Standard Operating Procedure (SOP) 9: Data QA/QC, Formatting and Archiving

Protocol for Monitoring Small Mammals

Central Alaska (CAKN) Inventory and Monitoring Network

Dylan T. Schertz, Melanie J. Flamme, Kumi L. Rattenbury and Josh H. Schmidt,

National Park Service
4175 Geist Road
Fairbanks, Alaska 99709

September 2018

U.S. Department of the Interior
National Park Service
Natural Resource Stewardship and Science
Fort Collins, Colorado

Please cite this Standard Operating Procedure as:

Schertz, D. T., M. J. Flamme, K. L. Rattenbury and J. H. Schmidt. 2018. Standard Operating Procedure (SOP) 9: Data QA/QC, Formatting and Archiving. Protocol for Monitoring Small Mammals in Central Alaska (CAKN) Inventory and Monitoring Networks. National Park Service, Fairbanks, Alaska. Available online at: https://irma.nps.gov/App/Reference/Profile/2219630.

The main Narrative for the Protocol for Monitoring Small Mammals in Central Alaska (CAKN) Inventory and Monitoring Networks is available from the NPS Inventory and Monitoring Central Alaska Network website (http://science.nature.nps.gov/im/units/CAKN/) and the Natural Resource Publications Management website (http://www.nature.nps.gov/publications/nrpm/

Standard Operating Procedure #9

Version 1.0 (September 2018)

Previous Version #	Revision Date	Author	Changes Made	Reason for Change	New Version #
Add rows as needed for each change or set of changes associated with each version.					

Contents

SOP 9.0 Introduction 4

SOP 9.1 Post Field Campaign Data Handling 4

SOP 9.2 Initial Quality Assurance/Control Procedures 5

SOP 9.3. Importing Field Data to the Master Small Mammal Monitoring IRMA Data Store 6

SOP 9.5 Final Quality Assurance/Control Procedures 6

SOP 9.6 Data Preparation for Analysis 7

SOP 9.7 Seasonal Close Out and Dataset Certification 7

SOP 9.8 Archiving Field Deliverables in IRMA 8

Creating a Field Deliverables Dataset Reference in IRMA 8

Link the Field Deliverables Reference as Product of the Master Small Mammal Monitoring Project Reference 10

SOP 9.9 Literature Cited 12

SOP 9.10 Appendix C: Dataset Certification Form 13

SOP 9.0 Introduction

SOP 9 covers data quality assurance and quality control (QA/QC) measures and formatting of the data for analysis and for final archiving of data to IRMA data store. See SOP 8 for details on proofing and filtering data for data quality assurance and quality control (QA/QC). For full protocols regarding Small Mammal analysis, see SOP 10 (Data Analysis) and Schmidt et al. (2017).

A general overview of data processing steps appears below with specific procedures described in more detail in the sections that follow. See SOP 8 for details.

1. Download all field data following the surveys and initiate the quality assurance/control (QA/QC) process.
2. Verify Park Observer data with hard copy data field data notebooks, field notes and mortality tags.
3. Edit, sort and proof data in Excel.
4. Import the proofed data into the Small Mammal Access database which contains all y previous years' survey data.
5. Perform quality control and assurance procedures.
6. Prepare data for analysis in program R and WinBUGS (Schmidt *et al.* 2017). If anomalies show up during analysis, another round of data verification, validation and analysis will be initiated.
7. Perform seasonal close out and dataset certification.

SOP 9.1 Post Field Campaign Data Handling

See SOP 8 for details on downloading and proofing data after field work is completed. Small mammal survey data should come from the field well organized according to SOPs 7 and 8. The first office task is to archive the files on the CAKN Network shared drive.

