Arctic Network
Inventory & Monitoring Program

U.S. Department of the Interior

Data Management
Standard Operating Procedure
NPS/ARCN/DMSOP-2007-5

Data Quality Assurance and Control

Quality assurance/quality control (QA/QC) guidance

Summary

This standard operating procedure provides quality assurance/quality control (QA/QC) guidance when working with Inventory and Monitoring Program information.

Contents

Data Collection

Before the data collection phase of a project begins, the data manager is responsible for providing the protocols/SOPs for data collection and storage to the project manager. Field sheets and field data recording procedures should be reviewed and approved by the data manager and documented in the protocol SOPs. The project manager, in turn, will ensure that field crews understand the procedures and closely follow them in the field. If training is necessary, the data manager will work with the project manager to provide that training. Field technicians are responsible for proofing raw data forms in the field, ensuring their readability and legibility, and verifying and explaining any unusual entries.

They are expected to understand the data collection forms, know how to take measurements, and follow the protocols.

Data Collection Methods

The following methods are suggested to reduce errors during field data collection:

Project-specific data sheets

When electronic data collection devices are not used, data will be recorded on paper data forms. The use of acid-free paper is strongly suggested to prevent fading and subsequent data loss. Some circumstances may warrant the use of paper and writing implements that can withstand moisture, dust, and other extreme environmental conditions. Standardized data sheets that identify the pieces of information to be recorded and display the data for efficient computer entry (i.e., reflect the design of the data entry interface) will help ensure that all relevant information is recorded and subsequent data entry errors are minimized. Data sheets should contain as much basic preprinted project information as possible and sufficient space for recording relevant data such as date, collectors names, weather conditions, etc. They should clearly identify all required information, using examples where needed to ensure that the proper data are recorded. Field personnel should adhere to the following guidelines:

- All information added to the data sheet must be printed and clearly legible
- If alterations to the information are necessary, the original information should be crossed out with a single line and the new information written next to the original entry. Information should never be erased and old information should not be overwritten
- Upon return from the field, original data sheets shall be reviewed and checked for legibility and completenes
- Two copies of the original field forms shall be produced, and each shall be checked to ensure that all original data is present and legible (i.e., no data cut off at the edges)
- Original forms shall be stored, and data entry will be completed using one set of the copies. Data entry notes and remarks, if necessary, shall be made only on the copies, and shall not be made on the original data entry forms
- The second set of copies shall be delivered to the Network Data Manager.

Handheld computers

Handheld computers offer a number of advantages over paper forms, including:

- Screen customization that can incorporate on-the-spot QA/QC checks
- Minimize the need for manual data entry and transcription; and
- Reduce the number of data entry errors.

Along with the advantages, hand-held units have the following negative characteristics

- Vulnerability to environmental contaminants such as heat, dust, and moisture
- Data must be downloaded daily to avoid potential loss of information
- System failures can result in lost data
- Short battery life
- An inoperable unit may necessitate the use of alternate data collection methods.

Data Entry

Though data entry may appear to be a reasonably straight forward process, many data related problems occur during this period. The Network preference is to enter data into a database specifically tailored for the project in question – Spreadsheets make very poor data entry systems, and should be discouraged for all but the most elementary applications. If data is recorded electronically in the field, data entry may in reality be data transfer. Procedures for electronic data transfer will follow the protocol specific standard operating procedures and/or the procedures specific to the equipment being used. For this reason, specific data transfer methodologies will not be discussed here.

The underlying hope is that data entry or data transfer will be 100% accurate. However, despite everyone's best efforts, errors will likely occur. One way to minimize data entry errors is to have the data entered by the person who collected it, and to enter the data shortly after collection. Any errors that do occur will be corrected during the data verification process.

General Procedures

The following measures help reduce errors during data entry:

Timeliness -- All data should be entered or downloaded into the project database as soon as possible, preferably no less than once a week. Do not delay data entry until all the project data have been collected. Downloaded data should be periodically backed up.

Design efficient data entry forms – A data entry application that resembles the field data forms can effectively reduce manual data entry errors due to the 1:1 correspondence of the attributes.

Devise a way to distinguish validated data from newly entered data – Data may be entered into an empty database, a temporary table, or simply flagged to distinguish data that has been checked. Regardless of the strategy, the process for validation must be clearly documented.

Build automated error checking features into the database. Data entry forms can incorporate auto-filled fields, range limits, pick lists, and spelling checks to reduce potential errors. Forms can also be designed to control access to the database (i.e., forms may be set for data entry only, which prevents accidental deletion or alteration of existing data) and control the sequence of data entry (i.e., certain fields require an entry before more information can be entered).

Organize data entry area – The workspace where data will be entered should be free of unrelated material, and should enable data entry staff to organize their forms and equipment. The same guidelines apply to the verification and validation phases.

