

Monroe, S. A., M. Dyer, S. Stumpf, C. Bliss, and C. Parker. 2016. Water quality monitoring protocol for streams and springs in the Southern Colorado Plateau Network. Natural Resource Report NPS/SCPN/NRR—2016/1298. National Park Service, Fort Collins, Colorado.

**Standard Operating Procedure #16: Data Quality Review:
Verification, Validation, and Certification, Version SCPN_
WATER_QUALITY_SOP16_20160901**

Contents

1 Tabular data 3

 1.1 Data verification 3

 1.2 Data validation 4

 1.3 Data certification 4

 1.4 Data modification 6

2 Geospatial data 6

3 Revision history 7

Standard Operating Procedure #16: Data Quality Review: Verification, Validation, and Certification

Version SCPN_Water_Quality_SOP016_20160901				
Revision History Log				
Previous version number	Revision date	Section and paragraph	Change(s) made	Approved by

Only changes in this specific SOP will be logged here. Record the previous version number and date of revision; identify sections and paragraphs where changes were made, who approved the revision, and the new version number. More detailed descriptions of the changes made and the reason for the changes is provided at the end of this SOP.

This Standard Operating Procedure (SOP) describes the procedures for verifying and validating the working project database for the SCPN Water Quality Monitoring Program. Data verification is the process of ensuring that the data entered into the database are true to the data on the field data sheets; data validation is the process of checking data for completeness, structural integrity, and logical consistency. Both quality assurance steps must be done after every field season's data have been entered into the database and before the data can be marked as certified.

To ensure that only quality data are included in reports and other project deliverables, the data certification step is a requirement for all tabular and spatial data. The steps that are required to ensure the data are ready to be certified are different. Both are described below.

1 Tabular data

The data manager and project lead cooperate to ensure the steps of data verification and validation are accomplished, and any errors or irregularities in the data identified, documented, and corrected, when appropriate.

1.1 Data verification

The goal of data verification is to ensure data are accurately transferred from the data sheets to the database. The water quality database application incorporates quality assurance strategies designed to improve data quality. Wherever possible, drop-down list options and value range tests have been incorporated into individual fields within the data tables help to minimize the potential for data entry errors and/or the transcription of erroneously recorded data.

It is still possible for input or transcription

errors to occur, however, so a visual review of the entered data is still considered necessary. However, because there is no way to automate this process, data verification is accomplished by a second person reviewing the data and comparing the database values against those on the field data-sheets. Ideally this task is performed by someone who participated in data collection, or who is familiar with the project and data, but not the same person who initially entered the data.

1.2 Data validation

Once the data is verified, the database will be validated to make sure the data make sense and are clean, complete, and useful. The process is amenable to automation, although a fair amount of human oversight is still required.

The Water Quality database is validated using R-scripts (R Core Team 2014) (Figure 16.1). These scripts first perform a systematic check of each table field looking for missing values. They then summarize the data in a variety of ways (Figures 16.2–16-3) that identify missing data, out of range values, and logical errors in the data that were not

captured by the value range restrictions and automated checks implemented in the database, or the verification process. These scripts also perform such tasks as summarizing the number of samples per park or sample site, the distribution, maximum, and minimum of continuous values, plotting the relationships between correlated values, confirming consistency of linear values, etc.

As errors and missing data are discovered, the data manager and project staff work together to correct and document any changes. Once approved, the annual dataset is marked as certified, and is ready to be used.

1.3 Data certification

Data certification is a benchmark in the project information management process that indicates that: 1) the data are complete for the period of record; 2) they have undergone and passed the quality assurance checks of verification and validation; and 3) they are appropriately documented and in a condition for analysis, archiving, posting and distribution as appropriate. Certification is not intended to imply that the data are completely free of errors or inconsistencies that

Southern Colorado Plateau Inventory and Monitoring Program



2012-2013 Core Parameters Water Quality Reports Validation

Compiled on 2016-02-09 09:51:25 by LMcCoy using R version 3.2.2 (2015-08-14)

- 1 Plotting RITS Along Stream Core Parameters

S02_2012-2013_WQ_CoreParameters_ReportValidation.Rmd

This document validates the completeness and correctness of the core parameters reported in the 2012-2013 water quality reports for BAND, CACH, GLCA, GRCA, and MEVE.

