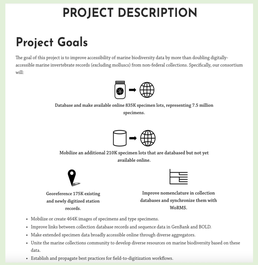
Scripps Oceanography's Pelagic Invertebrate Collection: Documenting Marine Biodiversity through *DigIn*: Digitization of Invertebrate Collections

Linsey M. Sala, Moira Décima, & Mark D. Ohman
Pelagic Invertebrate Collection, Scripps Institution of Oceanography, U.C. San Diego
https://scripps.ucsd.edu/pelagic-invertebrate-collection
http://www.digin-tcn.org/

The Scripps Pelagic Invertebrate Collection serves as a repository for both preserved specimens and associated data. Historically, the collection data associated with our unsorted parent net tow material have been digitally discoverable for the research and education community. However, we also house tens of thousands of individual sorted and identified specimens, some of which have come from our net tow collection material or from past and present researchers' specialized specimen collections (e.g. Martin Johnson's lobster phyllosoma larvae, Abraham Fleminger's copepods, Edward Brinton's euphausiids, etc.). The existence of this material has largely only been known by viewing removal tags within a net tow sample, by comparing comments in a line of data to a physical sample, communicating with our staff, or by special inquiry. In an effort to bring this reference specimen material into digitally discoverable territory, SIO-PIC joined SIO's Benthic Invertebrate Collection and a consortium of 18 other institutions across the United States to establish an NSF-funded Thematic Collection Network (TCN). This TCN, *Documenting Marine Biodiversity through Digitization of Invertebrate Collections* (*DigIn*) has set out to digitize and mobilize 835K lots, representing 7.5 million specimens. A core objective of this network is to assemble robust, vouchered marine invertebrate data and accompanying specimens to enhance research and education in marine biodiversity, ecology, systematics, oceanography and other disciplines.

DigIn Project Goals


PROJECT DESCRIPTION

Project Goals

The goal of this project is to improve accessibility of marine biodiversity data by more than doubling digitally-accessible marine invertebrate records (excluding mollusc) from non-federal collections. Specifically, our consortium will:

Database and make available online 835K specimen lots, representing 7.5 million specimens.

Mobilize an additional 210K specimen lots that are databased but not yet available online.

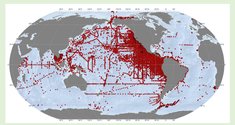
Georeference 175K existing and newly digitized station records.

Improve nomenclature in collection databases and synchronize them with WoRMS.

- Mobilize or create 464K images of specimens and type specimens.
- Improve links between collection database records and sequence data in GenBank and BOLD.
- Make extended specimen data broadly accessible online through diverse aggregators.
- Unite the marine collections community to develop diverse resources on marine biodiversity based on these data.
- Establish and propagate best practices for field-to-digitization workflows.

PIC Specific Goals

- **Digitize approximately 34,500 lots.**
- **Develop pathway to connect reference (sibling) specimens data to parent net tow collection event data.**
- Currently, ~146,000 parent net tow event based samples are digitally discoverable.



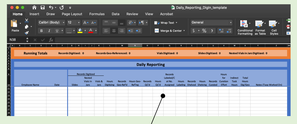
Reference Specimen Collection Areas

Lobster Early Life History (Phyllosoma and Puerulus) Collection

Copepod Fluid and Slide Collection

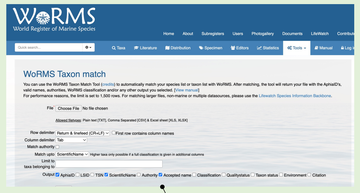
Digitization Team

Digitization Workflow



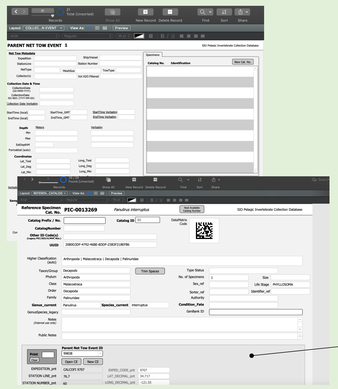
1) Digitizer 1 pulls sample and assigns physical space location (carriage, stack, bay, shelf, drawer, box).
2) Manual data matching to legacy data or direct data capture from labels.
3) Affix and scan in thermally printed, archival tag with assigned PIC Cat. No. and translated Datamatrix barcode.
4) Conduct data quality control check by 2nd Digitizer.
5) Return specimen lot(s) to assigned physical location(s).
6) Digitizers track daily progress for quarterly and annual reporting.

Taxonomy Validation



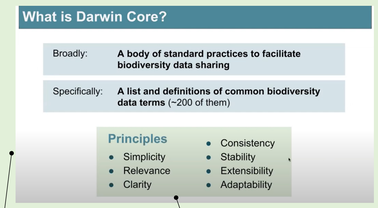
7) Batch edit produced data in GoogleRefine.
8) Validate and update taxonomy via World Registry of Marine Species portal (WoRMs).

Filemaker Pro Database Development



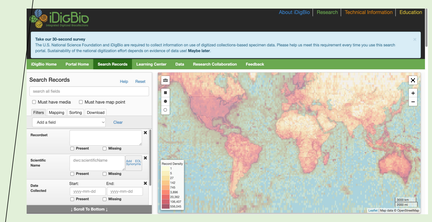
9) Import data (.csv) into Filemaker Pro.
10) Connect reference specimen data to parent net tow collection event data where possible.

Darwin Core Data Field (Term) Standardization



11) Map PIC's current specimen record fields to specified Darwin Core (DwC) terms related to a data record's:
- Geographic location
- Collection event details
- Taxonomic levels
- Preservation type, etc.

iDigBio Data Aggregator



12) Prepare (.csv) export in Filemaker Pro for data aggregators.
13) Serve data to DigIn project management for review.
14) Continue data quality control.
15) Import via IPT to data aggregators (e.g. iDigBio, GBIF, IvertEbase).
16) Facilitate specimen and data use for research and education!

Acknowledgements

We thank the past and present SIO PIC undergraduate Curatorial Assistant staff and volunteers that have been instrumental in digital data production and curation of our reference materials (Abby Lindemood, Nicolas Concha-Saiz, Jewliet Carrington, Thanh To, Hananielle LaRosa, Brenna O'Brien, Priscilla Ee, and Jim Beyster).

The numerous researchers who have deposited reference specimens and provided their valuable expertise to identify this material.

The CalCOFI sea-going and data management teams for collection and archival efforts to organize and maintain net tow sorted reference specimens (e.g., lobster phyllosoma and cephalopod paralarvae).

The Beyster Family Foundation for financial support of student apprenticeships that assist with executing our curation efforts and the National Science Foundation for support of the digitization effort.