https://doi.org/10.38750/2306959

Pinniped Monitoring for the San Francisco Bay Area Network: 1983 - 2024 - Data Release Report (DRR)

Kirk Sherrill (0009-0000-9082-4799)[1]

[1] San Francisco Bay Area Network, Building 1063, Fort Cronkhite, Sausalito, CA 94965

Abstract

The San Francisco Bay Area Network (SFAN) consists of five parks located in the state of California. A monitoring program was established in 1976 to collect pinniped occurrence data in two network parks. Pinnipeds were chosen for monitoring due to their importance as ecological indicators, to fulfill reporting requirements federally mandated by the Marine Mammal Protection Act of 1972, and because of their consistent use of SFAN parks for breeding and molting. The purpose of the monitoring program is to determine trends in species occupancy in order to inform management needs in SFAN parks. In 2024, the data were then processed for dissemination by the Inventory and Monitoring Division (IMD), complying with Executive Order 13642, Making Open and Machine Readable the New Default for Government Information. Processed tabular datasets have been created using the data provided from the parks. The SFAN_Pinniped_Disturbance dataset contains 58 columns and 4251 observations. The SFAN_Pinniped_ElephantSeal dataset contains 56 columns, 44013 observations, and 5 unique species. The SFAN_Pinniped_HarborSeal dataset contains 57 columns, 66126 observations, and 3 unique species. The SFAN_Pinniped_HarborSealRedShark dataset contains 43 columns and 20439 observations. The SFAN_Pinniped_Resights dataset contains 68 columns, 23919 observations, and 2 unique species.

Acknowledgements

Thank you to Abby Hobbs for developing the initial Pinnipeds DRR template being using for this DRR.

Data Records

Data Inputs

Pinniped monitoring was conducted by Sarah Codde and others in two SFAN parks from 1983 to 2024. The data package described by this DRR was compiled from five separate

datasets. Original data from the study are available from the NPS DataStore at https://doi.org/10.57830/2306963 (SFAN 2024).

Summary of Datasets Created

This DRR describes the data package *Pinniped Monitoring for the San Francisco Bay Area Network: 1983 - 2024 - Data Package* which contains a metadata file and five data files. These data were compiled and processed for dissemination by the National Park Service Inventory and Monitoring Division (IMD) and are available at https://doi.org/10.57830/2306945 (see Table 1).

Since the previous Pinnipeds data package with data through 2023 (https://doi.org/10.57830/2303620), this data package (https://doi.org/10.57830/2306945) includes monitoring data from field season 2024. In addition to the 2024 fields season data this data package includes records having not only 'Accepted' but also 'Provisional' and or 'Raw' data processing levels. Ratings of 'Accepted', 'Provisional' and 'Raw' are used to define the amount of data quality evaluation process that has been performed on the data. See section Data Quality Evaluation below for additional information on quality control processes and definitions.

Table 1. Pinniped Monitoring for the San Francisco Bay Area Network: 1983-2024 - Data Package: List of data files.

File Name	Description
SFAN_Pinniped_Disturbance.csv	Seal disturbance events survey data from 2000 to 2024.
SFAN_Pinniped_ElephantSeal.csv	Elephant seal survey data from 1997 to 2024.
SFAN_Pinniped_HarborSeal.csv	Harbor seal survey data from 2000 to 2024.
SFAN_Pinniped_HarborSealRedShark.csv	Harbor seal red fur and shark bite survey data from 2000 to 2024.
SFAN_Pinniped_Resights.csv	Seal resights survey data from 1983 to 2024.

Data Quality Evaluation

Within the context of an annual data lifecycle workflow, data are reviewed and validated through data verification and validation procedures implemented by the San Francisco Bay Area Inventory & Monitoring Network (SFAN). After surveys are completed, monitoring personnel perform data entry validation to confirm entered data matches what was recorded on original data forms. Fields are checked for accuracy and completeness, including correct site names, expected locations and standard date and time entries. Data are checked against field notes when appropriate. After data entry verification has been completed records are given a data processing level of 'Provisional'.

