Global Collection Dashboard

Using data we have to uncover data we don't.

The Field Museum

Janeen Jones
Sharon Grant
Kate Webbink
Pete Herbst
Rob Zschernitz
Rusty Russell

North American Axiell User Conference, *2017*

Tell me what you do and don't have digitized…

Grant Applications

Annual/Quarterly Board meetings

Research inquiries

Peer projects

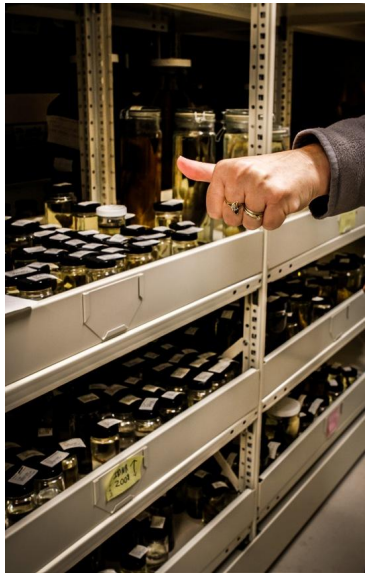
Data aggregators

…and tell me in the format I need it in.

By Collection Type
By Geographic Region
Over time
By Bio Region
Monthly estimates
By Research Project

So, what do we rely on? Estimates…

guestimates

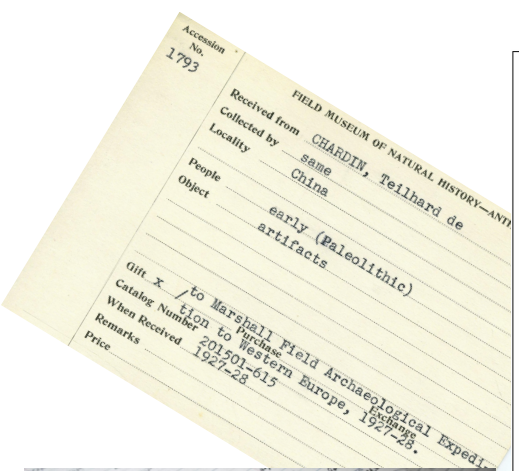


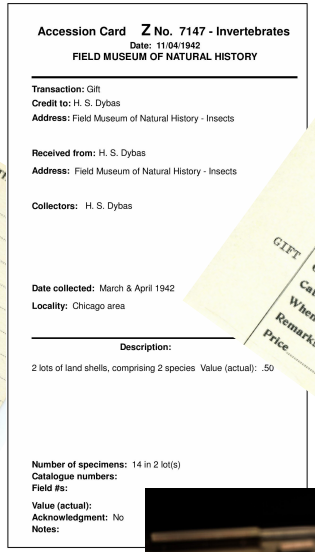
feelings

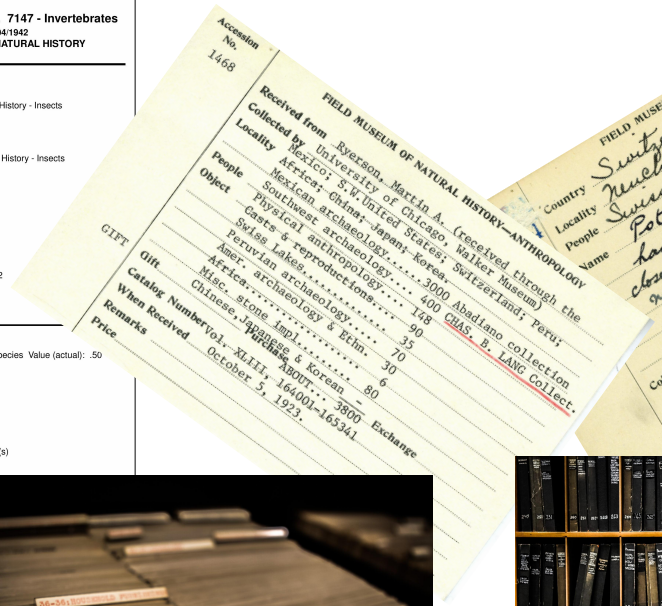
But we've done this before. Where's that data?

Acquisitions, Accessions, Invoices, Deeds of Gift...

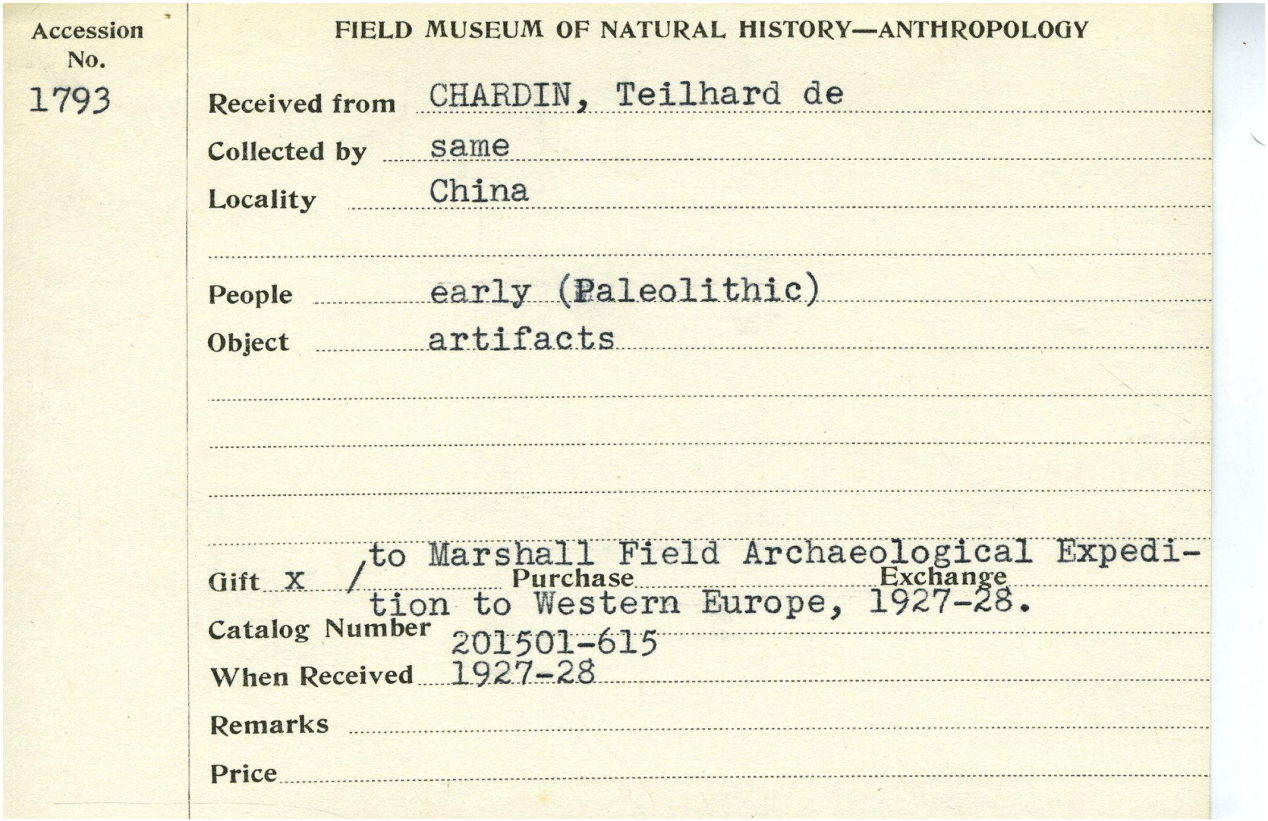
Acquisitions, Accessions, Invoices, Deeds of Gift…







Accession Cards



Accession No. 1793

FIELD MUSEUM OF NATURAL HISTORY—ANTHROPOLOGY

Received from CHARDIN, Teilhard de

Collected by same

Locality China

People early (Paleolithic)

Object artifacts

Gift x / Purchase Exchange to Marshall Field Archaeological Expedition to Western Europe, 1927-28.

Catalog Number 201501-615

When Received 1927-28

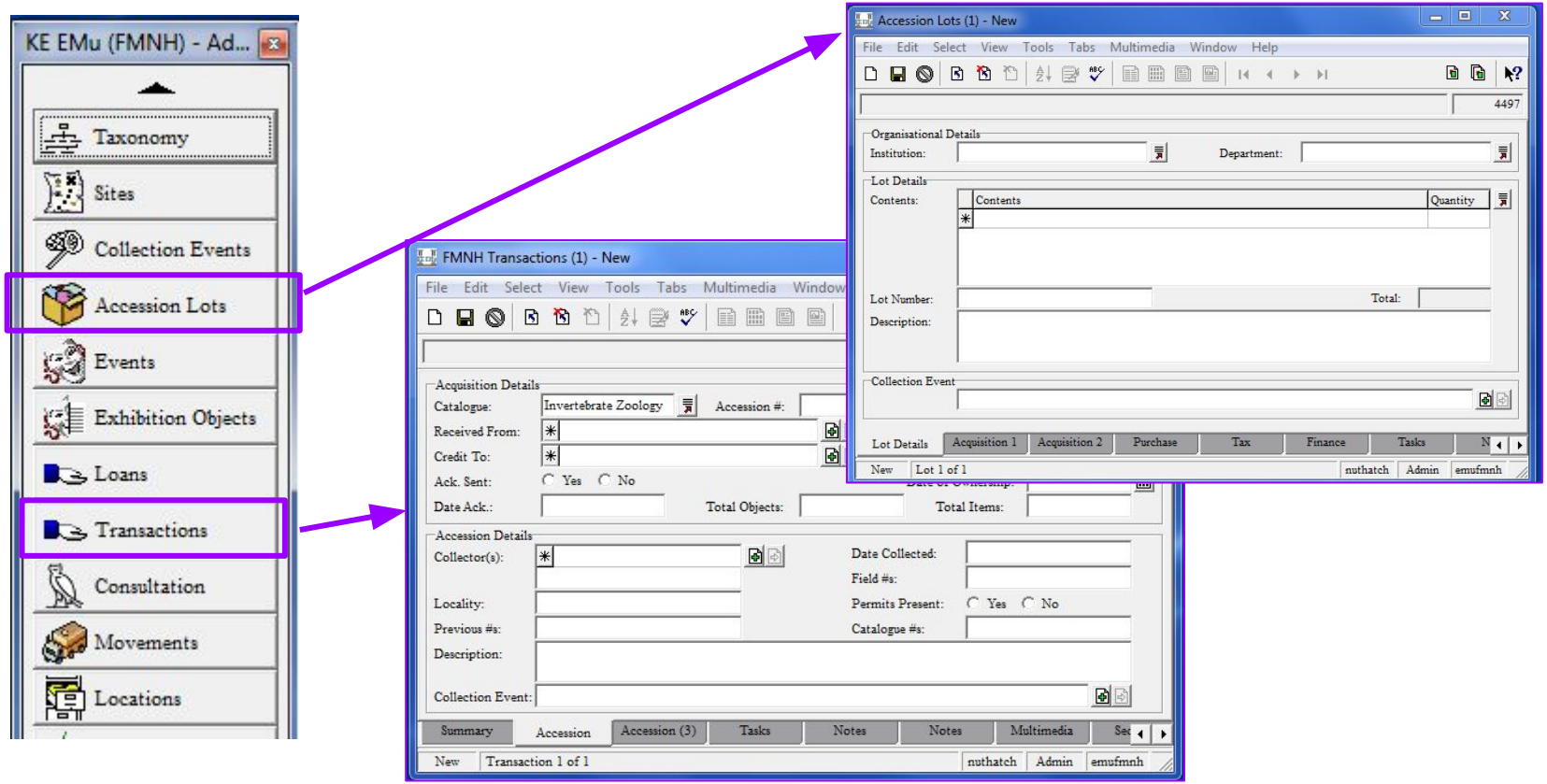
Remarks

Price

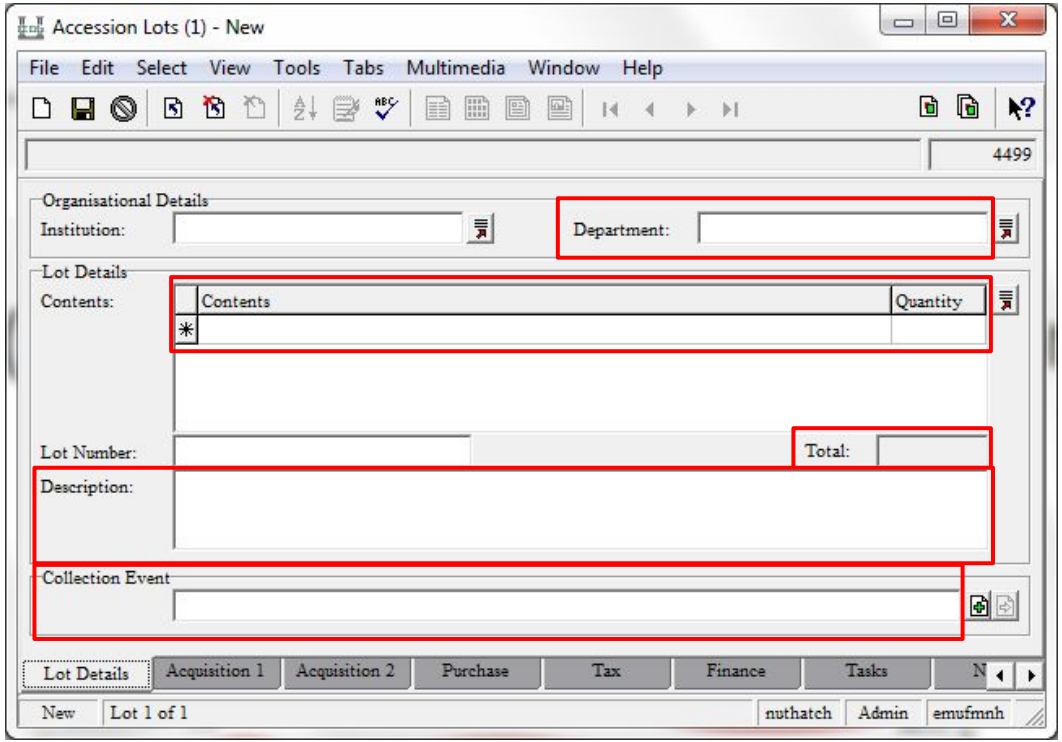
Historically -- on paper...

