



# Lessons from Engaging with Repositories and Cyberinfrastructure Initiatives

**Gary Motz, Jon Dunn**

INDIANA UNIVERSITY BLOOMINGTON

# Natural History Collections and Digital Repositories

# Imago

- Repository for data from two IU natural history collections
- Generalized DarwinCore schema
- Prototype awaiting replacement repository

The screenshot shows the Imago website header with the Indiana University logo and name. Below the header, there are navigation links for Home, About, and Help, and a search bar with a 'Go' button and an 'All' dropdown. A 'Back to search results' link is visible. The main content area displays the record ID 'IUPC-2018-00000066-00110' with an 'Open Access' badge. Under the 'Descriptions' section, there is a table with the following data:

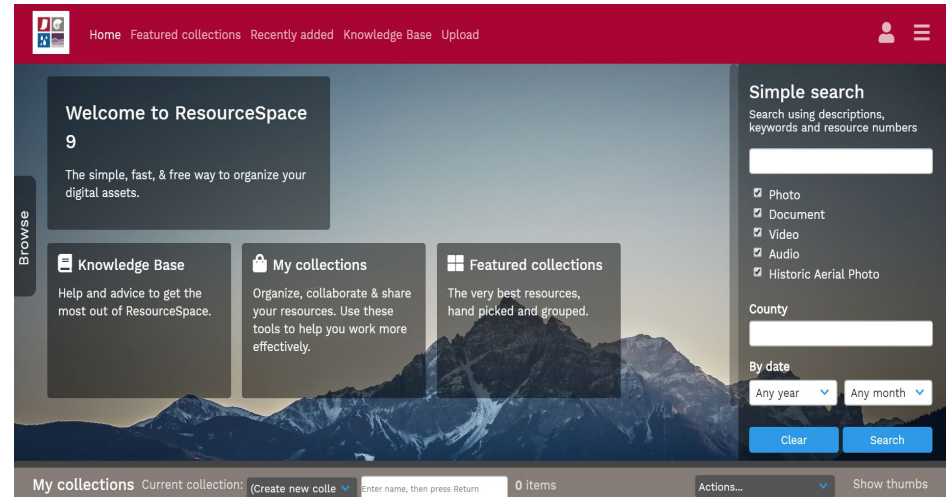
Attribute Name	Values
Identifier	<a href="http://purl.dlib.indiana.edu/iudl/paleontology/VAD8336-2018-00000066-00110">http://purl.dlib.indiana.edu/iudl/paleontology/VAD8336-2018-00000066-00110</a>
Rights	Attribution-NonCommercial 3.0 United States
Collection code	paleontology
Catalog number	IUPC-2018-00000066-00110
Other catalog numbers	10837
Basis of record	FossilSpecimen

To the right of the table, there is a 'Last modified:' label above a thumbnail image of a fossil specimen. Below the image is a 'Download the full-sized image' link and social media sharing icons for Facebook, Twitter, Email, and Print. At the bottom right, there is a 'Citations:' section with links for EndNote, Zotero, and Mendeley.



# CollectiveAccess & Resource Space Tandem (CARST)

- Repository for geological collections
- Over 2 million object records from several dozen collections
- Emphasis on data preservation and accessibility to public
- ABCD-EFG, ISO 19115, DarwinCore, and EML schema



# University Collections

- Novel upper-level administrative position to coordinate resources for all collections
- Goal to provide CMS, DAMS, and public-facing web accessibility for all

The screenshot shows the top navigation bar of the Indiana University Collections website. It features the IU Psi logo and the text "INDIANA UNIVERSITY" on the left, and a search icon on the right. Below the navigation bar, the main heading "Collections @ IU" is displayed. To the right of the heading are links for "CONTACT / NEWS / ABOUT / STRATEGIC PLAN". Below the heading is a horizontal menu with links for "FEATURED SPOTLIGHTS", "MUSEUMS AND GALLERIES", "LIBRARIES", "OTHER COLLECTIONS", and "ALL COLLECTIONS" (which is underlined). The main content area shows a breadcrumb trail "Home" followed by "ALL COLLECTIONS". Below this are six buttons arranged in a 3x2 grid: "University-wide", "Bloomington", "IUPUI", "East", "Kokomo", and "IUPUC".