Data Validation

Subsequent annual workflow steps are taken to perform formal and informal data validation quality control checks. Formal Data Validation procedures were performed on monitoring data through 2024, these validation checks are defined in appendix A. Records identified in the data validation routines were reviewed and corrected. Notes on any data validation steps taken outside of normal corrections are added to the respective narrative notes fields. After completion of data validation records are given a data processing level of 'Accepted'.

Data Quality Flagging

Data quality control (QC) flagging fields were added to the back-end schema in 2023. Flagging fields allow for formal designation via data flag codes and can be used when records have non-normal monitoring conditions (e.g. failed QC checks, adverse weather, pandemic, etc.). The data quality flags that have been realized through field season 2024 are defined in table 1.

In the datasets, the data QC flagging fields are denoted by inclusion of a 'QCFlag' value in the column name. QC flagging fields represent whole monitoring components or can be further broken down into finer scale sub monitoring components (e.g. Disturbance, Resight, Habitat, etc.). QC Notes fields accompany each QC flagging field and are similarly denoted by the inclusion of a 'QCNotes' value in a column name. This is where notes are added to provide additional information on the QC error and when applicable, notes on the data correction that was performed. The QC flagging fields and QC Notes fields by dataset and related monitoring components are defined in table 2.

Blank values in QC flagging fields denote that no QC flag has been applied to the record. Data that has completed data validation are given a data processing level of 'Accepted'.

Table 2. Realized data quality control flags.

Flag Code	Flag Definition	Analysis Interpretation
DBDO	Disturbance behavior data omission, data not collected	Not usable in a disturbance variable analysis
DEAD	Observation is for dead animal	No issues with use in analysis
DSSSU	Disturbance Source and Specific Source Updated post data collection	No issues with use in analysis
DSSU	Disturbance Specific Source Updated post data collection	No issues with use in analysis
DUP	Duplicate record, merged with another record	Do not use in analysis
MERG	Merged record with other duplicate record	No issues with use in analysis

Flag Code	Flag Definition	Analysis Interpretation
OBVUYN	Original binary value was updated to yes no value	No issues with use in analysis
ODEC	Original Data Entry value corrected from review of field sheet	No issues with use in analysis
STV	Suspect Time Value	Use with caution in analysis
TVSN	Time Value Set to Null, unknown time	Use with caution in analysis
Unable	Unable to survey all sites	Use with caution in analysis
UTP	Unknown Tag Position	No issues with use in analysis
VUQC	Value updated during Quality Control review	Use with caution in analysis

Table 3. Dataset quality control flag fields and monitoring component.

File Name [a]	Flag Field Name	Flag Notes Field	Monitoring Component
Disturbance	EventQCFlag	EventQCNotes	General monitoring event metadata
Disturbance	DisturbanceQCFlags	DisturbanceQCNotes	Disturbance at monitoring event
Disturbance	DisturbanceBehavQCFlag	DisturbanceBehavQCNotes	Disturbance behavior response at monitoring event
ElephantSeal	ElephantEventsQCFlag	ElephantEventsQCNotes	Elephant seal event metadata
ElephantSeal	EventQCFlag	EventQCNotes	General monitoring event metadata
ElephantSeal	SealCountQCFlag	SealCountQCNotes	Seal count

File Name [a]	Flag Field Name	Flag Notes Field	Monitoring Component
HarborSeal	PhocaEventsQCFlag	PhocaEventsQCNotes	Harbor seal event metadata
HarborSeal	EventQCFlag	EventQCNotes	General monitoring event metadata
HarborSeal	SealCountQCFlag	SealCountQCNotes	Seal count
HarborSealRedShark	EventQCFlag	EventQCNotes	General monitoring event metadata
HarborSealRedShark	SealCountQCFlag	SealCountQCNotes	Seal count
Resights	ResightEventsQCFlag	ResightEventQCNotes	Resights event metadata
Resights	EventQCFlag	EventQCNotes	General monitoring event metadata
Resights	ResightQCFlag	ResightQCNotes	Resights tagged data

Note: [a] The 'SFAN_Pinniped_' prefix and '.csv' suffix have been omitted from file names for brevity.

The data within the data records listed above have been reviewed by staff in the NPS Inventory and Monitoring Division to ensure accuracy, completeness, and consistency with documented data quality standards, as well as for usability and reproducibility. The *Pinniped Monitoring for the San Francisco Bay Area Network:* 1983 - 2024 - *Data Package* is suitable for its intended use as of the date of processing (2024-11-19).