1. Navigate to the Small Mammal monitoring root directory on the CAKN network drive: J:\Monitoring\Small_Mammals\CAKN_DENA_SmallMammals
2. Field survey data is organized in the CAKN data folder by year. Make a folder for the current survey year if it does not exist already. Example: J:\Monitoring\Small_Mammals\CAKN_DENA_SmallMammals_Data\(YEAR).
3. Create "Read Only" and "Working folders for archiving data. "Read Only" contains raw data and is a zip archive of survey files as they came from the field. "Working" contains proofed and edited data. See figure 9.1.
4. Retrieve emailed Park Observer: "Critter.csv," "TrackLogs.csv," and "GpsPoints.csv" (SOP 8).
5. All retrieved Park Observer data are imported into a consolidated an Excel file (SOP 8).
6. Consolidated Park Observer data in Excel are reconciled (verified) against field data notebooks and field notes to ensure every record is entered and duplicates are deleted (SOPs 8 and 9).

7. Incomplete or inaccurate data records in the consolidated Excel file are corrected. Every record in field data notebooks is checked off in red ink after proofing (SOP 8).
8. Consolidated data are filtered in Excel in various ways to ensure data make sense (validated, SOPs 8 and 9).
9. Once consolidated data are proofed, data are again spot checked for completeness and accuracy by a separate technician and the PI. The PI determines when the data are sufficiently proofed and considered final (SOP 8).
10. Proofed final data in Excel are appended to the Small Mammals Access database (SOP 8).
11. Field notes and filed data notebooks are signed and dated once data proofing is completed. The notebooks are scanned and saved as .pdfs and stored on the CAKN network folder: J:\Monitoring\Small_Mammals\Scanned Field Notebooks and Databooks (SOPs 8 and 9).
12. The proofed, consolidated Access database is renamed as CAKN_DENA_SmallMammalsYear_Final and saved to the CAKN network drive: J:\Monitoring\Small_Mammals\CAKN_DENA_SmallMammals_Data\FinalData (SOPs 8 and 9).
13. Photos are downloaded from all iPads, cameras and phones. Photos are culled for blurry and redundant images. Final photos are batch renamed (SOP 8). (CAKN_DENA_SmallMammalsYear). All photos are downloaded to the network folder by year: J:\Monitoring\Small_Mammals\CAKN_DENA_SmallMammals_Photos_Year (SOP 8).

Table 9.1 describes in more detail the structure and function of the paired field data directories:

Table 9.1. Example directory structure under the main loon monitoring network directory for raw field data files (read-only) and the same files available for processing.

Directory	Purpose
\data\Year \FieldData-ReadOnly	Directory for archiving raw field data. Tagged read-only and never modified. Add all field data files to a compressed (zip) archive and tag it and the directory read-only. If there are multiple datasets from multiple flights create logical subdirectories in the zip archive.
\data\Year \FieldData-Working	Same files as above but these working files may be modified, and processed for importing into the main Small Mammal database. Maintain a change log in a metadata file. If there are multiple datasets from multiple flights create logical subdirectories or a consistent file naming scheme.

Fully processed deliverables and associated metadata from the directories above are integrated into the main Small Mammals Access database, saved to the CAKN network drive and then archived in the IRMA data store during seasonal close-out at https://irma.nps.gov/DataStore/Reference/Profile/2216930.

SOP 9.2 Initial Quality Assurance/Control Procedures

Data verification and data validation are two processes used to ensure data quality assurance and control (QA/QC) (see SOP 8). Data verification checks that data entered into a secondary (e.g., electronic) format matches the source data (iPad/field notes, field data notebooks, mortality

tags), whereas data validation checks that data make sense. Although data entry and verification is conducted by personnel with a general familiarity of the data (e.g., survey observers), the validation process requires in-depth knowledge about the project and data collected (i.e., Project Manager, AKA Principal Investigator). Data editing to resolve errors can occur simultaneously with the initial verification process. The final round(s) of data verification and validation should be performed by the project manager. See SOP 8 for details on proofing and data validation for QA/QC requirements.

