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)

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

MORPHO SOURCE

oVert

NSF DBI-1661386

NSF DBI-1701714

MORPHO SOURCE *at a glance*

- Founded 2013

- Place for 3D data on vouchered specimens

- Supported by institutions and NSF

- Open and free (mostly)

Current Holdings

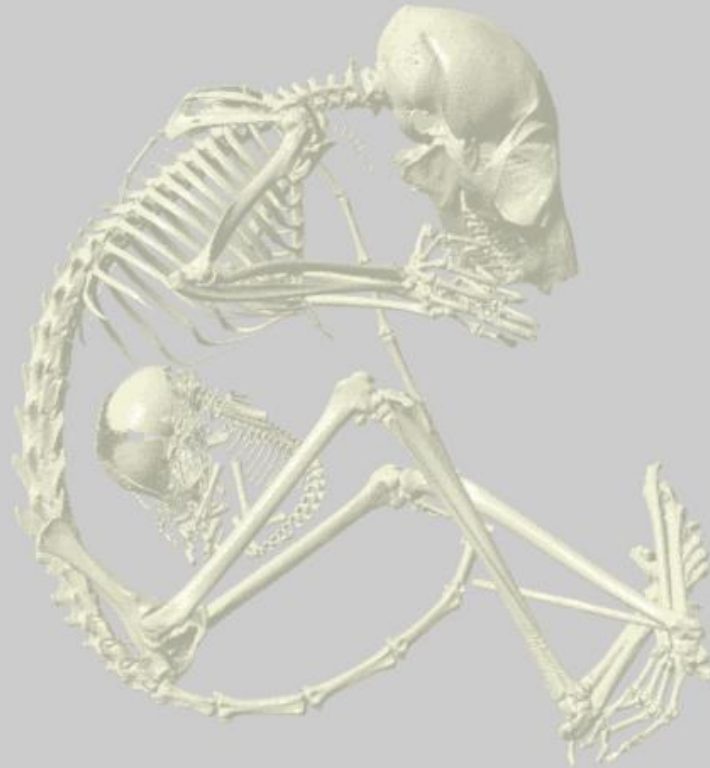44,743 3D media files

11,880 Specimens

260 Institutions

6,174 Users

at a glance

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
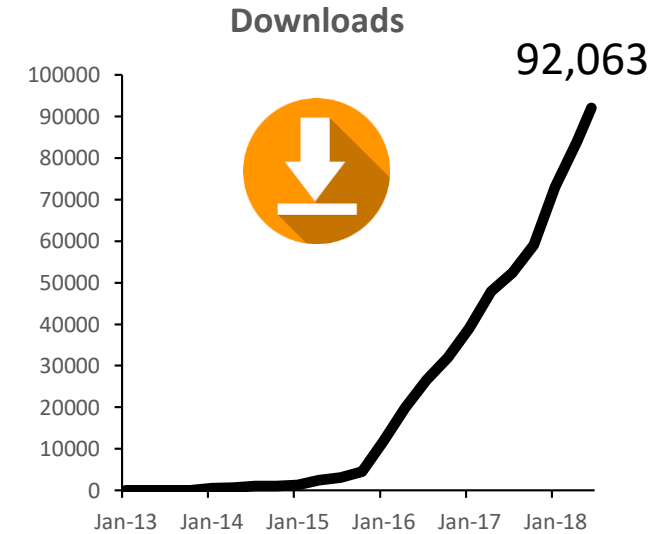
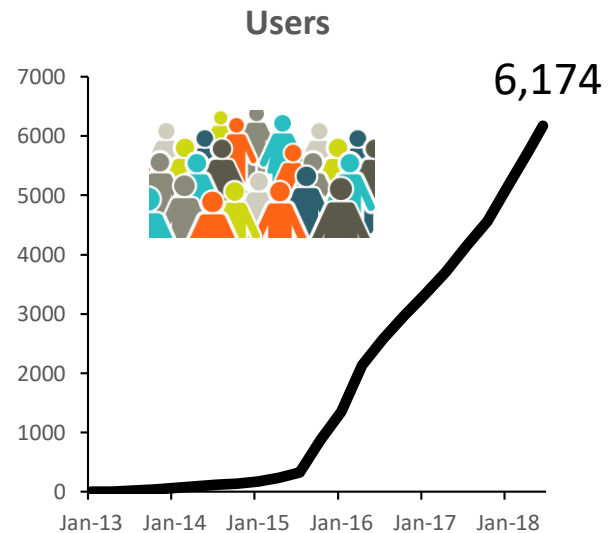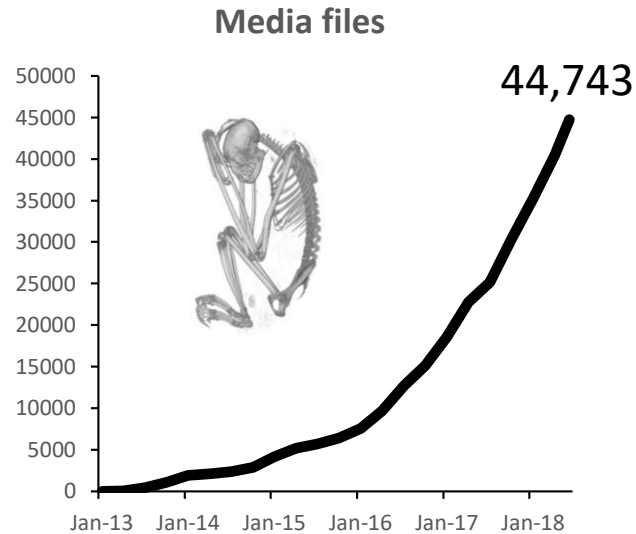*at a glance*

