# Migration from v0.10 to v1.0

## Summary

Datasets: 265

Observations: 661,018 (seabirds: 447,225)

Tracks: 4,468,981

Transects: 111,146

## Scope & Inclusion Criteria

Datasets are migrated to OBIS-SEAMAP based on the following criteria:

- Share level of AKN Level 5 (full data available)
- Have records in both track and transect tables

Datasets that were excluded based on share levels (240 datasets in total):

- 0 (No data, need to request or request in progress): 73
    - AMAPPS FWS Aerial 2020-2023 (19)
    - BarHarbor Whale watch (18)
    - EcoMon (10)
    - GOMCES (3)
    - ...
- 1 (Not to share due to sensitivity concerns): 2
    - Mayflower Wind Energy
    - Vineyard Wind 2018-2019
- 2 (Limited use: Only allowed in summaries and visualizations): 2
    - Cape Wind Aerial
    - Cape Wind Boat
- 6 (Unobtainable or nonexistant): 2
    - DE and Ches Bays USFWS 1990
    - EcoMon Sep 2012
- 7 (Part of another dataset): 2
    - Nantucket Aerial 2013
    - PIROP
- 9 (Acquired but not in the database): 157
    - Deepwater Wind Block Island (18)
    - Stellwagen Bank NMS (139)
- 99 (In progress): 2
    - BOEM HighDef NC2011 Camera
    - NJDEP aerial 2008

# Changes

- Each dataset has its own Dataset Page with an online map view on the OBIS-SEAMAP web site and can be downloaded.
    - The downloaded data has slightly different fields from those in the Catalog as it follows the standard OBIS-SEAMAP schema (OBIS-SEAMAP required fields added).
- For the fields that are accompanied with standardized codes, the field name ends with a suffix "_tx". The field name of the code ends with a suffix "_cd".
    - The field with "_cd" has a lookup table.
    - e.g. behavior_tx & behavior_cd => nwasc_lu_behavior_v1_0_0
    - The beaufort sea state has a lookup table but the field name is seastate_beaufort_nb. There is no free-text field.
    - The text field of the travel direction is travel_direction_tx but its code field is travel_direction (without the suffix "_cd"). This is because there is no lookup table.
- Standardized, categorical codes are uppercased.
    - e.g. CTS for Continuous time strip
    - e.g. F for female
- The Catalog is available in an ESRI file geodatabase and GeoPackage (SQLite-compatible).
- The Catalog's share levels follow the ones defined by OBIS-SEAMAP.
    - OBIS-SEAMAP uses Creative Commons rules.
    - Each dataset can be in public domain, CC-BY or CC-BY-NC.
- The data provider / contact information is provided in a separate table. The provider name is not in the observation table or the dataset table.
    - This structure change allows to list multiple contacts for the dataset.
    - Each contact is given a role, Primary contact, Secondary contact and Data entry.

# Standardization

## dataset_id

- Dataset ID in new catalog = Original ID + 200000
- Existing OBIS-SEAMAP-origin datasets keep their initial Dataset IDs (< 2000000)
- New datasets are assigned new Dataset IDs > 200000.

## obs_dt & obs_tm

- Date format yyyy-mm-dd (e.g. 2002-03-17)
- Time in 24-hour format (e.g. 05:13:50)
- Decimal places in time dropped
- Time beyond 24:00:00 => Date + 1 & time normalized (00:00:00 ~ 23:59:59)

## ln_transect

- Count in obs_count_intrans_nb => True (transect survey)
- Count in obs_count_general_nb => False (not transect survey)
- obs_count_intrans_nb & obs_count_general_nb  => obs_count

## animal_age_tx & life_stage_cd

- The field name is changed from animal_age to life_stage.
- Non-alphabet characters => space => lowercased
- a|ad|ads|adult|adults => A (Adult)
- j|juv|jv|juvs|juvenile|juveniles => J (Juvenile)
- sub|subad|subadult|subadults => S (Subadult)
- i|im|imm|immat|immature|immatures => I (Immature)
- Mix of above words plus yr|year|years|st|nd and/or calf => M (Mixed)
- Standardization rate: 17% => 44%

## sex_tx & sex_cd

- female => F (Female)
- male => M (Male)
- Standardization rate: 7% => 87%

## behavior_tx & behavior_cd

- Sdgsdfg
- Standardization rate: 13% => 77%

## travel_direction_tx & travel_direction

- Travel_direction does not have a lookup table as the codes are obvious and do not need descriptions.
- 348.75 ~11.25 => N
- 11.25 ~ 33.75 => NNE
- 33.75 ~ 56.25 => NE
- 56.25 ~ 78.75 => ENE
- 78.75 ~ 101.25 => E
- 101.25 ~ 123.75 => ESE
- 123.75 ~ 146.25 => SE
- 146.25 ~ 168.75 => SSE
- 168.75 ~ 191.25 => S
- 191.25 ~ 213.75 => SSW
- 213.75 ~ 236.25 => SW
- 236.25 ~ 258.75 => WSW
- 258.75 ~ 281.25 => W
- 281.25 ~ 303.75 => WNW
- 303.75 ~ 326.25 => NW
- 326.25 ~ 348.75 => NNW
- Standardization rate: 97%

## Flight_height_tx & flight_height_nb & flight_height_bin_min/max

- Unit (e.g. "meters", "meter", "m") dropped
- Single number => flight_height_nb
- Two numbers separated by "-" => flight_height_bin_min & flight_height_bin_max
- Number with ">" or ">=" or "+" (e.g. "10-20", ">=200", "200+") => flight_height_bin_min