Have two data entry technicians available for data entry. Although not required, when one technician reads the data off the field data forms and another enters them into the computer, the work is often faster and results in a lower error rate. If only one person is available, he/she should work at a slower pace to avoid errors. Like many monotonous tasks, data entry can be done in a personal rhythm that eases the work for some people.

Specific Procedures

The following specific procedures will help minimize data entry related problems:

1. *Field staff must be familiar with the data forms:* Forms that are new or staff that are not acquainted with existing forms present potential data entry challenges. Prior to introducing new forms, or expecting new staff to use existing forms, the project leader must familiarize field personnel with the data entry forms, and thoroughly explain all aspects of the forms.

2. *Standardized Value Coding:* Establish codes that will be used in specified fields to promote consistency. Only values contained in the code list for a field are acceptable. Null values should be identified with a predetermined value. For example, “N/A” may be used for values that are Not Applicable, and “None” may be used when no items were observed or when a measurement was not taken.

3. *Printing vs. Script or Handwriting:* Handwriting ranges widely in quality and legibility. To minimize difficulties standard “Block” printing is required.

4. *Establish Book-marking Procedures:* Data are entered in one logical set at a time--usually one complete field form. Each paper form should be initialed after its data are entered to avoid confusion about whether it has been processed. A colored pen is good for this purpose. Data entry should be interrupted only at logical stopping points -- identifying acceptable stopping points to prepare for interruptions or the end of the work day. The best stopping point is at the completion of the entry of any single, complete field form rather than in the middle of a logically single operation.

5. *Transcription Notes, Comments, and Questions:* Errors or questions about the data content can be recorded on separate, but associated data entry comment forms. Data entry comment forms must include a mechanism to relate the remarks back to the originating data.

6. *Print a hardcopy of the data:* A copy of all the entered data should be printed for later verification. The print should contain record identifying fields, and should be printed in the same order as it was entered to help reconcile field form and entered data. Verify that

ALL the data were printed (i.e., none is off the right margin) and are readable (font size and attributes).

7. *Data Entry Documentation:* The name of the operator and the date of entry should be indicated on a cover sheet or other suitable location. The identical information should be recorded at the top of the printout of the entered data. The field data and the printout are kept together from this point on for use in data verification.

Verification and Validation Procedures

Data contained in a database must be evaluated by verifying and validating it against the values present on the original field forms. As a rule, fewer problems are identified during the verification and validation process when preceded by effective quality assurance practices. Like the data entry phase, the process of verification and validation should follow a protocol specific Standard Operating Procedure (SOP) manual. Failing to follow SOPs for data entry, validation, and verification may render a dataset suspect.

The data manager and the project leader will work together to establish verification and validation procedures. The project manager will ensure that the procedures are followed.

Verification follows data entry and involves checking the accuracy of the computerized records against the original source, usually hard copy field records, and identifying and correcting any errors. When the computerized data are verified as accurately reflecting the original field data (i.e., the entered data matches the collected data), the paper forms can be archived and most data manipulation can be done on the computer.

Although data may have been correctly transferred from the field forms, the values contained in the dataset may lie outside the acceptable range of values. For example, entries of stream pH of 25.0 or a temperature of 95°C in data files raise doubt about their accuracy; and such entries almost certainly are incorrect, whether or not they were properly transcribed from field forms. The process of reviewing data for range and logic errors is *validation*.

Verification

The effectiveness of the following methods is correlated with effort -- the methods that eliminate the most errors are time consuming and costly, while the simplest and cheapest methods will not be as efficient at detecting errors.

Visual review at data entry – The data entry technician checks each record after it is input. The values recorded in the database are compared with the original values from the hard copy and any errors are corrected immediately. This method is the least complicated since no additional personnel or software is required. The reliability of this method depends wholly on the person keying the data and is generally the least reliable of the data verification methods.

Visual review after data entry – All records are printed following completion of data entry. The values on the printout are compared with the original values from the hard copy. Errors are marked and corrected in a timely manner. The reliability of this technique can be improved by using multiple people to perform the review. For example, one technician reads the original data sheets while the second checks the corresponding value on the data sheet.

Duplicate data entry – All data is entered as normal, then a predetermined number of records are randomly selected and re-entered into a replica of the primary database. The two record sets are compared and errors are corrected. The success of this method increases as the number of randomly selected values is increased.

Summary Statistics -- Simple statistical summaries using the entered data can be used to count the number of sampling sites, plots per site, or dates per sample. Other aspects of the dataset can be explored to find clues to errors. The more checks devised to test the completeness of the data, the greater the confidence that the data are completely verified.

Specific Procedures

In the following outline, a pair of people are working together is assumed. The checker needs a red and green fine-tipped marker for identifying errors and for indicating corrections. The reader also needs a marker (green is good). Straight edges make reading aligned tabular data much easier. A notepad should be kept handy for making notes that may be useful during validation.