http://bit.ly/MSGalagoMom

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

Media files — 44,743

Users — 6,174

Downloads — 92,063

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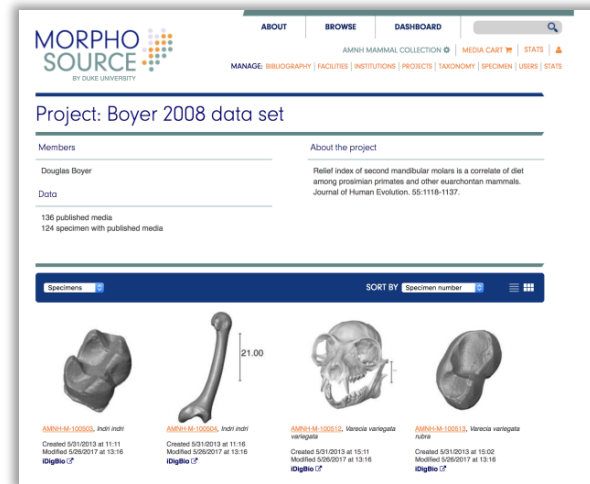
Challenges: Access and Preservation of Data

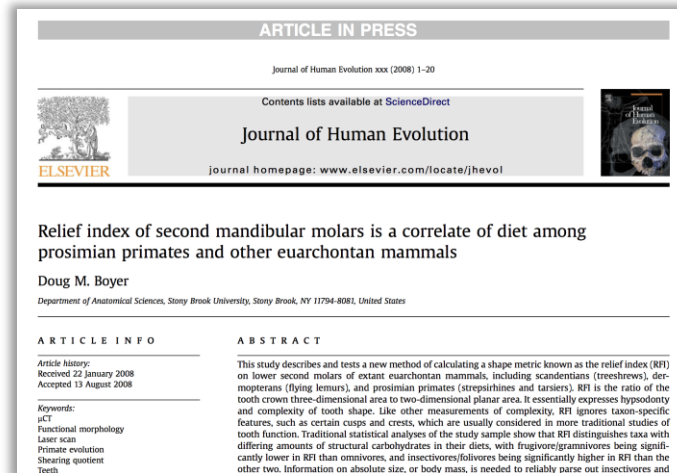Monitor data organization, appropriateness, and corruption

Keep data discoverable and accessible (no data silos!)

Challenges: Expanding User Needs
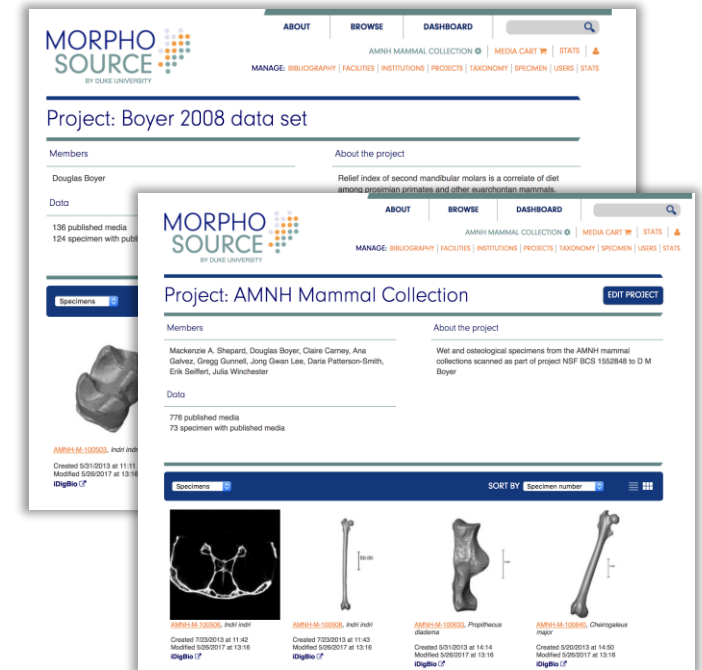
New forms of data management

Publication project…

…. Or collection project?

