



Smithsonian Institution

NATIONAL MUSEUM *of* NATURAL HISTORY

Digitization: Various Workflows and Challenges

William Moser, Katie Ahlfeld, Karen Reed, and Geoff Keel

NMNH, Department of Invertebrate Zoology

Marine Invertebrate Collections Digitization Workshop (4-5 February, 2019)

Key Resources



Smithsonian

- Authority Files
 - Taxonomy (WORMS, ITIS, quality control in-house database)
 - Digital Station Data

Will shorten digitization time with fewer introduced errors
- Data management staff
 - Extra quality control, proofing
 - Spreadsheet editing and import into database
 - Know when to automate (scripts to proof data)
- Collection supplies (often overlooked)
 - Funding source to purchase, replenish
 - Full curation with digitization takes more time

Challenges



Smithsonian

- Appropriate authority files (e.g. taxonomy, geography)
 - Without, slows down data capture and increases the likelihood of errors in data
(taxonomic combinations to check, legacy geographical names, partial geography)
 - Collect and digitize station data first, then digitize the collection lot
- Labels
 - Difficult to read handwriting
 - Denigrating paper or ink, fading
 - Scanning or imaging labels allows long-term preservation and digitation can be done off site
- Georeferencing, errors introduced by
 - legacy geographical names and changes in boundaries
 - georeferencing the middle of the ocean (“500 miles off of” or “500 miles East of”)
- Imaging (simple image vs scientifically useful)
- Physical Curation (new, minor upgrade, already done)
 - Determining the preservation fluid
 - Etching catalog number on glass slides (diamond pencil, Dremel tool, etc.)
 - Mixed lot (Bryozoa or sponges – 5 plus species per container)
 - Bulked lots (opening each individual vial)
 - Right size container for the specimen; number of vials per container



Smithsonian

Various types of Digitization Workflows

- Skeletal records (minimal data)
 - Data capture in spreadsheet, import into cataloging system
 - Pros- rapid creation of records; management footprint of collection (inventory)
 - Cons- later on, need to fill-out records; few data safeguards; easy to introduce errors
- Cataloging directly in database
 - With digitized station data and taxonomy (fast and efficient)
 - Create new locality/stn data records
 - Pros: records complete, tend to be fewer errors
 - Cons: time consuming
- Crowdsourcing
 - Ledgers, catalog cards, or labels have to be imaged and served online
 - Pros: Someone offsite digitizes records for free
 - Cons: Difficult to find appropriate cases; staff to proof-check; no timeline
- New collecting with tissue sampling, DNAs, etc.
 - Data imported quickly with templates
 - Curation/reconciling data with spms time-consuming and often problematic
 - IDs often general at first and must be updated