

OUT OF THE JAR AND INTO THE WORLD!

A CASE STUDY ON STORING AND SHARING VERTEBRATE DATA

SUSAN BORDA – DATA WORKFLOWS SPECIALIST

UNIVERSITY OF LIBRARY



UNIVERSITY OF
MICHIGAN

PROJECT BACKGROUND

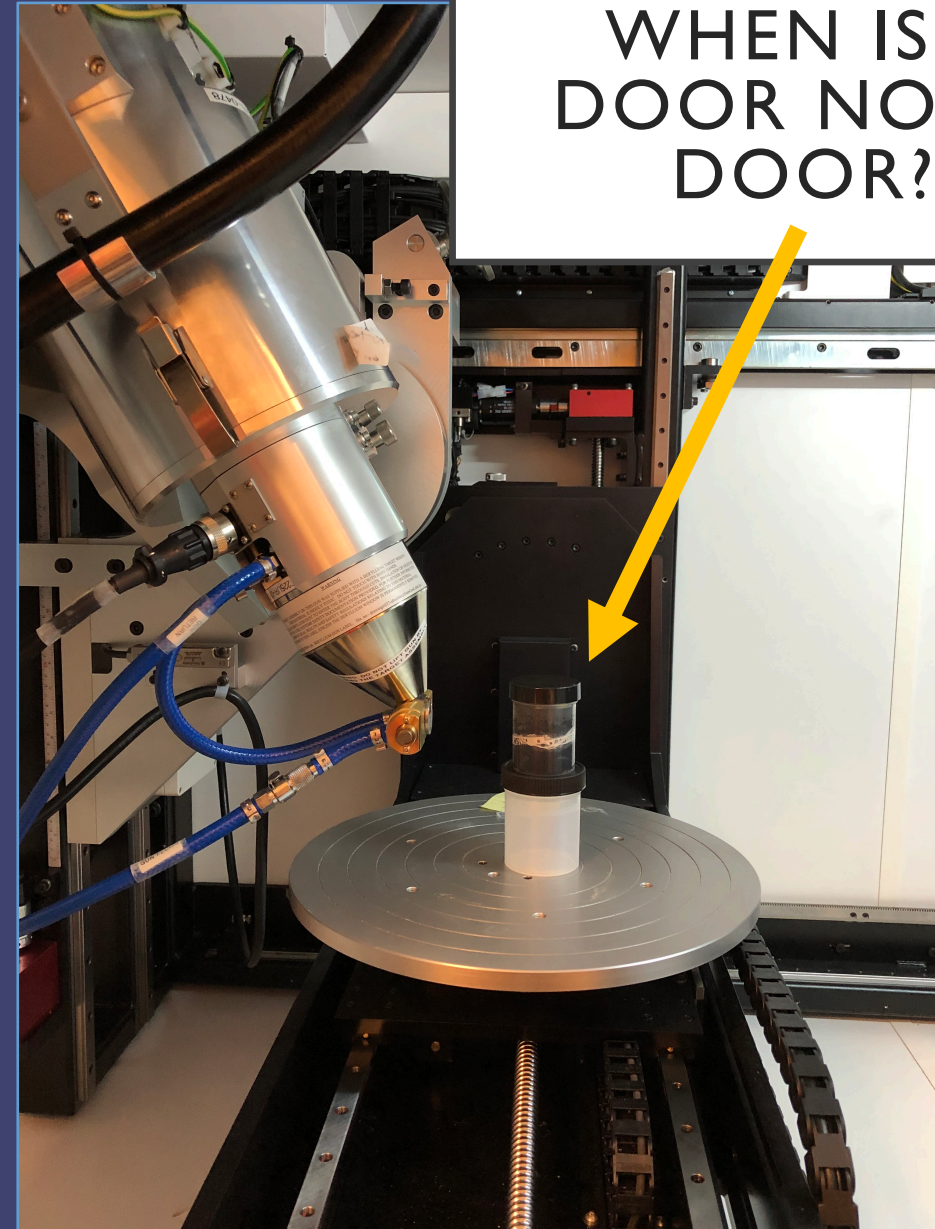
- In 2016, UMMZ became involved with oVert project.
 - Potentially producing 4,000-5,000 scans – 30+TB
- During oVert proposal preparation process, UMMZ ran into problems identifying long-term storage solution for CT data.
 - No college support beyond the length of the grant.
 - Proposed solutions were not cost effective.
- Contacted UM science librarian for assistance
 - Integrated library into data management plan