Media belong to one project

"Publication packet" works well for researchers

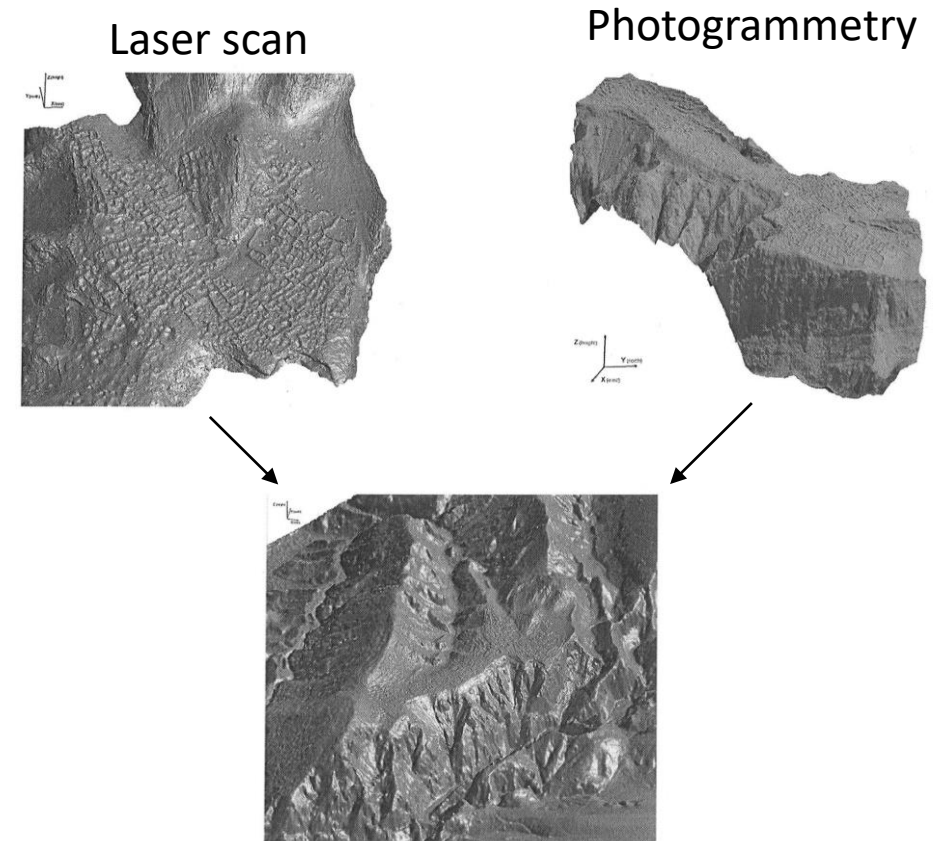**Works less well for museum curation purposes**

Challenges: Expanding User Needs

New forms of data management

New data types/use cases

3D media of objects that are
not biological specimens

Laser scan

Photogrammetry

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

Lambers et al., 2007

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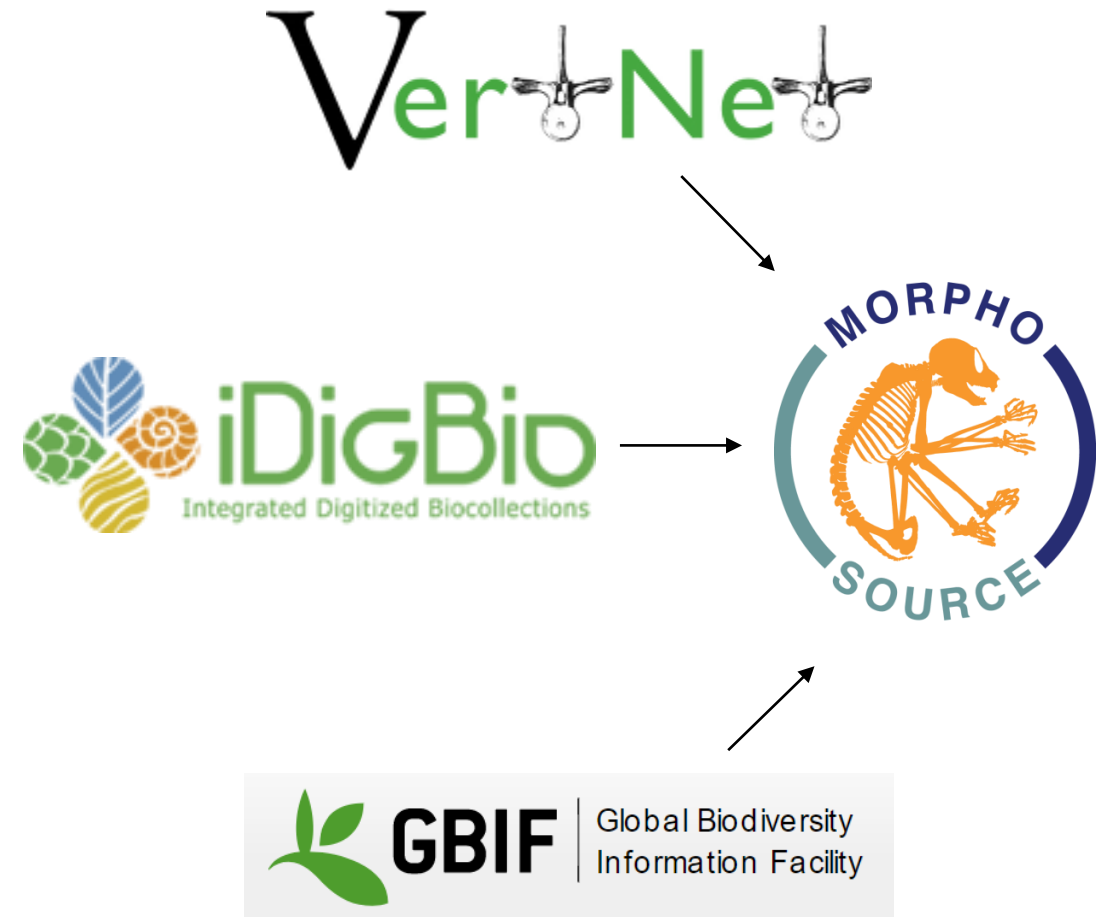
- Samvera/Fedora stack with library integration
- Renovating the core site data model