...but where is it in EMu?

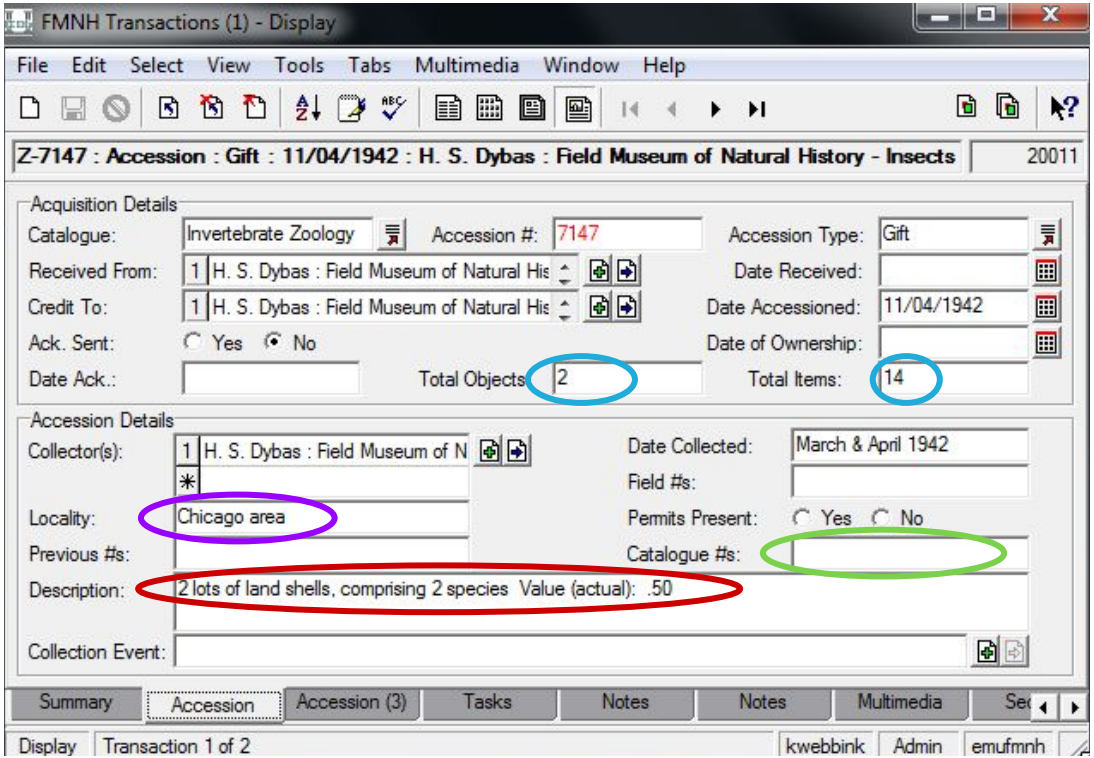
Where is that data in EMu?



Accession Lots records



Transaction (AllAcc2Tab)



Accession Card Z No. 7147 - Invertebrates

Date: 11/04/1942

FIELD MUSEUM OF NATURAL HISTORY

Transaction: Gift

Credit to: H. S. Dybas

Locality: Chicago area

Description:

2 lots of land shells, comprising 2 species Value (actual): .50

Number of specimens: 14 in 2 lot(s)

Catalogue numbers:

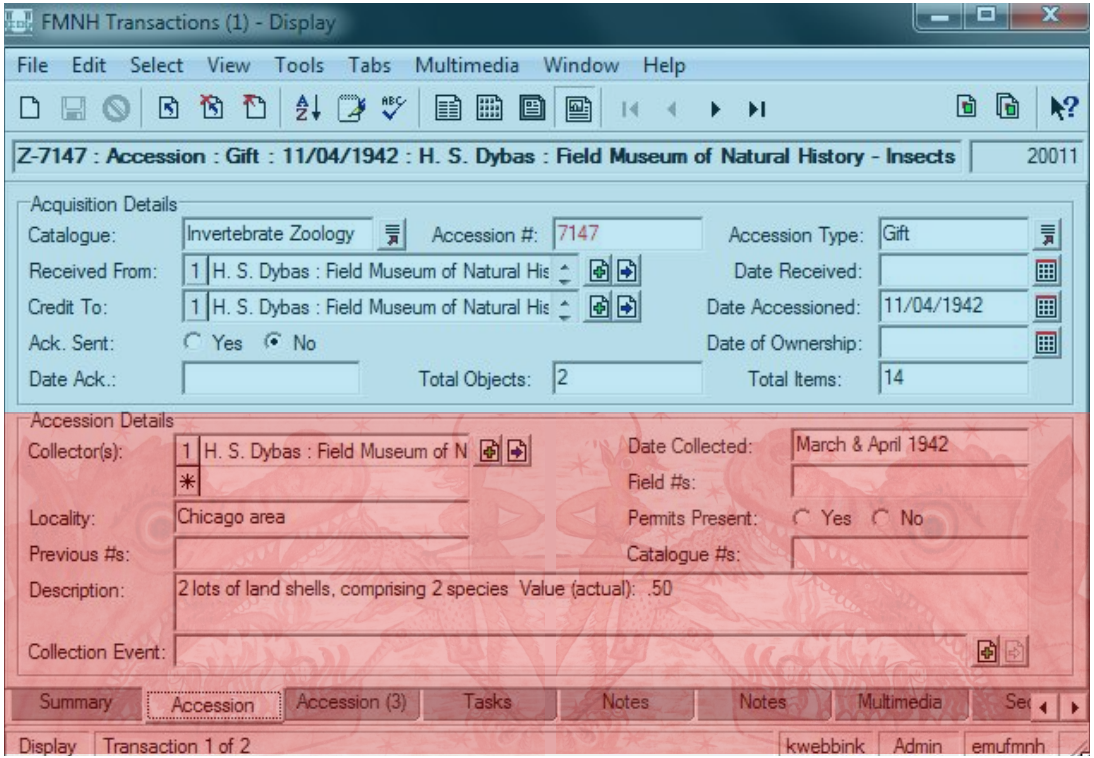
Field #s:

Value (actual):

Acknowledgment: No

Notes:

Transaction (AllAcc2Tab)



Accession Card Z No. 7147 - Invertebrates

Date: 11/04/1942

FIELD MUSEUM OF NATURAL HISTORY

Transaction: Gift

Credit to: H. S. Dybas

Locality: Chicago area

Description:

2 lots of land shells, comprising 2 species Value (actual): .50

Number of specimens: 14 in 2 lot(s)

Catalogue numbers:

Field #s:

Value (actual):

Acknowledgment: No

Notes:

So, we know where the original acquisition information lives in EMu....

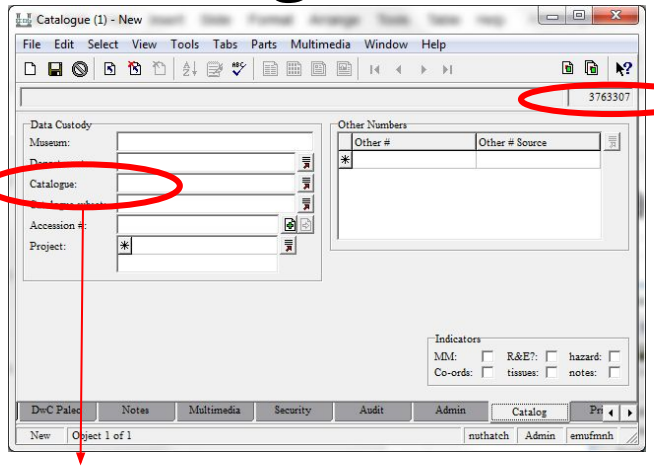
How do we determine what is backlog and what is digitized?

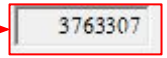
And if something is digitized, what does that mean?

- Is it rough data only?
- Is it full data?
- Has there been additional data added?

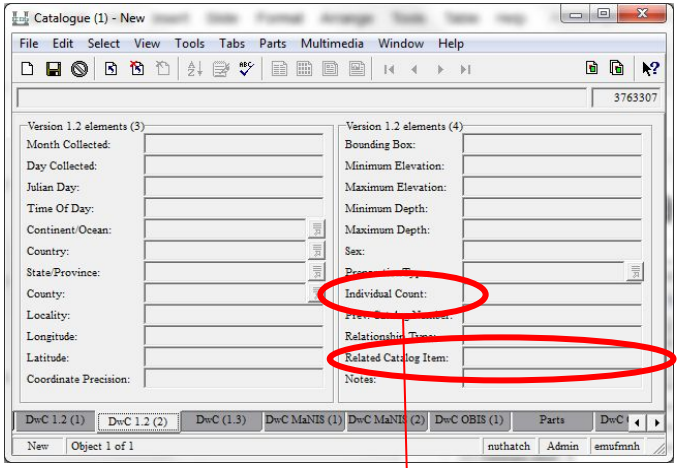
...For this information we need to turn to other modules.

Catalog

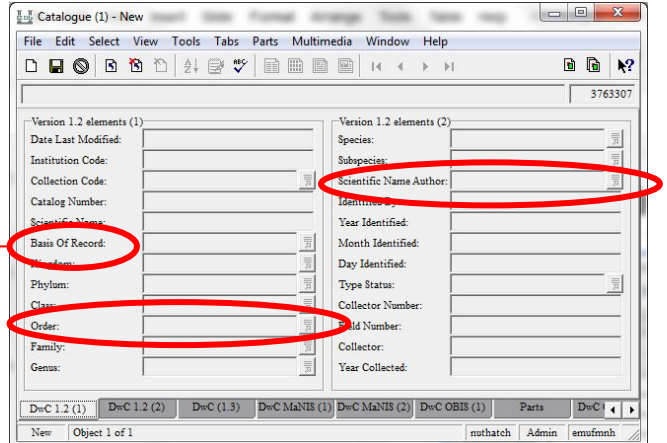

Catalogue (1) - New
File Edit Select View Tools Tabs Parts Multimedia Window Help
3763307
Data Custody
Museum:
Catalogue:
Accession #:
Project:
Other Numbers
Other #
Other # Source
Indicators
MM:
R&E?:
hazard:
Co-ords:
tissues:
notes:
DwC Palec
Notes
Multimedia
Security
Audit
Admin
Catalog
New
Object 1 of 1
nuthatch
Admin
emufmnh

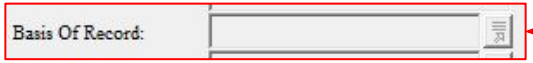

3763307

Catalogue:

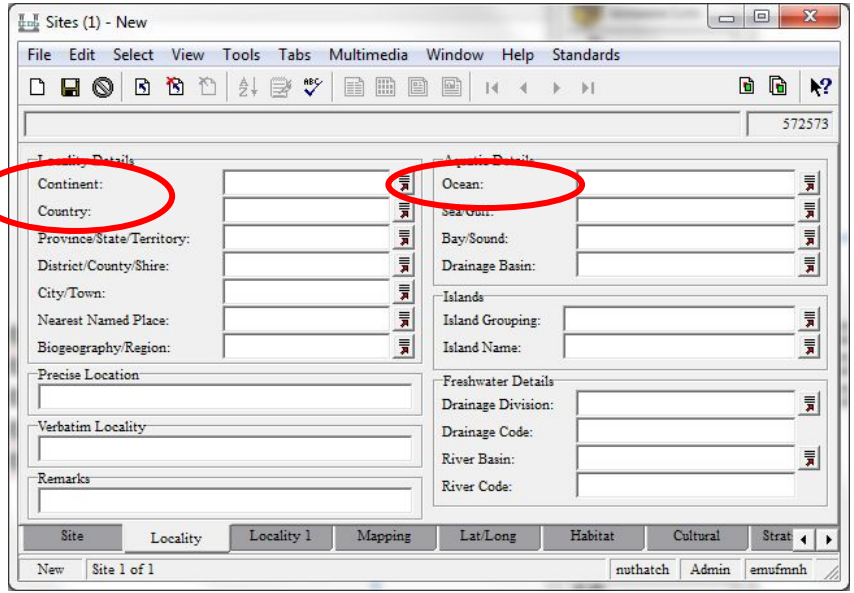

Catalogue (1) - New
File Edit Select View Tools Tabs Parts Multimedia Window Help
3763307
Version 1.2 elements (3)
Month Collected:
Day Collected:
Julian Day:
Time Of Day:
Continent/Ocean:
Country:
State/Province:
County:
Locality:
Longitude:
Latitude:
Coordinate Precision:
Version 1.2 elements (4)
Bounding Box:
Minimum Elevation:
Maximum Elevation:
Minimum Depth:
Maximum Depth:
Sex:
Individual Count:
Related Catalog Item:
Notes:
DwC 1.2 (1)
DwC 1.2 (2)
DwC (1.3)
DwC MaNIS (1)
DwC MaNIS (2)
DwC OBIS (1)
Parts
New
Object 1 of 1
nuthatch
Admin
emufmnh

Individual Count:


Catalogue (1) - New
File Edit Select View Tools Tabs Parts Multimedia Window Help
3763307
Version 1.2 elements (1)
Date Last Modified:
Institution Code:
Collection Code:
Catalog Number:
Basis Of Record:
Phylum:
Order:
Family:
Genus:
Version 1.2 elements (2)
Species:
Subspecies:
Scientific Name Author:
Year Identified:
Month Identified:
Day Identified:
Type Status:
Collector Number:
Collector:
Year Collected:
DwC 1.2 (1)
DwC 1.2 (2)
DwC (1.3)
DwC MaNIS (1)
DwC MaNIS (2)
DwC OBIS (1)
Parts
New
Object 1 of 1
nuthatch
Admin
emufmnh


Basis Of Record:

Sites (via ecollectionevents)



Fields used

(efmnhtransactions)

irn
AccCatalogue
AccTotalItems
AccTotalObjects
AccCount_tab
AccDescription_tab
AccAccessionDescription
AccGeography_tab
AccLocality

(ecatalogue)

PriAccessionNumberRef.CatCatalog
PriAccessionNumberRef.DarIndividualCount
PriAccessionNumberRef.irn
PriAccessionNumberRef.DarBasisOfRecord
PriAccessionNumberRef.CatItemsInv
EcbNameOfObject
DesMaterials_tab
DarOrder
DarScientificName
IdeTaxonRef_tab.ClaRank
IdeTaxonRef_tab.ComName_tab
DarRelatedInformation
CatProject_tab
DesEthnicGroupSubgroup_tab

(ecollectionevent)

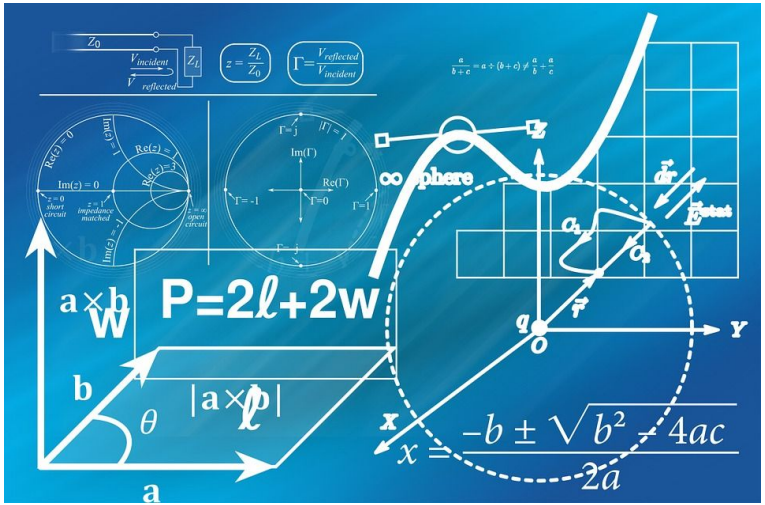
AccCollectionEventRef.ColSiteRef.LocContinent_tab
AccCollectionEventRef.ColSiteRef.LocCountry_tab
AccCollectionEventRef.ColSiteRef.LocOcean_tab

Now that we know where & what it is

How can we make it work for us?

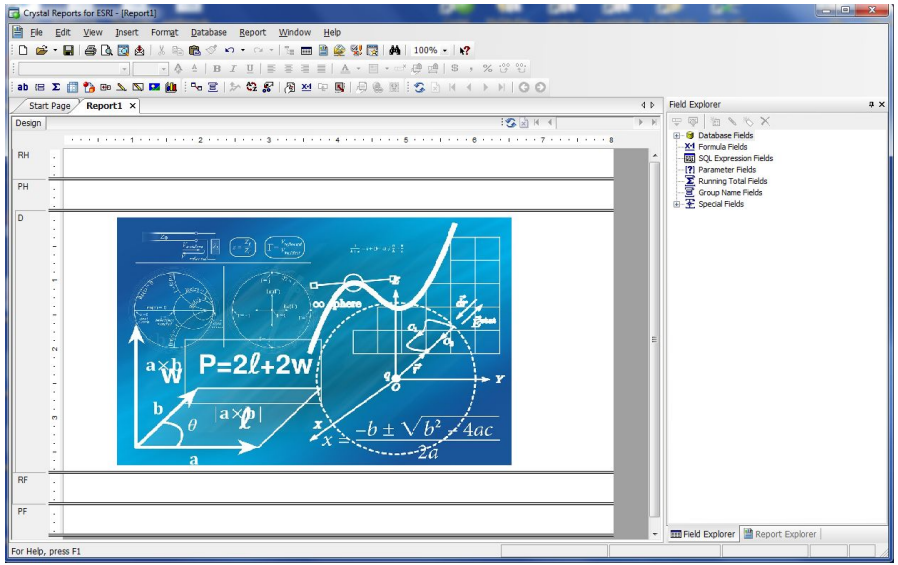
Now that we know where & what it is

How can we make it work for us?



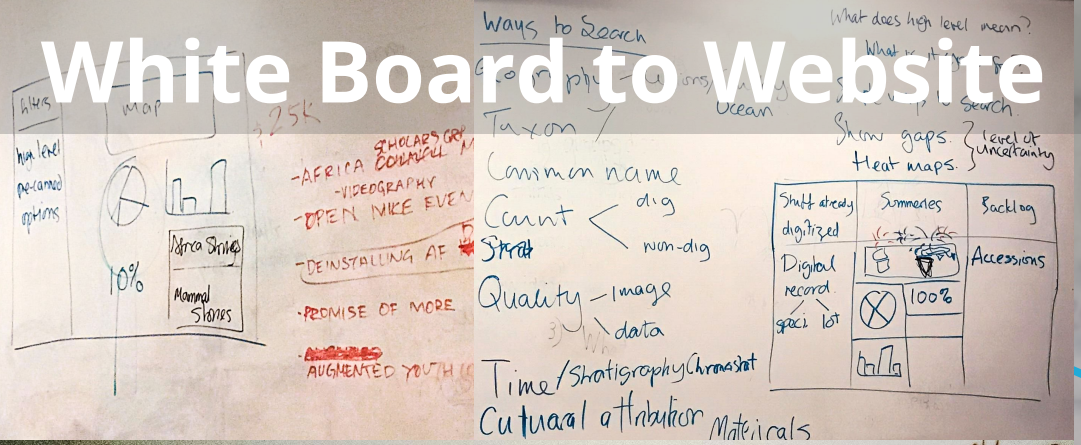
Summaries and formulas.

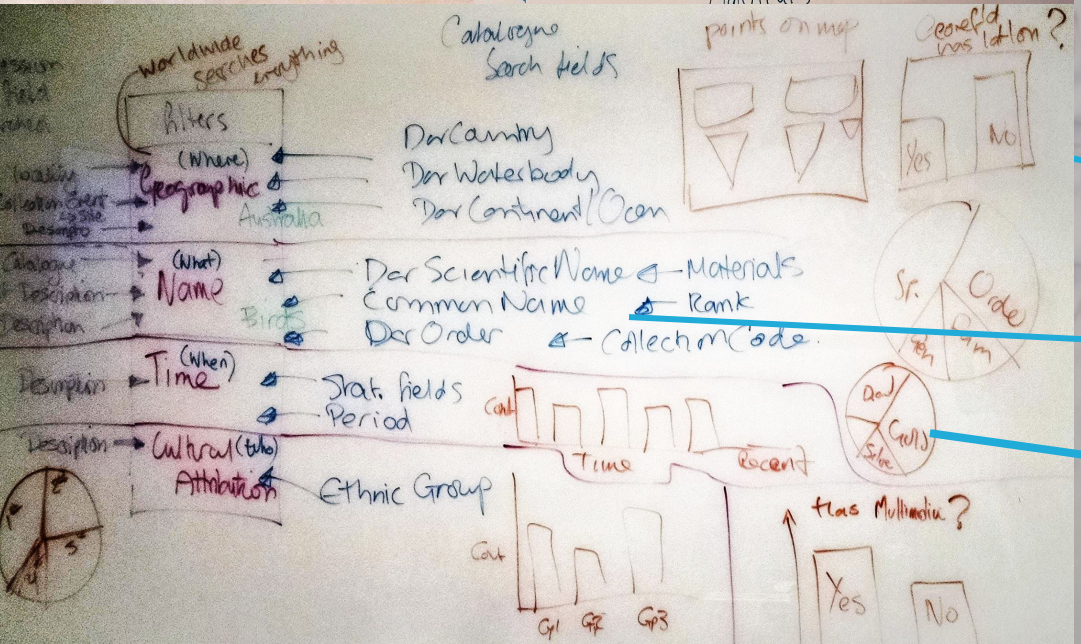
Could you do it with a report?

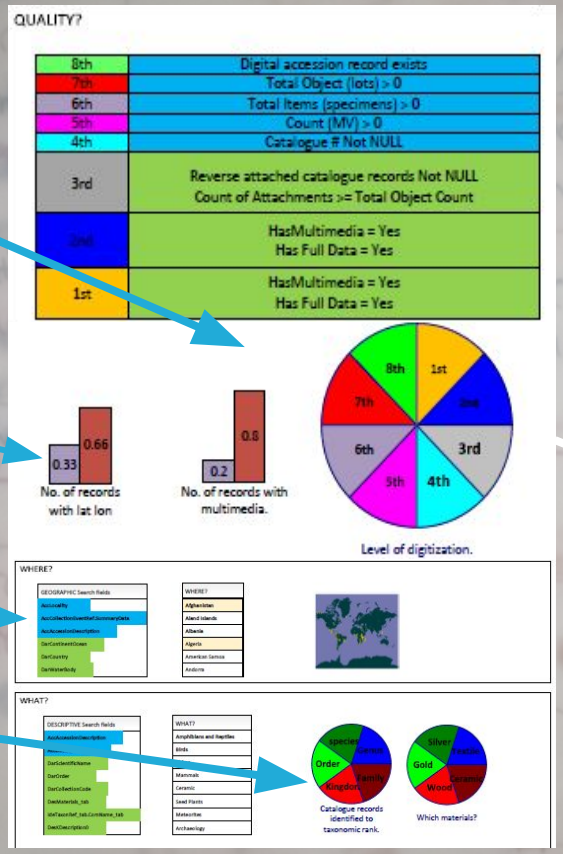


Sure, but in the best of all worlds wouldn't it be great if...

White Board to Website . . .







Searches

Where?

Pick a location

What?

Pick a thing

When?

