



MorphoSource's approach to describing, preserving and managing complex representational data

Julie M. Winchester / Product Manager & Lead Dev. Doug M. Boyer / MorphoSource Director Tim Ryan / Professor PSU Anthropology Tim McGeary / Duke Associate Librarian of IT

Part II: Integrating MorphoSource with other resources





- 3D data can be viewed as a microcosm for representational data more broadly
 - Direct representation of object, but...
 - High metadata load
 - Lack of easy software support
 - No best practices for linkages
 - Data produced by many people in many ways
- Solutions for 3D data can possibly be instructive for representational data more broadly





- Similarly, MorphoSource reflects challenges of representational data repositories
 - Supporting challenging type of data (3D)
 - Providing platform for individuals and groups
 - Archiving and serving access to data
 - Demonstrating value of data represented in archive
- Describing MorphoSource's journey and design to both seek feedback for self-improvement and hopefully say something about broader issues at play







- MorphoSource 1 (MS1) Development Timeline
 - How was the site built?
 - What stressors were experienced?
- MorphoSource 2 (MS2) State of the Refactor
 - Development
 - Tour of features
- Integrating MorphoSource with Other Resources



MS1 Development





Recently Published

A new genus and species of clingfish from the Rangitahua Kermadec Islands of New Zealand

See all project specimens Read the published article



Welcome

MorphoSource is a project-based data archive that allows researchers to store and organize, share, and distribute their own 3d data. Furthermore any registered user can immediately search for and download 3d morphological data sets that have been made accessible through the consent of data authors.

The goal of **MorphoSource** is to provide rapid access to as many researchers as possible, large numbers of raw microCt data and surface meshes representing vouchered specimens.

File formats include tiff, dicom, stanford ply, and stl. The website is designed to be self explanatory and to assist you through the process of uploading media and associating it with meta data. If you are interested in using the site for your own data but have questions about security or anything else contact the site administrator. Otherwise please download whatever data you need and check back frequently to see what's new.







MorphoSource 1 Development Timeline



Time



2013-2014: Initial Development



- February 13, 2013: First Code Contribution
- MS1 built on Collective Access
 - Open-source software for managing and publishing museum and archival collections
 - jQuery/PHP
 - <u>Significantly</u> customized to support 3D and media data model
 - Proof of concept





2015-2017: Maintenance and Scaling

- New Features
 - Media Cart
 - DOIs
- Scaling and Stressors
 - "Publication Packet" model
 - Media ownership
 - More institutional users and use cases
 - Aging UI



ZP







Aging technical stack (and UI) makes improvements more difficult

- Tech debt: Code accretion, lack of modularity, loose separation of powers
- Unable to update core framework
- jQuery/PHP: modern "web application" solutions difficult to impossible
- Server-side Meshlab a nightmare







2017-Now: Internal Dev & Triage

- Shifting to internal development (funding!)
 - More code contributions, more regular small-scale feature updates
 - New features: project dashboard UI revamp, file upload interruption protection, automated data reporting to institutions...
- Acceleration of scaling
 - oVert
- 2018: Beginning development of MS2









- MS2 is a Hyrax application
 - Open-source Samvera-powered
 repository solution
 - Ruby on Rails
 - Large, active community ecosystem
 - Relatively fresh tech stack









- Why Hyrax?
 - Standardize with existing Library repositories
 - Background job queueing
 - Robust multiple file upload, cloud plug-ins
 - Built-in derivative generation
 - User and group support
 - Storing records (media) in multiple collections (projects)
 - Modular design with solid design principles
 - Universal Viewer (+IIIF) for record preview by default
 - Google analytics
 - More suitable for automated deployment







MorphoSource 2 Development



- Starting with...
 - MS1
 - Appreciation for future needs
 - Vanilla Hyrax (significant customization)
 - Minimal working back-end AND front-end

- Rough development procedure (cyclical!)
 - Identify user needs via user stories
 - Implement back-end features
 - Make features accessible via front-end
- Logistics
 - Agile Sprints
 - Jira Issue Tracker





MS2 Data Model





Specimen Occurrence ID: a2770...



