

Growing Up Digital MorphoSource as a Case Study for Scaling a Digital Repository



Julie M. Winchester / Product Manager & Developer (Duke)
Doug M. Boyer / MorphoSource Director, Associate Prof. (Duke)
Tim McGeary / Associate University Librarian (Duke)
Tim Ryan / Associate Prof. (PSU)



DUKE UNIVERSITY
LIBRARIES



PennState



MORPHO SOURCE

at a glance

- Founded 2013
- Place for 3D data on vouchered specimens



MORPHO SOURCE

at a glance

- Founded 2013
- Place for 3D data on vouchered specimens
- Supported by institutions and NSF



Duke
UNIVERSITY



PennState



NSF DBI-1661386

oVert

NSF DBI-1701714



at a glance

- Founded 2013
- Place for 3D data on vouchered specimens
- Supported by institutions and NSF
- Open and free (mostly)

Current Holdings

44,743 3D media files

11,880 Specimens

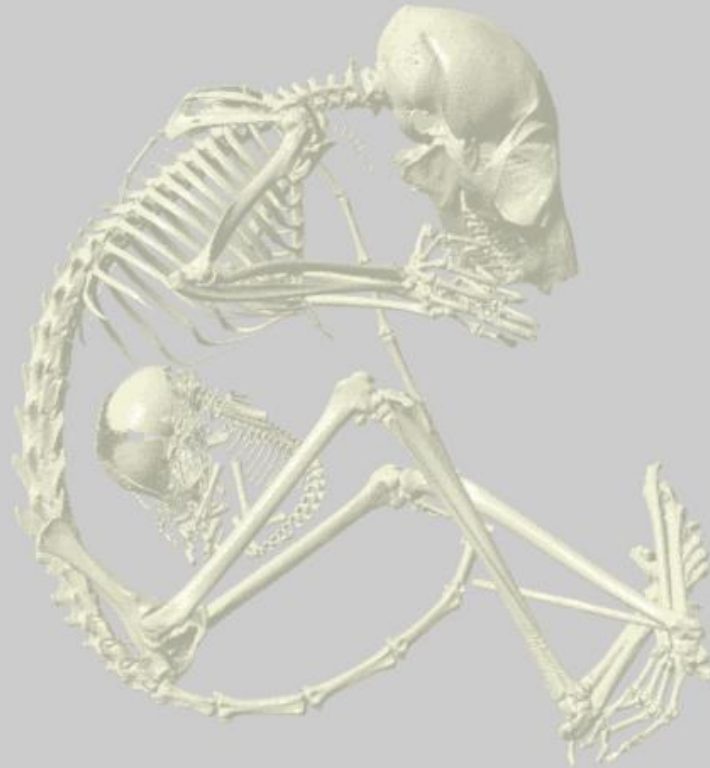
260 Institutions

6,174 Users

Using MorphoSource, you can...

- Browse and search 3D media by taxonomy, institution, project, and specimen description
- Upload, archive, and share 3D media from your research or collection
- Robustly control access to your media with strong privacy and sharing tools
- Track downloads/views and assign DOIs to media, enabling citation of media in publications

The screenshot shows the MorphoSource website interface. At the top, there is a navigation bar with links for ABOUT, BROWSE, and DASHBOARD. Below this, there is a search bar and a list of links including AMNH MAMMAL COLLECTION, MEDIA CART, and STATS. The main content area is titled "Getting Started" and features a "Find & Download Datasets" section with a "BROWSE" button and a search input field. To the right, there is a "Useful Info" section with links for "Information for Users", "Information for Contributors", "Terms", and "User Guide". A large image of a 3D model of a primate foot is displayed, with a caption: "foot of Daubentonia madagascariensis scanned at 38micron resolution at Duke Evolutionary Anthropology department's new high resolution microCt facility. Click here if you are interested in details on the facility". Below this, there is a "Recently Published" section with a list of projects, including "First North American Fossil Monkey and Early Miocene Tropical Biotic Interchange" and "Simbakubwa kutokaafrika (Hyaenodonta, 'Creodonta', Mammalia) from the earliest Miocene of Kenya and the evolution of".



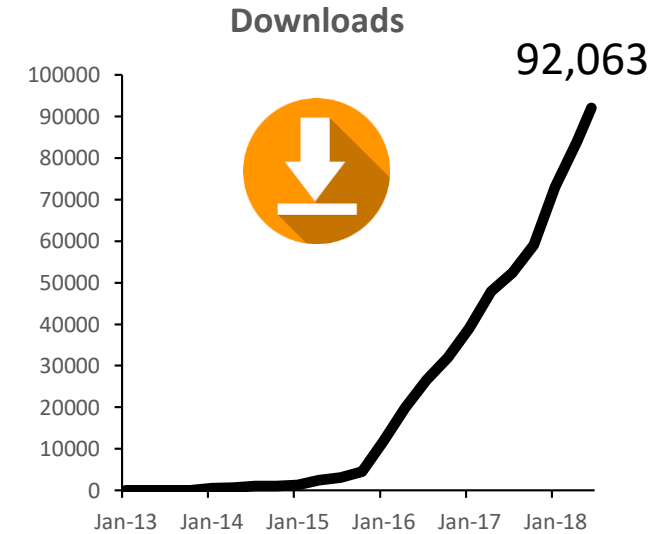
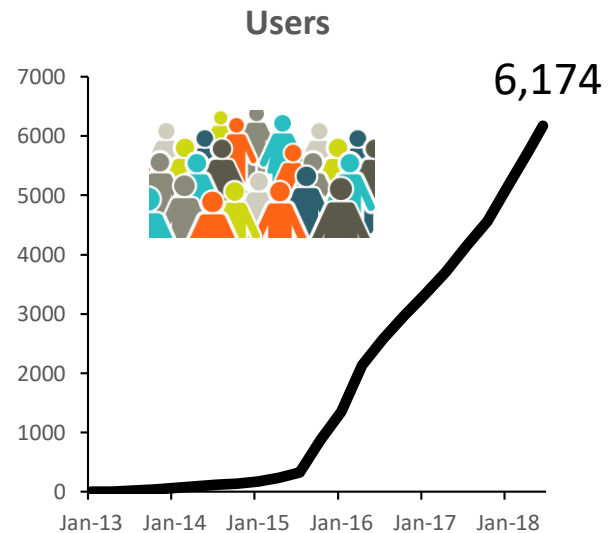
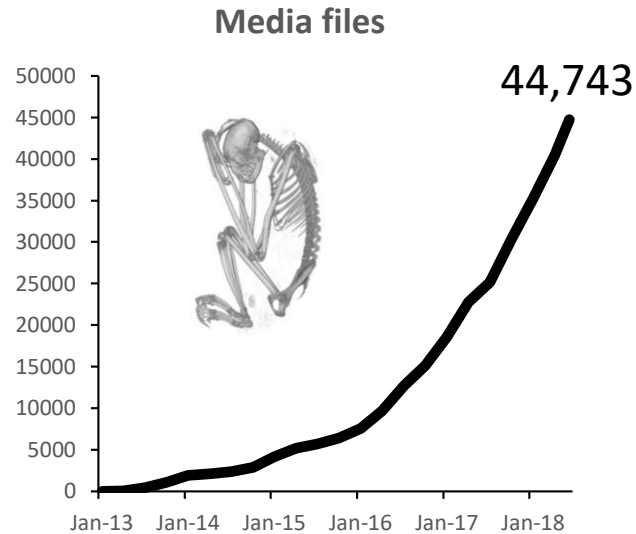
Natural History By-Catch