BACKGROUND - CONTINUED

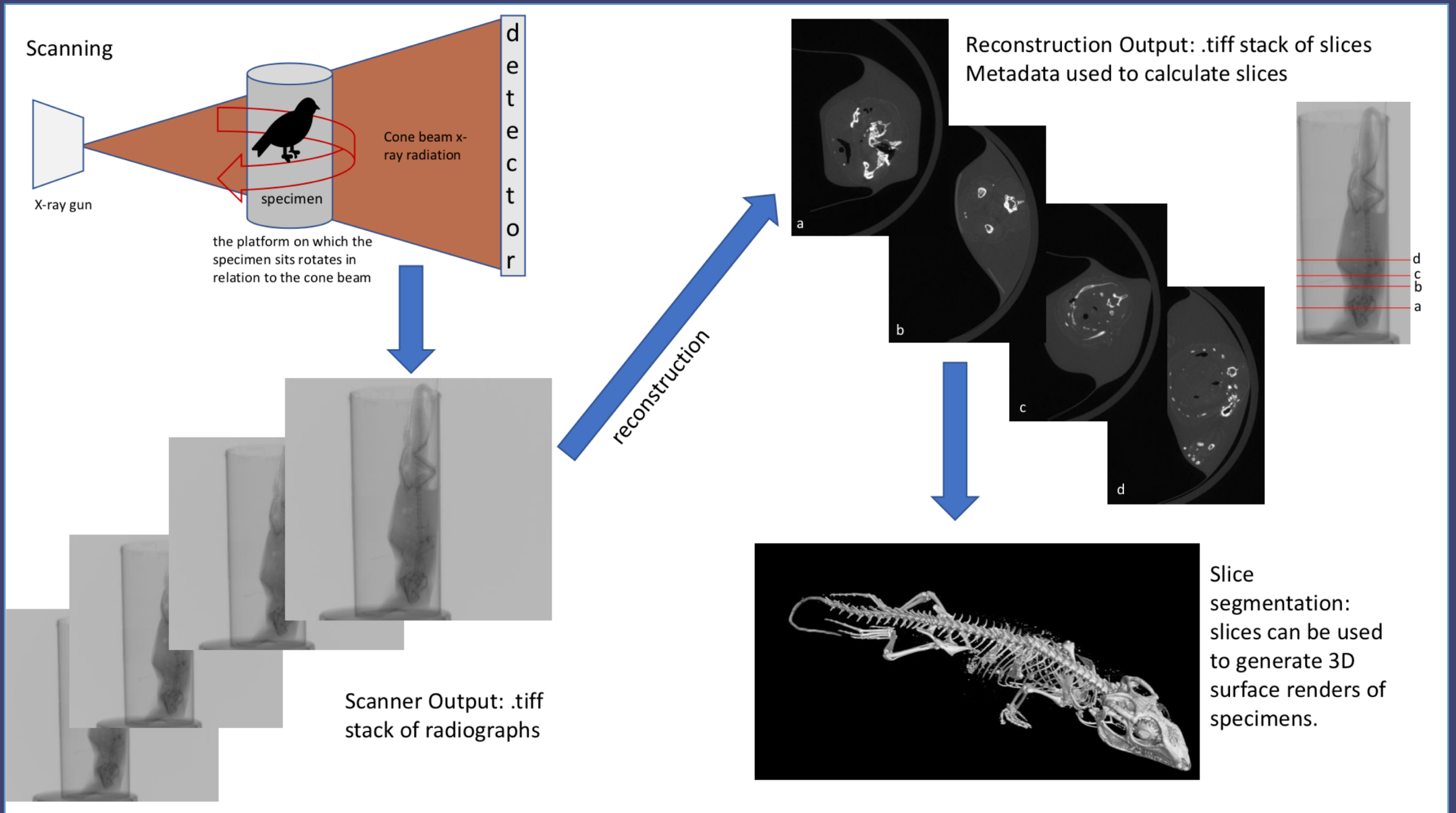
- In Summer 2018, UMMZ reached out to UM Library about their data needs.
 - Deep Blue and Deep Blue Data serve as UM repositories.
- Library and UMMZ created a workflow for transferring the data from UMMZ to the UM Library's data repository.
- This workflow involved copying the data, re-organizing it, pulling descriptive metadata from GBIF (Global Biodiversity Information Facility), creating a yaml file for ingest, ingesting the data, and returning DOI information.