The project manager is responsible for overseeing the initial verification of Small Mammal data (both physical and electronic sources). The purpose of verifying data is to reduce collection errors thereby enhancing accuracy of data and subsequent analysis and reporting. Physical data (e.g., field data notebooks, field notes, mortality tags) are verified for accuracy, legibility and completeness immediately after collection and are then compared with or appended to the data collected in Park Observer on the iPads before being uploaded into the small mammal Access database. Data are verified first by the observer while in the field (SOPs 7 and 8), and then by the project manager once the survey is completed (SOPs 8 and 9). The project manager is responsible for resolving any errors as soon as possible according to the procedures in this SOP.

The project manager(s) assembles all data following the surveys. He/she checks data notebooks, field notes, mortality tags and Park Observer data for legibility and completeness or delegates this to a technician who was involved with data collection. A minimum of two people verify and correct records on small mammal observations and metadata collected from each observer. Data from each observer will be compared and any discrepancies discussed with each observer who collected data. These data will then be compared to the Excel file. The project manager and experienced technician validate the data in the Excel file. Small mammal data may be updated in the verification and validation processes for QA/QC (see SOP 8).

SOP 9.3. Importing Field Data to the Master Small Mammal Monitoring IRMA Data Store

Once field data collection operations cease and initial quality assurance procedures are completed the information is transferred to the CAKN Small Mammal Access database to the IRMA Data Store https://irma.nps.gov/DataStore/Reference/Profile/2216930. See SOP 9.8 for more details uploading deliverables into the IRMA Data Store for file based management. Please reference the "Product Archival, Seasonal Closeout and Dataset Certification Procedures for the Arctic and Central Alaska Inventory and Monitoring Networks" (Miller 2017) for more details on certifying final data sets and uploading them to the IRMA Data Store in a file-based management system for archival and public access purposes.

SOP 9.5 Final Quality Assurance/Control Procedures

Primary tools used for assuring the quality of small mammals monitoring data include manual data verification and visualization checks and filtering and sorting to validate data. Analytical and statistical analyses will also reveal data quality deficiencies and trigger another round of data quality assurance checks. Validation and quality control efforts should ideally be completed by

mid-November for data from the previous summer's surveys. Once the data have been validated by a project manager, they can be certified and analyzed.

SOP 9.6 Data Preparation for Analysis

Analysis of small mammal data uses a spatial mark-recapture model presented by Gardner et al. (2009) to estimate the number of individuals, *N*, with activity centers within the sampling area, *S*. Density is defined as *N*/*S*. Analytical methods are detailed in Schmidt et al. (2017) and SOP 10. Validation is also an iterative process with analysis. If the estimates show anomalies or trends that do not make sense (basically following validation of analyzed data), the data will be re-verified, re-validated and re-analyzed. If results are reported from data that have not been fully validated (e.g., preliminary results before validation of the model estimates), a statement and explanation should be included to caution users regarding data quality. Program R and WinBUGS or OpenBUGS are the primary tools for analyzing the small mammal data. Formatted data used in analysis are organized in simple spreadsheets and can be produced from the full dataset spreadsheet.

SOP 9.7 Seasonal Close Out and Dataset Certification

Data certification is a benchmark in the monitoring project data life cycle that indicates that: a) the data are complete for the period of record; b) they have undergone and passed the quality assurance checks; and c) that they are appropriately documented and in a condition for archiving, posting and distribution as appropriate. Please reference the "Product Archival, Seasonal Closeout and Dataset Certification Procedures for the Arctic and Central Alaska Inventory and Monitoring Networks" (Miller 2017) for more details.

Seasonal close out is the best time to update SOPs and version their IRMA references if new data management tools or methods have been developed.

Procedure:

1. Verify the Small Mammal monitoring protocol and SOPs are all up to date and versioned in IRMA.
2. Verify QA/QC procedures have been performed to the satisfaction of the project manager.
3. Verify deliverables are documented and archived in IRMA with metadata (0).
4. Review data products and IRMA references for sensitive information and isolate them appropriately.
5. The Small Mammal monitoring program should have a single dataset reference in IRMA to which the latest certified snapshot of the monitoring database is attached. If this reference does not exist the program or data manager should create it. Ensure it is tagged 'sensitive' and 'operations'. The program and data manager fill out the reference's metadata.
    a. The program manager fills out a data certification form and signs it (Appendix C). This form is scanned and uploaded to the dataset IRMA reference above.

b. A human readable copy (SQL script) of the main sheep monitoring database is uploaded to the loon monitoring program's dataset reference and versioned. This copy is the certified database.