Outline

How is MorphoSource responding to emerging challenges?

- Current and future challenges
- Two-pronged response
 - Updates to the v1.0 site
 - Rebuilding site as v2.0, a digital data preservation stack

Challenges: Scaling



- Approximately 100% growth in data size over previous year
- Increases load on infrastructure and has unique pressures on system (user interface, etc.) design

Challenges: Access and Preservation of Data



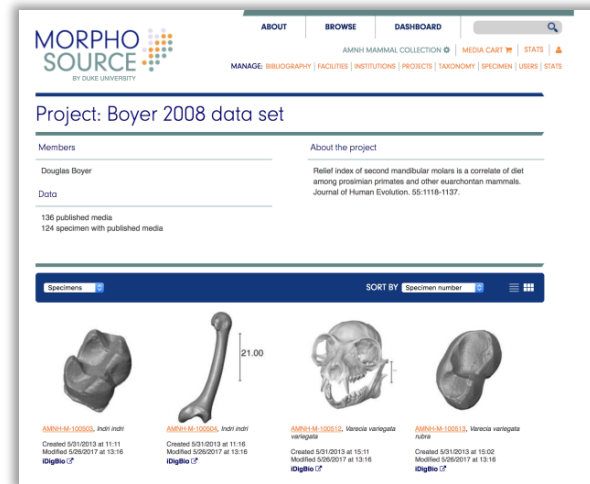
Monitor data organization, appropriateness, and corruption



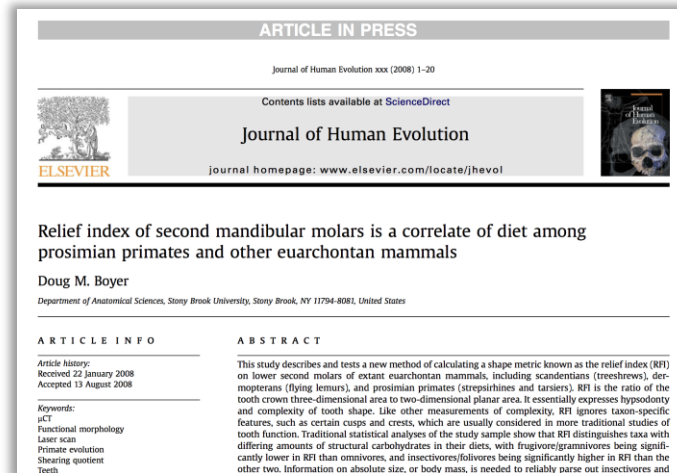
Keep data discoverable and accessible
(no data silos!)

Challenges: Expanding User Needs

New forms of data management

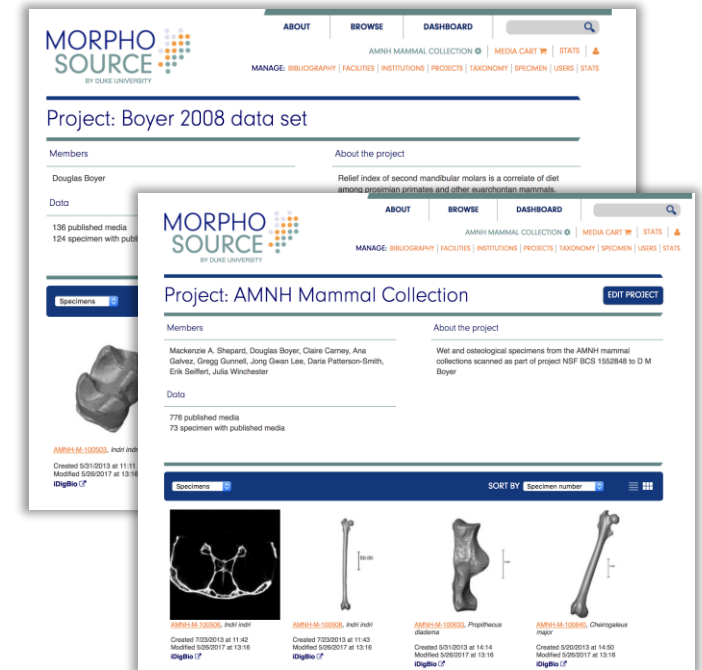


Media belong to one project



“Publication packet” works well for researchers

Publication project...



... Or collection project?

