

Increasing Discoverability: A Challenge in Managing Microfossil Bulk Collections

Nicholas Drew

Department of Paleobiology, Smithsonian National Museum of Natural History

Drewn@si.edu

Microfossil collections include the remains and associated material of single-celled or otherwise extremely small organisms and are valuable in their use for biostratigraphy and ancient climate analysis. These collections pose a somewhat unique situation whereby a single "specimen" can be divided across multiple collections. The taxonomically organized entities are handled similarly to other taxonomy based collections. However, the bulk material from which those taxonomic specimens were removed is retained for future study and in some instances may have data that provides further context to multiple specimens in other collections. This leads to massive amounts of data and years of research work that are locked up in endless drawers of vials, jars, and bags.

How do you begin to unlock the data held within these collections so it can be discovered and utilized by the larger research community?

We start with a principle that any little step is always a step closer.

The Smithsonian Micropaleontology Collections

The Smithsonian Micropaleontology Collections represent one of the largest collections of microfossils in the world, comprised of bulk material and slides of foraminifera, ostracodes, and radiolarians collected and worked on by US Geological Survey and Smithsonian scientists.

21,000 Primary
Type Slides

200,000 Faunal, Taxonomic
and Secondary Type Slides

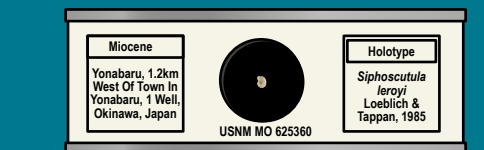
283 m³ of Bulk
Sample Material

The Extended Specimens of Micropaleontology Collections



Bulk samples including drill cores, rock samples, or jars of sediment.

Processed samples utilizes mesh sieves of various porosity to further break down bulk material.



Single species slides are identified specimens and make up our taxonomic and type collections.



Faunal slides showcase the diversity and abundance of microfossils from a locality.

Establishing Baselines

Collating Existing Data

Numerous inventories, transcription efforts, and an access database served as the initial data sets that were collated to create a single file.

Defining Whats Data is Most Valuable

By reducing the types of data that were incorporated to only that information that would have the most value in discovering the material for any future potential research or cataloging efforts.

- Location and Locality Numbers
- Geologic Age and Formation
- Preparations (i.e. rock sample, core, vial with sieved matrix, picked slide)
- Other Numbers from previous inventories or filing systems

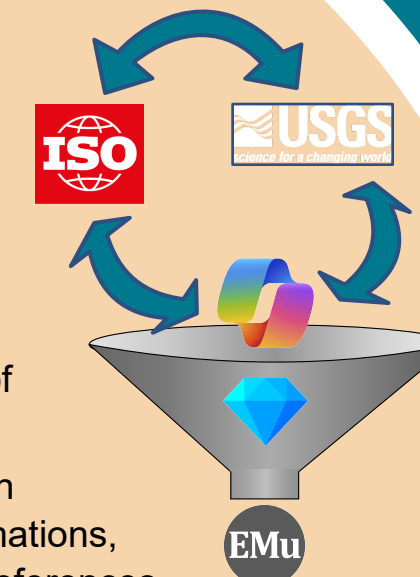
Preparing to Import

Cleaning and Validating Data

Collated data is cleaned, standardized, and discrepancies are resolved through OpenRefine.

Existing data is enhanced using online databases

- National Geologic Map Database - flesh out geologic age and lithostratigraphy
 - ISO 3166 country lists - standardization of recognized countries and divisions
 - AI - translation and expansion of unknown locations, formations, and references.



New catalog numbers are assigned and data is imported into Emu in batches. Further data refinement will be part of future projects.

Collection Connections

The extended specimen record is assembled through linking records in the Relationship tab.

This field can also be utilized to identify entities of the specimen held at other institutions as they are discovered.

If it's Digital, it's Accessible

Data needs to be correct not complete to be discoverable.

Digital ↔ Physical

Barcodes and Updates

Barcodes allow quick assignment of catalog numbers to a specimen or a lot of specimens and track specimens while providing a visual representation of the holes in our digitization efforts.



During this pass, the type of preparation(s) is noted and the record is updated. Every instance where the collection is used presents an opportunity to slowly build up the digital specimen record further.

Lots of Projects, Lots of Problems

Next steps will be to go through and catalog the specimens that had been skipped in previous inventory efforts and further refine the record data.

As the collection is continuously curated and refined, data is used to begin associating specimens across the collections.

Collection Improvements

Associated benefits of the first pass digitization efforts of the bulk collections has been a surface level clean-up of the collections. As labels are being added, exceptionally poor housing was replaced with new vials, bags, and boxes. Specimens are rearranged in the drawers to clearly show the barcode and specimen information.