Media MS ID: M1771





Event

















🕅 MorphoSource		Q 47 179 UN
1. Initial Information	2. Physical Object	3. Media
BACK START OVER		
We need to gather some initial information about y or MRI media, we consider reconstructions to be der upload projections.	our media. Please select all that apply conc rived from parent projection media, even th media.	erning raw or derived status. Please note for CT lough in most cases we suggest users do not
My media is derived from some parent media. Se	elect this even if you do not plan to upload the	e parent media.
There is media already on MorphoSource that is	derived from my media.	
т	IS ACCURATELY DESCRIBES MY MEDIA	



MS2 Media Submission



MorphoSource			Q 4 7	¥79 UN
1. Initial Information	2. Physical Object	3. Me	edia	START OVER
Biological/Cultural >				
We assume yo better de	ur media file(s) represents a scribed as a biological specin	physical object. Is this physical object. Is this physical object. Is this physical object of the second second	sical object bject?	
BIOLOGICAL	SPECIMEN OBJECT	CULTURAL HERITAGE	OBJECT	





iological/Cultural	Search (Create Obio	2. Physical Obje	ect	3. Media	START OVER
	Search/create Obje				
Searc	h for a biological spe	ecimen already in l	MorphoSource or iD	igBio, or select Ci	reate New
		Biologica	l Specimen.		
Occurence ID	Institution Code	Collection Code	Catalog Number	Genus	Species
Lorem ipsum	Lorem ipsum	Lorem ipsum	Lorem ipsum	Lorem ipsum	Lorem ipsum
	SEARCH		CREATE NEW	BIOLOGICAL SPEC	IMEN





Morphosodree	-			
1. Initial Informatio	on 2	. Physical Object	3. Media	START OVER
Biological/Cultural > Sea	rch/Create Object	Organization		
If this object is manage currently pre	d by an organization sent in the list. If th	n, either select it from the li e object is not managed by	st or create a new organiza an organization, select No	tion record if it is not Institution.
	Select Institu	tion	\sim	
SEL	ECT ORGANIZATION	CREATE NEW ORGANIZATIO	ON NO ORGANIZATION	





2. Physical Object e Object > Organization > ate with this biological specia axonomy keywords here SELECT TAXONOMY	3. M Taxonomy men by searching below. C	ledia Or select Create No	START OVER
e Object > Organization > ate with this biological specir axonomy keywords here SELECT TAXONOMY	Taxonomy men by searching below. C CREATE NEW TAXONOMY	Dr select Create No	ew Taxonomy.
ate with this biological specin axonomy keywords here SELECT TAXONOMY	men by searching below. C CREATE NEW TAXONOMY	Dr select Create No	ew Taxonomy.
axonomy keywords here SELECT TAXONOMY	CREATE NEW TAXONOMY		
SELECT TAXONOMY	CREATE NEW TAXONOMY		





