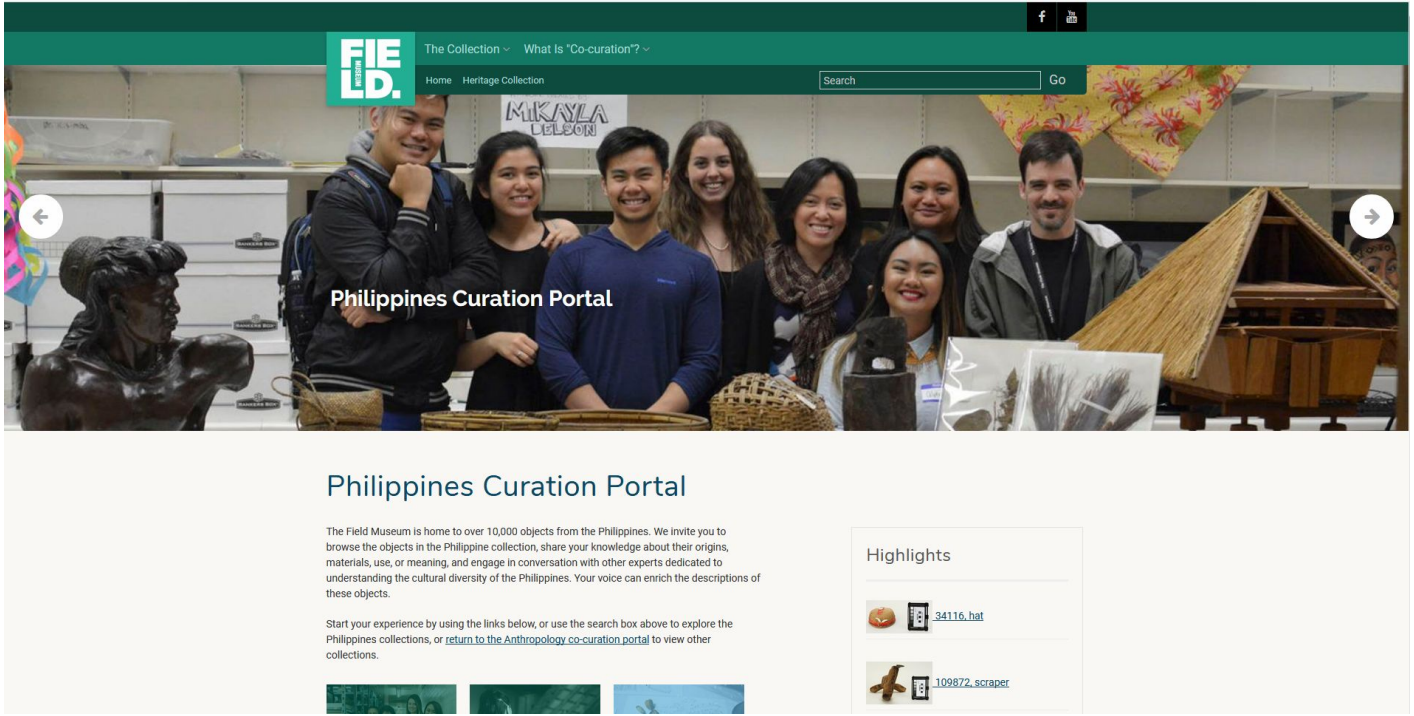











The Collection
What Is "Co-curation"?
Home
Heritage Collection
Search
Go
MIKAYLA NELSON
Philippines Curation Portal
Philippines Curation Portal
The Field Museum is home to over 10,000 objects from the Philippines. We invite you to browse the objects in the Philippine collection, share your knowledge about their origins, materials, use, or meaning, and engage in conversation with other experts dedicated to understanding the cultural diversity of the Philippines. Your voice can enrich the descriptions of these objects.
Start your experience by using the links below, or use the search box above to explore the Philippines collections, or return to the Anthropology co-curation portal to view other collections.
Highlights
34116, hat
109872, scraper