Usage Notes

The columns TimeZone, Type, BasisofRecord, GeodeticDatum, DecimalLatitude, DecimalLongitude, VerbatimCoordinates, VerbatimCoordinateSystem, and VerbatimSRS were added on behalf of the Inventory and Monitoring Division during data processing and follow Simple Darwin Core guidelines.

Acquiring the Data Package

This data package is available for download from the NPS DataStore at https://doi.org/10.57830/2306945 and can be directly imported into R data frames using the NPSutils package (Baker et al. 2023) and the following code:

```
devtools::install_github('nationalparkservice/NPSutils')
NPSutils::get_data_packages(REFERENCE2306945)
dat <- NPSutils::load_data_package(REFERENCE2306945)
list2env(dat, envir = .GlobalEnv)
```

Methods

Data Collection and Sample Processing Methods

Monitoring focuses on the two numerically dominant pinniped species that breed, haul out, and molt in the San Francisco Bay Area Network (SFAN) parks: the Pacific Harbor Seal (Phoca vitulina richardii) and Northern Elephant Seal (Mirounga angustirostris). Ancillary data on other pinniped species are collected concurrently, but they are not the primary monitoring targets. The sample design and methods for this program were developed so that the data could be integrated with other regional surveys, allowing for the results to be interpreted in a regional context.

Monitoring is performed at two SFAN parks: Point Reyes National Seashore (PORE) and Golden Gate National Recreation Area (GOGA). Pinnipeds at GOGA are limited to haul out sites on islands within San Francisco Bay. Pinnipeds use both terrestrial and marine habitats of PORE. Haul out and pupping sites occur throughout the park, but are limited mostly to remote beaches, estuaries, or rocky shorelines. All primary sites where seal breeding occurs within SFAN parks are included in the current study design. The topographic diversity of this coastal zone provides a broad range of substrates upon which seals haul out: tidal mud flats, rocky intertidal areas, offshore tidal ledges, and sandy beaches. Most sites are accessed via trails from park roads. Counting locations are typically off trail and not in public access areas or in view of the public. The exceptions are Drakes Beach (DB1 and DB2), the Lifeboat Station, South Beach, Duxbury Reef, Point Bonita, and Bolinas Lagoon, which are in public access areas.

Harbor Seal Monitoring: Shore-based harbor seal surveys are conducted from standardized observation points using binoculars and/or spotting scopes a minimum of twice per week at each site during the breeding and molting seasons, which run from March 1 to May 31 and June 1 to July 31, respectively. Volunteers and park biologists survey each site, weather and logistics permitting. Each survey lasts two hours, if possible, with seal counts occurring every half-hour. Each subsite is surveyed separately, comprising a grand total for the site on each half-hour count. Regional surveys are coordinated with other monitoring organizations outside of the study area and occur bi-weekly during a five-day window (Thursday - Monday) between March 1 and July 31. Disturbance monitoring is

carried out concurrently with the seal population monitoring. All disturbances and potential disturbances that occur during the survey period are recorded.

Northern Elephant Seal Monitoring: The survey period for the breeding season extends from December 1 to March 31. Surveys for population size, distribution, productivity, and phenology are conducted using binoculars and/or spotting scopes a minimum of two times per week at all sites during the peak breeding season. This frequency is warranted to capture the maximum counts and changes in phenology. One count is conducted per survey, with surveys typically performed on the same days each week (i.e., Tuesday and Friday), but that schedule is weather dependent. In addition, during the breeding season, since adult females remain in the same location for at least one month, tags are resighted at each site once every two weeks at a minimum.

Dye-Marking, Tagging and Resighting Tags: Northern elephant seal tag information is collected both during scheduled surveys and opportunistically. Between 300 and 500 weaned northern elephant seal pups are tagged each year. Data on tagging and tag resighting are submitted as a component of the annual report to the National Marine Fisheries Service (NMFS). Tag resight data within and across seasons may be used to understand mechanisms of dispersal, colony expansion, and movements between PORE breeding sites. During a single season, temporary dye marks may be applied to tagged and/or dominant males to aid in tracking individual movements.

