# Monitoring Protocol for the Denali Caribou Herd, Denali National Park and Preserve, Alaska

## Standard Operating Procedure (SOP) # 3

## Data Management

## Version 1.0 (January 2017)

Layne Adams
Alaska Science Center, U.S. Geological Survey
4210 University Drive, Anchorage, AK 99508

## Revision History Log:

| Prev. Version # | Revision Date | Author | Changes Made | Reason for Change | New Version # |
|---|---|---|---|---|---|
| | | | | | |
| | | | | | |
| | | | | | |
| | | | | | |
| | | | | | |
| | | | | | |

Data management for the DENA caribou monitoring program will proceed through an ongoing life cycle as shown in Figure 1. Major stages of the life cycle include:

- Data collection
- Data entry and importing
- Quality control
- Dataset certification
- Data analysis and reporting
- Dataset documentation, archival and publication
- Standard operating procedures revision.

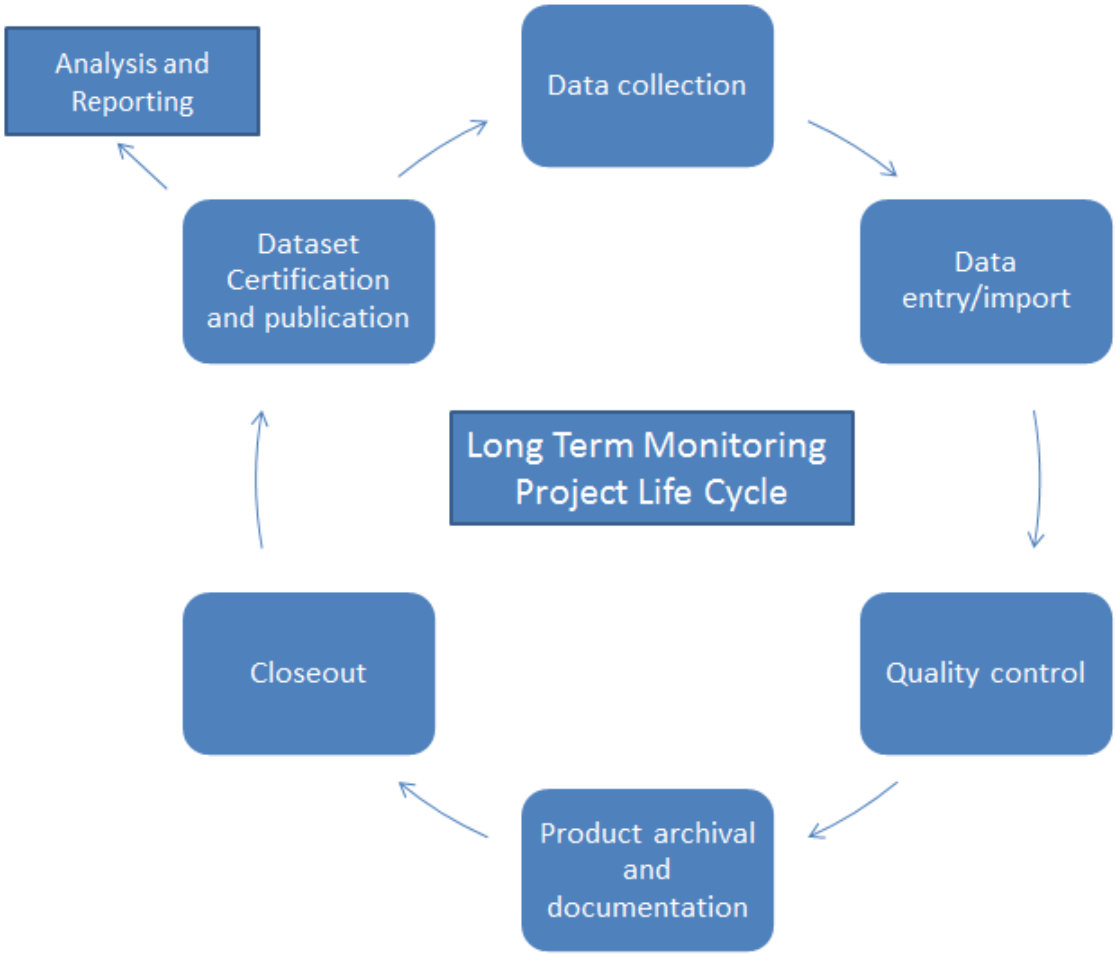



Figure 1. Long term monitoring project life cycle.