- Number with "<" or "<=" (e.g.  "<200") => flight_height_bin_max
- Standardization rate: 98% (single number: 89%)

## association_tx & association_cd

- 'cetaceans' or 'dolphin' included or type_id = 2 => 11 (Associated with cetaceans)
- 'fish', 'fish school', 'fish shoal' included or type_id = 4 => 10 (Associated with fish shoal)
- 'fishing vessel', 'bofi', 'bolo', 'botd'included => 26 (Associated with fishing vessel)
- 'floating litter' included => 15 (Associated with floating litter)
- 'floating weed' included => 17 (Associated with sea weed)
- Standardization rate: 1%
- (Many records have a particular species code in association_tx)

## Distance_to_animal_tx & distance_to_animal_nb & distance_bin_min/max

- Unit (e..g. "meters", "meter", "m") and "from ship" dropped
- Single number => distance_to_animal_nb
- Two numbers separated by "-" => distance_bin_min & distance_bin_max
- Number with ">" or ">=" or "+" (e.g. "10-20", ">=200", "200+") => distance_bin_min
- Number with "<" or "<=" (e.g.  "<200") => distance_bin_max
- Standardization rate: 93% (single number: 76%)

## angle_from_observer_tx & angle_from_observer_nb

- Original angle_from_observer_nb => angle_from_observer_tx
- Number in angle_from_observer_tx between 0 and 360 => angle_from_observer_nb
- Number not between 0 and 360 (e.g. 999) => null
- Standardization rate: 99.5%

## perpendicular_distance_nb

- Numbers in distance_to_animal_nb and no numbers in angle_from_observer_nb => perpendicular_distance_nb
- Standardization rate: 33%

## visibility_tx & visibility_cd

- Assume small numbers like 0.2 are in nautical miles and large numbers like 250 are in meters.
- '0-300m' or less than 0.162 or 100 ~ 300 => 0 (0-300m)
- '300-500m' or 0.162 ~ 0.27 or 300 ~ 500 => 1 (300-500m)
- '500m-1km' or 0.27 ~ 0.54 or 500 ~ 1000 => 2 (500m-1km)
- '1-3km' or 0.54 ~ 1.62 or 1000 ~ 3000 => 3 (1-3km)
- '2 nm', '3-5km', '3-5' or 1.62 ~ 2.7 or 3000 ~ 5000 => 4 (3-5km)
- '3 nm', '4 nm', '5 nm or above', 'unlimited' or 2.7 ~ 10 => 5 (unlimited)
- Standardization rate: 23%

## glare_tx & glare_cd

- '0' or 'none' => 0 (None)
- 'low', 'minimal', 'slight' included but 'moderate' not included => 1 (Slight/Low)

- 'Bright', 'heavy', 'high', 'significant' included => 2 (Bright)
- Standardization rate: 32%

### weather_tx & weather_cd

- '1/8 CLOUD COVER; No rain', '2/8 CLOUD COVER; No rain', '3/8 CLOUD COVER; No rain', 'Clear to partly cloudy (0-50% cloud cover)' => 0 (< 50% cloud cover)
- 'clear', 'sunny', 'excellent' but not include 'partly' => 0 (< 50% cloud cover)
- '4/8 CLOUD COVER; No rain', '5/8 CLOUD COVER; No rain', '6/8 CLOUD COVER; No rain', '7/8 CLOUD COVER; No rain', 'COMPLETE COVER; No rain' => 1 (> 50% cloud cover)
- 'mostly cloudy', 'mainly cloudy', 'mostly coudy', 'overcast' but not include 'rain', 'fog', 'partly' => 1 (> 50% cloud cover)
- '4/8 CLOUD COVER; Patchy fog', 'COMPLETE COVER; Patchy fog', 'Fog (patchy)', 'patchy fog' => 2 (patchy fog)
- 'Fog (solid)', 'thick fog' => 3 (solid fog)
- 'Snow' => 7 (snow)
- Standardization rate: 8%

### obs_position_cd

- obs_position_cd does not have a corresponding free-text field.
- 'left front of aircraft', 'lf' => LF (Left front)
- 'lr' => LR (Left rear)
- 'port' => 'PT (Port)
- 'rf', 'right front of aircraft' => RF (Right front)
- 'rr' => RR (Right rear)
- 'starboard' => SB (Starboard)
- Standardization rate: 100%

## OBIS-SEAMAP datasets that would be added to the Catalog in the future

- SEUS Right Whale EWS Aerial Surveys 2018-19
- Beacon Wind Digital Aerial Wildlife Surveys for BOEM Lease Area OCS-A 0520, Equinor Wind US LLC, March 2021 – OctOBIS-SEAMAP - Explore Marine Megavertebratesober 2021
- Beacon Wind Digital Aerial Wildlife Surveys for BOEM Lease Area OCS-A 0520, Equinor Wind US LLC, December 2019 – November 2020
- Aerial digital survey of ecological studies for offshore wind energy development in Mid-Atlantic Bight
- Digital video aerial surveys of seabirds and marine megafauna in the Gulf of Maine from May 2023 to January 2024
- Empire Wind Digital Aerial Wildlife Surveys for BOEM Lease Area OCS-A 0512, Empire Offshore Wind LLC, July 2024 to May 2025
- SEFSC Southeast Atlantic Marine Mammal Assessment Survey 2011 (Not published)
- SEFSC Mid-Atlantic Cetacean Habitat Survey 2006 (Not published)
- SEFSC Mid Atlantic Tursiops Survey Winter 2002 (Not published)
- SEFSC Mid Atlantic Tursiops Survey Winter 2005 (Not published)
- SEFSC Mid Atlantic Cetacean Survey 2002 (Not published)