Works less well for museum curation purposes

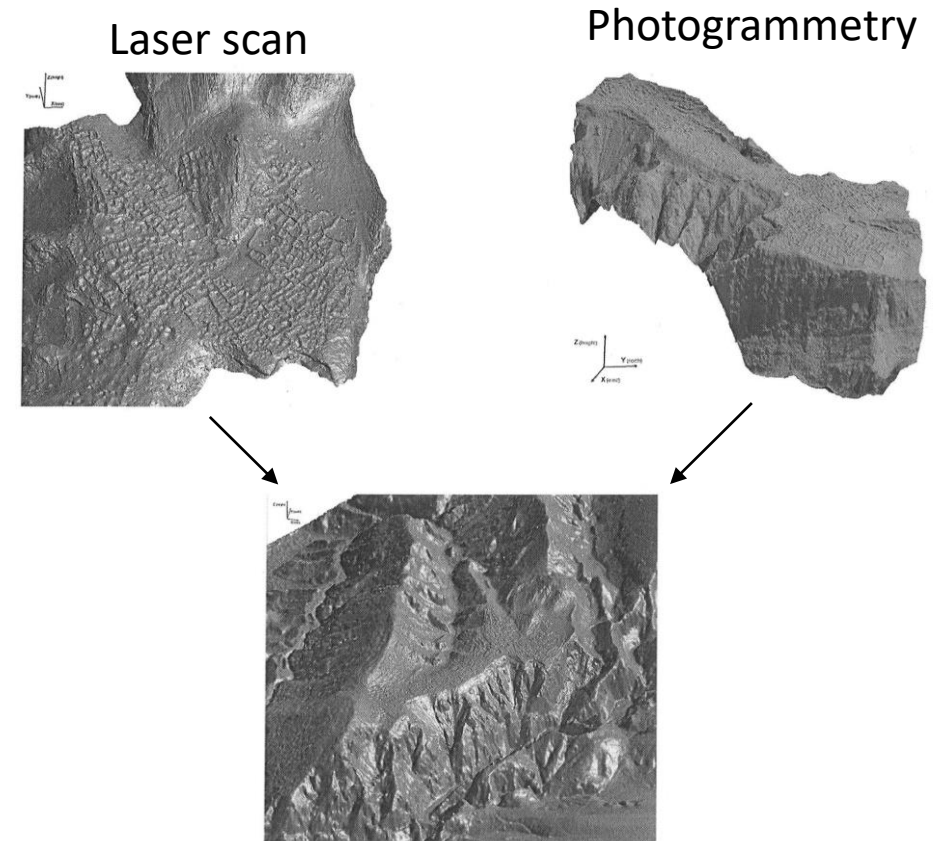
Challenges: Expanding User Needs

New forms of data management

New data types/use cases



3D media of objects that are not biological specimens



Multi-modal images (e.g., archaeological or paleontological site models)

Responding to Selective Pressures



Making necessary modifications to current site

- Resolving consistent taxonomy
- Comprehensive recordset-level data reporting



Rebuilding site as digital data preservation stack using library/information science repository tools

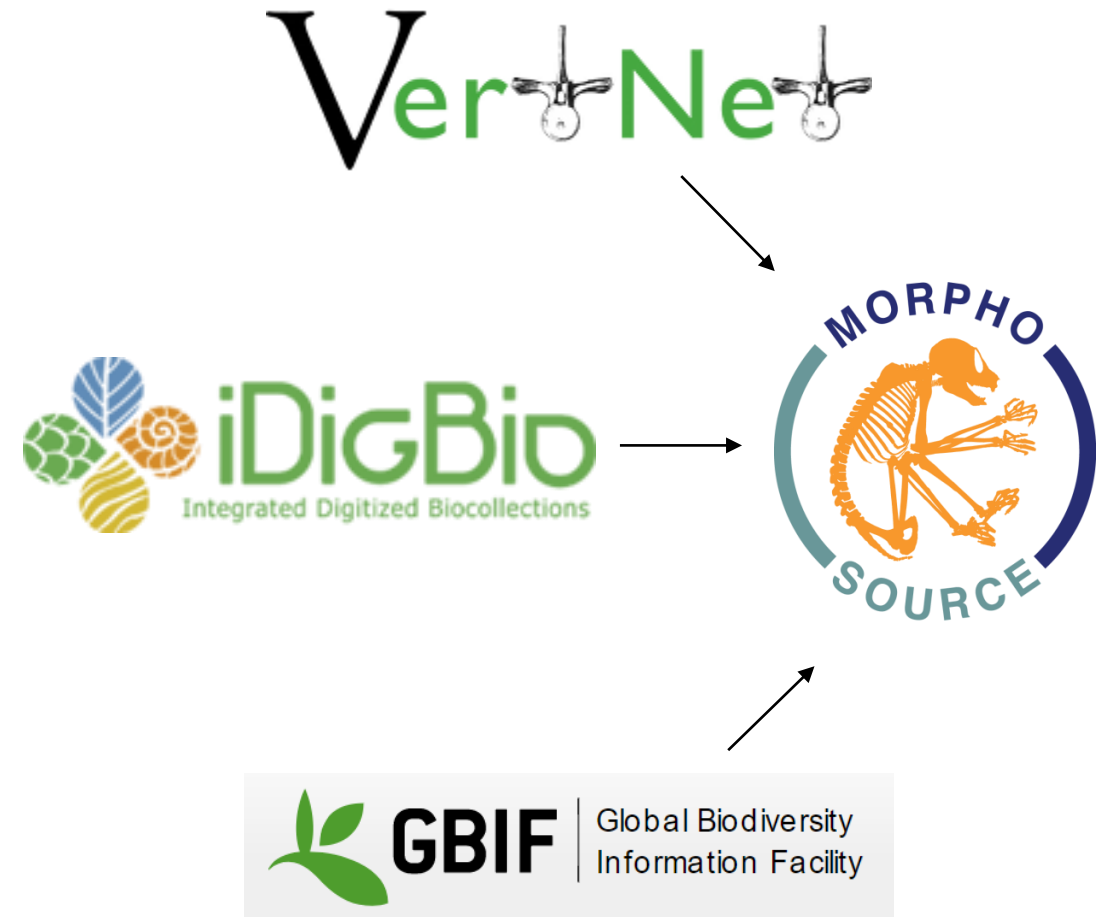
- Samvera/Fedora stack with library integration
- Renovating the core site data model

MorphoSource 1.0: Resolved Taxonomy

Method for externally resolving taxonomy

- Consistency
- Address persistent rank-level issues (e.g., Synapsida and Mammalia are both orders)

Consistency from voting



MorphoSource 1.0: Resolved Taxonomy

Method for externally resolving taxonomy

- Consistency
- Address persistent rank-level issues (e.g., Synapsida and Mammalia are both orders)