Pick a time

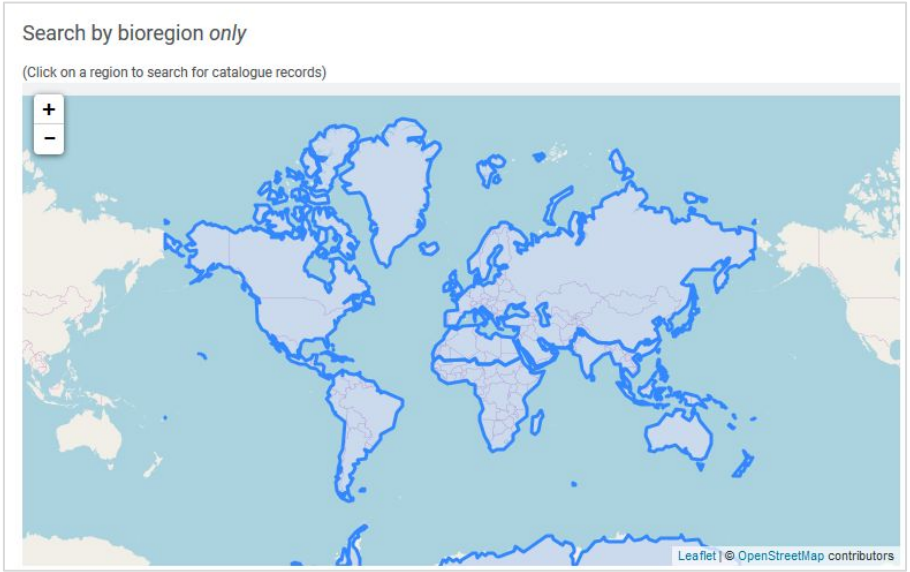
Who?

Pick a culture

Institution? (***very*** slow if run on its own)

Pick an institution

Search



Searches - Where

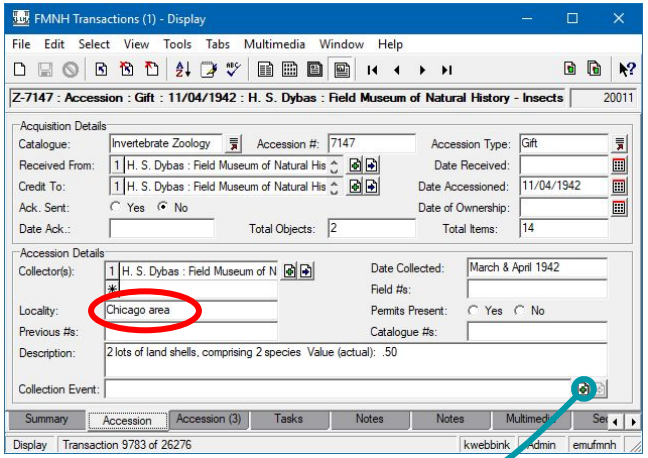
Where-related fields

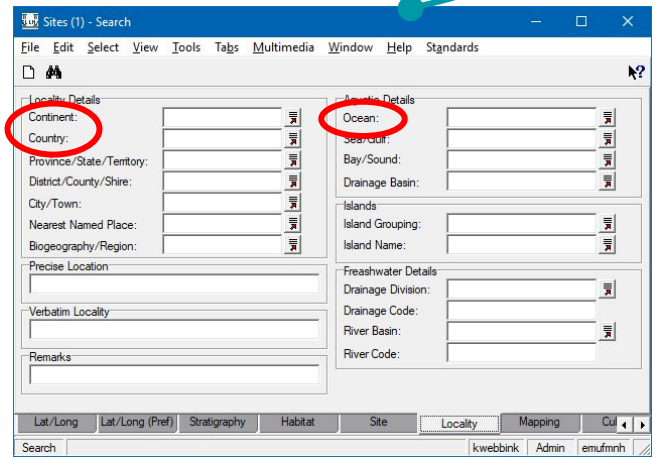
...in backlog records:

- Acc**Geography**_tab
- Acc**Locality**
- AccCollectionEventRef.ColSiteRef.Loc**Continent**_tab
- AccCollectionEventRef.ColSiteRef.Loc**Country**_tab
- AccCollectionEventRef.ColSiteRef.Loc**Ocean**_tab

...& in catalog records:

- **DarLatitude**
- **DarLongitude**
- **DarCountry**
- **DarContinent**
- **DarContinentOcean**
- **DarWaterBody**



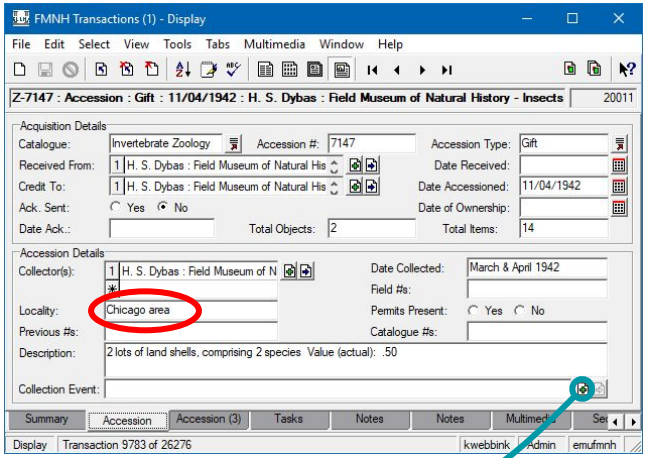


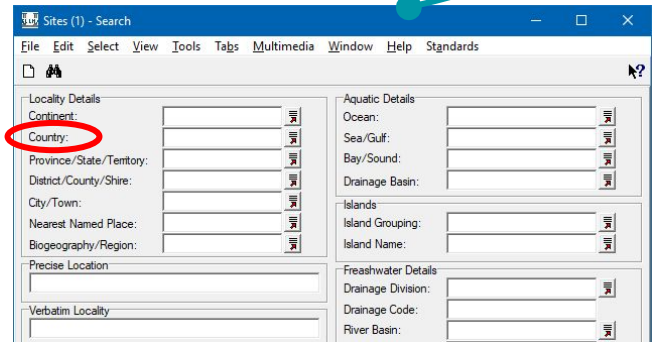
Searches - Where [by bioregion]

Bioregions *...include Where-related fields*

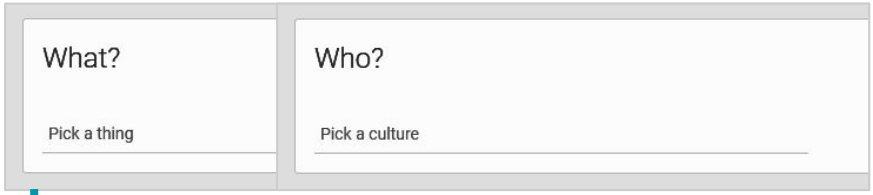
	A	B	C
1	**EcoRegions Type**	**EcoRegions Environment**	**Countries**
2	Tropical and Subtropical Moist Broadleaf	Afrotropical	Cameroon
3	Tropical and Subtropical Moist Broadleaf	Afrotropical	Central African Republic
4	Tropical and Subtropical Moist Broadleaf	Afrotropical	Côte d'Ivoire
5	Tropical and Subtropical Moist Broadleaf	Afrotropical	Democratic Republic of
6	Tropical and Subtropical Moist Broadleaf	Afrotropical	Equatorial Guinea
7	Tropical and Subtropical Moist Broadleaf	Afrotropical	Gabon

*...mapped to WWF "**Terrestrial Ecoregions of the World**" data & shapefiles*



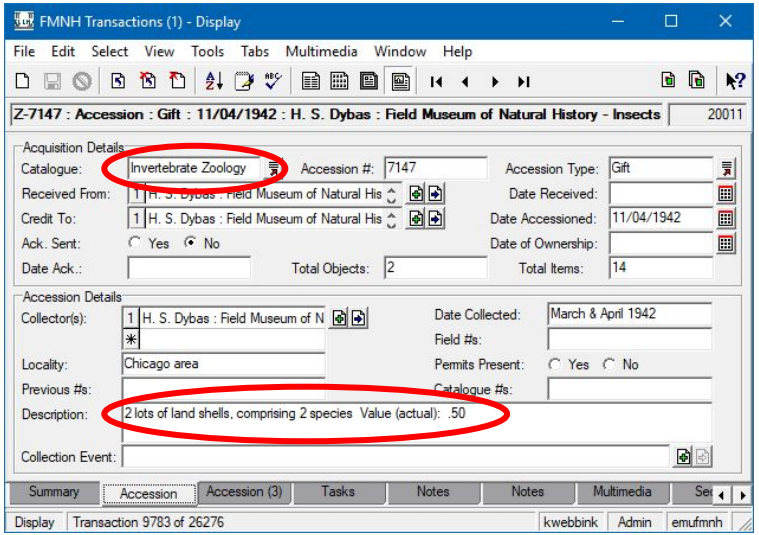


Searches - What & Who



What- & Who-related fields in backlog records:

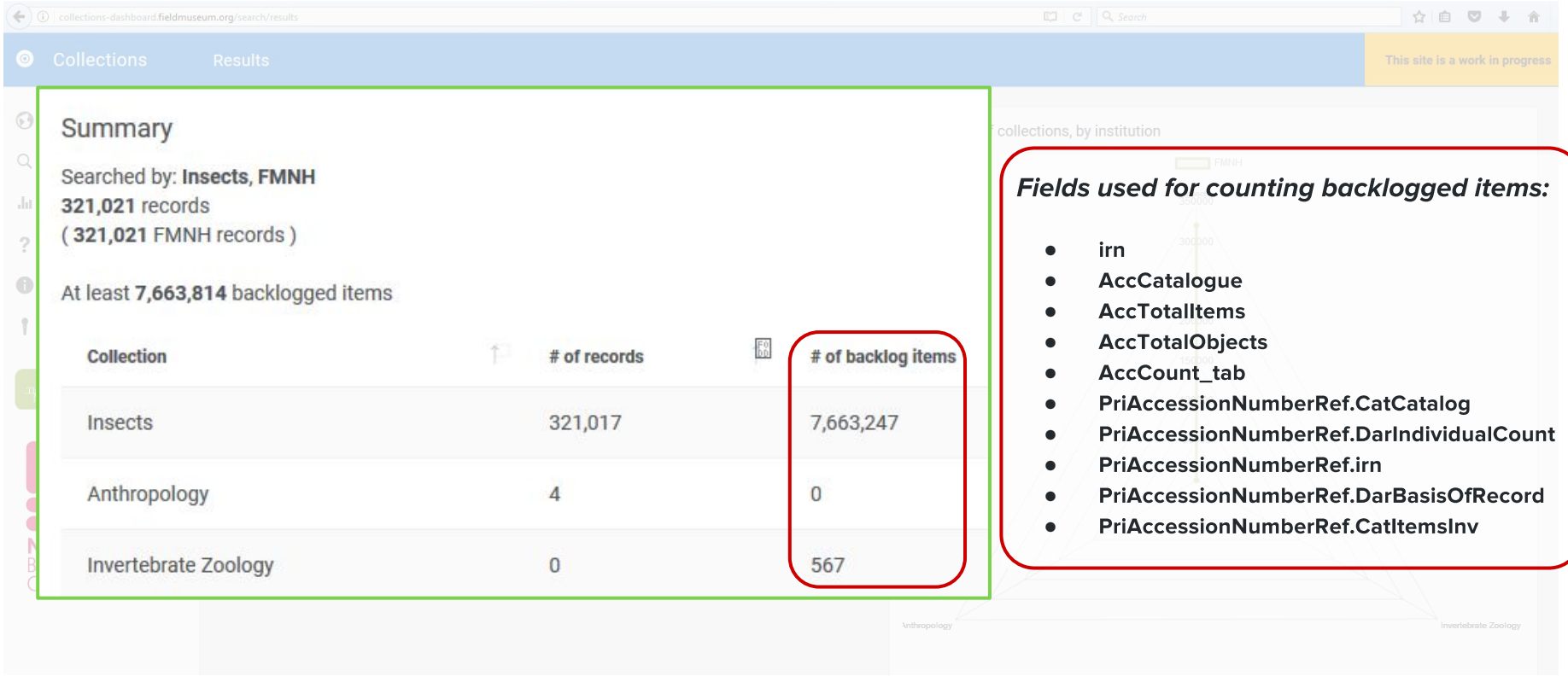
- AccCatalogue
- AccDescription_tab
- AccAccessionDescription



...& in catalog records:

- DesEthnicGroupSubgroup_tab
- CatProject_tab
- DarOrder
- DarRelatedInformation
- DarScientificName
- DesMaterials_tab
- EcbNameOfObject
- IdeFiledAs_tab (to be added)
- IdeTaxonRef_tab.ClaRank
- IdeTaxonRef_tab.ComName_tab

Summary Results - Counts



Fields used for counting backlogged items:

- **irn**
- **AccCatalogue**
- **AccTotalItems**
- **AccTotalObjects**
- **AccCount_tab**
- **PriAccessionNumberRef.CatCatalog**
- **PriAccessionNumberRef.DarIndividualCount**
- **PriAccessionNumberRef.irn**
- **PriAccessionNumberRef.DarBasisOfRecord**
- **PriAccessionNumberRef.CatItemsInv**

Answers - How “good” are your records?