ROOMS AND ROOMS
OF SHELVES OF JARS



WHEN IS A
DOOR NOT A
DOOR?



UMMZ simplified specimen scanning process. (Ramon Nagesan, personal communication)

UMMZ

UMMZ DEPOSIT PROCESS

DBRRDS/
Library

Converted Nikon
scanner output

UMMZ-mammals-100000
UMMZ-herps-100000
UMMZ-mollusks-10000

.TIF/vgi/xtek

Specify
(cataloged
records)

LSA - IT
IPT

• VertNet
• iDigBIO
• GBIF

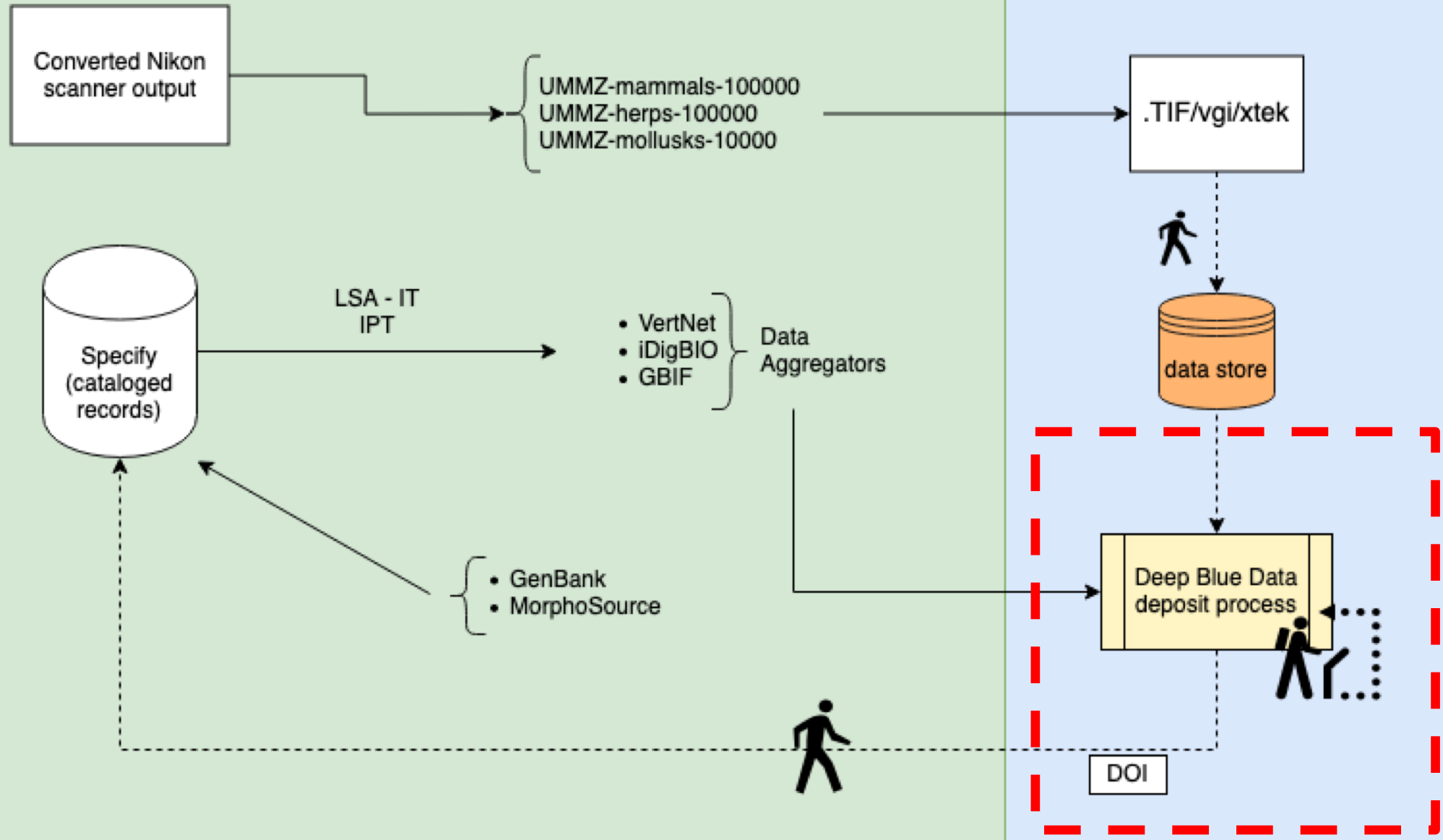
Data
Aggregators

data store

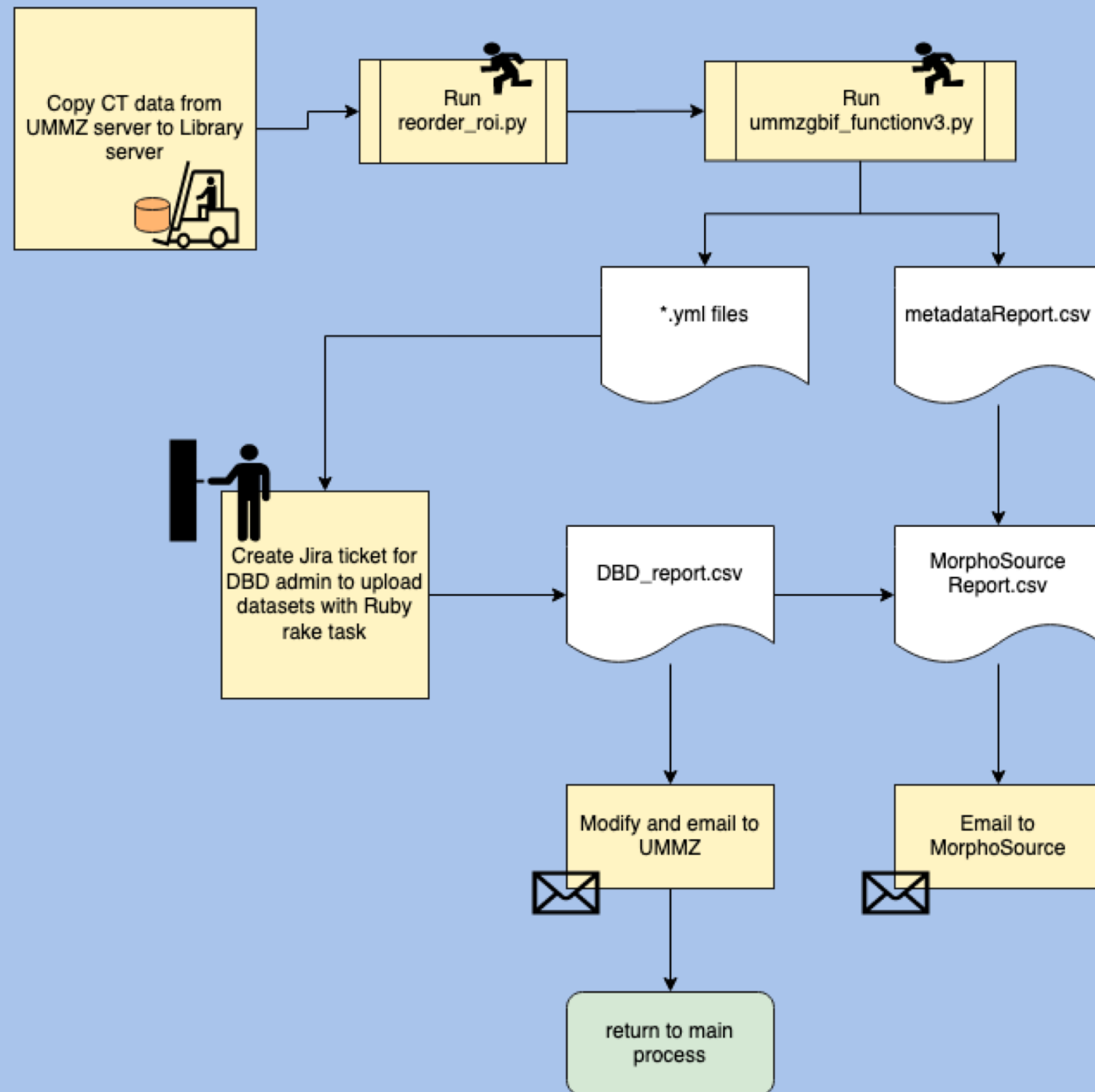
• GenBank
• MorphoSource

Deep Blue Data
deposit process

DOI



DEEP BLUE DATA DEPOSIT PROCESS



SINGLE DEPOSIT (ONE DOI):