Each of these topics is covered in the sections below with specific guidance provided in standard operating procedures.

## NPS Standards

### *Data Quality*

Director's Order #11 mandates a high level of information quality from the service's scientific programs:

*"The NPS will ensure that information it releases to the public or utilizes in management decisions will be developed from reliable data sources that provide the highest quality of information at each stage of information development. The NPS's methods for producing quality information will be made transparent, to the maximum extent practicable, through accurate documentation, use of appropriate internal and external review procedures, consultation with experts and users, and verification of the quality of the information. The NPS will also keep users informed about corrections and revisions.*

*Information will be developed only from reliable data sources based on accepted practices and policies, utilizing accepted methods for information collection and verification. To ensure high quality, information should be peer reviewed at an appropriate level and at appropriate times. Data will be reproducible to the extent possible. Influential information will be produced with a high degree of transparency about data and methods. The information should include all pertinent information to allow the public to understand the park's legislative authorities, mission, activities, organization, strategic plan, performance plan, and performance accomplishments."*

*Adherence to the data management procedures in this monitoring protocol will ensure compliance with DO#11.*

#### *Freedom of Information Act*

The Network will comply with all FOIA restrictions regarding the release of data and information as instructed in NPS Director's Order #66.

#### *Records Management*

Direction for managing physical and digital materials is provided in NPS Director's Order 19: Records Management (2001) and its appendix, NPS Records Disposition Schedule (NPS-19 Appendix B, revised 5- 2003) ( http://www.nps.gov/policy/ DOrders/DOrder19.html ). Director's Order NPS-19 states that all records of natural and cultural resources and their management are considered mission-critical records, that is, necessary for fulfillment of the NPS mission. NPS-19 further states:

"*Mission critical records are permanent records that will eventually become archival records. They should receive the highest priority in records management activities and resources and should receive archival care as soon as practical in the life of the record."*

### Core Functional Model

Data generated by the DENA caribou monitoring program are managed according to the Core Functional Model described in Figure 2.