Consistency from voting

>97% of all specimens categorized

Taxonomy

Start Browsing by Phylum

[Arthropoda](#) (43)

[Chordata](#) (21936)

[Echinodermata](#) (1)

[Mollusca](#) (27)

[Tracheophyta](#) (12)

MorphoSource 1.0: Resolved Taxonomy

Method for externally resolving taxonomy

- Consistency
- Address persistent rank-level issues (e.g., Synapsida and Mammalia are both orders)

Consistency from voting

>97% of all specimens categorized

Taxonomy

Browsing By

[Eukaryota](#) (22021) › [Animalia](#) (22009) › [Chordata](#) (21936)

Refine

[Actinopterygii](#) (2093)

[Amphibia](#) (768)

[Aves \(and fossil outgroups\)](#) (470)

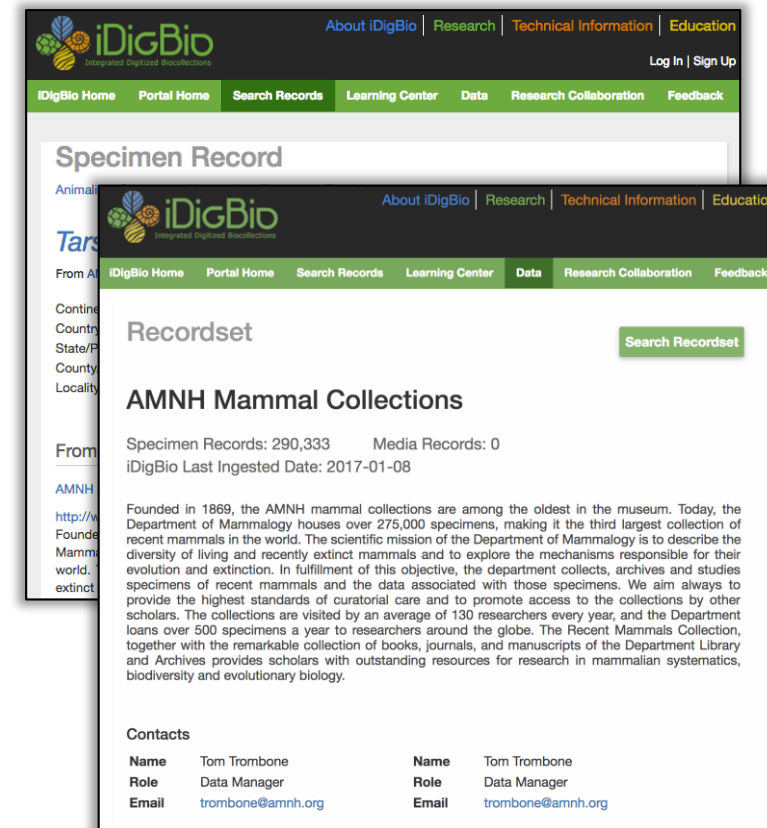
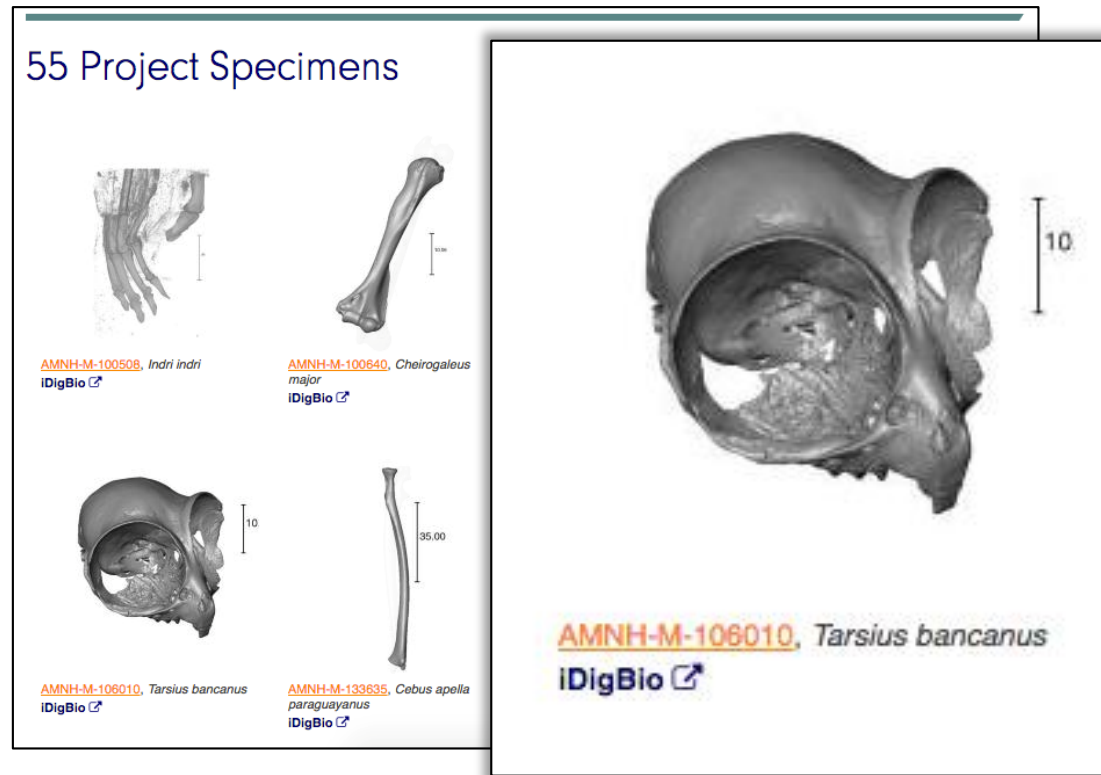
[Chondrichthyes](#) (65)

[Mammalia \(and fossil outgroups\)](#) (17858)

[Reptilia](#) (681)

[Sarcopterygii](#) (1)

MorphoSource 1.0: Data Reporting



- Specimens link to iDigBio records, which are reported as a recordset (collection)

MorphoSource 1.0: Data Reporting

Reports per recordset

- Audubon Core description of media metadata, including links to media page
- All media downloads with user demographics
- Download request reports with request free text

The screenshot shows the MorphoSource website header. On the left is the logo "MORPHO SOURCE BY DUKE UNIVERSITY". To the right are navigation links: "ABOUT", "BROWSE", "DASHBOARD", "AMNH MAMMAL COLLECTION", "MEDIA CART", "STATS", and "MANAGE: BIBLIOGRAPHY | FACILITIES | INSTITUTIONS | PROJECTS | TAXONOMY | SPECIMEN | USERS | STATS". Below the navigation is a search bar and the text "Data reporting for MorphoSource media".

