Digital Assets Management via MCZbase at the Harvard Museum of Comparative Zoology

Brendan Haley, Paul Morris, Linda S. Ford & James Hanken
Quick History/Overview of MCZbase

• Installed in 2008 as sole collections data management system for MCZ, including Special Collections; ten collections migrated by 2011.
• Third, originally identical instance of Arctos platform; subsequently managed and enhanced independently at MCZ.
• Data types include primary voucher specimen information (e.g., Darwin core), plus cryo samples, gene sequences, audio and video recordings, primary literature, fieldnotes and images (2D and 3D).
• Links to data aggregators: iDigBio, GBIF, EOL, etc.
• Dynamic linking with GenBank, BHL, special projects and other museum records.
• 2,204,308 records as of 21 January 2020.
Under the Hood

- MCZ storage and MCZbase servers (development and production) maintained by Harvard FAS Research Computing (FASRC).
- Database housed on Dell PowerEdge R440 server running Oracle 11g; web-based front-end application runs in Coldfusion11 on separate virtual machine.
- FASRC storage is “highly available networked storage application” (Lustre Network Storage).
- MCZbase software: GitHub used for collaborative development; Zenodo used for archiving software versions.
- Full DB export archived biweekly on FASRC storage.
- Snapshots of MCZbase dataset (Darwin Core Archives produced by IPT) will be deposited to Harvard Dataverse (Harvard University Library Systems).
- Specimen/occurrence data, media metadata and metadata for discovery are shared through MCZ's IPT instance.
- DB bug reports done via Bugzilla; DB project management done via Redmine.
Operational Assumptions of the MCZ

• Responsible for own material (historical and future)
  – Depository of record for specimens, metadata and augmented data
  – Responsible for tracking specimen usage

• Collaborate robustly with research and specialized projects
  – Encourage enhancement of MCZ data
  – Tether enhanced data back to original specimen data

• Integrate seamlessly among internal, external within Harvard, and external resources
  – Increases discoverability and usability of data
  – Allows dynamic linking with open software as new resources emerge
  – Ensures conservation of original material, plus data consistency and integrity
Cyrophthalmaus solentensis voucher MCZ:IZ:135079 18S ribosomal RNA gene, partial sequence

GenBank: KJ857519.1

Accession: KJ857519.1

Definition: Cyrophthalmaus solentensis voucher MCZ:IZ:135079 18S ribosomal RNA gene, partial sequence.

References:

Author: Dreszer, T.B., Radja, T. and Giribet, G.
Title: Cyrophthalmaus solentensis sp. nov. (Cyrophthalmaus, Sironidae), a new endemite benthic harpacticoid species from Croatia, with an application of confocal laser microscopy to illustrate gonditalla in Ophiiones

Journal: Breviora 543 (2014) in press

Remark: Publication Status: Available-Online prior to print

Sequencing technology: Sanger dyelex sequencing

Features:

Location/Qualifiers:

/organism="Cyrophthalmaus solentensis"

/seq_region="genomic DNA"

/specimen_voucher="MCZ:IZ:135079"

/seq_region="holotype of Cyrophthalmaus solentensis"

/locus="IZ:135079"

/locus_class="naval"

/product="18S ribosomal RNA"
DAM & DB Handling

• Internally through MCZbase
  – Leveraging the media module to track metadata, provenance, file location and integrity (md5/hash)
  – Independent full archival DAM (e.g., Fedora, DSpace) under review to enhance current system

• Fully integrated DAM and collections data management (will not change with enhancement)
Media Handling (2D, 3D, etc.)

• Criteria for handling depends on desired functionality
• May be linked internal, external (w/in Harvard) and external
• 2D
  – Internal: specimen images, x-rays, audio, video and digital microscope images (Keyence)
  – External (w/in Harvard): DRS for ledger scans and fieldnotes; SlideAtlas for microslide scans (open in Girder)
  – External: BHL for publications
• 3D
  – Internal: 3D PDFs of microCT and CT scans
  – External: MorphoSource, Aves 3D
Collection/Department Handling

- Centralized collection DB structure
- Museum-wide protocols recognize idiosyncrasies of different disciplines
  - Facilitate management and unified addressing
  - Ensure scalability
- Internal consistency allows for common-point connection
  - Increases efficiency, response time, unified data sharing (e.g., IPT) and shared goals and initiatives
Loading slide holder into 10-capacity autoloader of Huron scanner
• Chaerephon plicatus, wrinkle-lipped free-tailed bat
• Collected in Indonesia
• Slides prepared in February 1910
Welcome to SlideAtlas

Users: Sign in Register Guest

Girder: a data management platform

What is Girder?

Girder is a free and open source web-based data management platform developed by Kitware as part of the Resonant data and analytics ecosystem. What does that mean? Girder is both a standalone application and a platform for building new web services. It's meant to enable quick and easy construction of web applications that have some or all of the following requirements:
EMBRYO—Free tail bat (Choerophorus plicatus?)

LENGTH—13.4 mm

AGE—

FIXATION—Formalin

SECTIONS—Transverse 10 µ

STAIN—Alun cockineal and orange G.

HISTORY—

REMARKS—Traniers 1608 A. B.
Coll. Owen Bryant £500, Java.

PUBLICATIONS—