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)
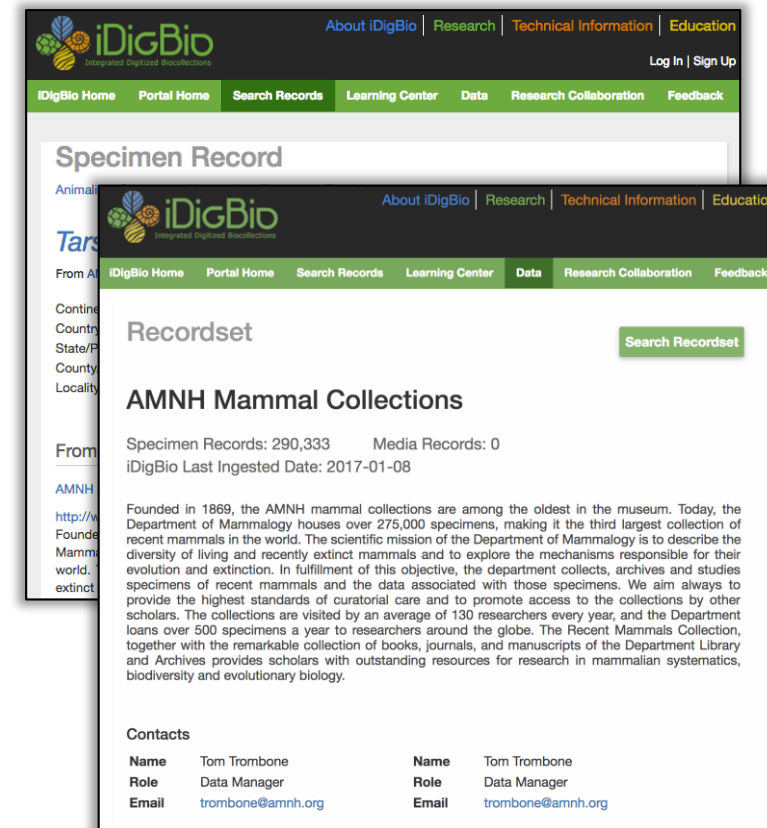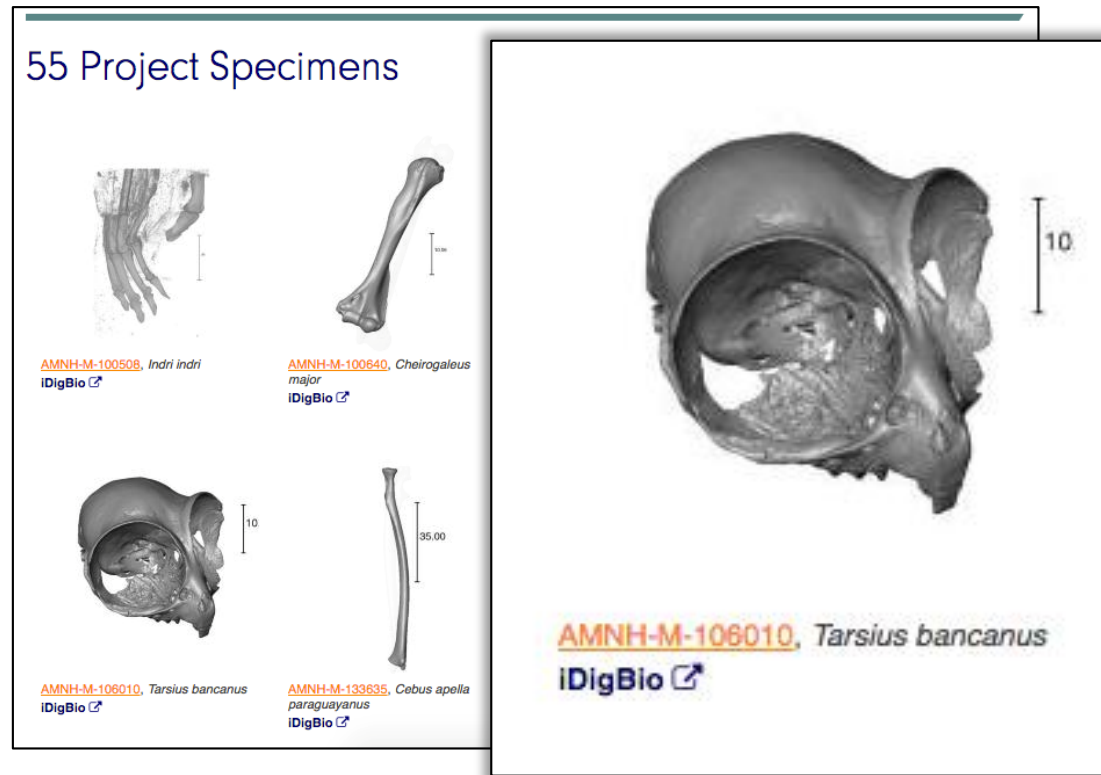Amphibia (768)
Aves (and fossil outgroups) (470)
Chondrichthyes (65)
Mammalia (and fossil outgroups) (17858)
Reptilia (681)
Sarcopterygii (1)