IPT - Hosted by VertNet (842a2bb5-d705-4d6c-8401-abf3ca28c05d)

Recordset	Media	Downloads	Download Requests	Pub Date
AMNH Herpetology Collections (7ce9b7d0-a8da-4528-bbe3-2c4f40719cea)	CSV EML	CSV EML	CSV EML	Thu, 24 May 2018 12:33:01 -0400
AMNH Mammal Collections (cb790bee-26da-40ed-94e0-d179618f9bd4)	CSV EML	CSV EML	CSV EML	Thu, 24 May 2018 12:06:01 -0400
AUM Fish Collection (cab6d230-499b-4a21-8cc2-2750e14e92f8)	CSV EML	CSV EML	CSV EML	Thu, 24 May 2018 12:33:37 -0400
Biodiversity Research and Teaching Collections - TCWC Vertebrates (79dfdec6-3e24-489c-a7ce-85dce52bc3f9)	CSV EML	CSV EML	CSV EML	Thu, 24 May 2018 12:32:30 -0400
BYU Herpetology Collection (3872f27e-cf4e-40bd-b91b-ba7b723a88e5)	CSV EML	CSV EML	CSV EML	Thu, 24 May 2018 12:31:07 -0400
CM Herps Collection (045aa661-f985-4203-80ff-98daafdf377)	CSV EML	CSV EML	CSV EML	Thu, 24 May 2018 12:30:50 -0400
CM Vertebrate Paleontology Collection (71b8ffab-444e-43f9-9a9c-5c42b0aa5eb)	CSV EML	CSV EML	CSV EML	Thu, 24 May 2018 12:34:52 -0400
CUMV Fish Collection (Arctos) (4f1ecde59-9f59-44eb-ab6f-4a50b4ed85cf)	CSV EML	CSV EML	CSV EML	Thu, 24 May 2018 12:33:24 -0400
ISM Ornithology Collection (6e6bec70-a148-49cb-9f97-e64e2f1ae5b7)	CSV EML	CSV EML	CSV EML	Thu, 24 May 2018 12:30:58 -0400
LACM Vertebrate Collection (2d853a6d-50ec-4931-8e91-48fc2491fdee)	CSV EML	CSV EML	CSV EML	Thu, 24 May 2018 12:30:49 -0400
LSUMZ (LSU MNS) Fishes Collection (6d9d4f55-4cb4-4c8f-acc7-b465eb5f703c)	CSV EML	CSV EML	CSV EML	Thu, 24 May 2018 12:33:01 -0400
MVZ Herp Collection (Arctos) (b3976394-a174-4ceb-8d64-3a435d66bde6)	CSV EML	CSV EML	CSV EML	Thu, 24 May 2018 12:32:58 -0400