INDIANA UNIVERSITY

SEARCH

**Collections @ IU** CONTACT / NEWS / ABOUT / STRATEGIC PLAN

FEATURED SPOTLIGHTS | MUSEUMS AND GALLERIES | LIBRARIES | OTHER COLLECTIONS | ALL COLLECTIONS

Home  
ALL COLLECTIONS

University-wide | Bloomington

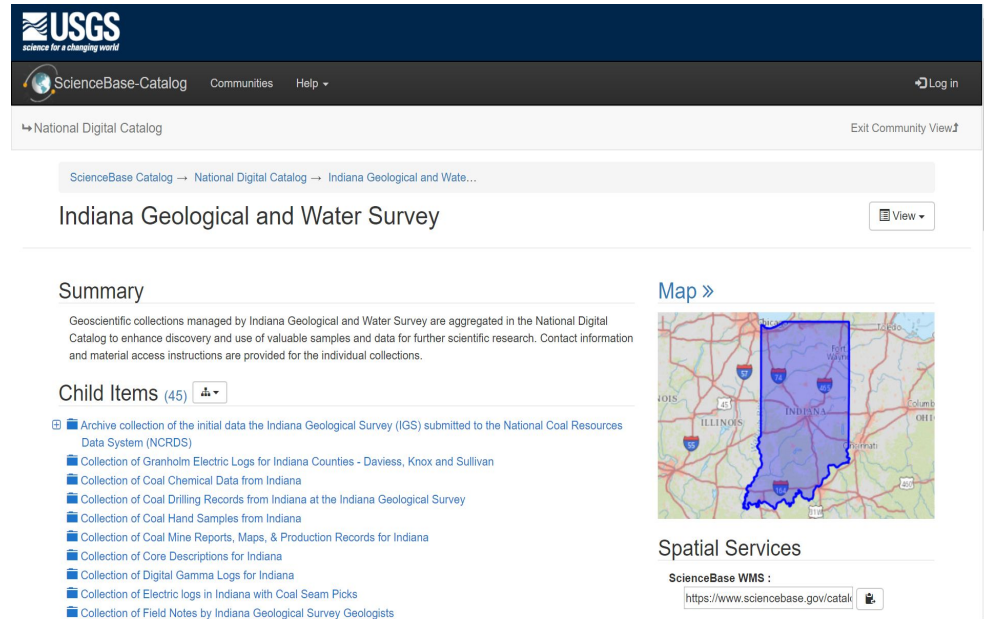
IUPUI | East

Kokomo | IUPUC



# USGS National Digital Catalog

- Repository for US geological collections data
- Federated collections-level discoverability for state surveys and USGS science centers
- Currently manual entry of metadata; duplicative efforts



The screenshot displays the USGS National Digital Catalog interface. At the top, the USGS logo and "ScienceBase-Catalog" are visible. The main content area shows the "Indiana Geological and Water Survey" page. A breadcrumb trail indicates the path: ScienceBase Catalog → National Digital Catalog → Indiana Geological and Water... A "View" button is present. The "Summary" section states: "Geoscientific collections managed by Indiana Geological and Water Survey are aggregated in the National Digital Catalog to enhance discovery and use of valuable samples and data for further scientific research. Contact information and material access instructions are provided for the individual collections." Below this, the "Child Items (45)" section lists various data collections, including "Archive collection of the initial data the Indiana Geological Survey (IGS) submitted to the National Coal Resources Data System (NCRDS)", "Collection of Granholm Electric Logs for Indiana Counties - Daviess, Knox and Sullivan", "Collection of Coal Chemical Data from Indiana", "Collection of Coal Drilling Records from Indiana at the Indiana Geological Survey", "Collection of Coal Hand Samples from Indiana", "Collection of Coal Mine Reports, Maps, & Production Records for Indiana", "Collection of Core Descriptions for Indiana", "Collection of Digital Gamma Logs for Indiana", "Collection of Electric logs in Indiana with Coal Seam Picks", and "Collection of Field Notes by Indiana Geological Survey Geologists". To the right, a "Map »" link is followed by a map of Indiana with a blue outline. Below the map, the "Spatial Services" section includes a "ScienceBase WMS" link: "https://www.sciencebase.gov/catalog".