All Species Pinniped Monitoring – Point Reyes Headlands: Point Reyes staff and volunteers, under direction of the PORE Marine Ecologist, conduct year-round counts of all pinniped species at all locations on the Point Reyes Headlands. Winter surveys overlap with and are coordinated with counts of northern elephant seal breeding sites. The park's all species monitoring program compliments the harbor seal and northern elephant seal monitoring programs by documenting the northern elephant seal molt, harbor seal pupping, and harbor seal molt on the Point Reyes Headlands. The program additionally collects data on California sea lions, Steller sea lions, and northern fur seals. The monitoring program contributes to a long-term dataset for the Point Reyes Headlands dating back to 1980. The SFAN Inventory and Monitoring (I&M) program provides data management support for the all species monitoring program. All data collected at the Point Reyes Headlands are entered into the pinniped monitoring database. I&M data management staff assists PORE with data entry, data maintenance, and data summaries.

Data Management, Analysis, and Reporting: The SFAN uses a Microsoft Access relational database to store Pinnipeds monitoring data. The database is modeled after the Natural Resource Database Template (NRDT), an application developed by the National Park Service (NPS) I&M Program. The data schema utilizes the concept of sample events which are entered into an event table. Collected event data includes who, what, where, and when metadata information. The database is used to record counts of pinnipeds, natural and anthropogenic disturbances for harbor seals, and tagged northern elephant seals. There are two main data collection protocols, one for harbor seals and the other for northern

elephant seals. The primary information collected by both programs is the number of pinnipeds observed at a specific time and location.

Data Entry, Verification, and Editing: Data entry is performed by field staff with data entry being done in the Pinniped Access database front end forms. The data entry forms resemble the field data sheets as closely as possible. Data entry and editing are discrete operations which cannot be combined. Separate entry and editing forms are provided for each pinniped monitoring component (e.g. Elephant Seal, Harbor Seal, Tag Resight, Disturbance, etc.). For data entry verification, field staff must 100% cross check entered data in the database against the paper datasheets completed in the field. In addition, after the field staff have completed their review, the Project Lead/Network Data Manager should independently check a random 10% of the records for accuracy. Following NPS Inventory and Monitoring data certification guidance (see https://irma.nps.gov/DataStore/Reference/Profile/2227397), upon completion of the data entry verification process, the survey events for the field season are given a data processing level of 'Provisional'.

The harbor seal monitoring program undergoes a higher level of data scrutiny because it is primarily collected by volunteers and is more subject to difficult field conditions, such as fog and high winds, that may compromise the accuracy of the data. In addition, data collected by volunteers are given data reliability ratings based on the amount of experience and evaluation by project staff. During the season, and again at the end of the season, each survey should be evaluated and assigned a yes/no for high quality rating, based on factors such as poor visibility, incomplete survey, and poor observer quality.

Validation QA/QC Queries: As part of the annual workflow after data entry verification has been completed, a series of data validation quality assurance quality control (QAQC) queries are run from the Pinnipeds data entry front end. These queries help identify missing, out of range, or illogical data across the monitoring components. The project lead reviews each validation query and will either correct the record via data update or make note on why the record is not passing the QAQC check.

After reviewing and addressing any data issues, each validation query is defined as 'Done' for the field season. When all validation queries have been reviewed, an annual QAQC data validation results pdf report is exported. The data validation report documents completion of the annual validation process and outlines steps taken during the review process for each data validation routine. Upon completion of the data validation process, the survey events for the field season are given a data processing level of 'Accepted'.

The methodology protocol supporting this project is Adams D, Press D, Hester M, Nevins H, Roberts D, Becker B, Jensen H, Flynn E, Koenen M, Allen S. 2009. San Francisco Bay Area Network Pinniped Monitoring Protocol: Version 3.02. Natural Resource Report. Natural Resource Report NPS/SFAN/NRR—2009/170. National Park Service. Fort Collins, Colorado. It is available at https://irma.nps.gov/DataStore/Reference/Profile/663657.