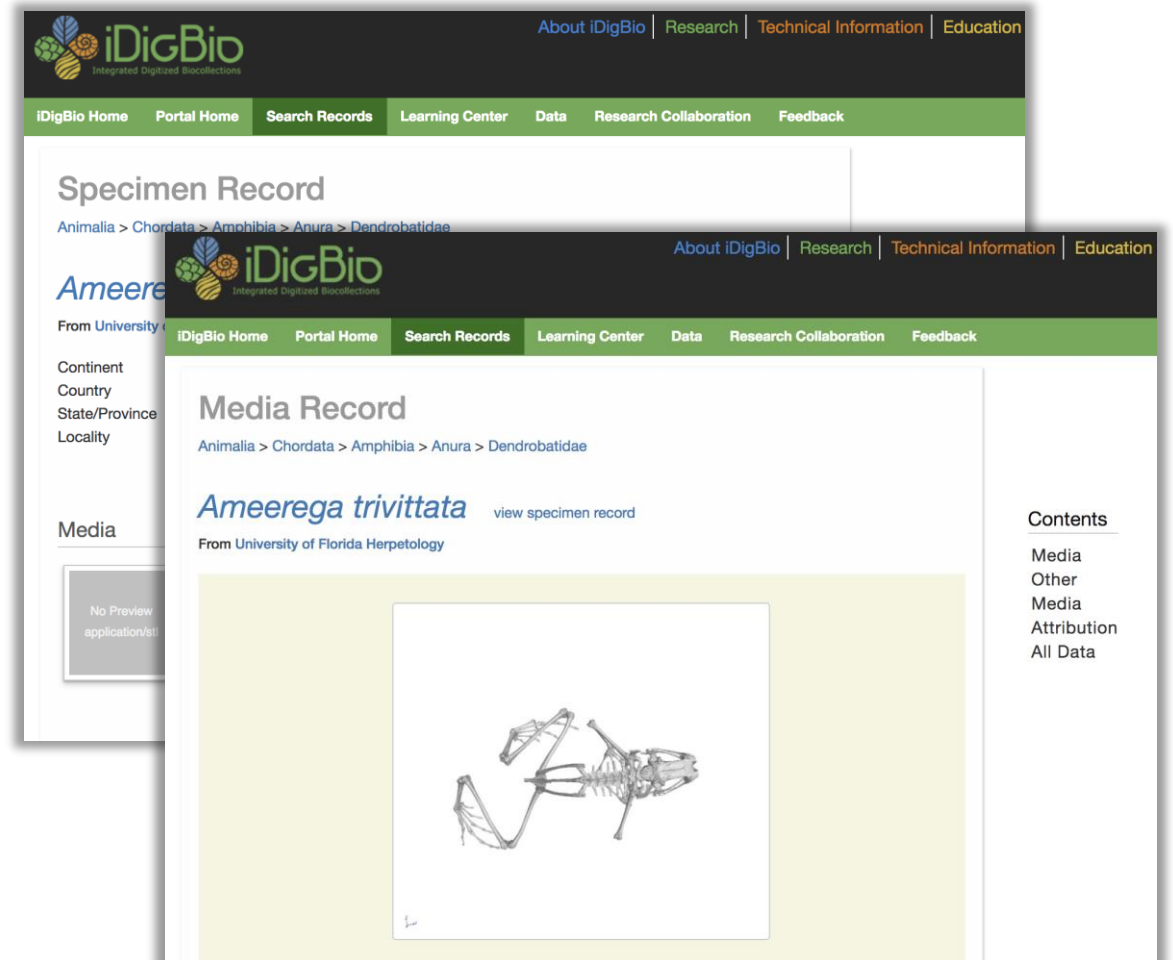
MorphoSource 1.0: Data Reporting

Reports per recordset

- Audubon Core description of media metadata, including links to media page
- All media downloads with user demographics
- Download request reports with request free text

RSS Feed

- Enables reporting of MorphoSource media to iDigBio
- Recordset reporter must download media CSV, include in IPT



MorphoSource 2.0: Hyrax Repository

Hyrax

- Samvera/Fedora technology stack
- Digital data preservation software
- Can support most standardized metadata schema through RDF
- Fixity checking



Duke Library Governance

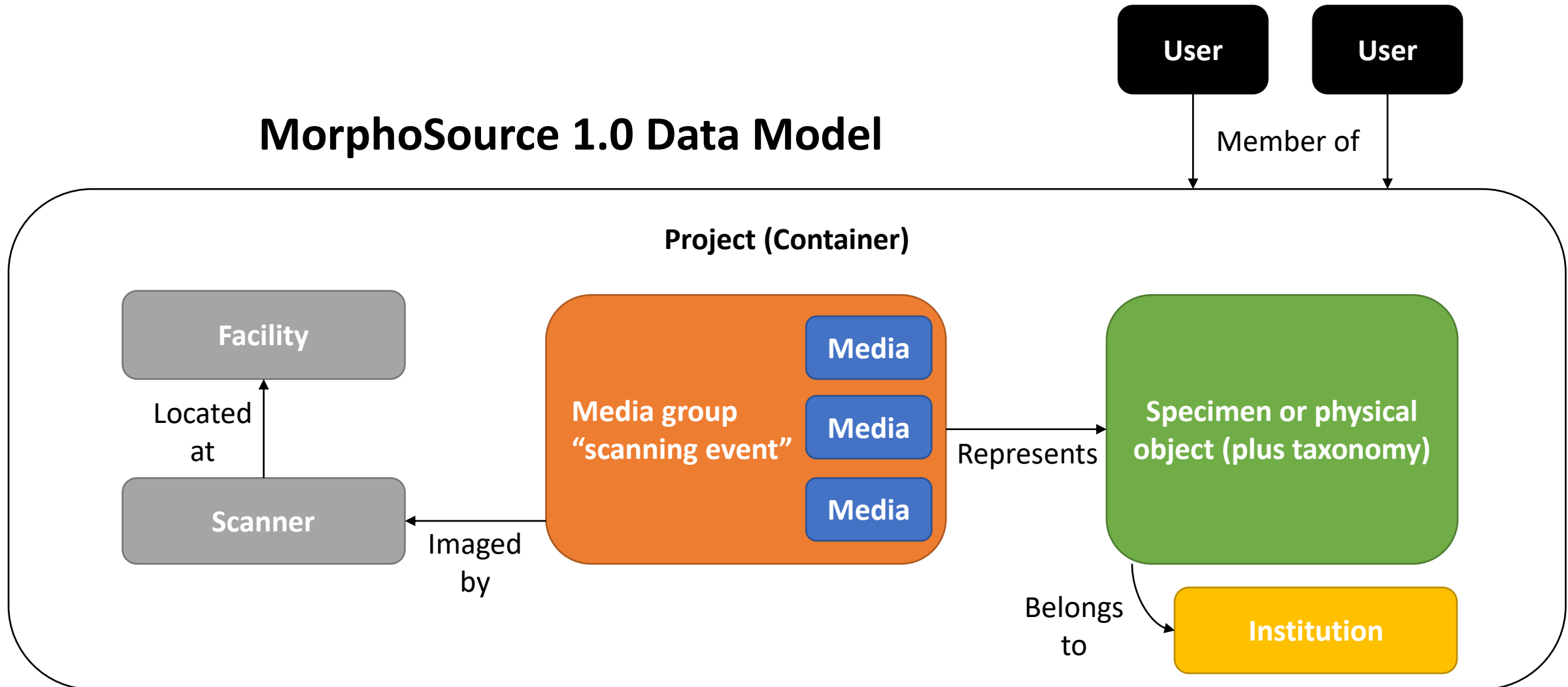
- Integrating library technology (e.g., Hyrax)
- Support in resources and personnel