SOP 9.8 Archiving Field Deliverables in IRMA

Field data files should be archived on IRMA in a private Dataset reference with each deliverable from the data deliverables schedule. (See narrative.) Please reference the "Product Archival, Seasonal Closeout and Dataset Certification Procedures for the Arctic and Central Alaska Inventory and Monitoring Networks" (Miller 2017). for more details on certifying final data sets and uploading them to the IRMA Data Store in a file-based management system for archival and public access purposes.

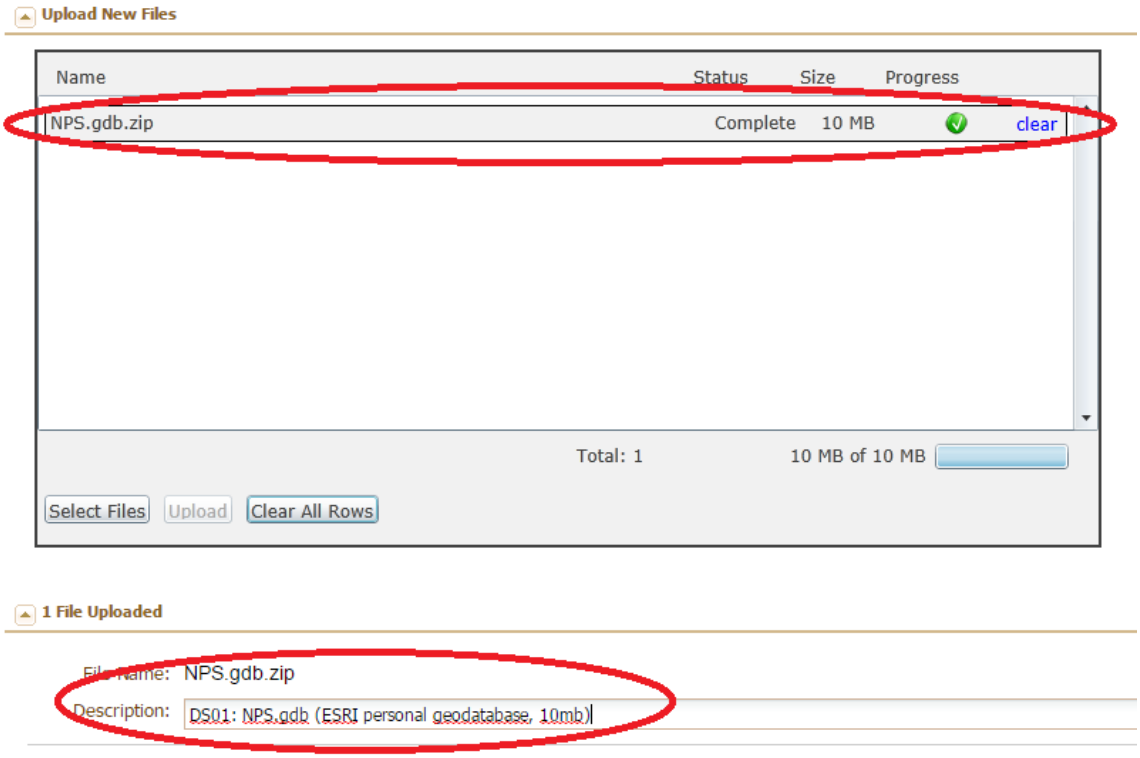
Creating a Field Deliverables Dataset Reference in IRMA

1. Create a new IRMA Dataset Reference and title it informationally similar to '2016 Central Alaska Network Small Mammal Survey Dataset'.
2. Fill out as much metadata as possible.
3. **Ensure the Dataset is tagged as 'Sensitive' under the Information Resource Evaluation tab with the justification of 'Commercial Value'.** Also tick the 'Operations' justification as we don't want raw, unprocessed data available to the public (certified data will be available minus the sensitive data through the main loon monitoring dataset in IRMA).