# Cyberinfrastructure Needs for Digitized Collections

- Imago / CARST / IUCollections / USGS NDC
  - 2 servers, <0.5 FTE dev, <100GB storage for 100,000+ media
    - Automated, distributed backup of archived records and media
  - 2 servers, 3.5 FTE dev, >700TB storage for 2,000,000+ media
    - Manual backup of records and media to on-site tape
  - 2 servers, 5 FTE, <unknown> storage/media requirements
  - 1 server, 0.4 FTE dev, ....
- Modest requirements, substantial hurdles

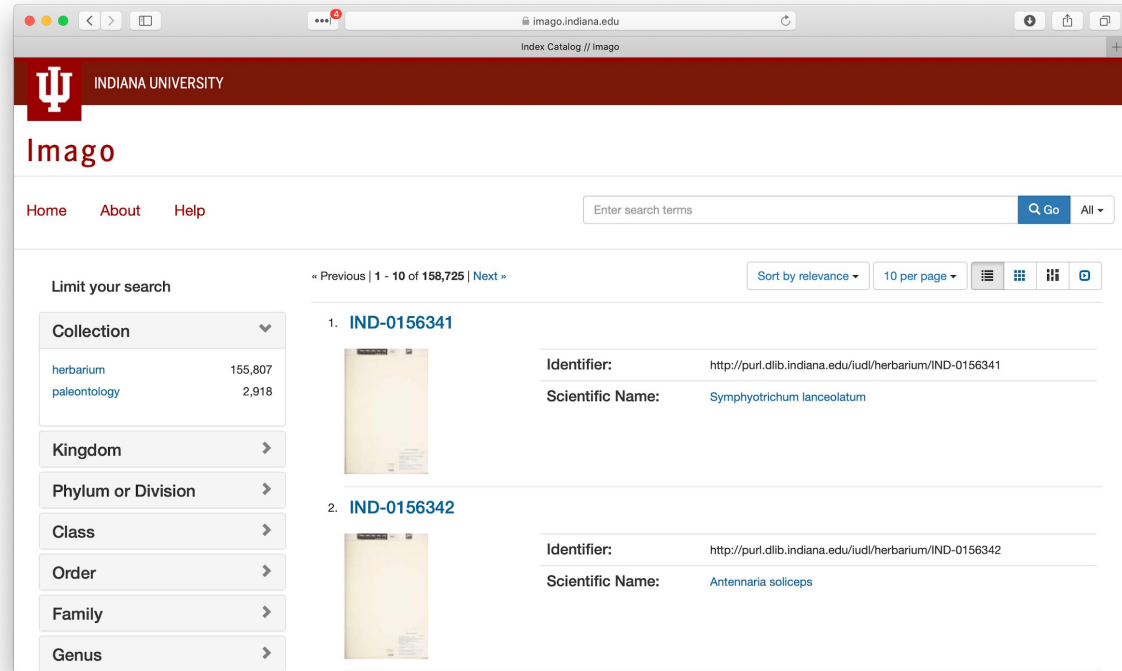


# **IU Libraries Repository Services**



# Imago

- Repository for natural history collections data
- Based on Sufia (Hyrax predecessor), Samvera, Fedora
- Goal to inform future work on digital collections and research repositories