1. *Compare entered data with field forms:* Two people work together to compare the entered data with the field form data. On person reads the original data (field forms) while the other individual "verifies" that the entered data matches the original. The three most common errors that will be detected are: duplicated records (entered twice); missing records (inadvertently skipped during entry); and, misspellings (wrong number or code). When an error iis found, the correction to be made is highlighted in red on the printout, NOT on the original data sheets.

2. *Certify verification*: After verifying the data from each field sheet, the reader should date and initial the original field form at the top (or where indicated), stating that verification was done. The reading and checking is continued until all the data sheets in a data set are compared. Thereafter, an original set of data sheets with completion marks (both entry and verification) and a set of printouts with needed corrections marked in red are available.

3. *Correct errors:* Locate problematic errors one-at-a-time and make appropriate corrections. Do not rely on "search and replace" features – they may inadvertently introduce "other" undesirable problems. As each correction is made, a second mark (green) will be added to the red mark on the printout signifying that the record was corrected. When all identified errors are corrected in the computer file, the printout is inspected again for any corrections that were missed (red without green check). Finally, the printout is initialed and dated at the top to indicate that all errors were corrected. The

printout is saved with the original field form to document that this portion of the verification process was completed.

4. *Summary analyses*. Entered data can be summarized to identify duplicate or omitted entries. For example, the number of records, the number of sampling sites, samples per site, or sites per date can be viewed. If the results differ from the known number from the field forms, additional investigation would be necessary. The same question can be posed in different ways; differences in the answer provide clues to errors. The more checks you can be devised to test the completeness of the data, the greater is the confidence that the data are completely verified.

Validation

General step-by-step data validation instructions are not useful because each data set has unique measurement ranges, sampling precision, and accuracy. Nonetheless, validation is a critically important step in the certification of the data, and will be accomplished by the Project Leader after verification is complete. Certain components of data validation can be built into data entry forms (e.g. range limits), while additional data validation can be accomplished during verification. As a general rule, invalid data commonly consist of slightly misspelled species names or site codes, the wrong date, or out-of-range errors.

The following general methods may be used to validate data, though specific procedures for data validation depend upon the specific monitoring protocol.

Data entry application programming – Data entry forms can be designed to reject unacceptable values. The simplest validation during data entry is range checking. •

Outlier Detection –Values that are unusually large or small with respect to the majority of collected data. Database, graphic, and statistical tools can be used for ad-hoc queries and displays of the data. Histograms, line plots, and basic statistics reveal possible logic and range errors. Such exploratory techniques identify obvious outliers, however the challenge in detecting outliers is in deciding how unusual a value must be before it is rejected. Data quality assurance procedures should not aim to eliminate outliers. Eliminating data contamination may be a better way to describe this quality assurance goal. When an outlier is detected, attempts should be made to determine if data contamination is responsible. Some values may appear unusual but prove to be quite valid after thorough investigation.

Catalog the error types found in each data set – When particular validation errors are found, it is important to catalog them in an error log for that data set. Notes on the error(s) should include a description, how detected, and how corrected. Simple, generic errors and more esoteric and cryptic errors must be documented. This list of errors is a valuable reference for the next validation session and ultimately for building formal validation procedures into the data entry process and other automated, post-entry error-checking routines.

Summary Statistics -- Simple statistical summaries can be used to assess the validity of the data. The more thoroughly a dataset is tested the more certain we will be that the data are valid.

Modify field data forms to avoid common mistakes – After validating a dataset, and equipped with a catalog of validation errors and exploratory data results, field data forms and field procedures should be evaluated. Often minor changes, small annotations, or adding check boxes to a field form remove ambiguity about what to enter on the form. In fact, repeated validation errors suggest that the field form --not the field crew-- is usually the cause. Repeated validation errors can also mean that protocol(s) or field training is faulty, which must be recognized and corrected.

Rectify ALL instances of Invalid Data – A vitally important, but easily overlooked validation step involves correcting eroneous, or invalid data at the source. This involves retrieving the original record and any copies to correct and annotate the errors that were found and fixed in the digital files. Without annotating the original field forms, the digital and paper records will not be consistent. If this step is not performed and documentation, *all* data may be rendered suspect.

Review and Communication

QA/QC review is required prior to communicating/disseminating data and information. Only data and information that adhere to NPS quality standards will be released.

Brunt (2000) identifies a series of checks that should be performed during the QA/QC process and associates each with four distinct project segments: design; acquisition; metadata; and, archive. Documentation from these checks coupled with metadata will be used to notify end users, project managers, and network management of data quality. A descriptive document for each data set/database will provide information on the specific QA/QC procedures applied and the results of the review. Descriptive documents or formal FGDC-compliant metadata will document quality for spatial and non-spatial data files posted on the Internet.