1. Initial Information 2. Physical Object 3. Media START OV ological/Cultural > Search/Create Object > Organization > Taxonomy > Everything Else Finish creating a new biological specimen by filling in the following fields and selecting Create Biological Specimen. SPECIMEN IDENTIFIERS AND EXTERNAL LINKS COLLECTIONS PHYSICAL OBJECT BELONGS TO Vouchered? Image: Select Image: Se				
blogical/Cultural > Search/Create Object > Organization > Taxonomy > Everything Else Finish creating a new biological specimen by filling in the following fields and selecting Create Biological Specimen. SPECIMEN IDENTIFIERS AND EXTERNAL LINKS Vouchered? Vees Select Vouchered? Reg Select Collection Code Res Select Collection Code Res Select Collection Code Catalog Number Courrence ID Occurrence ID Cother identifier External Object URL Collection Code Sex @ Select Seription @ Select Seription @ Select Surre Of Record @ Digbio Digbio Dist Created Digbio BIBLIOGRAPHIC CITATIONS FOR THIS PHYSICAL OBJECT	1. Initial Information	2. Physical Obj	ect 3. Media START	
SPECIMEN IDENTIFIERS AND EXTERNAL LINKS Vouchered? ? Ris Select Vouchered? ? Ris Select Collection Code Recommended Collection Code Collections Catalog Number Recommended Courrence ID Recommended Courrence ID Other Identifier External Object URL Other Identifier External Object URL PHYSICAL OBJECT TAGS Sex ? Select Select Nunc armet dolor ac odio Sex ? Select Select Select Source Of Record ? Delgo BibliOGRAPHIC CITATIONS FOR THIS PHYSICAL OBJECT	ological/Cultural > Search	/Create Object > Organizatio	n > Taxonomy > Everything Else	
SPECIMEN IDENTIFIERS AND EXTERNAL LINKS Vouchered? Veachered? Veachered?<		Finish creating a new biolo following fields and selecti	ogical specimen by filling in the ng Create Biological Specimen.	
Vouchered? Reg Select ADD Institution Code Reg Select ADD Collection Code Recommended Collection Code ADD Catalog Number Recommended Catalog Number Organizational Collection Coccurrence ID Recommended Occurrence ID Occurrence ID Other Identifier(s) Other Identifier Image: Collection Code Image: Collection Code GENERAL DETAILS Enternal Object URL PHYSICAL OBJECT TAGS ADD Sex @ Select Select Select Source Of Record @ DigBio DigBio BIBLIOGRAPHIC CITATIONS FOR THIS PHYSICAL OBJECT	SPECIMEN IDENTIFIERS AND EXTE			
Institution Code @ Reg Select Collection Code @ Recommended Collection Code Catalog Number @ Recommended Collection Code Catalog Number @ Recommended Catalog Number Occurrence ID @ Recommended Occurrence ID Other Identifier Image: Select External Object URL External Object URL Description @ Select Is Type Specimen @ Select Source Of Record @ DigBlo Data Created @ Data Created	Vouchered? 2 Reg	Select		
Collection Code Recommended Calalog Number Recommended Catalog Number Catalog Number Occurrence ID Recommended Occurrence ID Recommended Occurrence ID Recommended Occurrence ID Cercopithecoids of West Africa Other Identifier Image: Cercopithecoids External Object URL External Object URL PHYSICAL OBJECT TAGS Enter Tag ADD Lorem ipsum Sit Amet Nunc amet dolor ac odio Source Of Record I DigBio Data Created I Data Created I	Institution Code 🖉 Reg	Select 🗸	Search Collections A	DD
Catalog Number Catalog Number Catalog Number Catalog Number Occurrence ID Recommended Occurrence ID Occurrence ID Other Identifier Image: Control of the co	Collection Code ② Recommended	Collection Code	American Museum of Natural History Organizational Collect	ion)
Occurrence ID Recommended Occurrence ID Other Identifier(s) External Object URL GENERAL DETAILS Description Description Sex Select Is Type Specimen Select Source Of Record DigBio BiBLIOGRAPHIC CITATIONS FOR THIS PHYSICAL OBJECT	Catalog Number (2) Recommended	Catalog Number	Cerropitheroids of West Africa	• x
Other Identifier(s) Other Identifier External Object URL External Object URL PHYSICAL OBJECT TAGS Description Enter Tag Sex @ Select Is Type Specimen @ Select Source Of Record @ IbigBio Data Created @ Data Created Bata Created @ Data Created @	Occurrence ID (2) Recommended	Occurrence ID	Winchester 2016: Dental tonography of extant cerconitheroids	x
External Object URL External Object URL PHYSICAL OBJECT TAGS Description Description Image: Select Is Type Specimen Select Source Of Record ? DigBio Date Created ? Physical Object TAGS Enter Tag ADD Lorem ipsum Nunc amet dolor ac odio BibliogRaphic Citations For This Physical Object	Other Identifier(s)	Other Identifier	Whenester 2010, Dental topography of extant cereophilecolas	~
GENERAL DETAILS Description ? Sex ? Select Is Type Specimen ? Select Source Of Record ? IbigBio Date Created ? PHYSICAL OBJECT TAGS Enter Tag ADD Lorem ipsum X Dolor X Sit Amet X Sapien etiam X Nunc amet dolor ac odio X BIBLIOGRAPHIC CITATIONS FOR THIS PHYSICAL OBJECT	External Object URL 👔	External Object URL		
GENERAL DETAILS Description ? Secription ? Select Is Type Specimen ? Select Source Of Record ? IbigBio Date Created ?			PHYSICAL OBJECT TAGS	
Description ? Sex ? Select Is Type Specimen ? Select Source Of Record ? IbigBio Date Created ? BIBLIOGRAPHIC CITATIONS FOR THIS PHYSICAL OBJECT	GENERAL DETAILS		Enter Tag A	DD
Sex @ Select Nunc amet dolor ac odio X Source Of Record @ IDigBio BIBLIOGRAPHIC CITATIONS FOR THIS PHYSICAL OBJECT	Description 👔		Loreminsum X Dolor X Sit Amet X Sanien etiam	
Sex @ Select Sel				
Sex ② Select Is Type Specimen ③ Select Source Of Record ③ iDigBio BIBLIOGRAPHIC CITATIONS FOR THIS PHYSICAL OBJECT				
Is Type Specimen ② Select Select Select BIBLIOGRAPHIC CITATIONS FOR THIS PHYSICAL OBJECT	Sex 🕐	Select 🗸		
Source Of Record ② iDigBio BIBLIOGRAPHIC CITATIONS FOR THIS PHYSICAL OBJECT	Is Type Specimen 🕑	Select 🗸		
Data Created	Source Of Record 😢	iDigBio	BIBLIOGRAPHIC CITATIONS FOR THIS PHYSICAL OBJECT	
Data created	Data Created 👔	Date Created		
	The operation of the second seco			~