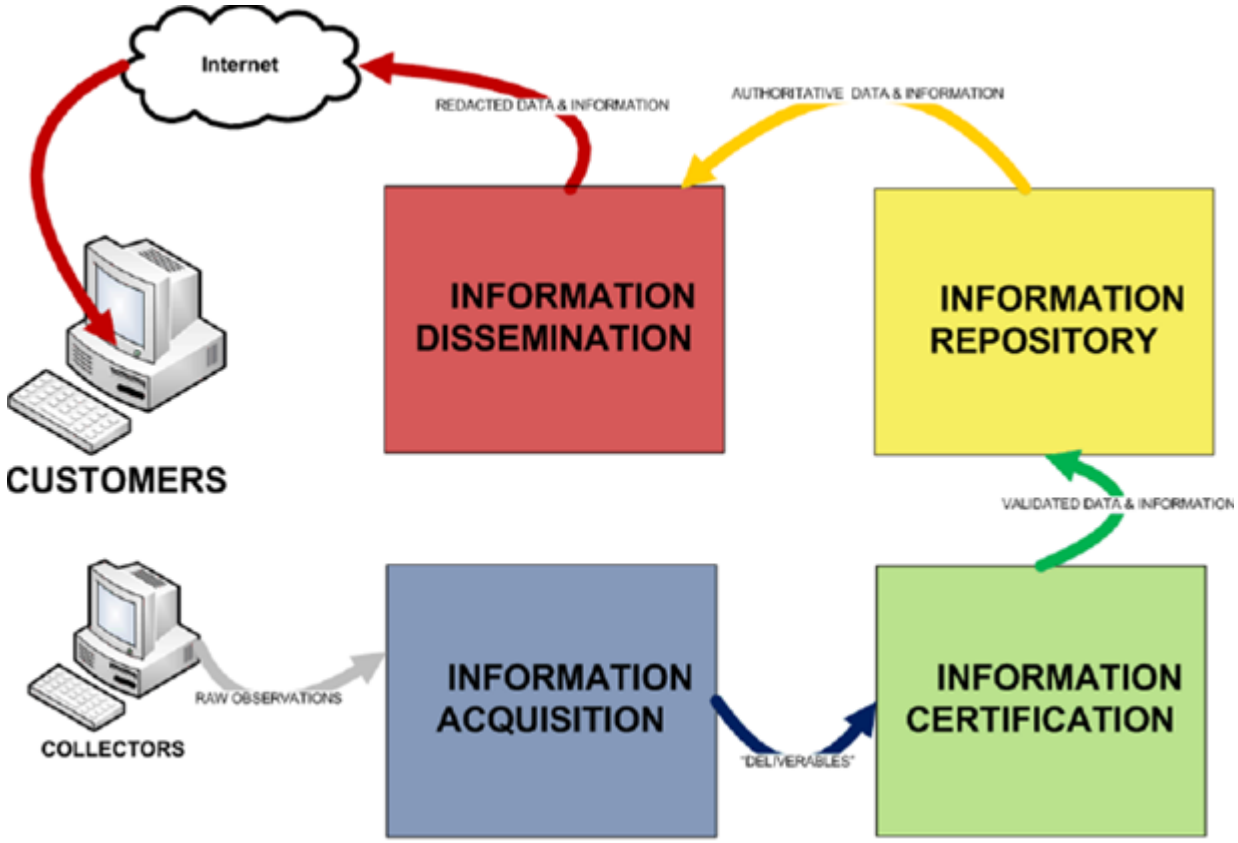



Figure 2. The CAKN Core Functional Model overview.

Dissemination services (red, Figure 2) provide all deliverables to all customers using Internet web services through the NPS Integrated Resource Management Applications (IRMA) Data Store. Repositories (yellow) store the certified inventory and monitoring products called for in the protocols. Certification processes (green) assure repositories have the highest quality data and that sensitive items are restricted to authorized users. Data acquisition processes (blue) led by park staff and cooperators include a wide array of tasks, ranging from collecting raw data to composing reports on long-term trends.

### Data Sensitivity Statement

The NPS is directed to protect information about the nature and location of sensitive park resources under one Executive Order and four resource confidentiality laws:

- Executive Order No. 13007: Indian Sacred Sites
- National Parks Omnibus Management Act (NPOMA; 16 U.S.C. 5937)
- National Historic Preservation Act (16 U.S.C. 470w-3)
- Federal Cave Resources Protection Act (16 U.S.C. 4304)

- Archaeological Resources Protection Act (16 U.S.C. 470hh)

When any of these regulations are applicable, public access to data can be restricted. If disclosure could result in harm to natural resources, the records may be classified as 'protected' or 'sensitive'. The NPS recognizes the following resources as sensitive:

- Endangered, threatened, rare, or commercially valuable National Park System resources
- Mineral or paleontological sites
- Objects of cultural patrimony
- Significant caves

**The CAKN caribou monitoring protocol defines the collection of sensitive data and requires special handling procedures.**

The justification for this declaration of sensitivity is that the data deals with 'rare, threatened, endangered, or commercially valuable species'. Network staff must ensure that all references to protected information are removed or obscured in any reports, publications, maps, or other public forum. Following the standard for Freedom of Information Act (FOIA) requests, the Network will segregate the non-releasable information and where practical will not withhold associated releasable information. Metadata for that dataset and the IRMA citation must also be updated to reflect that the associated data is sensitive. Network staff must ensure that all references to protected information are removed or obscured in any reports, publications, maps, or other public forum.
The project leader is responsible for maintaining the security of data released to any NPS dataset catalog and ensuring confidentiality agreements are in place before sharing data with any non NPS entity.

## System Architecture

The DENA caribou monitoring data management system has several components. Files related to day to day operations as well as raw and working data files reside on the USGS Principal Investigator's computer which is backed up weekly (?) to the USGS Alaska Science Center's computer network system. There are four primary sources of data that encompass the DENA caribou protocol: 1) captures of female caribou to install, replace, or remove radiocollars; 2) radiotracking to assess the survival and locations of collared caribou; 3) investigations of caribou mortalities; and 4) surveys conducted to estimate natality, age/sex composition and population size. Raw and processed data, certified datasets and related data products such as reports and analyses are archived with metadata in the IRMA Data Store.

## Deliverables, and Roles and Responsibilities

The caribou monitoring program creates and maintains specific data deliverables that are described in **Error! Reference source not found.**. Deliverables are integrated with the caribou monitoring databases at which time they are documented with metadata and archived as private Dataset

references in the IRMA Data Store. While data products may reside in the CAKN shared drive for convenience while processing for quality and integrating with master databases, the single authoritative source for certified and documented data products is always the IRMA Data Store. In the event that a data product ever diverges from that in the IRMA Data Store the latter version should be considered authoritative.