UMMZ-HERPS-246849:

SKULL - RAW.ZIP
WHOLEBODY-RAW.ZIP
SKULL-RECON.ZIP
WHOLEBODY-RECON.ZIP

TITLES

FOUR DEPOSITS (FOUR DOI_s):

UMMZ-HERPS-SKULL-246849-RAW
RAW.ZIP

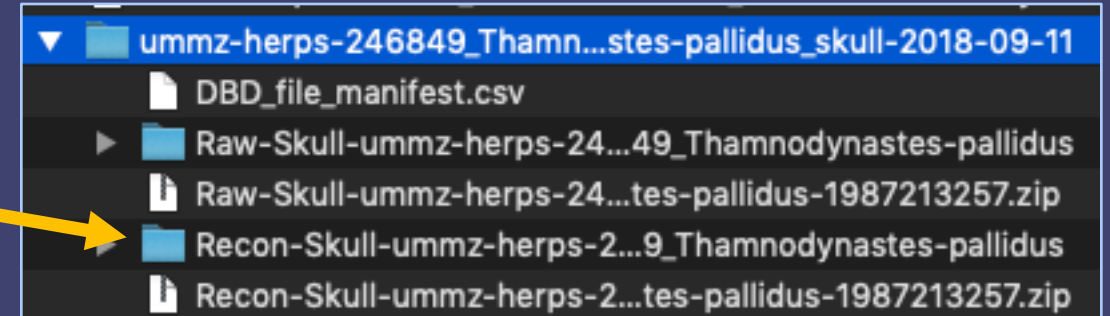
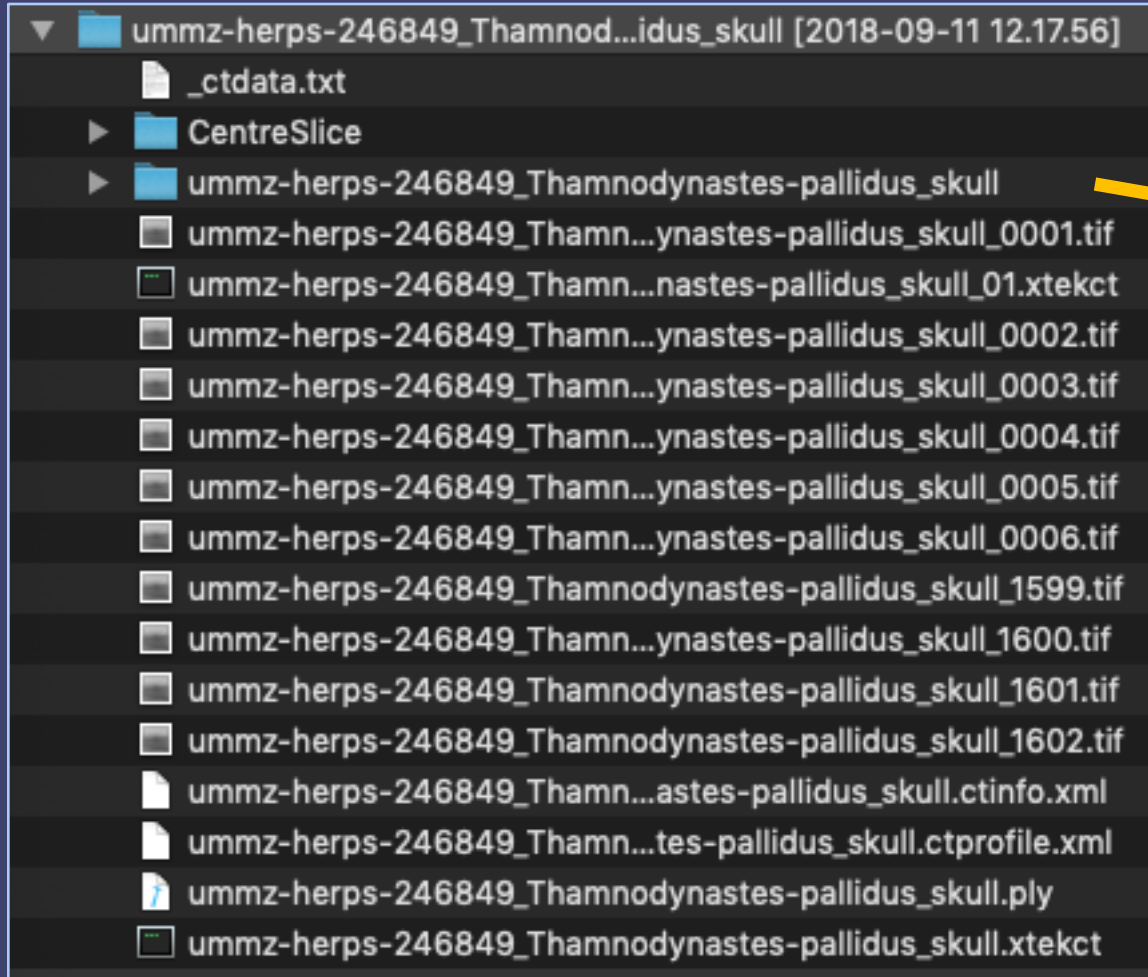
UMMZ-HERPS-SKULL-246849-RECON
RECON.ZIP