COMPLETENESS RANKING
1 (low) through **9 (high)**

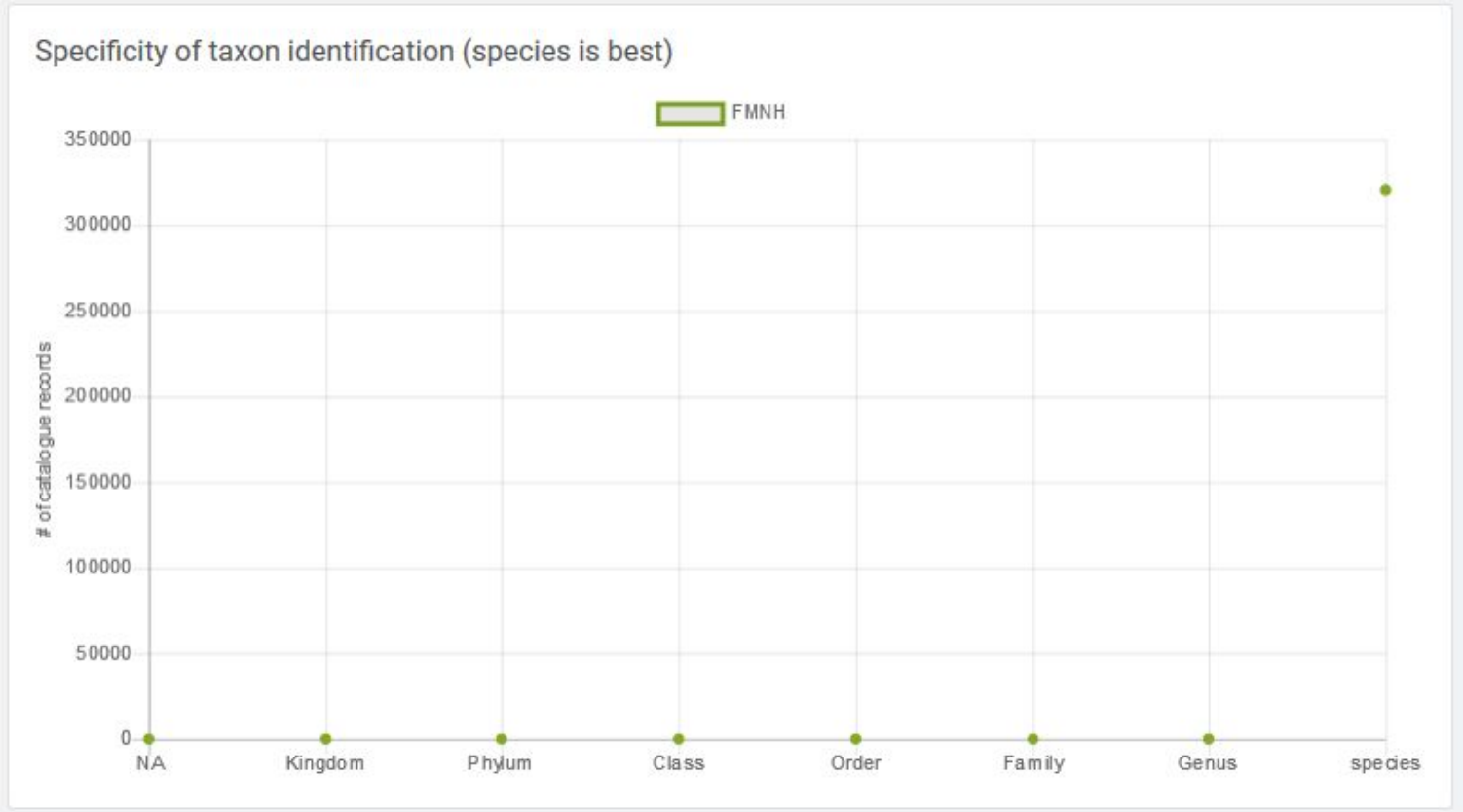
1. **Digital accession record exists**
2. **Total Object (lots) > 0**
   ***OR* Total Items (specimens) > 0**
3. **Locality of Accession record Not Null**
4. **Catalogue # Not NULL**
5. **Reverse attached catalogue records Not NULL**
   (Total Count - Count of Attachments > 0)

6. **Has Digital Catalogue record**
7. **Has Partial Data***
8. **HasLatLon = Yes OR HasMultimedia = Yes**
9. **HasLatLon = Yes AND HasMultimedia = Yes**
   AND* Has Full Data*

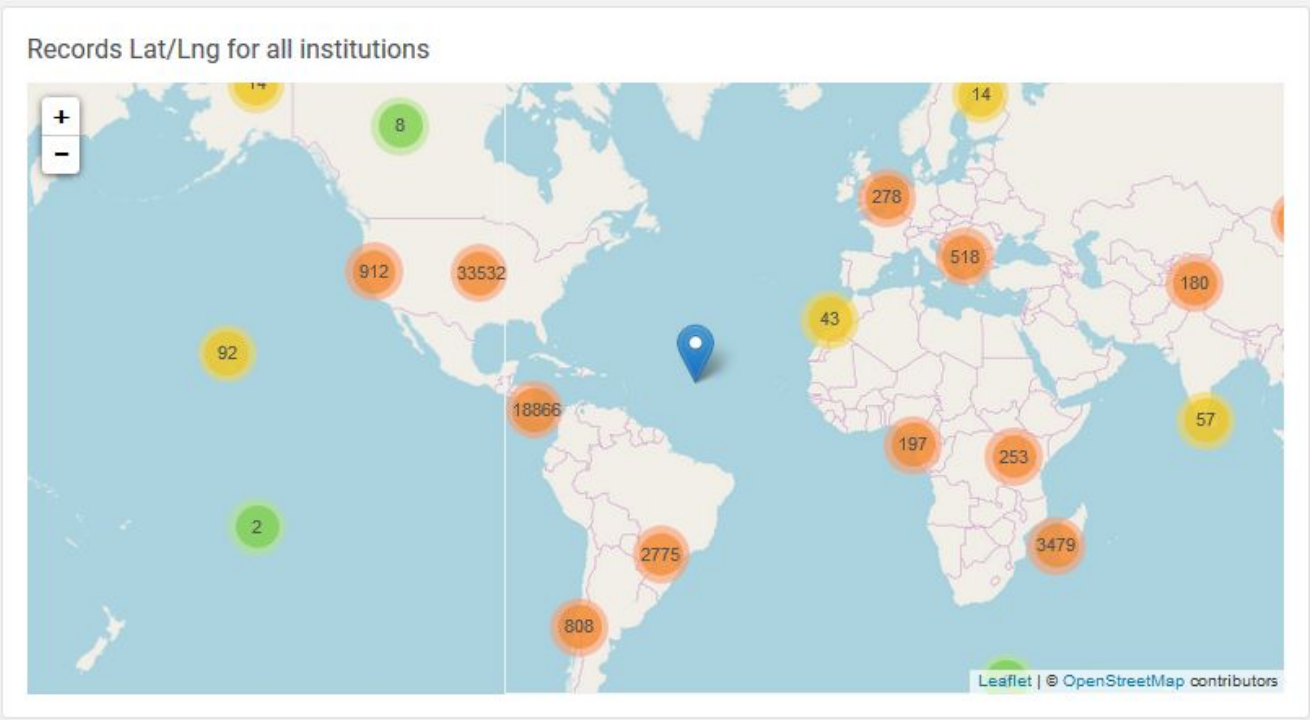
* **Partial Data** = Has 3 or 4 of the following
** **Full Data** = Has all 5 of the following

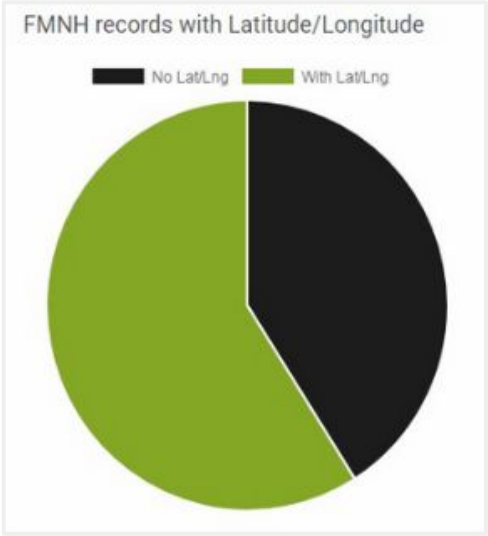
TaxonRank	= *Family, Genus, species, subspecies or variety*
State/Province	= *Not NULL*
Collector	= *Not NULL*
YearCollected	= *Not NULL*
DarCatalogNumber	= *Not NULL*

Taxonomic Completeness



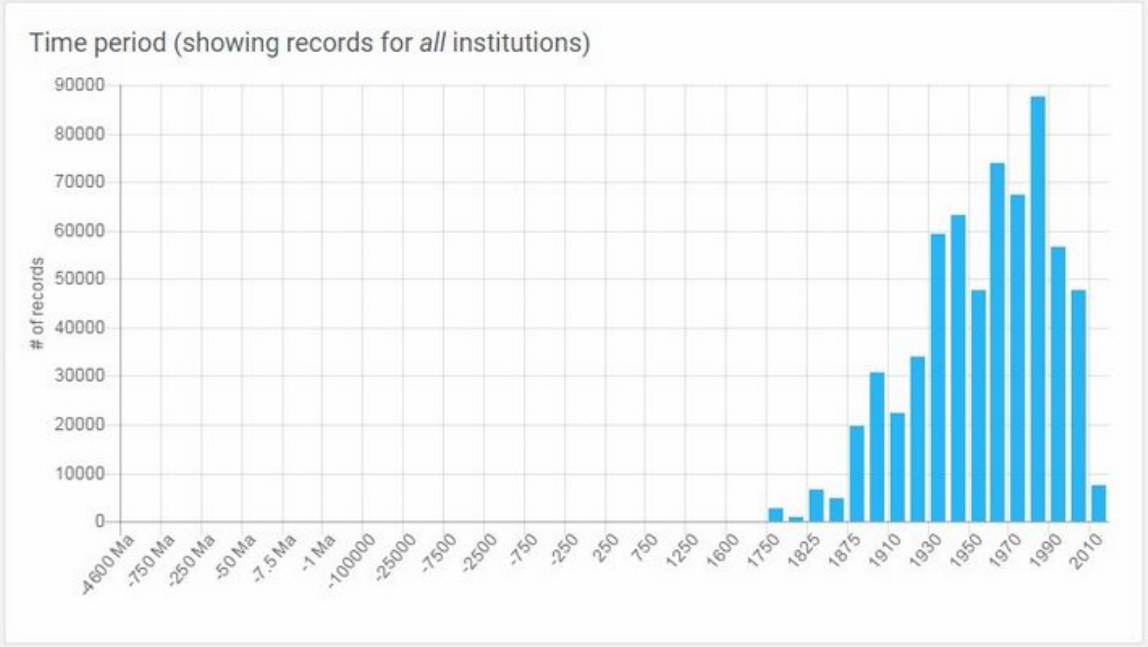
Supplementary data (georeferenced, imaged)





Dates of Collection

Explanation of ranking
This chart displays time of existence/creation (this may or may not be date collected). Before the 18th century, time along the x-a



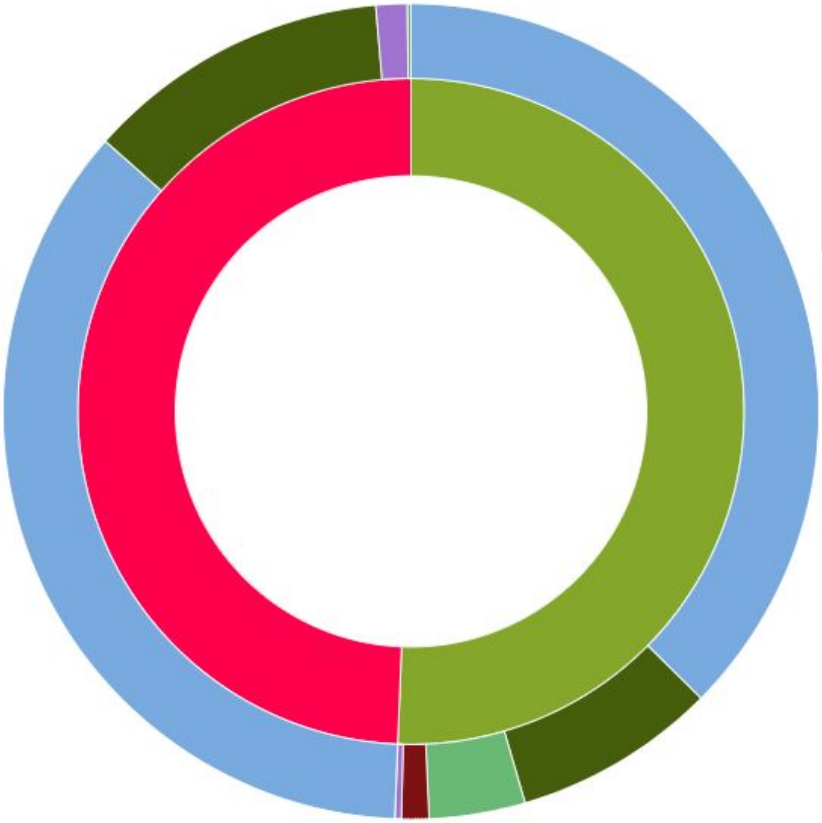
Explanation of ranking

This chart displays time of existence/creation (this may or may not be date collected). Before the 18th century, time along the x-axis is grouped in increments analogous to geological time periods (ages, periods, epochs, etc). Afterwards, it is grouped loosely in decades.