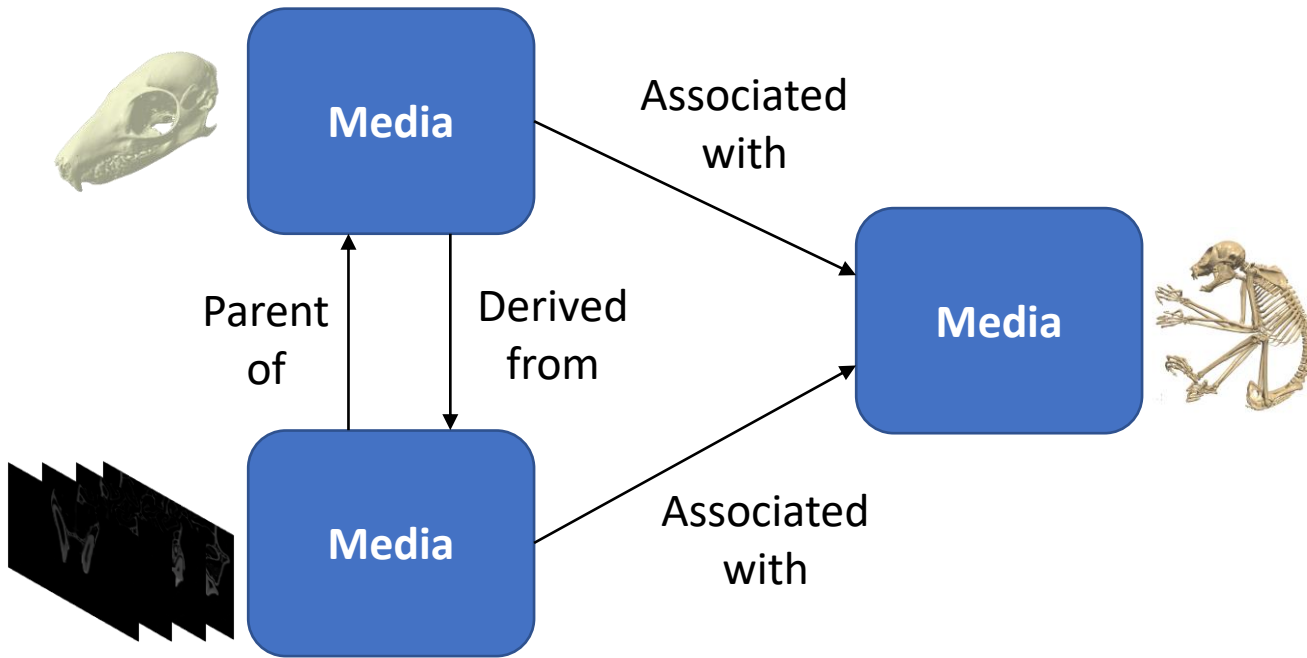
Record with metadata

MorphoSource 2.0: Updating data model

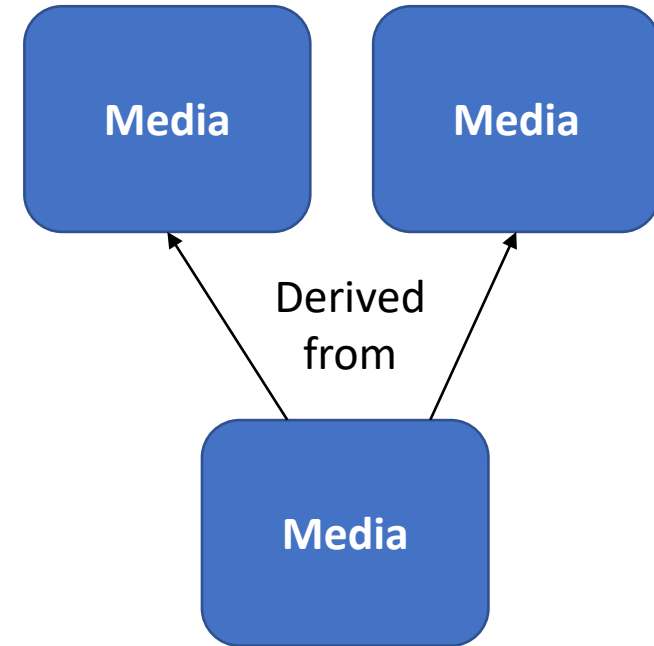
MorphoSource 1.0 Data Model



MorphoSource 2.0: Updating data model

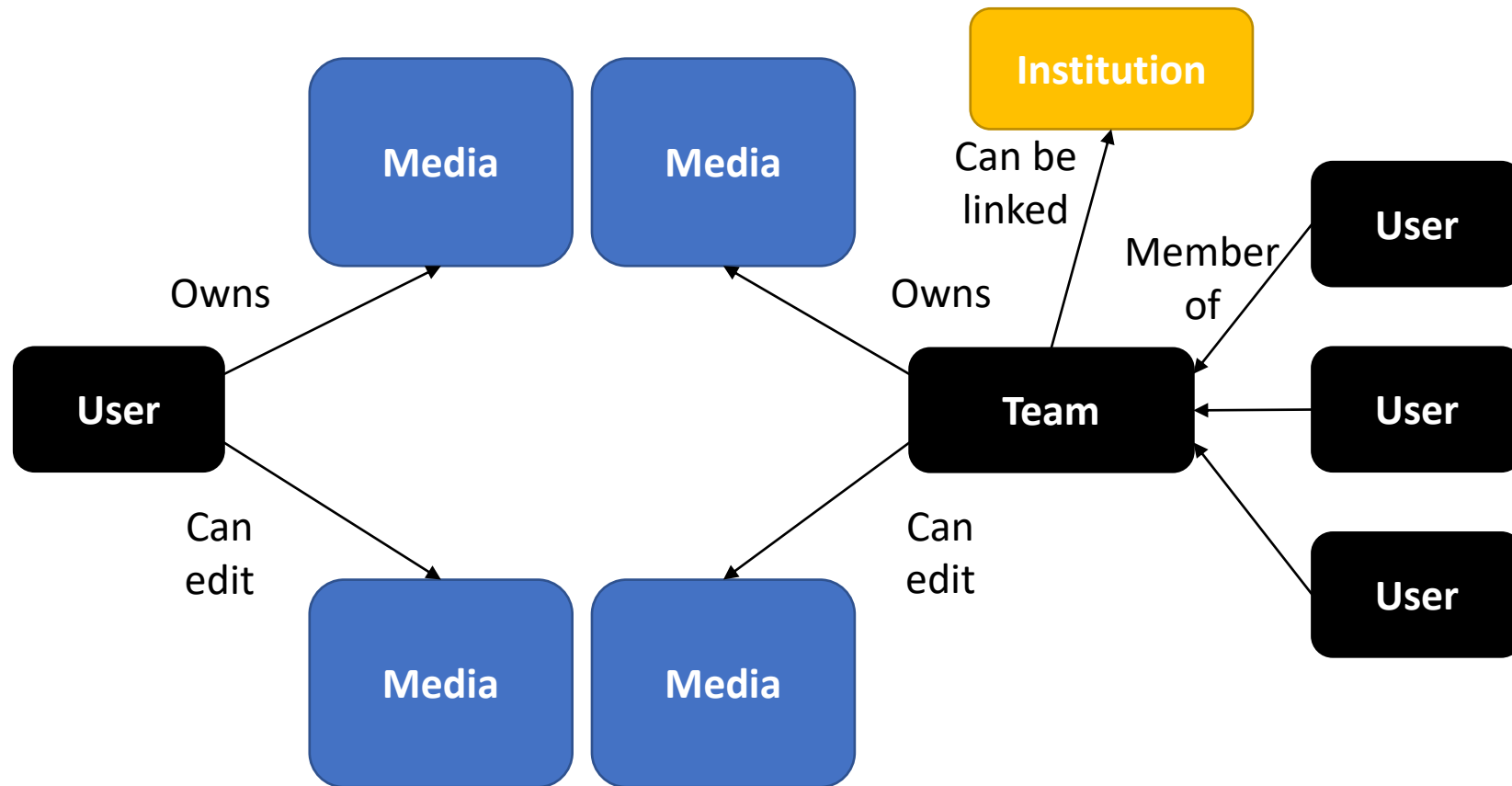


Enables multi-modal media



No more media groups. Instead, media have flexible relationships with other media