### Master Database

Field deliverables are collected, processed for quality and then integrated into the respective Excel spreadsheets. The data are situated behind the USGS firewall and access is limited to the Principal Investigator. The master working database is used for assessing and improving data quality until such time that it is certified annually. Those certified data are transferred to the park and network and a certified copy is archived in the IRMA Data Store and the master dataset's reference in IRMA updated with metadata and access information.

### Quality Control and Assurance

Data entry and quality control operations occur primarily using GIS and database applications to visually inspect spatial and tabular data. Database and statistical scripts change and improve frequently. We will store them in GitHub, a public source code repository (https://github.com/NPS-ARCN-CAKN/DENA-Caribou), where they are readily accessible, documented and versioned.

### Certification

Certification is the last step in the data management lifecycle and indicates that the monitoring dataset is ready for analysis. Certification does more than indicate that a set of records is complete it also indicates that the related protocol, standard operating procedures and data products are properly documented and archived in good order.

Certification communicates to the data consumer that they may be confident that the data is of excellent quality:

- The monitoring protocol and all standard operating procedures are complete and up to date
- Each data record is traceable to the protocol version and SOP version that guided its collection.
- The data being published has undergone quality control and assurance procedures and has been deemed as free of errors as reasonably possible and to the best abilities of the scientist or their designees for data management.
- All dataset closeout activities have been performed. Such activities include archiving all raw, intermediate and finished data processing files with metadata or written documentation in the IRMA Data Store
- Products are linked to the main monitoring project's IRMA reference in a logical and traversible way that provides for discovery of such products from the main project reference.

- The scientist in charge of the data has signed that the dataset is of sufficiently high quality that analyses may be performed on it and the results published.

## Reporting

### *Master IRMA Project Reference*

The NPS IRMA Data Store is the primary organizational repository and publication platform for information related to the DENA caribou monitoring program. A single IRMA project reference (https://irma.nps.gov/DataStore/Reference/Profile/2221183) serves as the master reference for the Vital Sign. All deliverables and raw data products are archived in IRMA where possible and linked to the master project reference. This architecture guarantees that discovering any individual product leads to discovery of all the other products (Figure 3).