UMMZ-HERPS-BODY-246849-RAW
RAW.ZIP

UMMZ-HERPS-BODY-246849-RECON
RECON.ZIP

Data Model Iterations

DATA - BEFORE AND AFTER RE-ORG



DATA SEPARATED BY REGION OF INTEREST (ROI) – SKULL OR WHOLE BODY

SOURCES OF METADATA

Scanning process metadata

[XTekCT]

Name=ummz-herps-246849_Thamnodynastes-pallidus_skull

InputSeparator=_

OutputSeparator=_

InputFolderName=

OutputFolderName=ummz-herps-246849_Thamnodynastes-pallidus_skull_01

VoxelsX=1416

VoxelsY=914

VoxelsZ=654

VoxelSizeX=0.0150099083890765

VoxelSizeY=0.0150099083890765

VoxelSizeZ=0.0150099083890765

OffsetX=0.190600817656717

GBIF

```
{ "endOfRecords": true, "count": 1, "results": [ { "key": 1987213257, "datasetKey": "8e479b8f-ac16-4bed-a1b3", "publishingOrgKey": "6ea87510-0561-11d8-b851-b8a03c50a862", "installationKey": "cfbb523a-88cfce32b11d64", "publishingCountry": "US", "protocol": "DWVC_ARCHIVE", "lastCrawled": "2020-02-08T13:14:55.927+0000", "lastParsed": "2020-02-10T21:37:29.562+0000", "crawlId": 155, "extensions": {}, "basisOfRecord": "PRESERVED_SPECIMEN", "individualCount": 1, "establishmentMeans": "NATIVE", "taxonKey": 2454765, "kingdomKey": 1, "phylumKey": 44, "classKey": 358, "orderKey": 715, "familyKey": 4819481, "genusKey": 2454765, "acceptedTaxonKey": 2454765, "scientificName": "Thamnodynastes Wagler, 1830", "acceptedScientificName": "Thamnodynastes Wagler, 1830", "kingdom": "Animalia", "phylum": "Chordata", "order": "Squamata", "family": "Dipsadidae", "genus": "Thamnodynastes", "genericName": "Thamnodynastes", "taxonRank": "GENUS", "taxonomicStatus": "ACCEPTED", "decimalLongitude": -72.249633, "decimalLatitude": -3.577267, "stateProvince": "Loreto", "issues": ["COORDINATE_ROUNDED", "GEODETIC_DATUM_ASSUMED_WGS84", "CONTINENT_INVALID"], "modified": "2018-05-24T00:00:00.000+0000", "lastInterpreted": "2020-02-10T21:37:29.562+0000", "license": "http://creativecommons.org/licenses/by/4.0/legalcode", "identifiers": [], "media": [], "facts": [], "relations": [], "geodeticDatum": "WGS84", "class": "Reptilia", "countryCode": "PE", "country": "Peru", "identifier": "eb3f06f1-2d04-4406-baa0-3fb85af835c3", "recordNumber": "RAB 1816", "higherGeography": "PERU, LORETO", "locality": "YANASHI", "language": "en", "gbifID": 1987213257, "collectionCode": "herps", "occurrenceID": "eb3f06f1-2d04-4406-baa0-3fb85af835c3", "type": "PhysicalObject", "disposition": "in coll", "catalogNumber": "246849", "recordedBy": "JOANNA LARSON AND CREW", "occurrenceStatus": "present", "institutionCode": "UMMZ", "higherClassification": "OPHIDIA | COLUBRIDAE | THAMNODYNASTES"} ], "facets": [] }
```