MorphoSource 2.0: Updating data model

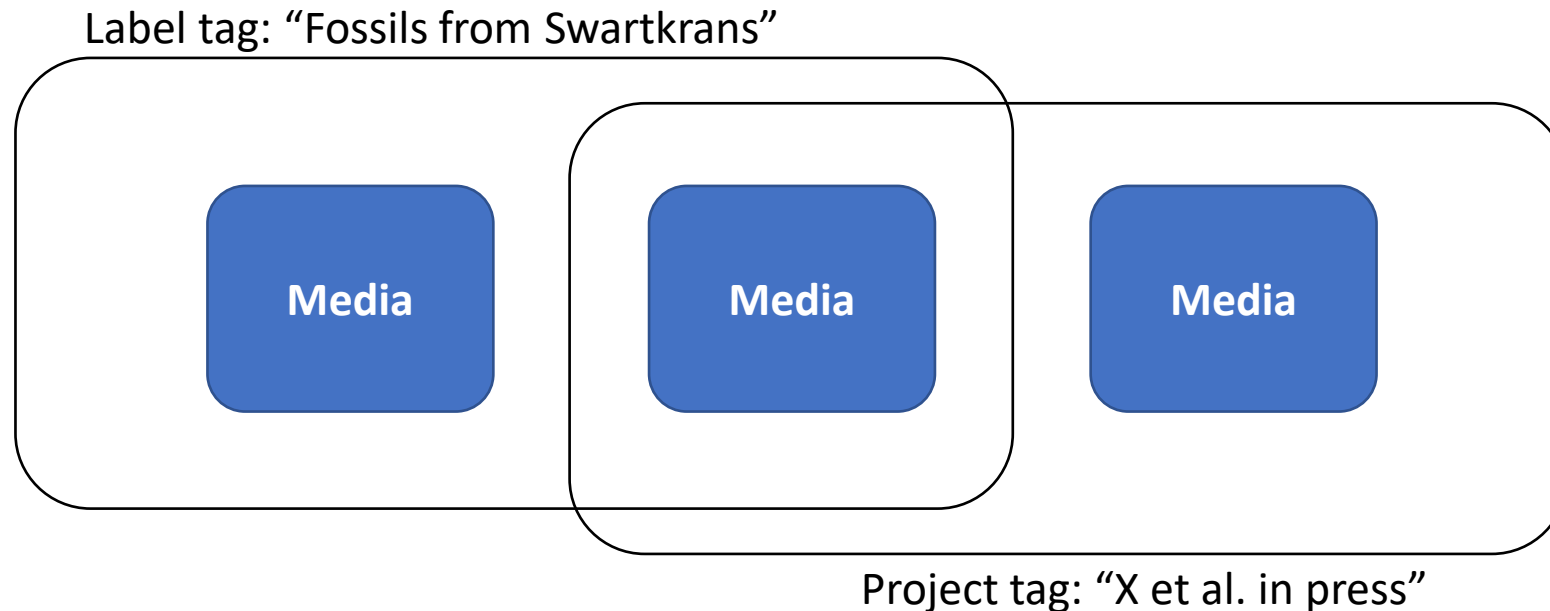


Users and teams both:

- Own media
- Can grant access to and share media
- UI: Have public profile page showing published media

Users and teams own media, not projects. Teams represent groups of users that can be linked to institutions.

MorphoSource 2.0: Updating data model



Projects as tags, not containers. Different tag types for publication, lesson plans, etc. Can be simple labels or for batch edit/granting access

Conclusions

- MorphoSource is an online repository where researchers, curators, and the general public can find, download, upload 3D media of physical objects
- Like organisms, repositories must respond to changing conditions
 - Designing systems for unanticipated challenges is (unsurprisingly) hard
 - Proof of concept → Scaled repository
- Example of restructuring a repository in response to evolving needs

Acknowledgments

For support & funding

- NSF DBI-1661386 (MorphoSource) and NSF DBI-1701714 (oVert)
- Duke University Trinity College of Arts & Sciences
- Duke Shared Materials Instrumentation Facility
- Duke Biology IT Center
- Duke Libraries

For discussion leading to development of concepts

- J. Jernvall, A. Evans, G. Evans, D. Blackburn, J. Blundell, G. Motz, H. Little, R. Snyder, E. Stanley, J. Thostenson, J. Sessa, M. O’Leary, M. Uhen, G. Yapuncich, J. Maisano, E. Stanley, R. O’Leary, S. Grant, K. Webbink, L. Witmer, T. Ryan, C. Grant, C. Ross, J. Richtsmeier, G. Nelson
- CS3DP (J. Moore, H. Scates-Kettler, A. Rountrey), LIB3DVR (N. Hall, R. MacDonald)
- iiiF (S. Snyderman)

For work developing and populating MorphoSource

- Seth Kaufman, Maria Passarotti (Whirl-i-gig, Inc.)
- Alex Thompson, Kevin Love, Dan Stoner (iDigBio)
- Technicians & students: Mackenzie Shepard, Darbi Griffith, Ana Galvez, Miles Schaeffer, Mercedes Zapata-Garcia, Shane Daly, Jong Gwan Lee, Sunghoon Liu, Ksenia Sokolova, Anne Driscoll, Kevin Vo, Annie Lott, Callie Crawford, Barbara Zhao, and many, many more