MS2 Media Submission



					•	
1. Initial Information	2. Phys	sical Object	3.	Media		
Device > Image Capto	ure > Image Processir	ng > Upload	File > Everything Else	2		
	Finish creating new ı se	media by fillin electing Creat	ng in the following field e Media.	s and		
GENERAL DETAILS			COLLECTIONS MEDIA BELO	NGS TO		
Media Modality ? Required	Select	\sim	Search Collections			ADD
Media Type ⑦ Required	Select	\sim				
Element Or Part Of Object ③	Element Or Part Of Object		American Museum of Natural H Mammalogy Collections	istory	Institution	al Collection
Side ?	Select	\sim	Cercopithecoids of West Africa			/ ×
Orientation ③	Select	\sim	Winchester 2016: Dental topogr	raphy of extant cerco	oithecoids	×
Short Description ⑦	Short Description					
Full Description ⑦	Full Description		MEDIA TAGS			ADD
OWNERSHIP, PERMISSIONS, ANI	DACKNOWLEDGEMENTS		Lorem ipsum X Dolor	X Sit Amet	X Sapie	n etiam X
Data Managed By ?	Julie Winchester		Nunc amet dolor ac odio 🗙			
Download Permission ?	Select	\sim				
Usage License ?	Select	\sim				
Copyright Statement ?	Select	\sim				
Custom Terms Of Lise 2	ATTACH		FILE OBJECT DETAILS			







Live Demo

MS2 Teams and Organizations



Q Go All -

English 🗸 🐥 Jocelyn Triplett 🚔 🗸

Search

MorphoSource

Enter search terms

Collection Details

6

117 KB

of-fossil-primates/

Durham, NC

https://lemur.duke.edu/discover/division-

=

2020-01-21

2020-01-21

III III 🖸

Public

Public

Date Added Visibility

2020-01-21 Public

Total items

Location

Related C

URL

Size







Integrating MorphoSource with other resources

Julie M. Winchester / Product Manager & Lead Dev. Doug M. Boyer / MorphoSource Director Tim Ryan / Professor PSU Anthropology Tim McGeary / Duke Associate Librarian of IT



Linked (Open) Data & RDF



Linked Data (Berners-Lee, 2006)

- Use URIs (uniform resource identifier) as names for things
- Use HTTP URIs (AKA URLs) so people can look things up
- HTTP URIs should provide useful information when looked up
 - Machine readable
 - Standardized (RDF)
- Each URI resource should have links to other URI resources
- If resource is freely available, Linked Open Data



Tim Berners-Lee, Inventor of the World Wide Web



Linked (Open) Data & RDF

Duke



Linked Open Data Cloud (www.lod-cloud.net) Legend Cross Domai



MorphoSource Integration



- Does MorphoSource meet those goals?
 - Absolutely not!
- Part of the reason for this workshop, to move forward together
- Simple goals belie complexity and nuance of implementation







Federation Characteristics

- Restricted data and use case domain
- Centralized authority, fewer groups for buy-in
- High ability to dictate uniformity
- Deep but narrow