Secure | https://plantidtools.fieldmuseum.org

PLANT IDENTIFICATION TOOLS

FIELD GUIDES, SPECIMENS, PHOTOS, AND OTHER RESOURCES

FIELD GUIDES

RAPID REFERENCE

LIVE PLANT PHOTOS

VPLANTS

VISIT

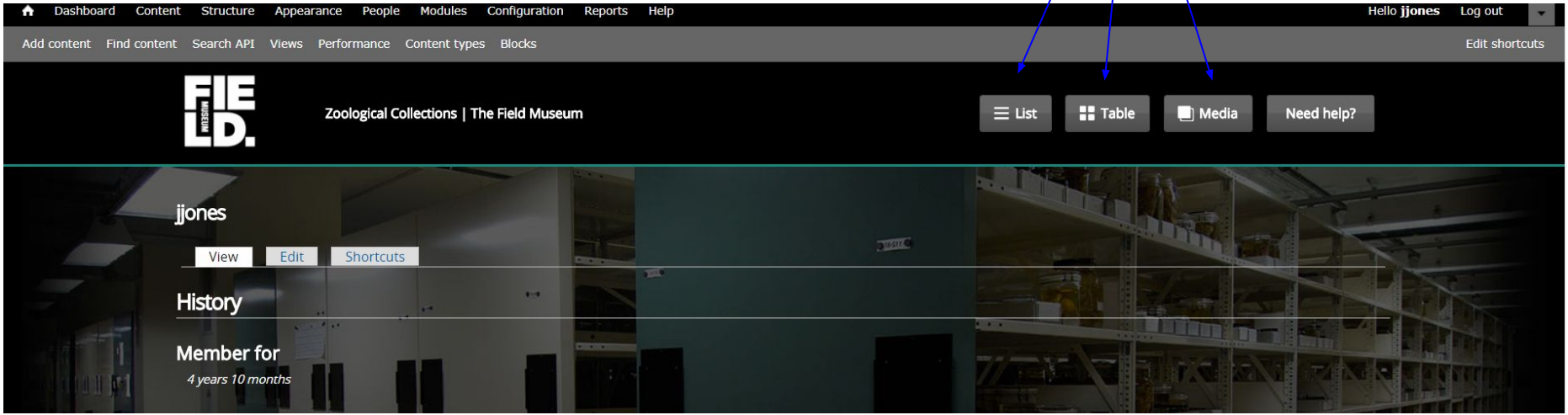
1400 S. Lake Shore Drive
Chicago, IL 60605-2496

HOURS

9am - 5pm
open every day except Christmas

Drupal menus to access other parts of site (where permissions are given) including homepage text, posting of policies etc.

Buttons to get to views (lists, tables, media)



https://artsandculture.google.com/partner/the-field-museum

Google Cultural Institute Field Museum

Google Arts & Culture

HOME EXPLORE NEARBY PROFILE

The Field Museum

Chicago, United States

One of the world's leading natural history museums, The Field Museum is home to more than 30 million artifacts and specimens, exciting exhibitions, and more than 150 scientists, conservators, and collections staff. The Field Museum inspires curiosity about life on Earth while exploring how the world came to be and how we can make it a

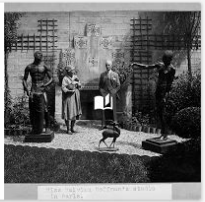
READ MORE

1 story

STORY

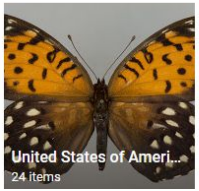
Rethinking the Sculptures of Malvina Hoffman

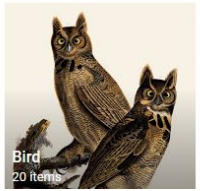
The Field Museum

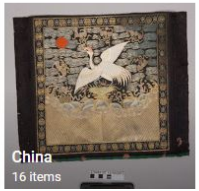


In this collection

VIEW ALL







Takeaways *(in no particular order)*

- Put data out there
- Standardize the data
  - Publicly visible dirty data's a good motivator...
- How to cope with teases?
  - &/or with willing-but-unable / unwilling-but-able data-owners
- [insert other issues/questions/shaming about reporting data here]