The screenshot shows the Imago website interface. At the top, there is a red header with the Indiana University logo (Ψ) and the text "INDIANA UNIVERSITY". Below the header, the word "Imago" is displayed in a large, bold, red font. A navigation menu includes "Home", "About", and "Help". A search bar is located on the right side of the page, with the text "Enter search terms" and a "Go" button. Below the search bar, there are options for "Sort by relevance" and "10 per page". The main content area displays a list of search results. The first result is "1. IND-0156341" and the second is "2. IND-0156342". Each result includes a thumbnail image of a specimen, an "Identifier" field with a URL, and a "Scientific Name" field. The first result's scientific name is *Symphyotrichum lanceolatum* and the second is *Antennaria soliceps*. On the left side of the search results, there is a "Limit your search" section with a dropdown menu for "Collection" (showing "herbarium" with 155,807 items and "paleontology" with 2,918 items) and several other filters: "Kingdom", "Phylum or Division", "Class", "Order", "Family", and "Genus", each with a right-pointing arrow.



# Current IU Libraries Repository Environments

- Digital collections
  - Format-specific repository services
    - Image Collections Online (Fedora/local application)
    - Media Collections Online (Fedora/Samvera/Avalon)
    - Archives Online (Fedora/XTF/local application)
    - Etext platforms (XTF)
  - Online exhibit platforms (Omeka)
- University research
  - IUScholarWorks institutional repository (DSpace)
    - Research publications, presentations, and data



# Digital Collections Repository Environment

- Focus on library special collections and archives
- Service and technical ownership in Library Technologies
- Collaboration with IUPUI, University Information Technology Services, IU Office of the Bicentennial (President's Office) to expand to libraries/archives university-wide
- Reliant on standards to enable scalability of systems and support models: EAD, MODS, IIIF
- Heavy focus on audio/video driven by IU Media Digitization and Preservation Initiative (MDPI)
- Consolidation on Samvera/Hyrax



# Research Repository Environment

- IUScholarWorks
  - Current DSpace-based repository for research publications, papers, presentations, data
  - Format-agnostic but standardized metadata (Qualified Dublin Core)
- DataCORE
  - Hyrax-based research data repository currently under development
  - Based on UM's Deep Blue Data work
  - Goal: more flexible metadata structures, greater interoperability with other systems via APIs
- Service ownership in Scholarly Communication (Public Services); technical ownership in Library Technologies



# Repository Storage Infrastructure

- Mandate to use central IT-provided storage resources
  - SAN or CAS for small files / immediate access (Enterprise Systems)
    - Hitachi SAN, Hitachi Content Platform
  - HSM for large files / archival (Research Technologies)
    - IBM HPSS, IBM enterprise tape
- Fedora managed by central IT; soon application hosting as well



# Challenges for Libraries

- Staffing/prioritization
- Cross-institutional collaboration
- Scaling of service models
  - Beyond libraries/archives
  - Across disciplines
- Growing fuzziness around collections vs. research data
  - Multi-modal imaging
  - “Collections as data”
  - Researcher output, annotations
- Finding common goals, interests, and incentives across collections, researchers, library, cyberinfrastructure providers



# Challenges for Natural History Collections

- Staffing/prioritization
- Cost of supporting cyberinfrastructure/storage
  - Centralized campus CI / local department resources
- Software development
  - Cost, turnover, specialization, etc.
- Metadata and data management training!
- Recognition of a need to collaborate and minimize reinvention of the wheel



# Solutions for Natural History Collections, Libraries, and Institutional Resources

- Commitments to long-term synergy and collaboration
- Communication and recognition of shared priorities and needs
- Strategies for sustainable development
- Community-led movement for metadata schema alignment and adoption
- Reliance on national resources
- Well-documented practices for integration with communities of practice
- Workshops, like this, that engage diverse stakeholders from both large and small institutions





# Thank you!



Gary Motz  
garymotz@indiana.edu



Jon Dunn  
jwd@indiana.edu



**INDIANA UNIVERSITY BLOOMINGTON**