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
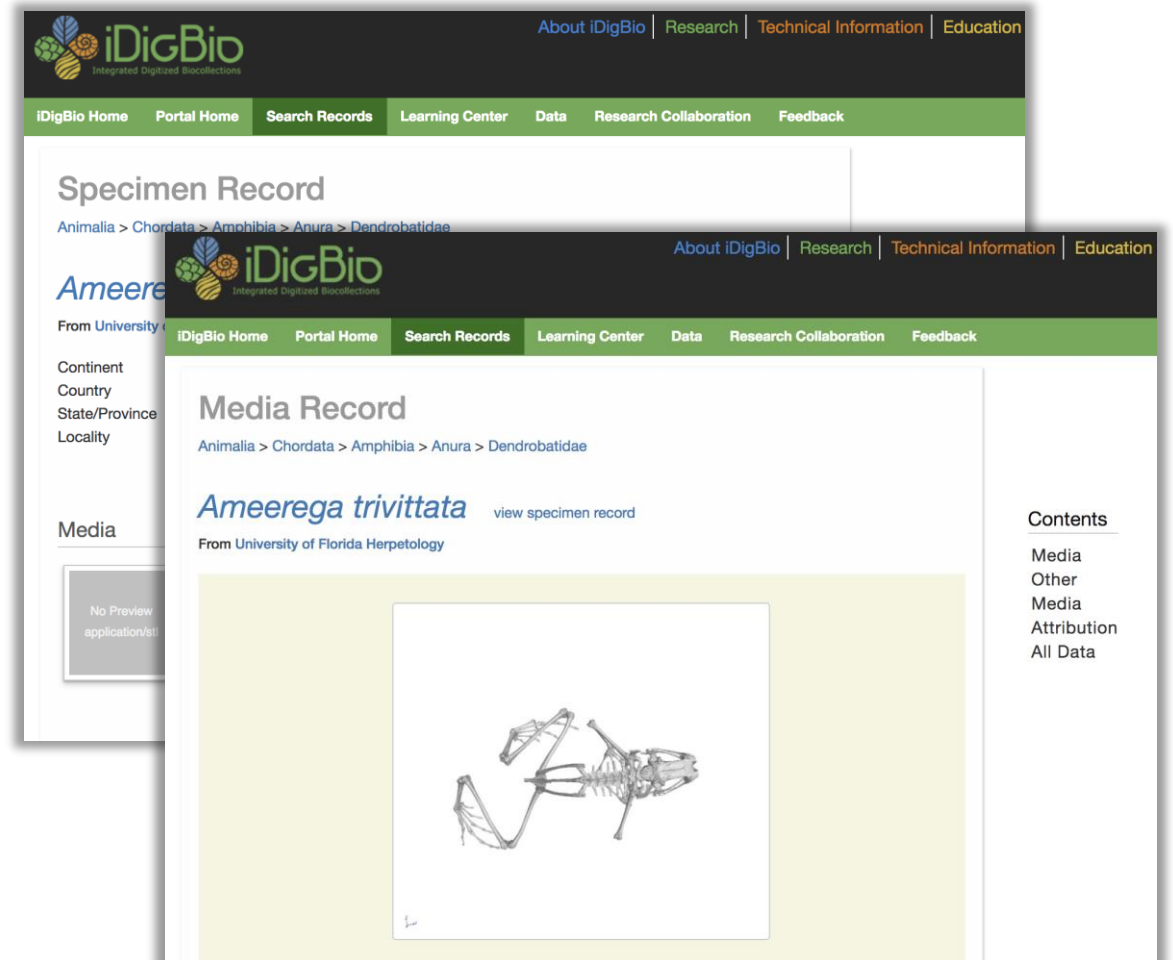- Download request reports with request free text