Data Processing

- Rows were ordered by Start Date and Start Time.
- Columns were added to all datasets for TimeZone, Type, BasisofRecord, GeodeticDatum, DecimalLatitude, DecimalLongitude, VerbatimCoordinates, VerbatimCoordinateSystem, and VerbatimSRS to add additional context.
- NAs were formatted in newly added columns of all datasets.
- Extra white space was removed from cells in all datasets.
- The column type for the TimeZone column was set as character in all datasets.
- The format of date and time columns was standardized as yyyy-mm-dd and HH:MM:SS in all datasets.
- All datasets were formatted to retain blank cells instead of autopopulating with NAs.
- Columns were ordered within each dataset.
- Cleaned CSV files were exported.
- Temporary copies of the datasets were formatted for use in EML generation by adding a coded variable to avoid parsing errors in columns that were blank in at least the first 1000 rows.
- Temporary copies of the cleaned CSV files for EML generation were exported.

Metadata Processing

Metadata in the data package is in Ecological Metadata Language (EML) and was produced using the NPSdataverse (Baker et al. 2024b) package EMLeditor (Baker and Patterson 2024). Each of the five files contained some columns that were partially or fully blank, resulting in parsing issues during EML creation. In order to produce valid EML metadata, versions of these files with a flag code or date in each of the partially and empty columns were used to run the EML script. The codes were chosen as the placeholder value because they are possible values for these columns. The files included in the data package do not contain these placeholder values but instead reproduce the columns exactly as they exist in the original database. Data and metadata were run through quality control functions using the NPSdataverse packages QCKit (Baker et al. 2024a) and DPchecker (Baker and Wright 2024) before publication.

Code Availability

The scripts used to create this report and associated datasets are available on the NPS GitHub Repo https://github.com/NPS-SFAN/Pinnipeds_DataPackage. Scripts were prepared using packages available in the NPSdataverse (Baker et al. 2024b).

References

Adams, D and Press, D and Hester, M and Nevins, H and Roberts, D and Becker, B and Jensen, H and Flynn, E and Koenen, M and Allen, S. 2009. San francisco bay area network pinniped monitoring protocol: Version 3.02. National Park Service, Fort Collins, Colorado; Available at https://irma.nps.gov/DataStore/Reference/Profile/663657.

Baker, R. L., J. DeVivo, and J. Patterson. 2023. NPSutils: A collection of functions to read information from the NPS DataStore v0.3.0.

Baker, R., Q. Issac, and J. Patterson. 2024a. QCkit v0.1.4.

Baker, R., P. Judd, and J. DeVivo. 2024b. NPSdataverse: Tools and packages for data and metadata manipulation.

Baker, R., and J. Patterson. 2024. EMLeditor v0.1.3.

Baker, R., and S. Wright. 2024. DPchecker v0.3.4.

Hobbs, A. 2024. Pinniped monitoring for the san francisco bay area network: 1997-2023 - processing script. National Park Service, Fort Collins, Colorado; https://doi.org/10.57830/2303621.

K. Sherrill and S. Codde. 2024. Pinniped monitoring for the san francisco bay area network: 1983 - 2024 - data package. National Park Service, Fort Collins, Colorado; https://doi.org/10.57830/REFERENCE2306945.

National Park Service. 2016. Certification guidelines for inventory and monitoring data products. National Park Service, Fort Collins, Colorado; Available at https://irma.nps.gov/DataStore/Reference/Profile/2227397.

SFAN. 2024. San Francisco Bay Area network pinniped monitoring database. National Park Service, Fort Collins, Colorado; Available at https://irma.nps.gov/DataStore/Reference/Profile/2306963.

Appendix A. Data quality validation checks

qa_a101_UnverifiedEvents - Shows records that have not been marked as verified (i.e. DPL Status = 'Raw') qa_f101_ESeal_InvalidSeasonDates - Finds elephant seal records where the date does fall into the correct season (breeding or molt) qa_f102_HSeal_CheckStartTime - Shows records where the difference in minutes between the first subsite count in a survey and the start time of the survey is >15 mins (typically should not be significantly different).

qa_f111_ESeal_InvalidResightSeasonDates - Finds resight records where the date does not fall into the correct season (breeding or molt)

qa_f111_HSeal_CheckEndTime - Shows seal count records that have a count time that is later than the survey end time.