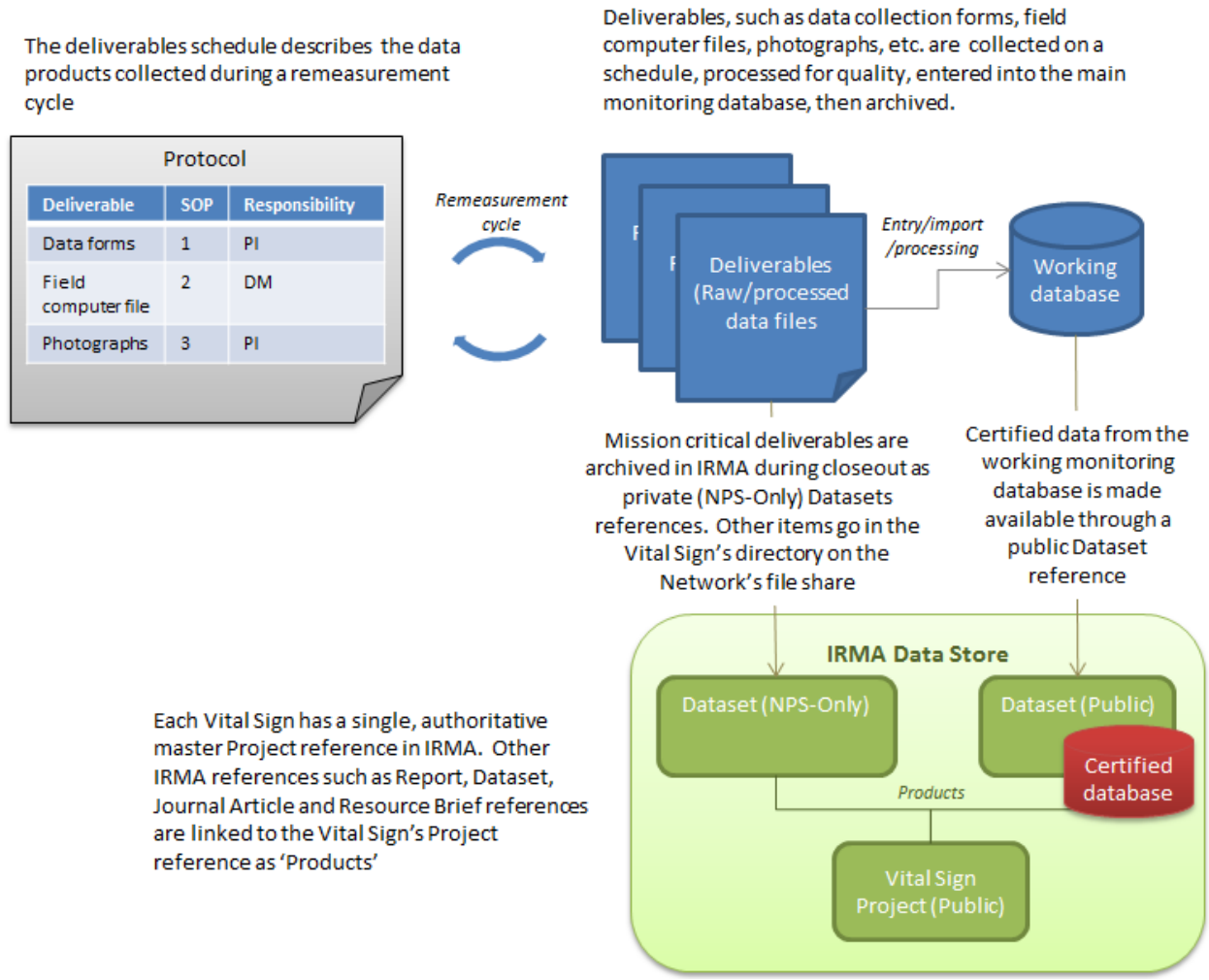



Figure 3. Overview of data collection and archival process. Data quality improves as it moves from collection of raw data deliverables through quality control and database integration procedures and into a series of well-documented IRMA Data Store references. All Data Store references are linked to a single, authoritative Vital Sign master Project Reference.

### *Protocols and Standard Operating Procedures*

The monitoring protocol and standard operating procedures are stored and versioned in IRMA and linked to the master IRMA project reference. The national I&M office maintains a directory of monitoring protocols and SOPs and this protocol will integrate with their system.

### *Website*

The public caribou monitoring web profile (https://science.nature.nps.gov/im/units/cakn/vitalsign.cfm?vsid=49) will publish the releasable contents of the DENA caribou Vital Sign's master Project reference using IRMA's web services. The website pulls data from the master Project reference for display. It is the joint responsibility of the PI and the data manager to keep the system of related IRMA references up to date.

### *Reports*

Reports are published in the form of NPS Natural Resource Reports (NRR), Natural Resource Data Series (NRDS) Reports (http://nature.nps.gov/publications/nrpm/index.cfm) or professional journal articles. Data are published with reports and archived in tandem in the IRMA data store.

All reports providing statistical results will also publish the dataset with metadata on which they were based, as well as link to the certified original dataset, i.e. 'publish the data with the report'.

### *Resource Brief*

The I&M resource brief is the primary outreach document for the DENA caribou monitoring program. The principal investigator revises and publishes the resource brief after new findings to reflect current understanding of the resource's trend and implications for park management. The resource brief is archived in the IRMA Data Store and linked as a product of the Vital Sign master Project reference.

## Product Archival and Documentation

NPS Director's Order 19 states that all records of natural and cultural resources and their management are considered mission-critical records (necessary for fulfillment of the NPS mission) and must be permanently archived.

Deliverables are not considered delivered until they exist in the IRMA data store with complete metadata. IRMA is the only authoritative repository for mission critical I&M products.

Data archiving is done in concert with seasonal close out and data certification outlined in this SOP. All data are archived in the following ways:

- Within our SQL Server database
- Within CAKN's database backup files
- In DENA Caribou monitoring directory on the CAKN shared drive
- In the IRMA Data Store.

The network data manager and project manager(s) develop metadata for archived data products. Data deliverables that are ancillary to published products but valuable to in-house operations are archived in the DENA caribou monitoring directory on the CAKN shared network drive.