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
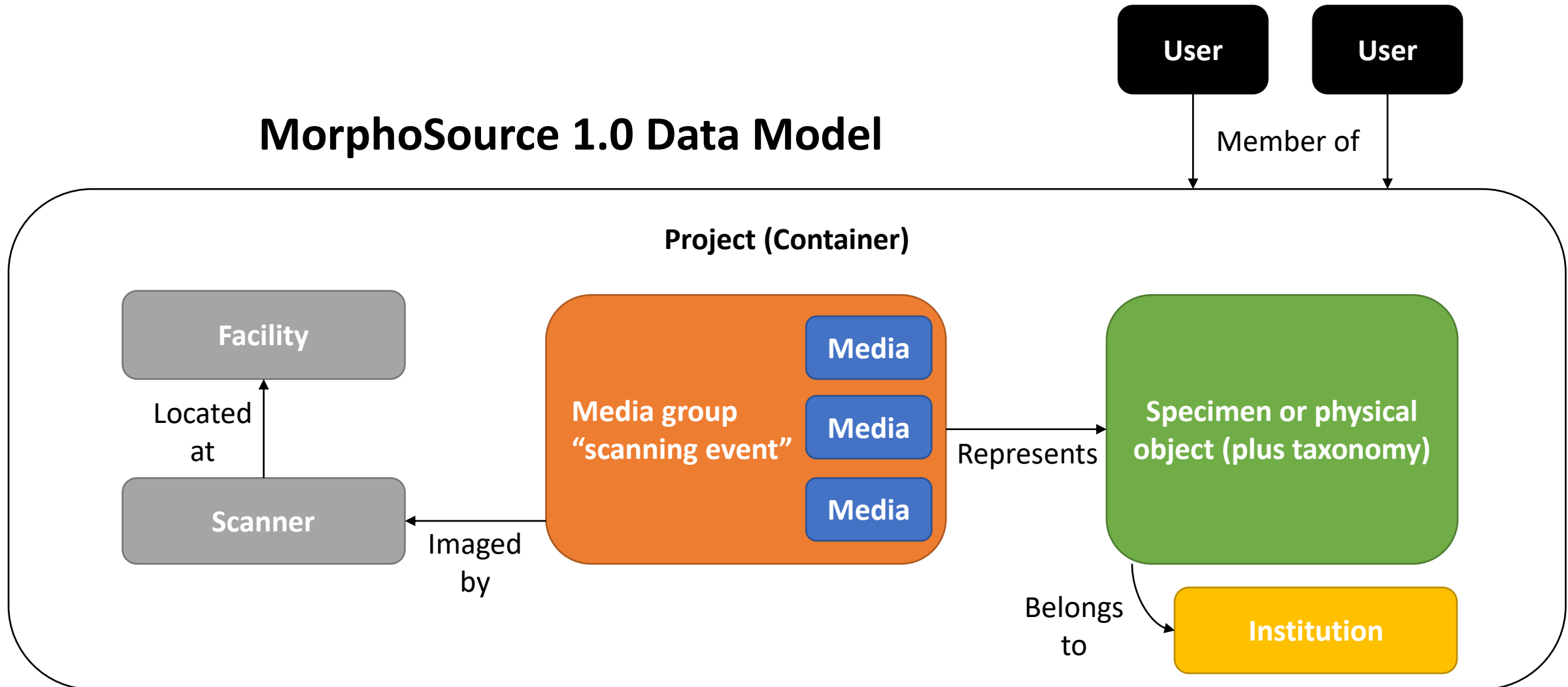- Fixity checking

Duke Library Governance

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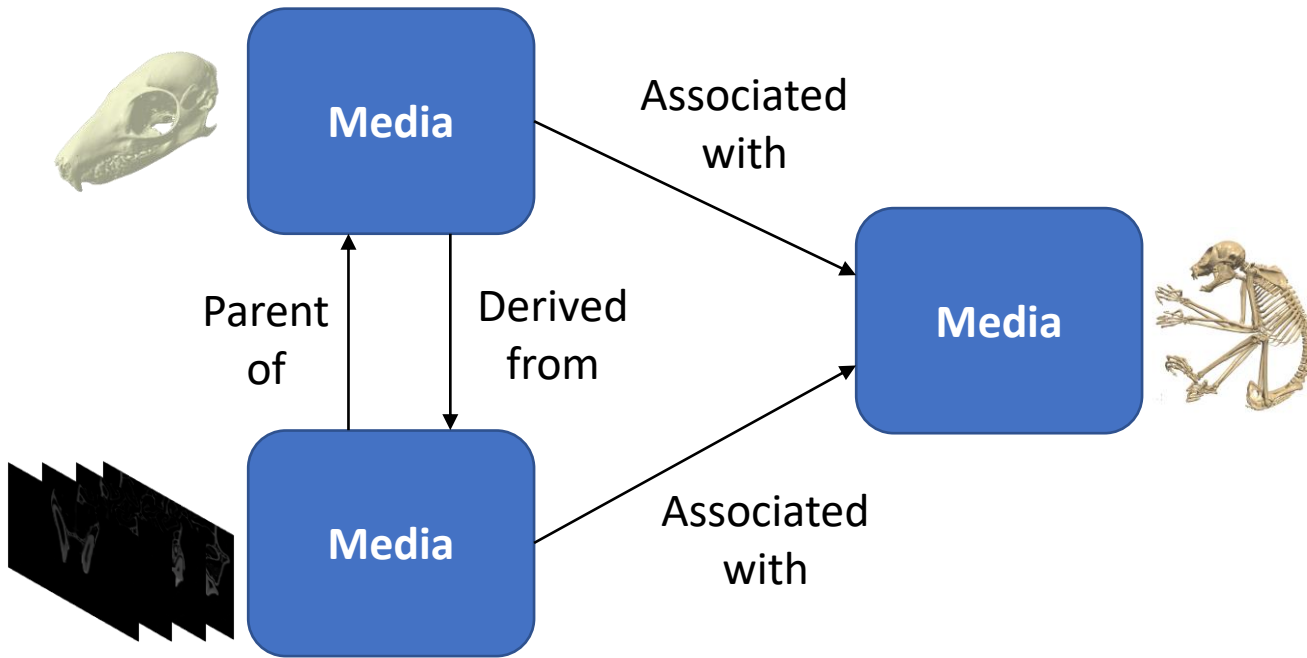- Support in resources and personnel