qa_f122_HSeal_CheckEndTime2 - Shows records where the difference in minutes between the last subsite count in a survey and the end time of the survey is >60 mins for TP or >30mins for other sites (typically should not be significantly different).

qa_f131_HSeal_DupCount - Shows duplicate seal count records (same observation time, species, maturity code, and live/dead status) for a survey event.

qa_f142_HSeal_SubsiteCounts - Checks that all subsites are counted for each survey: Shows where there is a difference between the number of counts and the maximum number of times a subsite is counted.

qa_f152_HSeal_PupDates - Shows the dates when harbor seal pups were observed outside their normal date range (before Mar 15 or after May 31). Double-check for erroneous dates. Updated to account for leap year logic 10/7/2024.

qa_f162_HSeal_PoorQualityCounts - Finds records with 'High Quality Count' marked as False, but a reason is not provided for the poor quality count.

qa_f202_Regional_Site_CheckStartTime - Shows records where the difference in minutes between the first subsite count in a survey and the start time of the survey is >15 mins (typically should not be significantly different).

qa_f211_Regional_Site_CheckEndTime - Shows seal count records that have a count time that is later than the survey end time.

qa_f231_Regional_Site_DupCount - Shows duplicate seal count records (same observation time, species, maturity code, and live/dead status) for a survey event.

qa_f252_Regional_Site_PupDates - Shows the dates when harbor seal pups were observed outside their normal date range (before Mar 15 or after May 31). Double-check for erroneous dates. Updated to account for leap year logic 10/7/2024.

qa_f262_Regional_Site_PoorQualityCounts - Finds records with 'High Quality Count' marked as False, but a reason is not provided for the poor quality count.

qa_q101_HSeal_OrphanRed - Shows records in tblPhocaSealCount that do not have a corresponding record in tblSealCount..

qa_q201_Regional_Site_OrphanRed - Shows records in tblPhocaSealCount that do not have a corresponding record in tblSealCount.

qa_r101_HSeal_DisturbanceDuplicates - Shows records that have the same disturbance response entered more than once for the same survey event.

qa_r112_HSeal_DisturbanceSource - Finds records in the disturbance table that has a specific source entered that is not in the lookup table.

qa_r122_Hseal_DisturbanceFWCountNullPup - ‘Flush in Water’ records with null “pups remain” where “count before” does not equal “remain” + “flush”.

qa_r132_Hseal_DisturbanceFWCountPup - ‘Flush in Water’ records with “count before” does not equal “remain” + “pups remail” + “flush”.

qa_r152_Hseal_DisturbanceFlush - Shows disturbance records with a “F” response type indicating flush towards water (not into water) that has a value >5 for “Flush” (indicating > 5 individuals flushed into water).

qa_r201_Regional_Site_DisturbanceDuplicates - Shows records that have the same disturbance response entered more than once for the same survey event.

qa_r212_Regional_Site_DisturbanceSource - Finds records in the disturbance table that has a specific source entered that is not in the lookup table.

qa_r222_Regional_Site_DisturbanceFWCountNullPup - ‘Flush in Water’ records with null “pups remain” where “count before” does not equal “remain” + “flush”.

qa_r232_Regional_Site_DisturbanceFWCountPup - ‘Flush in Water’ records with “count before” does not equal “remain” + “pups remail” + “flush”.

qa_r252_Regional_Site_DisturbanceFlush - Shows disturbance records with a “F” response type indicating flush towards water (not into water) that has a value >5 for “Flush” (indicating > 5 individuals flushed into water).

qa_s102_Eseal_ResightDuplicateTags - Finds new tag records in the resight table where the left tag and right tag are the same.

qa_s112_Eseal_ResightDuplicateLTag - Identifies if duplicate new left flipper tags have been entered in the resight table.

qa_s122_Eseal_ResightDuplicateRTag - Identifies if duplicate new right flipper tags have been entered in the resight table.

qa_s132_Eseal_ResightMissingLTag - Finds left flipper tag records on weaned elephant seals that were never entered as new tags.

qa_s142_Eseal_ResightMissingRTag - Finds right flipper tag records on weaned elephant seals that were never entered as new tags.