Table 1 Quality assurance and quality control procedures that are associated with data design, data acquisition, metadata development and data archival phases in a comprehensive data management system (From Brunt, 2000).

Quality assurance and quality control (QA/QC)	Design	Acquisition	Metadata	Archive
Check that data sheets represent experimental design	X			
Check that measurement units are defined on the data sheet	X			
Check that attribute names meet project standards	X			
Check that date, site, and coded values meet project standards	X			
Check that attribute names and descriptions are provided	X			
Check that data are complete		X		
Check that data entry procedures were followed		X		
Check that data include time, location, and collector(s)		X	X	X
Check that measurement data are within the specified range		X		
Check that data values or codes are represented correctly		X		
Check that data are formatted correctly for further use		X	X	X
Check that data table attribute names are reasonable		X	X	X
Check that data table design reflects experimental design		X	X	X
Check that values for each attribute are represented one way		X	X	X
Check that errors and corrections are recorded		X	X	X
Check that metadata are present			X	X
Check metadata for content (accuracy and completeness)			X	X
Check that data dictionary is present and accurate			X	X
Check that measurement units are consistent		X	X	X
Check that data and metadata are complete				X

Feedback

Quality assurance procedures will need to be periodically reviewed and revised to improve the quality and effectiveness of the QA/QC program. For this reason, quality checks should not be performed with the sole objective of eliminating errors as the results may also prove useful in improving the overall data collection process. The Network intends to evaluate each project by conducting post-project reviews with cooperators, project leaders, Network staff, and others familiar with the project to identify notable project successes as well as elements that could be improved. For projects that are anticipated to continue for long-periods, similar assessments will be performed with cooperators, project leaders, Network staff, and others familiar with the project to identify notable successes as well as elements that warrant a mid-project correction.

Audits and Quality Control

Periodic data audits and quality control checks help ensure that data quality procedures specified are effective and/or are being utilized. Periodic checks may include verification of the following:
Data collection and reporting requirements have been met;

- Data collection and reporting procedures are being followed
- Verification and validation procedures are being followed
- Data file structures and maintenance is clear, accurate and according to plan
- Revision control of program documents and field sheets is adequate
- Calibration and maintenance procedures are being followed
- Seasonal and temporary staff have been trained in data management practice

- Metadata collection and construction for the program is complete; and
- Data is being archived and catalogued appropriately for long term storage.

The results of quality assessments should be documented. The project manager and coordinator are responsible for ensuring that non-conformities in data management practices are corrected.

Data Backup and Archiving

Losing data at any stage of a project is frustrating, but to lose some or all of a data set after it has passed a comprehensive QA/QC process is inexcusable. To prevent this from happening, the following minimal steps shall be followed

- Field forms shall be systematically organized and stored on a regular basis. For example, data may be organized according to sampling date, with all forms from each date combined and sealed in folders
- Working copies of original field forms shall be made shortly following data verification and stored separately from the originals. One set of copies shall be forwarded directly to the Network
- Electronically acquired data shall be downloaded from the collection device on a regular basis, preferably on a daily or weekly basis. Dedicated data loggers will utilize more extended download periods
- Following data verification, copies of electronic data sets shall be burned to CD's or other removable storage media, and a copy of the electronic data shall be forwarded to the Network.

References

Dieffenbach, Fred. 2005. QA / QC and Field Collection Standards. National Park Service. Woodstock, Vermont.

About This Standard Operating Procedure

Version: 1.0
Status: Draft
Publication Date: October 30, 2007
Author(s): Scott D. Miller, Data Manager, Arctic Network Inventory & Monitoring Program.
Abstract: This standard operating procedure provides quality assurance/quality control (QA/QC) guidance when working with Inventory and Monitoring Program information.
Suggested Citation: NPS-ARCN (2007).Data Quality Assurance and Control: Quality assurance/quality control (QA/QC) guidance . NPS/ARCN/DMSOP-2007-5 Version 1.0, Arctic Network-Inventory and Monitoring Program, National Park Service. Fairbanks, Alaska.

Revision History

This standard operating procedure was adapted from Sonoran Desert Network, by Debbie Angel, the Acadia Water Quality Data Collection Protocol, by William Gawley and 'QA / QC and Field Collection Standards' by Fred Dieffenbach, Northeast Temperate Network

Version	Version Date	Revised By	Changes
1.0	20071029	S. Miller	Original

This table reflects changes to this document. Version numbers will be incremented by one (e.g., Version 1.3 to Version 2.0) each time there is a significant change in the process and/or changes are made that affect the interpretation of the data. Version numbers will be incremented after the decimal (e.g., Version 1.6 to Version 1.7…1.10….1.21) when there are changes to grammar, spelling, or formatting, or minor modifications in the process that do not affect the interpretation of data.