Caribou monitoring program data are maintained on a network file share and SQL Server database located in the Fairbanks Administrative Center (FAC) in Fairbanks, Alaska. The database is backed up off-site. The NRSS and FAC use enterprise-level business continuity processes to maintain data security.

### *Metadata*

Documentation is a critical component of I&M goals of preserving information over long periods of time. Metadata is 'data about data' and should provide all the information a future data consumer should know in order to use a dataset. Metadata describes the attributes of any data resource as well as background information needed to understand and interpret it. Metadata formats that are used by the DENA caribou monitoring protocol include Federal Geographic Data Committee (FGDC, 1998) for spatial data and Ecological Metadata Language (https://knb.ecoinformatics.org/#external//emlparser/docs/index.html) for tabular data. These formats are text based and reflect current industry standards.

### *Line of Sight*

Line of Sight is the informal name given to the I&M program's philosophy that every data record should be inextricably and traversibly linked to the standard operating procedure and protocol version that guided its collection. This metadata chain is maintained through metadata columns in the main monitoring database coupled with metadata maintained in each protocol and standard operating procedure's IRMA reference. A data consumer will reliably be able to access a data record and discover the standard operating procedure version that defined its collection and processing and vice versa.

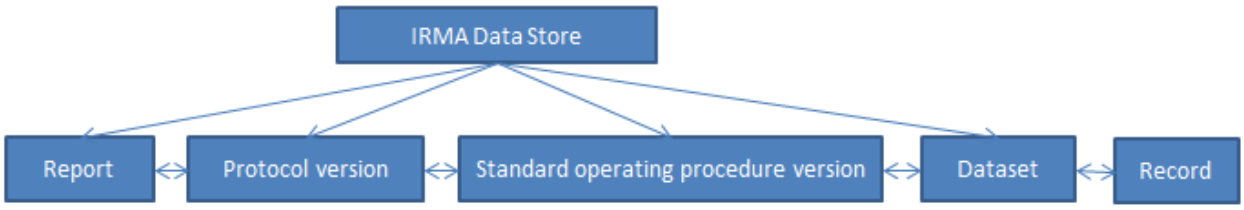



Figure 3. Each component informs and is informed by other components. Discovery of one component leads to discovery of all other components in a formalized data management chain of custody.

## Data and Information Security

Digital products are stored in the IRMA Data Store, SQL Server databases on I&M servers and/or local servers and the Network file server in descending order of preference. Under no circumstances will any mission-critical information permanently reside anywhere but on I&M shared network

drives or database servers. Project materials must be secured and available on I&M servers at all times. The data manager, in conjunction with IT staff at FAC and with the Network support office in Ft. Collins, CO is responsible for backups, security and longevity of monitoring data.

## Entering and Using Caribou Data

As described throughout the Monitoring Protocol for the Denali Caribou Herd, data necessary to meet the monitoring objectives come from 4 sources: 1) captures of female caribou to install, replace, or remove radiocollars; 2) radiotracking to assess the survival and locations of collared caribou; 3) investigations of caribou mortalities; and 4) surveys conducted to estimate natality, age/sex composition and population size. Data from captures, radiotracking, and mortality investigations are recorded on specific datasheets in the field, and then archived by the principal investigator. These data are regularly compiled into 3 databases maintained currently as EXCEL spreadsheets. These databases include: 1) records of the initial capture and status (alive, dead, cause of death) of each radiocollared female caribou (DCMORT.xlsx); 2) body masses and other physical measurements obtained with each capture of each caribou (DCCOWPHYS.xlsx); and 3) radiotracking location data (DCCOWRT.xlsx). Currently, data from surveys are maintained in original files or spreadsheets specific to each survey, and summarized in annual reports.

**DCMORT.xlsx:** This database is designed to be active and useful for the ongoing caribou studies rather than an archival database populated only with completed data. Because this database serves as the record of caribou currently being monitored and is therefore crucial for managing project activities such as radiotracking and recollaring caribou, it is maintained by the project leader. Data entries are reviewed and compared against the original datasheets and the database is regularly crosschecked against other information (e.g., radiotracking datasheets, morts to be recovered, etc.) to ensure accuracy.