MorphoSource 2.0: Updating data model

Record with metadata

User

User

MorphoSource 1.0 Data Model

Member of

Project (Container)

Facility

Located at

Scanner

Imaged by

Media group "scanning event"

Media

Media

Media

Represents

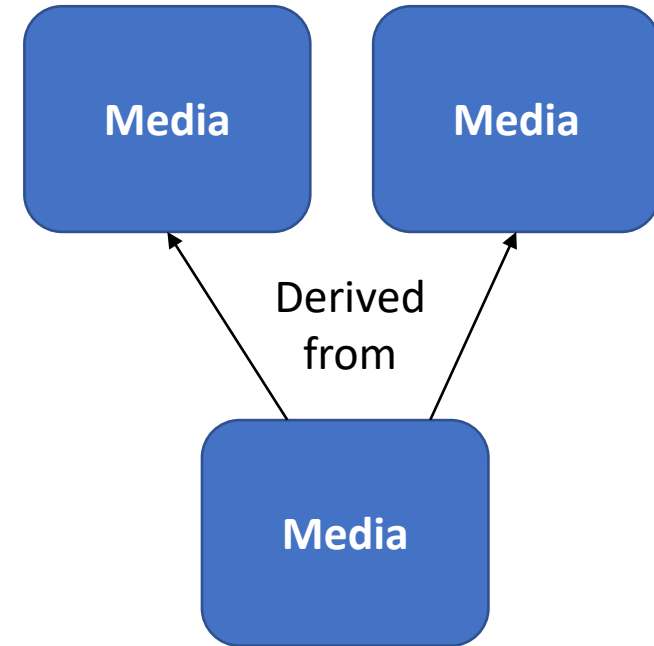**Specimen or physical object (plus taxonomy)**