Confederation Characteristics

- Broad and varied domain
- Decentralized authority, buy-in from many groups
- Community-wide solutions
- Need for translation/normalization
- Wide but shallow



Integration of 3D data resources

- Connecting existing data to/from MorphoSource
- Making MorphoSource 3D archival software more available

Integration with wider data world

- Connecting MorphoSource 3D data to...
- Specimen repositories
- Taxonomy resources
- Other data archives



Integration of 3D data resources



- Metadata-only MorphoSource records
 - Data linkages
- MorphoSource-compatible 3D data archive instances
 - Automated deployment strategy
 - Turnkey solutions
 - Design complexity



Data Linkages



- Use case
 - MorphoSource user has data in web-accessible storage
 - Wants to make these data available in MorphoSource...
 - ... but for many possible reasons, does not want to re-store that data in MorphoSource

URLS

File Store

- Solution
 - Metadata-only MorphoSource records that point to remote resource



MorphoSource.org







Requirements for basic version

- HTTP URL for individual media (mesh, CT stack, etc.)
- File binary accessible at HTTP URL
 - Ideally for routing users to download
 - At minimum, for derivative generation and fixity checks
- Machine readable metadata
- All of the above persistent and reliably accessible













Ingest

 URL is specified for ingest
 MS GETs remote file and metadata
 Automated file characterization and derivative generation

• Fixity Checking

4. On a regular basis, metadata and file integrity are verified

5. Automated process for error reporting

- Access
 - 6. Users informed of remote nature
 - 7. Routed to remote file as appropriate





- Pros and cons
 - Pro: sustainability
 - Pro: community investment
 - Con: fragile
 - Con: bespoke, at least to start with



MorphoSource Instances







MorphoSource Instances



- Advantages of this work outside the MorphoSource domain
- Automated deployment and/or containerization could be used to make biodiversity repository technology more portable and easy to use
- Enables use across a diversity of terrains, from cloud to local hardware







XNHM 3D Data





Containerization Strategy



Containerization: Encapsulating a software application (or software application components) into its own environment, so it can run uniformly and consistently on any infrastructure.

Procedure

- Start with component application modules
- Wrap each module in a single container
 - Docker
- Connect containers
 - Docker compose
- Orchestrate containers in an automated way
 - Kubernetes





Design Complexity



- XNHM has own MorphoSource instance
- Indexes instance-local media in MorphoSource.org
- Should XNHM be able to edit that media from MorphoSource.org?
 - Balkanization versus increasingly complex dynamics
- How about for researcher-scanned media of XNHM?





Integration with Wider Data World

Examples of work previously completed

- Automated data reporting
- iDigBio specimen integration
- API

MorphoSource





The find command takes the following parameters:



Integration with Wider Data World

Possible Future Directions

- Finish MS2 (so we can move on to new things!)
- Not just APIs, but standardized machine-readable metadata at URLs
- Connecting media to specimen repositories
- Better importation from specimen and taxonomy resources
- Priorities largely determined by feedback from this community!



Acknowledgements



Dave Blackburn, Ed Stanley, Jon Blundell, Gary Motz, Eric Delson, Gil Nelson, Natalia Lopez Carranza, Jessica Maisano, Ruth O'Leary, Holly Little, Sharon Grant, Kate Webbink, Larry Witmer, Claudia Grant, Callum Ross, Hilmar Lapp, Jennifer Moore, Adam Rountrey, Cody Thompson, Zach Randall, Nathan Hall, Ingrid Daubechies, Joshua Wisor, Catherine Early, Matthew Gage, Greg Watkins-Colwell, Nelson Rios

Internal Test Users: Mackenzie Nieto-Aguilar, Arianna Harrington, Ethan Fulwood, Bernadette Perchalski, Paul Morse, Gabe Yapuncich, Boyer Lab Group



NSF DBI-1701714



NSF DBI-1661386 NSF DBI-1661132

MorphoSource/Duke Team

Primary Developers: Jocelyn Triplett, Simon Choy



Additional Development: Ryan Baumann, Jim Coble



DevOps: David Chandek-Stark, Jack Hill

Project Support: Will Sexton, Ed Gomes

Additional Duke Support: Charley Kneifel, Sophia Lafferty-Hess, Tracy Futhey, Mark DeLong





All MorphoSource contributors and users!