...But can we do this with
more than one institution?...

...you bet yer sweet bippy.

Comparisons between institutions and collections



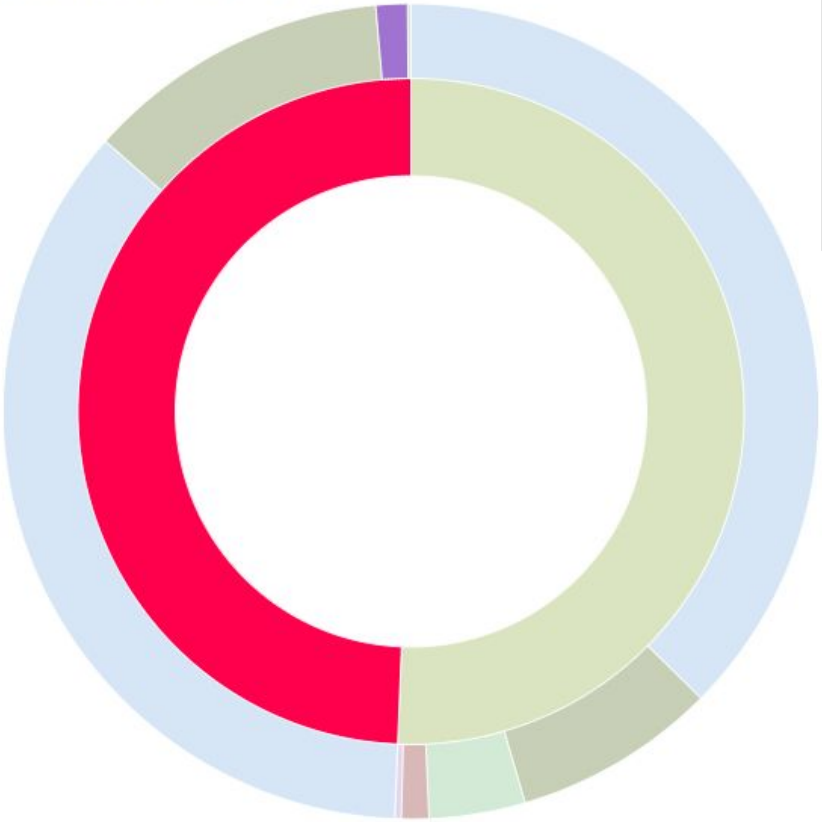
Global collections

(Mouse over the ring segments to see details. The outer ring represents collections within an institution. The inner ring represents institutions.)

Legend

- FMNH
- Naturalis
- IZ
- VZ
- Botany
- Paleo
- Anthropology

Comparisons between institutions and collections



Global collections

(Mouse over the ring segments to see details. The outer ring represents collections within an institution. The inner ring represents institutions.)

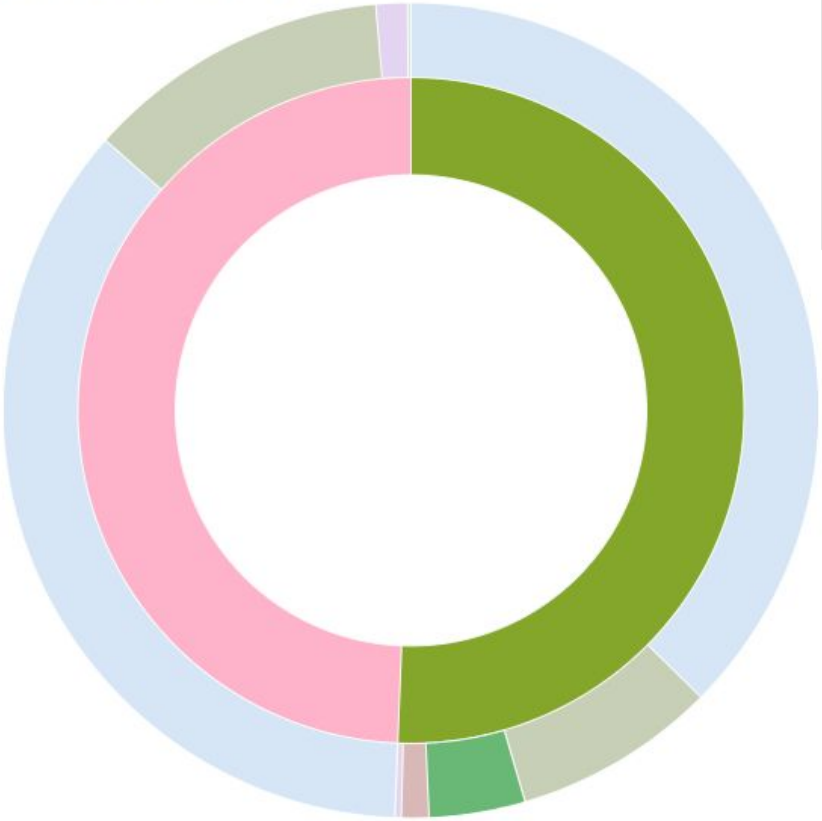
Naturalis | Paleo

is 1.24% of the total global collections

Legend

- FMNH
- Naturalis
- IZ
- VZ
- Botany
- Paleo
- Anthropology

Comparisons between institutions and collections



Global collections

(Mouse over the ring segments to see details. The outer ring represents collections within an institution. The inner ring represents institutions.)