The database includes a single record for each female caribou captured as a part of the monitoring effort. Following initial capture, an entry is made that includes several static fields to designate the individuals identifying number (COWNO), its inclusion in the age-structured sample (CS), birth year (CBY), the aging criteria used for this individual (CAC), the date of initial capture (DOC), and the individual's current age (AGE5_16 for 2016 biological year). The individual's current radiofrequencies (CURFREQ) and capture date (RECDATE) are also entered but are updated each time the caribou is recaptured, and cleared once the caribou is dead or removed from the sample. Periodically as radiotracking is completed, caribou that are alive are denoted under STATUS. When a caribou dies or its collar is removed, fields related to the timing of death (LSTALIVE, FRSTMRT, DOD) and probable cause (COD, STATUS) are completed. At that point the record is complete serving to specify the period the individual was included in the study and its fate.

DCMORT.xslx FIELD STRUCTURE/DEFINITIONS

COWNO: Five-digit animal number (e.g., 95602)

CURFREQ: Current radiofrequency for caribou that are alive (e.g., 4.570)

CS: Age-structured cow sample. 1 = in sample; 0 = not in sample; -9 if not assigned yet.
CBY: Cow birth year (e.g., 1996).
CAC: Ageing criteria: 1 = known-age (captured as calf); 2 = cementum analysis
DOC: Original date of capture (yyyymmdd)
RECDATE: Most recent recapture date for caribou that are alive (yyyymm)
Note: useful for determining caribou that are due for recollaring.
LSTALIVE: Last date located alive (yyyymmdd; -9 if alive)
FRSTMRT: First date mortality was detected (yyyymmdd; -9 if alive)
DOD: Estimated date of death; usually the midpoint between LSTALIVE and FRSTMRT, unless other information is available to determine the date of death (yyyymmdd; -9 if alive).
COD: Cause of death. See current codes below. BLANK if alive.
STATUS: Updated periodically as radiotracking is completed (e.g. ALIVE LT SEP 2016). Once animal dies or is removed from radiocollared sample, a brief descriptor and date are added (e.g., dead-bear (5/09).
AGE5_16: Age on May birthday of surviving cows. Updated each year to keep track of age structure of radiocollared sample.

Cause of Death Codes (COD):

-ACCIDENT:
-AVALANCHE:
-BEAR:
-CAPTURE:
-DELTA: 5 caribou calves that were captured in March 1993 that ended up being with the Delta Herd.
-DROP: 1 caribou that lost her collar shortly after capture in September 1986 because the nuts were not tightened.
-FAIL: Radiofailures.
-OTHER: Nonpredation deaths from otherwise undetermined causes.
-REMOVE:
-ROADKILL: 1 caribou found dead next to Parks Highway
-UNKNOWN: No evidence of a particular cause of death.
-UNKPRED: Likely predation deaths but specific cause not determined.
-WOLF:
-WRINE: Wolverined-caused deaths.

**DCCOWPHYS.xlsx**: The purpose of this database is to consolidate information on mass and other physical measurements of female caribou taken during each capture. This information is useful for assessing changes over time in body mass and skeletal size of caribou relative to changes in caribou abundance and environmental factors. On occasion, no measurements are taken during a capture due to health issues with the individual or a difficult location; nothing is entered for these captures. All data entries are reviewed and compared against the original datasheets. Static information on each

caribou (COWNO, CS, CBY, CAC, CAGE) are crosschecked against DCMORT.xlsx to ensure accuracy. Measurement fields are sorted and examined for outliers, typos, and other obvious data entry errors.

DCCOWPHYS.xslx FIELD STRUCTURE/DEFINITIONS

COWNO: 5 digit animal number (e.g., 95602)
DOC: date of capture (yyyymmdd)
BIOYR: Biological year starting 1 May (e.g., 15 Nov 1999 is 1999; 15 Mar 2000 is 1999)
MO: Month of capture (1 – 12; March = 3 etc.)
SPR: Spring ? Spring captures (Mar, Apr) = 1; Fall captures (Sep, Oct, Nov) = 0
CS*: Age-justified sample? Yes = 1 , No = 0, Not entered = -9
CBY*: Cow birth year (e.g., 1996); Not entered = -9
CAC*: Ageing criteria: Known-age = 1, Cementum analysis = 2
CAGE*: Cow age. (CAGE = BIOYR – CBY)
WGTKG: Actual weight in kg (Total weight in lbs - tare divided by 2.2) to nearest 0.1 kg.
SKUL: Skull length to nearest 0.5 cm.
MAND: Mandible length to nearest 0.5 cm.
METL: Metatarsus length to nearest 0.5 cm.
HFL: Hindfoot length to nearest 0.5 cm.
LAL and RAL: Left and right antler lengths to nearest 0.5 cm. If antler is shed or no pedicel, enter '0'. If data is missing, enter '-9'. If antler is broken, -9 and make note in comments.
LAP and RAP: Left and right antler points.
*Do not enter these fields from the capture data sheet or radiotracking sheet. Use list of cows generated from Layne's DCMORT.XSLX file only.