YAML FILE FOR INGEST

```
---
:user:
:visibility: open
:email: sborda@umich.edu
:ingerster: 'fritx@umich.edu'
:source: DBDv2
:works:
:depositor: sborda@umich.edu
:in_collections:
- 05741r77z
:owner: 'ummz-herp-data@umich.edu'
:authoremail: 'ummz-herp-data@umich.edu'
:creator:
- 'University of Michigan Museum of Zoology'
:title:
- 'Computed tomography voxel dataset for ummz:herps:246849-THAMNODYNASTES PALLIDUS-Skull'
:referenced_by:
- 'For more information on the original UMMZ specimen, see: https://www.gbif.org/occurrence/1987213257'
:methodology: 'This dataset was created at the University of Michigan Museum of Zoology using a procedure involving computed tomography'
:keyword:
- 'Animalia'
- 'Chordata'
- 'Reptilia'
- 'OPHIDIA'
- 'COLUBRIDAE'
- 'THAMNODYNASTES PALLIDUS'
- '1987213257'
- 'computed tomography'
- 'X-ray'
- '3D'
:description:
- 'Scan of specimen ummz:herps:246849 (THAMNODYNASTES PALLIDUS) - Skull. Reconstructed Dataset includes 654 TIF images (each 1416 x 1416 x 1416) and 654 NII images (each 1416 x 1416 x 1416)'
- 'Scan of specimen ummz:herps:246849 (THAMNODYNASTES PALLIDUS) - Skull. Raw Dataset includes 1601 TIF images (each 1416 x 1416 x 1416) and 1601 NII images (each 1416 x 1416 x 1416)'
:rights_license:
- https://creativecommons.org/licenses/by-nc-sa/4.0/
:date_coverage:
- '2018-09-11'
:subject_discipline:
- 'Science'
:language:
- 'English'
:curation_notes_admin:
- 'UMMZ Batch Ingest'
:doi: 'mint_now'
:filenames:
- Recon-Skull-ummz-herps-246849_Thamnodynastes-pallidus-1987213257.zip
- ummz-herps-246849_Thamnodynastes-pallidus_skull_0654.tif
- Raw-Skull-ummz-herps-246849_Thamnodynastes-pallidus-1987213257.zip
- ummz-herps-246849_Thamnodynastes-pallidus_skull.ply
```


IN DEEP BLUE DATA

https://deepblue.lib.umich.edu/data/concern/data_sets/9880vr066?locale=en

Description Scan of specimen ummz:herps:246849 (THAMNODYNASTES PALLIDUS) - Skull. Raw - Dataset includes 1601 TIF images (each 1416 x 914 x 1 voxel at 0.0150099083890765 mm resolution, derived from 1601 scan projections), xtek and vgi files for volume reconstruction.

Scan of specimen ummz:herps:246849 (THAMNODYNASTES PALLIDUS) - Skull. Reconstructed - Dataset includes 654 TIF images (each 1416 x 914 x 1 voxel at 0.015010 mm resolution, derived from 1601 scan projections), xtek and vgi files for volume reconstruction.






Keyword [Animalia](#) [Chordata](#) [Reptilia](#) [OPHIDIA](#) [COLUBRIDAE](#) [THAMNODYNASTES PALLIDUS](#) [1987213257](#) [computed tomography](#) [X-ray](#) [3D](#)

Date coverage 2018-09-11

Citations to related material For more information on the original UMMZ specimen, see: <https://www.gbif.org/occurrence/1987213257>

DATA MODEL - FINAL

Files (Count: 5; Size: 12.4 GB)

	Title	Original Upload	Last Modified	File Size	Access	Actions
	Raw-Skull-ummz-herps-246849_Tham...7.zip	2019-11-21		10.8 GB	Open Access	Select an action ▼
	ummz-herps-246849_Thamnodynastes...l.ply	2019-11-21		91.7 MB	Open Access	Select an action ▼
	ummz-herps-246849_Thamnodynastes...0.tif	2019-11-21		7.63 MB	Open Access	Select an action ▼
	Recon-Skull-ummz-herps-246849_Th...7.zip	2019-11-21		1.49 GB	Open Access	Select an action ▼
	ummz-herps-246849_Thamnodynastes...0.tif	2019-11-21		2.47 MB	Open Access	Select an action ▼