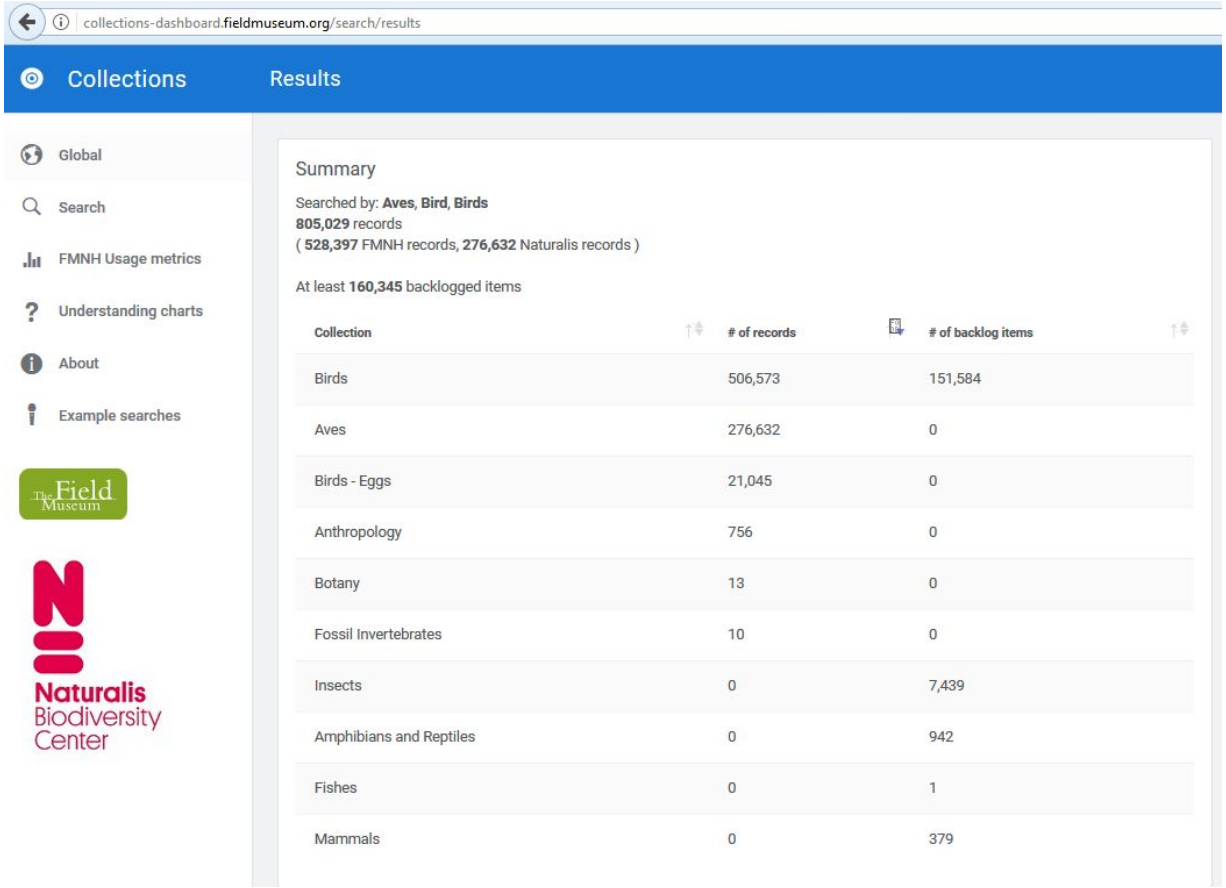
FMNH › Botany

is 3.81% of the total global collections

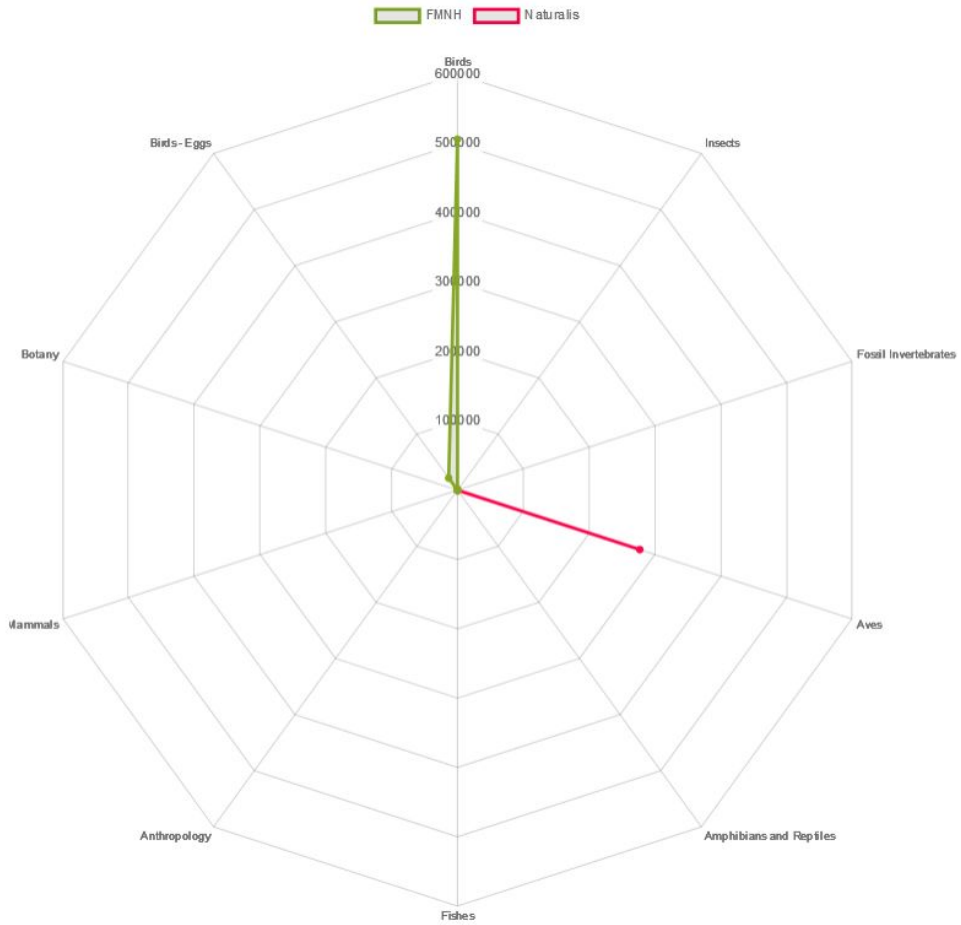
Legend

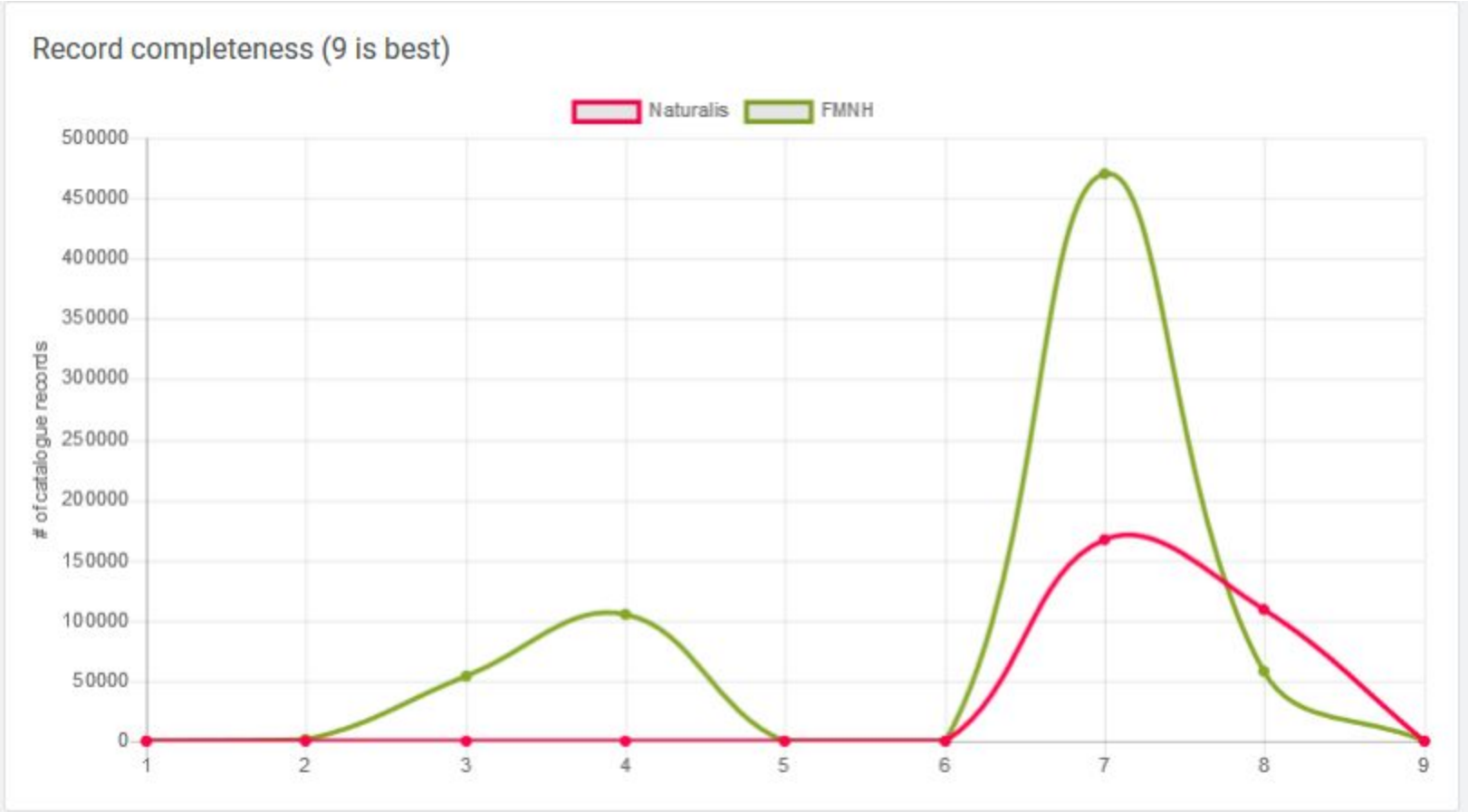
- FMNH
- Naturalis
- IZ
- VZ
- Botany
- Paleo
- Anthropology

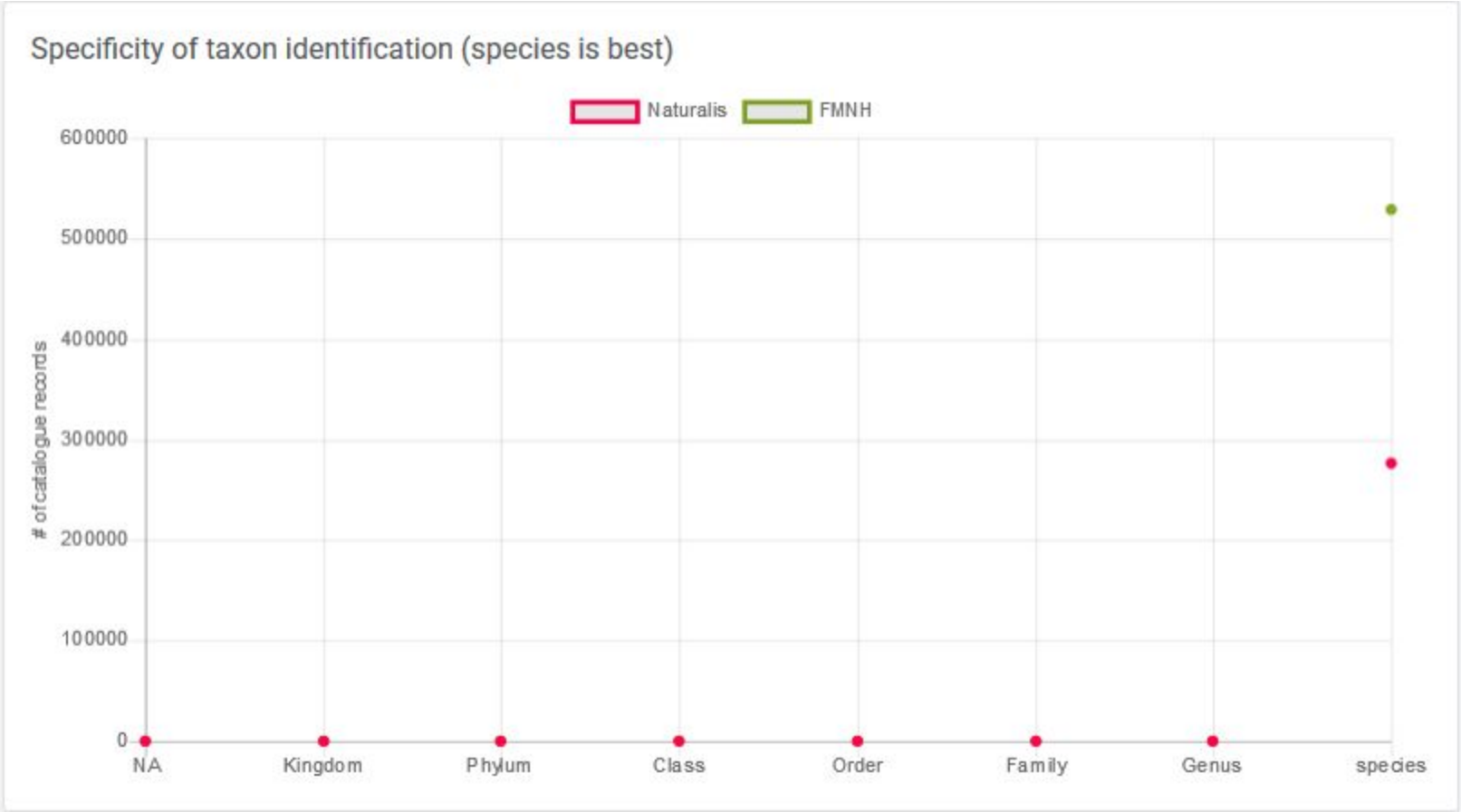
Results of a search for Birds (and other bird search terms) for both Institutions.



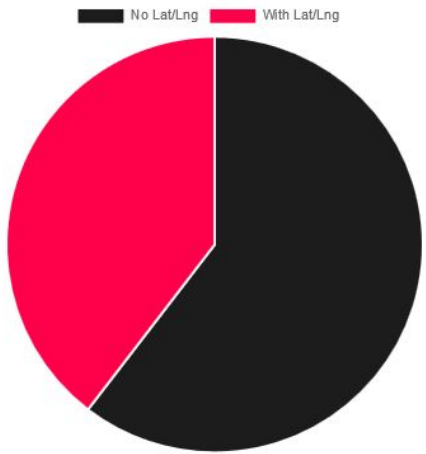
Breakdown of collections, by institution



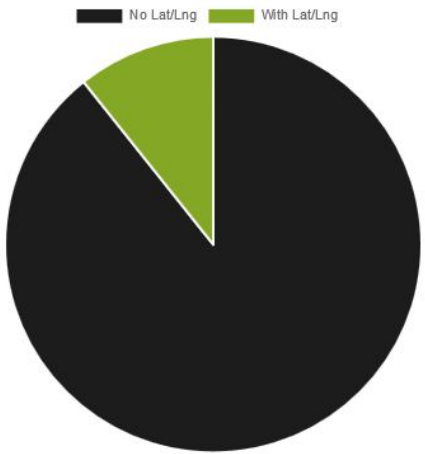

Record completeness (9 is best)
Naturalis
FMNH
# of catalogue records
500000
450000
400000
350000
300000
250000
200000
150000
100000
50000
0
1
2
3
4
5
6
7
8
9


Specificity of taxon identification (species is best)
Naturalis
FMNH
600000
500000
400000
300000
200000
100000
0
# of catalogue records
NA
Kingdom
Phylum
Class
Order
Family
Genus
species

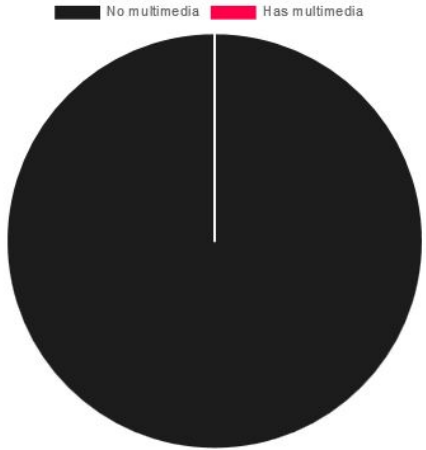
Naturalis records with Latitude/Longitude