**DCCOWRT.xlsx:** The data incorporated here is used to assess the seasonal distribution of caribou from the periodic radiotracking that occurs throughout the year. Given that purpose, this database includes 1 location from each active caribou that was located during each of the 8 standard radiotracking bouts. Radio contacts where caribou were heard but not located, located as a mortality, or excess radiolocations during a bout are not included; therefore this database is not a complete record of all radiocontacts.

Methods for acquiring coordinates associated with locations have changed over the course of the study and therefore, should be used to assess caribou distribution at a course scale only. From September 1986 to approximately June 1995, locations were mapped on 1:63,360 USGS quads (occasionally 1:250,000) and later digitized to determine coordinates. These locations were probably very accurate in regions with many obvious terrain features and when mapped by experienced personnel with intimate knowledge of the geography of the study area. Locations could be off substantially in flat, monotonous terrain or when mapped by inexperienced personnel. Since June 1995, coordinates have been acquired from GPS units in the radiotracking aircraft. Locations were degraded by selective availability until about 2000. Generally, all these locations were collected using the WGS84 datum.

Although, radiotracking schedules have varied somewhat during the course of the study, 8 radiotracking bouts (RTBOUT) were identified in 2008 in an attempt to standardize the radiotracking effort and to identify periods when caribou have characteristically been located. Variation in the period that locations were acquired for each bout are included below. In 2012, RTBOUT 1 was changed to listen-only to assess survival of individuals at the breakpoint between biological years, but not to actually locate individuals, whereas radiolocation has occurred for all other bouts throughout the year.

Data are entered from the original radiotracking datasheets as each biological year is completed. Once the data is entered, information used to designate an individual (COWNO, CS) is checked against DCMORT.xlsx. Locations are screened to identify any outliers or other obvious data entry errors. Locations are plotted via GIS, scrutinized to identify locations that are out of place (e.g., outside of normal caribou range for the given season, located at high elevation, etc.), and compared against the descriptive locations to identify and correct any coordinate errors.

DCCOWRT.xslx FIELD STRUCTURE/DEFINITIONS

BIOYR: Biological year starting 1 May (e.g., 15 Nov 1999 is 1999; 15 Mar 2000 is 1999).
RTBOUT: Radiotracking bout 1-8 (see description below)
DATE: (yyyymmdd)
COWNO: 5 digit animal number (e.g., 98782)
CS: Age-justified sample? Yes = 1, No = 0
LOCATION: Brief descriptive location (e.g., Turtle, U E Fork; included since May 2008)
LAT: Latitude in decimal degrees (e.g. 63.2012)
LON: Longitude in negative decimal degrees (e.g., -149.4526)

RTBOUT as follows (total and interquartile range figured on Sep 1986-Apr 2008 data):

1. End of Apr/early May (end of biological year): 26 April – 11 May (interquartile range [IR] = 5-9 May).
2. Mid-May (mid-calving): 11 – 24 May (IR = 14-15 May).
   NOTE: This is the only case where 2 bouts overlap (11 May in bouts 1 and 2), but the overlap does not occur in the same year. All 11 May bout 2 locations in 2006. Not an issue because we are unlikely to use bout 1 data.
3. Late May/early Jun (late/post-calving): 25 May-29 Jun (IR = 28 May-6 Jun).
4. Late Jul/early Aug (midsummer): 9 Jul-29 Aug (IR = 25 Jul-9 Aug).
5. Late Sep (rutting season): 22 Sep-15 Oct (IR = 24-26 Sep)
6. Late Nov (early winter): 5 Nov-15 Dec (IR = 18 Nov-2 Dec)
7. Late Jan/early Feb (mid-winter): 7 Jan-11 Feb (IR = 24 Jan-11 Feb)
8. Mid-Mar (late winter): 6-31 Mar (IR = 12-18 Mar)