Belongs to

Institution

MorphoSource 2.0: Updating data model
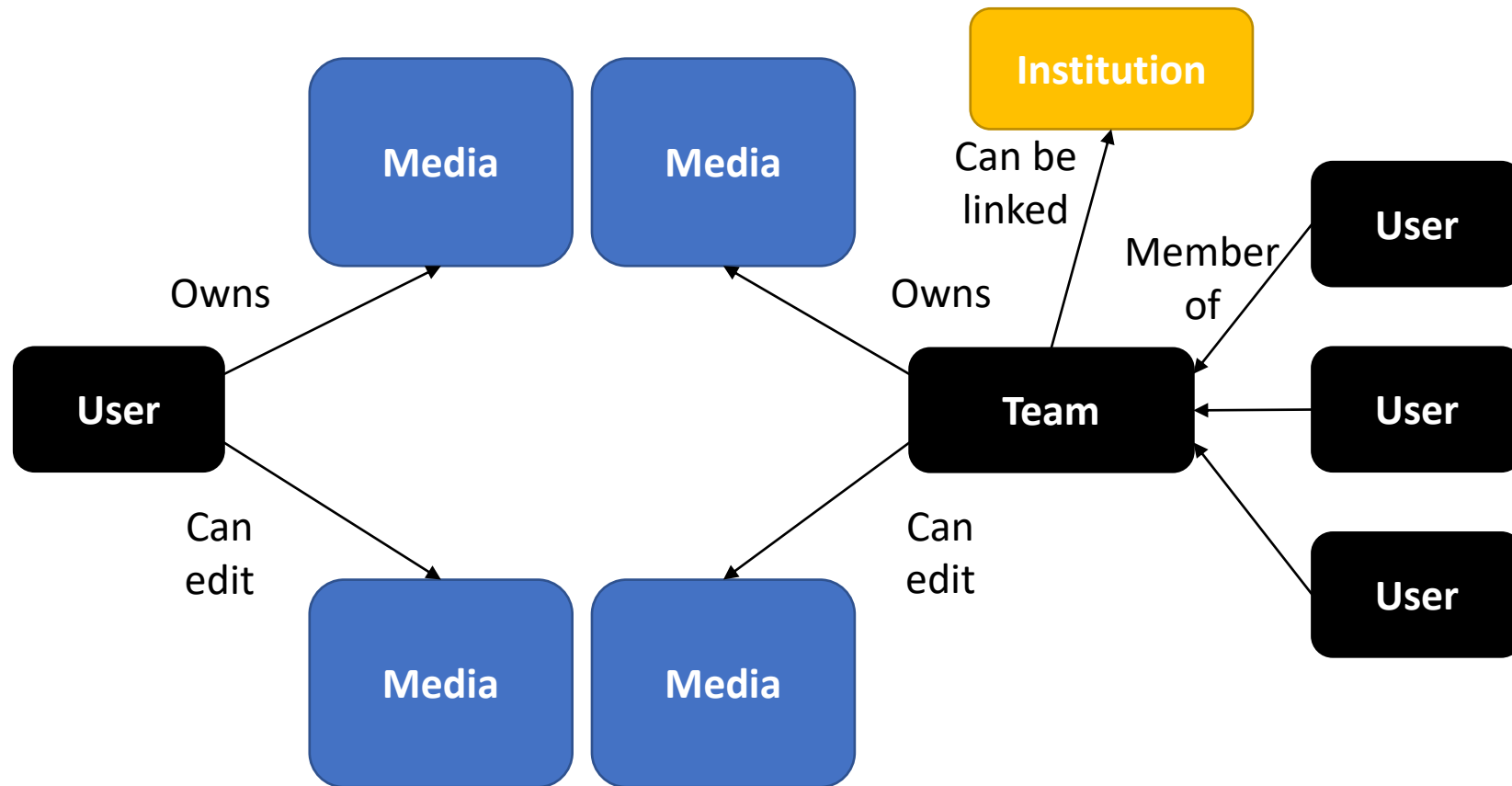

Enables multi-modal media

<u>No more media groups.</u> Instead, media have flexible relationships with other media

MorphoSource 2.0: Updating data model

Media **Media**

Institution

Can be linked

Owns

User

Owns

Member of

Team

Can edit

User

Can edit

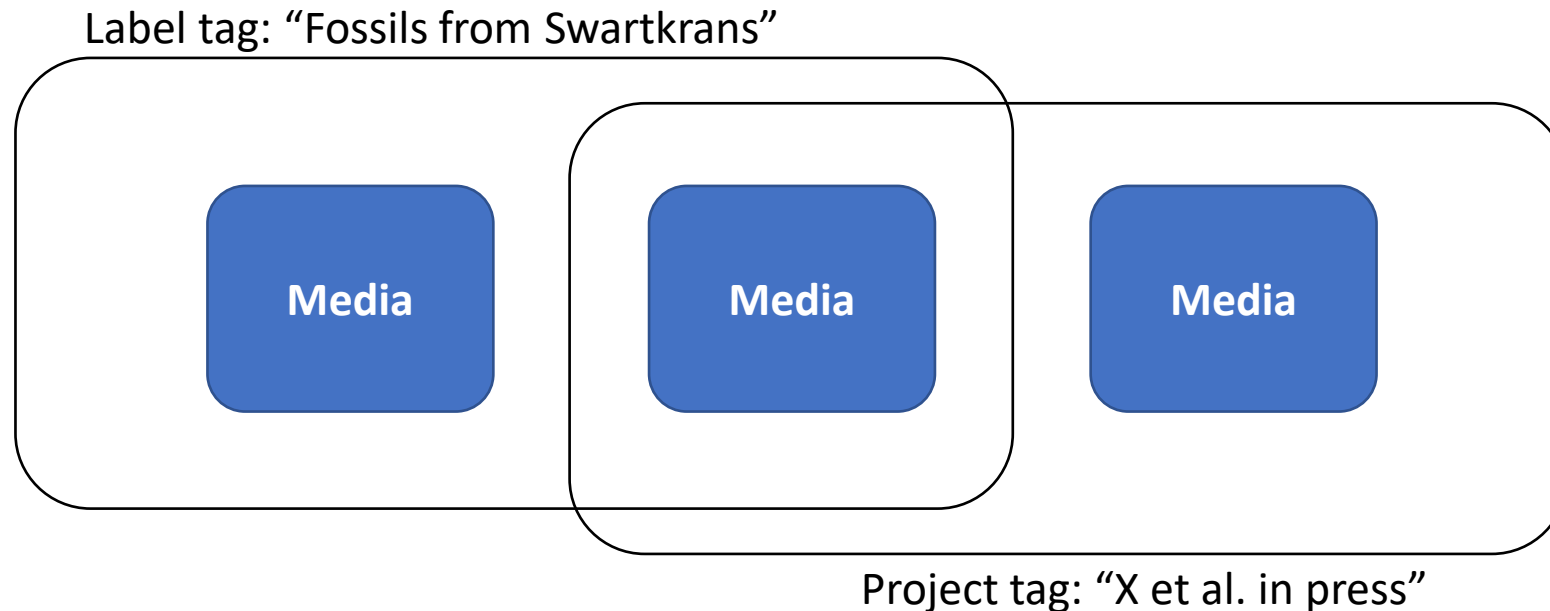**User**

Media **Media**

User

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, <u>not projects</u>. Teams represent groups of users that can be linked to institutions.

MorphoSource 2.0: Updating data model

Label tag: "Fossils from Swartkrans"

Media   **Media**   **Media**

Project tag: "X et al. in press"

Projects as tags, <u>not containers</u>. 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