



# The Long Journey

# The Road Behind

Full Production - Data Migration, Forms, and in use.	Catalog Numbers	Agg Publishing	Web Portal
Ichthyology (Feb 2014)	229,094	Y	Y
Ichthyology Teaching	557	N	N
Ichthyology Uncataloged Transaction Items	1,945	N	N
Herpetology (April 2014)	178,347	Y	Y
Mammalogy (May 2014)	34,179	Y	Y
Invertebrate Paleontology (Oct 2014)	306,489	Y	Y
Paleobotany (Jan 2016)	58,885	Y	Y
Paleobotany Modern Pollen Fields (Jan 2016)	12,838	N	N
Paleobotany Fruit & Seeds (Jan 2016)	2,723	N	N
Paleobotany Modern Leaf (Jan 2016)	6,629	N	N
Vertebrate Paleontology (April 2016)	421,391	Y	Y
Invertebrate Zoology (April 2016)	601,320	Y	Y
Ornithology Skins & Skeletons (Mar 2017)	51,525	Y	Y
Ornithology Eggs & Nests (Mar 2017)	13,371	N	N
Ornithology Sounds (Mar 2017)	28,790	N	Y
Genetic Resource Repository (Jun 2018)	63,878	N	Y
Total Records	2,011,961		

# Genetic Resource Repository

Project goal was to establish a data pipeline for Specify data to flow from donor collection databases to the GRR database. Simply stated, we want the data published in the GRR to match the data in the respective donor collection database.

## Solution:

1. Leverage Specify DataExporter and Schema Maps in donor databases to make pipeline upgrade safe.
2. SQL Scripts to update GRR from the DataExporter builds.
3. Cron is our friend
  1. Cron job for DataExporter update.
  2. Cron job for SQL scripts
  3. Cron job for Web Portal Updates

Specify 6.7.01

[Welcome](#)
[Data](#)
[Trees](#)
[Reports](#)
[Interactions](#)
[Statistics](#)
[Query](#)
[Workbench](#)
[BatchEdit](#)
[Attachments](#)

[Create Mapping](#)  
 Create New Mapping  
 Import Schema  
 Mappings  
 GRR  
 IPT  
 Public Web Portal  
 Record Sets  
 0133  
 0256  
 Expeditions  
 geo\_2016-4-15\_17:31  
 loan students  
 QualifierFixMe  
 tax1\_2016-4-15\_13:10  
 tax2\_2016-4-15\_13:23  
 tax3\_2016-4-15\_15:33  
 tax4\_2016-4-15\_16:40  
 tax5\_2016-4-15\_16:48  
 tax6\_2016-4-15\_16:7

Collection Object  
 Accession  
 Col Obj Attribute  
 Collecting Event  
 Collection Object Attachments  
 Created By Agent  
 DNA Sequences  
 Date Cataloged  
 Date Cataloged (Day)

Determinations  
 Current  
 GUID  
 ID Confidence  
 ID Date  
 ID Date (Day)  
 ID Date (Month)  
 ID Date (Year)  
 Identified by

Schema Concept	Not Operator	Criteria	Sort	Show	Allow Null
Current	<input type="checkbox"/> Yes		<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
UFID	<input type="checkbox"/> Contains		<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Phylum	<input type="checkbox"/> Contains		<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Class	<input type="checkbox"/> Contains		<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Order	<input type="checkbox"/> Contains		<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Family	<input type="checkbox"/> Contains		<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Genus	<input type="checkbox"/> Contains		<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Species	<input type="checkbox"/> Contains		<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Subspecies	<input type="checkbox"/> Contains		<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Locality	<input type="checkbox"/> Contains		<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Latitude	<input type="checkbox"/> =		<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Longitude	<input type="checkbox"/> =		<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Geography		Geography [Formatted]			<input type="checkbox"/>
Preparations		Preparations [Aggregated]			<input type="checkbox"/>
Collectors		Collectors [Aggregated]			<input type="checkbox"/>
Date Collected	<input type="checkbox"/> =	Start Date	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Series
  Search Synonyms
  Distinct
  Count

Welcome GRR

Invertebrate Zoology | IZ | OMTAdmin

```
Untitled-1
1 ALTER TABLE uf_invertebratezoology.grr
2 MODIFY Preparations VARCHAR(512);
3
4 CREATE INDEX Text6IDX
5 ON uf_invertebratezoology.collectionobjectattribute (Text6);
6
7 CREATE INDEX PrepGUIDIDX
8 ON uf_invertebratezoology.grr (Preparations);
9
10 UPDATE
11 uf_grr.collectionobject co,
12 uf_grr.collectionobjectattribute coa,
13 uf_invertebratezoology.grr grrexport
14 SET
15 co.fieldNumber = grrexport.UFID,
16 coa.text1 = grrexport.Locality,
17 coa.text2 = grrexport.Geography,
18 coa.text3 = grrexport.Subspecies,
19 coa.text4 = grrexport.Genus,
20 coa.text5 = grrexport.Species,
21 coa.text7 = grrexport.Collectors,
22 coa.text8 = grrexport.Date_Collected,
23 coa.text10 = grrexport.Phylum,
24 coa.text11 = grrexport.Class,
25 coa.text12 = grrexport.Order,
26 coa.text13 = grrexport.Family,
27 coa.number1 = grrexport.Latitude,
28 coa.number2 = grrexport.Longitude
29 WHERE co.name = 'FLMNH_Invertebrate Zoology'
30 AND instr(grrexport.Preparations, coa.text6)
31 AND co.CollectionObjectAttributeID = coa.CollectionObjectAttributeID
32 ;
33
34 ALTER TABLE uf_invertebratezoology.grr
35 DROP INDEX PrepGUIDIDX;
36
37 ALTER TABLE uf_invertebratezoology.collectionobjectattribute
38 DROP INDEX Text6IDX;
39
```

# Kept It Simple

1. The taxon and geography trees are not used in the GRR database, so we do not incur the overhead and expense of updating them.
2. DataExporter adds a layer of abstraction for the complex ADT's in the donor database which makes the pipeline less brittle from schema changes with version updates.

**FM** FLORIDA  
MUSEUM.®

**UF** | UNIVERSITY *of*  
FLORIDA

**FM** FLORIDA  
MUSEUM.®

**UF** | UNIVERSITY *of*  
FLORIDA