4. Upload the data deliverables from the Deliverables Schedule above, one by one. Tag them with the date and deliverable ID in the filename and especially the holding

description. An example would be 'DS08: Survey data sheets'.



5. Remove all permissions except for the project manager and data manager (to reduce the risk of accidental publication of sensitive data).
6. Activate the reference.
7. Note the ReferenceCode.
8. Open the main Small Mammal monitoring IRMA reference at: https://irma.nps.gov/DataStore/Reference/Profile/2216930.
9. Navigate to the Products section and click Manage Products.
10. The Manage Products form appears. Use the form's tools and the ReferenceCode noted above to search for, and link to the dataset created above.
11. Save and exit the Manage Products form.
12. Verify the dataset appears as a product of the main sheep monitoring reference.
13. Activate the reference.
14. Enter the IRMA reference code into the Surveys table of the sheep monitoring database for the associated survey.

Your finished reference should be well documented, marked internal, active and sensitive and have a list of holdings that matches the list of data deliverables in the Small Mammal monitoring protocol narrative. If done well the annual closeout procedure will be little more than a glance at this reference (see example from Sheep Protocol below):

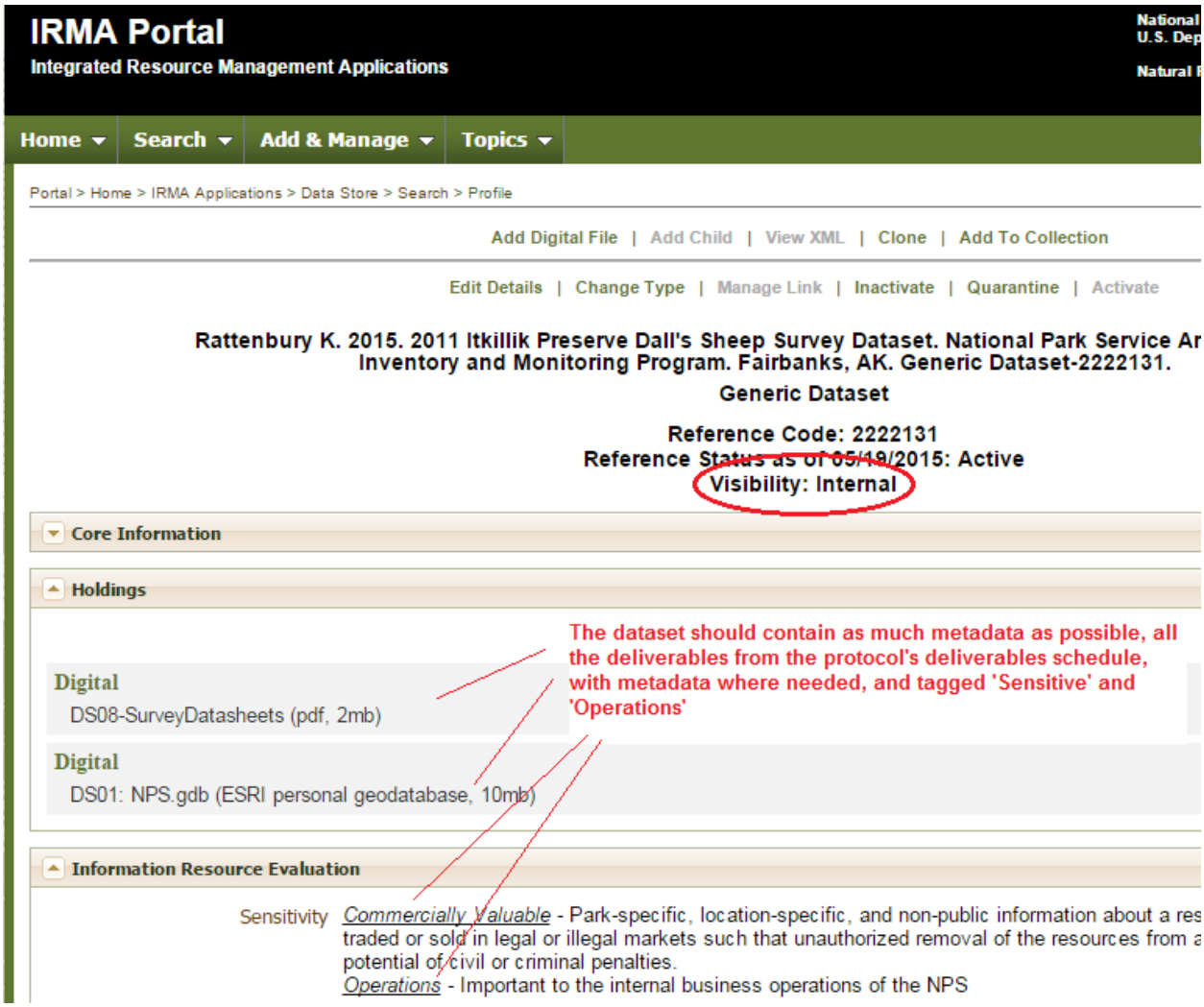


Figure SOP11-9. Example of a sheep survey dataset IRMA reference. The example is incomplete but shows how annual monitoring deliverables should be archived. Having everything listed in one place will simplify annual closeout and data certification procedures.

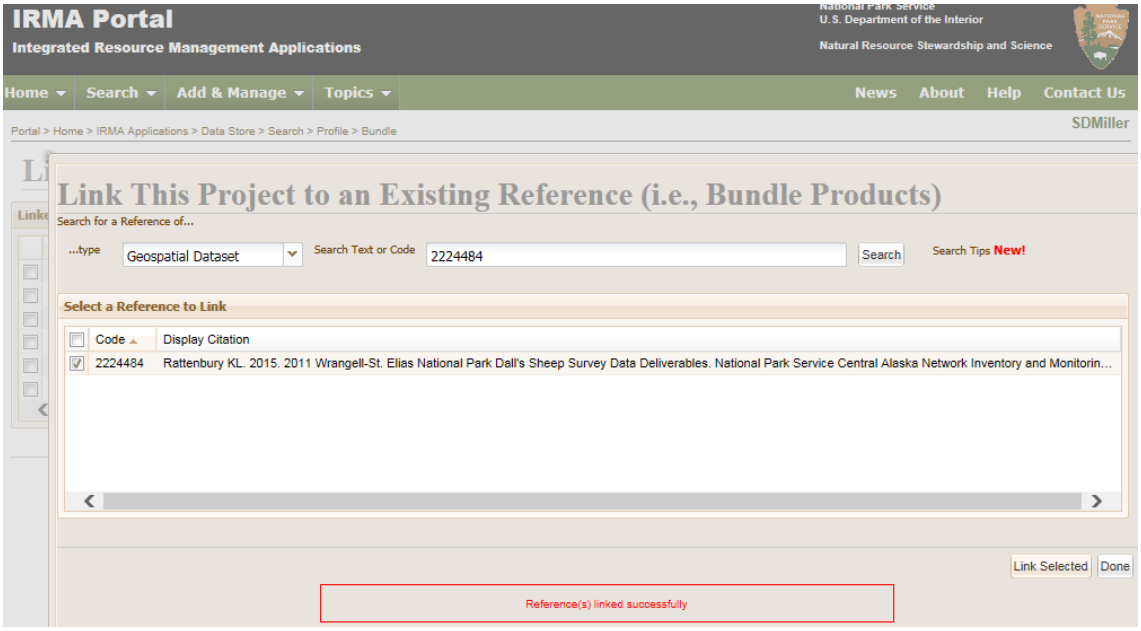
Link the Field Deliverables Reference as Product of the Master Small Mammal Monitoring Project Reference

The seasonal data deliverables Dataset references in the IRMA Data Store must be linked as a product of the master Small Mammal monitoring Project reference in the IRMA Data Store. Please reference the "Product Archival, Seasonal Closeout and Dataset Certification Procedures for the Arctic and Central Alaska Inventory and Monitoring Networks" (Miller 2017). for more details on certifying final data sets and uploading them to the IRMA Data Store in a file-based management system for archival and public access purposes.