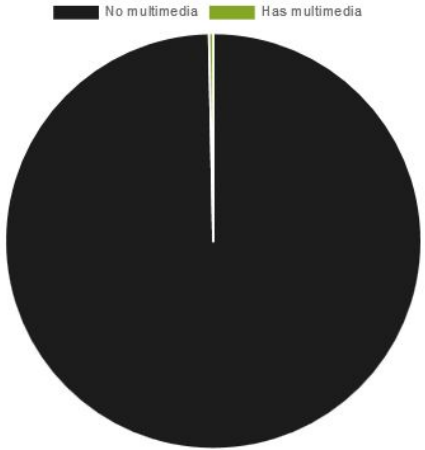
FMNH records with Latitude/Longitude

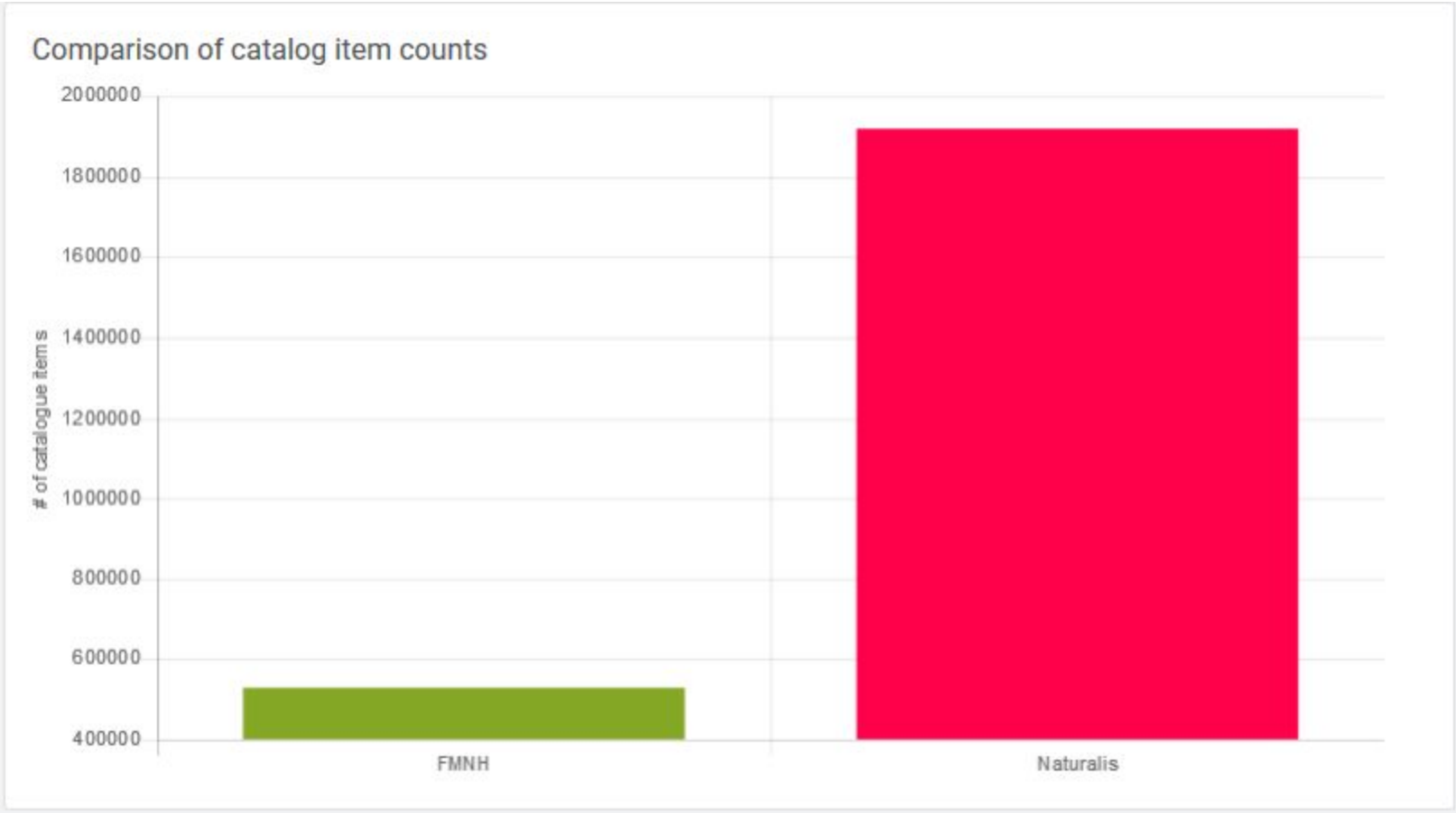


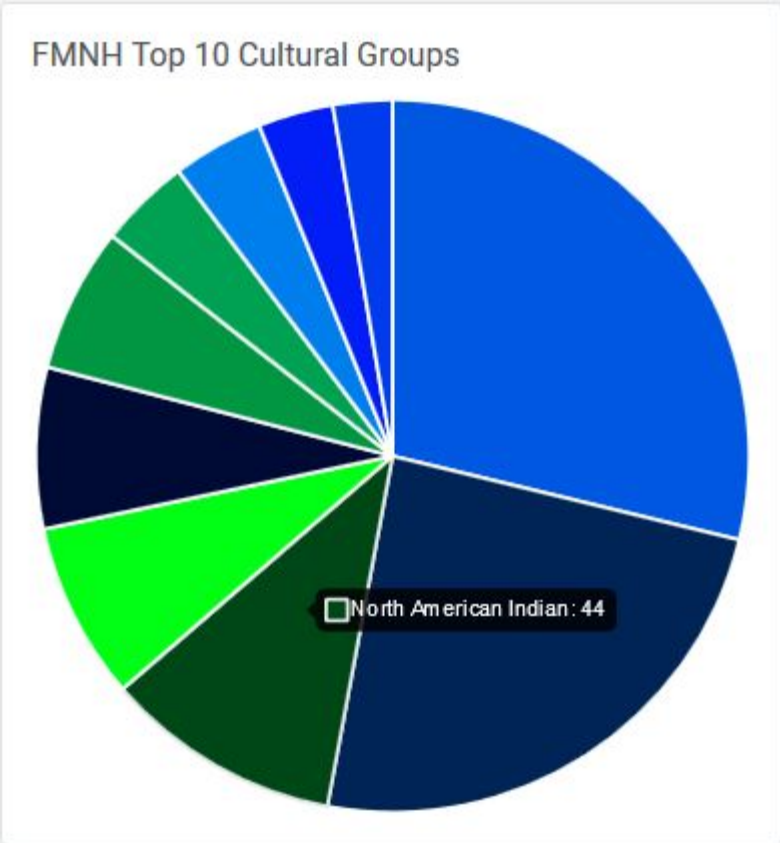
Naturalis records with multimedia



FMNH records with multimedia



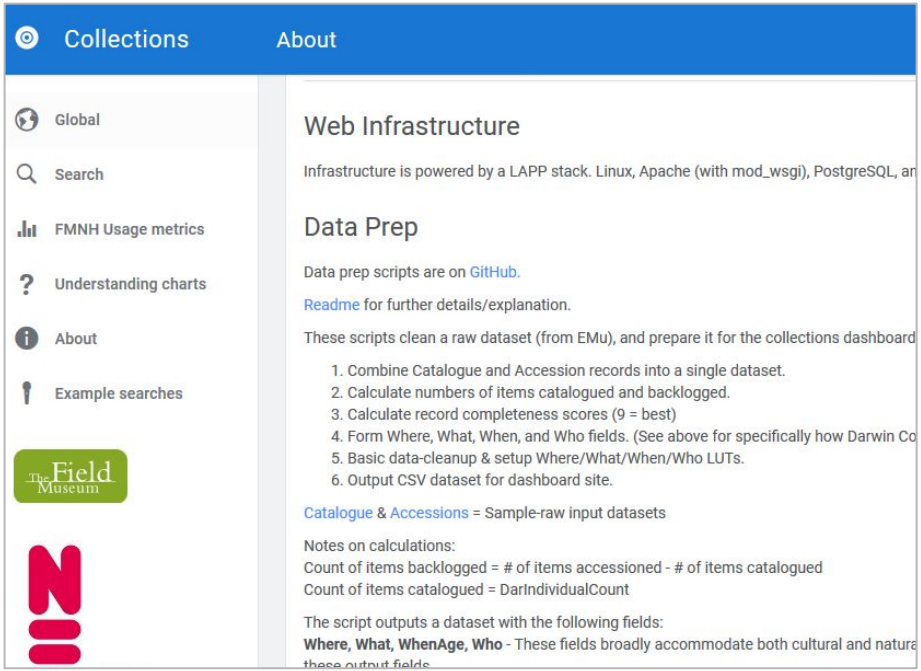

Comparison of catalog item counts
2000000
1800000
1600000
1400000
1200000
1000000
800000
600000
400000
# of catalogue items
FMNH
Naturalis



756 records are returned from Anthropologic collections, so the breakdown of represented cultural groups chart appears in results.

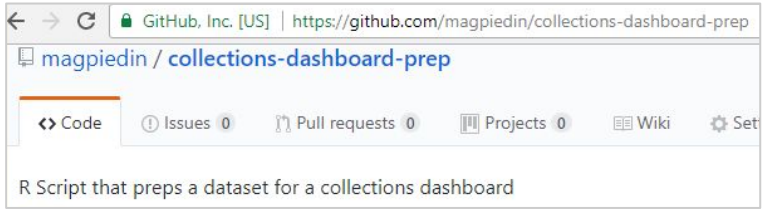
Did we show our work? What did we do?

…see the “**About**” & “**Understanding charts**” pages
http://collections-dashboard.fieldmuseum.org/



…see “**Web Infrastructure**” section
for web-development information.

…see “**Data prep**” for links to R-scripts
https://github.com/magpiedin/collections-dashboard-prep



Why do this at all?

Big Data Digitized...

At time of writing there are over:

- **849 million** occurrence records in the Global Biodiversity Information Facility (GBIF) portal (gbif.org)
- **104 million** on the iDigBio site (idigbio.org)
- **71 million** in the Atlas of Living Australia (ala.org.au)
- **20 million** in VertNet (vertnet.org).

...Big Data Not Yet Digitized.

In 2000 there were an estimated **2.5 billion(ish)** specimens in the world's natural history collections.

(Bruce A. Stein, Lynn S. Kutner, Jonathan S. Adams - 2000 - Nature)

Today's **"Backlog"** is:

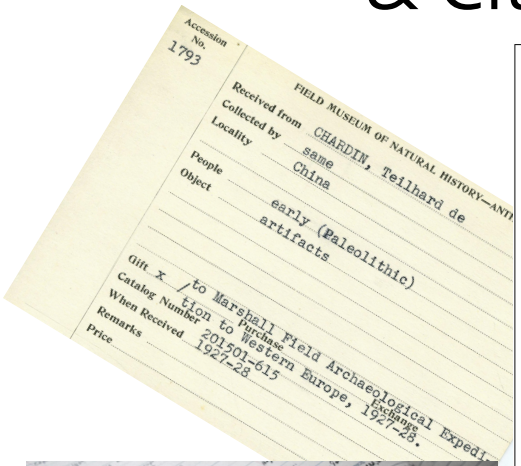
2.5 billion(ish) - 850,000,000 = **1.65 billion(ish)**

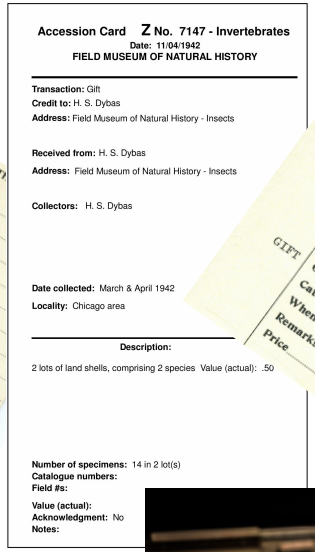
Taxonomic bias in biodiversity data and societal preferences

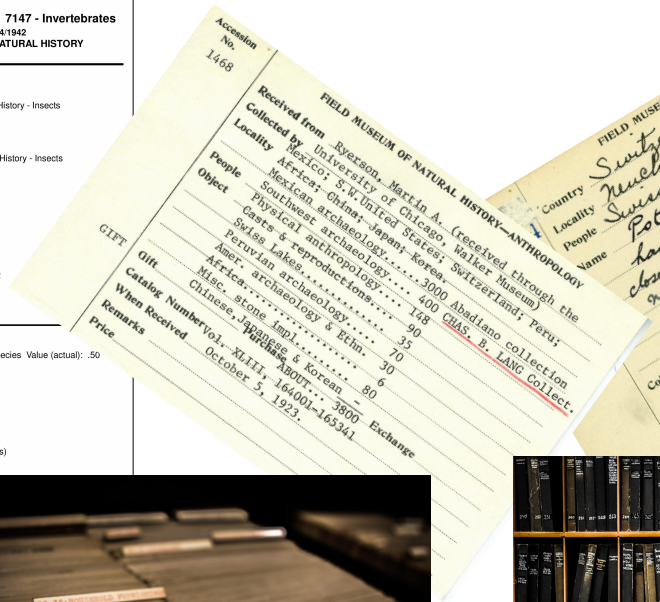
"Studying and protecting each and every living species on Earth is a major challenge of the 21st century. Yet, most species remain unknown or unstudied, while others attract most of the public, scientific and government attention. Although known to be detrimental, this taxonomic bias continues to be pervasive in the scientific literature, but is still poorly studied and understood. ... **Our results show that societal preferences, rather than research activity, strongly correlate with taxonomic bias, which lead us to assert that scientists should advertise less charismatic species and develop societal initiatives (e.g. citizen science) that specifically target neglected organisms**..."

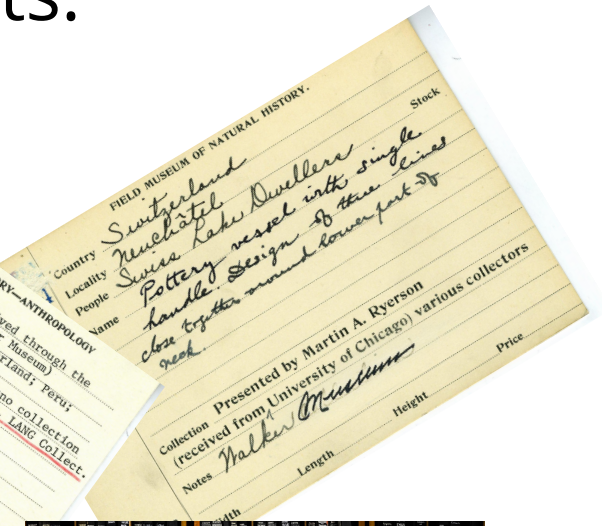
Troudet J, Grandcolas P, Blin A, Vignes-Lebbe R, Legendre F. Taxonomic bias in biodiversity data and societal preferences. Scientific Reports. 2017;7:9132. doi:10.1038/s41598-017-09084-6.

Medium lends itself to quick imaging & Citizen Scientist efforts.









Tell me what you do and don't have digitized…

Grant Applications

Annual/Quarterly Board meetings

Research inquiries

Peer projects

Data aggregators

…and tell me in the format I need it in.

By Collection Type
By Geographic Region
Over time
By Bio Region
Monthly estimates
By Research Project

Links

GitHub

Readme

Dashboard

More about the Dashboard development