LESSONS LEARNED

- Determine the sustainability of the source for specimen metadata, to use iDigBio, or GBIF. The project team decided to switch from iDigBio to GBIF mid-project because of stability concerns, which meant adjustments to the script that captured the metadata.
- Some issues are only found when code is applied. In the case of noticing that GBIF records were not showing the second epithet of the scientific name had to change the GBIF API end-point to /Fragment.

```
record_results = requests.get(gbif_baseurl + 'occurrence/' + str(item['key']) + '/fragment')
```

- Thoroughly review test uploads into DBD for completeness, example the “Citations to related materials” field from the *.yml file was not being uploaded to the work
- Code can always benefit from additional error handling.
- Should have nailed down the data model at the very beginning. the project team spent a lot of time working through various data models.

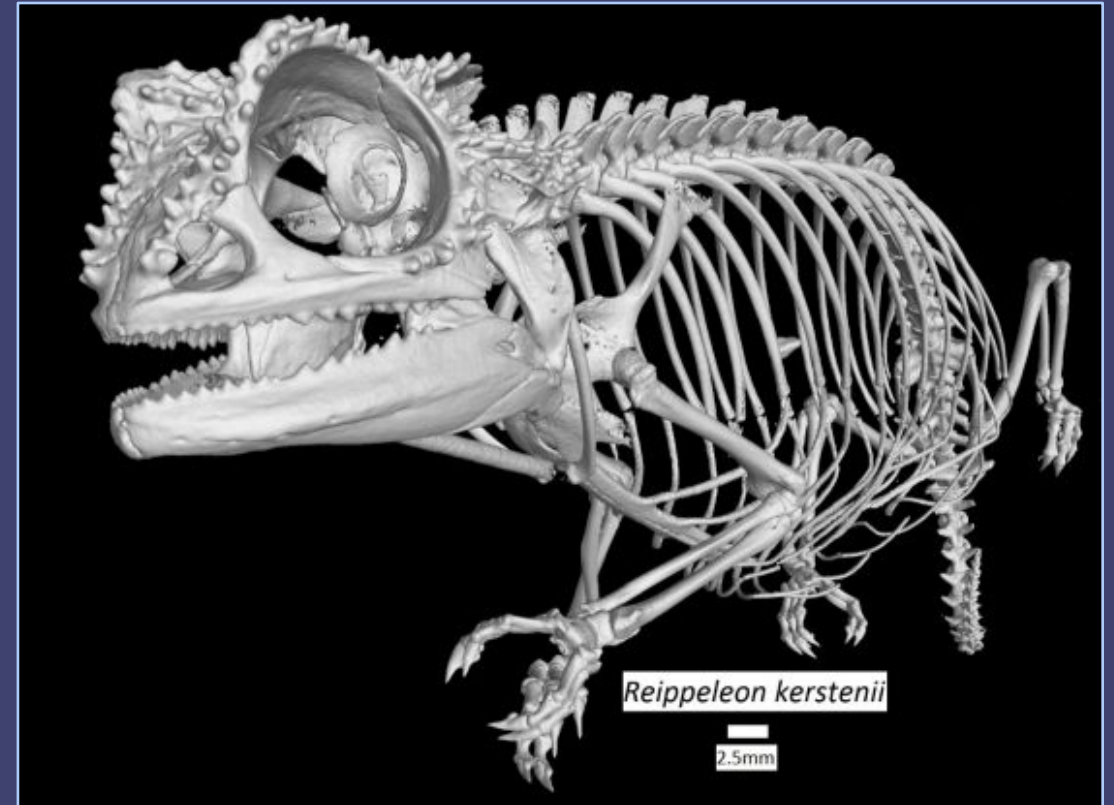
INFO

Susan Borda – sborda@umich.edu

Code: <https://github.com/mutanthumb/ummzgbif>

Deep Blue Data repository -
<https://deepblue.lib.umich.edu/data/>

Museum of Zoology -
<https://lsa.umich.edu/ummz>



<https://lsa.umich.edu/ummz/ct-scanner.html>