Procedure:

1. Using a browser, navigate to the seasonal deliverables Dataset reference created in the section above.

2. Record the reference's reference code.
3. Navigate to the main Small Mammal monitoring IRMA reference at (the example below is for Sheep): https://irma.nps.gov/App/Reference/Profile/2216930.
4. Ensure you have editing privileges.
5. Scroll to the Products section.
6. Click 'Manage Products'
7. Click 'Create New Link'.
8. Search for the reference to attach using the reference type and reference code from above.



9. Check the correct reference and click Link Selected.
10. Click Done and back out of the remaining forms.
11. Check that the deliverables Dataset exists as a product of the master Project reference.
12. The deliverables dataset reference should not appear on the Small Mammal monitoring web profile on the CAKN website because it was tagged 'Operations' and 'Sensitive'. Verify that it does not appear publicly at http://science.nature.nps.gov/im/units/arcn/vitalsign.cfm?vsid=44.

SOP 9.9 Literature Cited

Gardner B, Royle JA, Wegan MT (2009) Hierarchical model for estimating density from DNA mark-recapture studies. Ecology 90:1106-1115.

Miller, S. D. 2017. Product archival, seasonal and dataset certification procedures for the Arctic and Central Alaska Inventory and Monitoring Networks, DMSOP-1. Nature Resource Report NPS/ARCN-CAKN/NRR-2017/xxxx. National Park Service, Fairbanks, Alaska.

Schmidt, J. H., E. A. Rexstad, C. A. Roland, C. L. McIntyre, M. C. MacCluskie, M. J. Flamme. 2017. Weather-driven change in primary productivity explains variation in the amplitude of two herbivore population cycles in a boreal system. *Accepted with revisions, Oecologia*.

Walsh, P., J. Reynolds, G. Collins, B. Russell, M. Winfree, and J. Denton. 2010. Application of a double-observer aerial line-transect method to estimate brown bear population density in southwestern Alaska. *Journal of Fish and Wildlife Management* 1:47-58.

SOP 9.10 Appendix C: Dataset Certification Form

Please reference the "Product Archival, Seasonal Closeout and Dataset Certification Procedures for the Arctic and Central Alaska Inventory and Monitoring Networks" (Miller 2017). for more details on certifying final data sets and uploading them to the IRMA Data Store in a file-based management system for archival and public access purposes. Data certification is a benchmark in the project data life cycle that indicates that:

- The monitoring data are complete, up to date and ready for analysis and publication
- The data have passed quality assurance checks
- The data are appropriately documented
- All aspects of the monitoring program that generated the data, including the protocol, standard operating procedures and data management systems are in order, documented and up to date.

Certification is not intended to imply that the data are completely free of errors or inconsistencies which may or may not have been detected during quality assurance reviews. This certification form must be uploaded to the monitoring dataset's IRMA reference in scanned pdf format.

Certification date:

_____

Certified by:

_____

Title:

_____

_

Affiliation:

_____

Project title:

_____

Range of dates for certified data:

_____

Description and scope of data being certified:

_____

_____

____ This certification refers to data in accompanying files. Check all that apply, and indicate file names to the right:

_____ Database file(s): _____

_____ Spatial data file (s): _____

_____ Geodatabase file(s): _____

_____ Other (specify): _____

Is there any sensitive information in the certified data which may put resources at greater risk if released to the public (e.g., spotted owl nest sites, cave locations, rare plant locations)?

_____ No _____ Yes

Details: _____

_____

_____

_____

9) Description of data processing, quality assurance measures, and results and summary of quality assurance reviews, including details on steps taken to rectify problems encountered during data processing and quality reviews.