Load the required packages.

```
library(RODBC)
library(ggplot2)
library(dplyr)
library(tidyr)
library(lazyeval)
library(xlsx) #Load the xlsx package
library(htmlwidgets)
library(DT)
options(width=100, scipen=999) #keeps text from wrapping
```

Figure 16.1. Sample R Validation report.

```
datatable(rews, filter="top", options = list(pageLength = 100))
```

Show entries

Search:

	Loc_Name	Period	Metric	Count	Median	Range
	All	All	All	All	All	All
1	BANDCAP01	2009-2011	Water temperature (C)	16	11.5	0.1-23.6
2	BANDCAP02	2009-2011	Water temperature (C)	3	11.8	11.2-18.9
3	BANDRIT01	2009-2011	Water temperature (C)	21	9.1	0.2-21.8
4	BANDRIT02	2009-2011	Water temperature (C)	18	8.1	1.6-17.6
5	BANDRIT03	2009-2011	Water temperature (C)	7	15.3	0.2-18.4
6	CACHBLA01	2009-2011	Water temperature (C)	1	16.3	16.3-16.3
7	CACHCHI01	2009-2011	Water temperature (C)	10	9.2	0-27.7
8	CACHTSA01	2009-2011	Water temperature (C)	15	8.6	0.1-23.4

Figure 16.2. Example of data summary table in R.

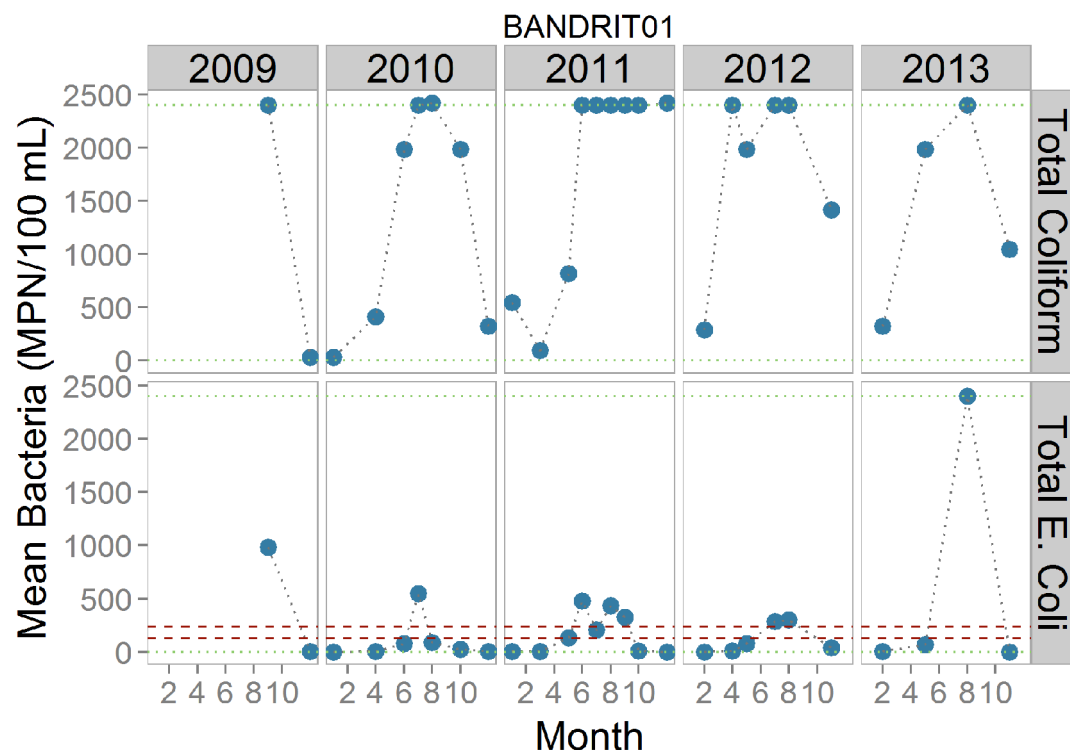


Figure 16.2. Example of a graphical summary of the data in R.

may or may not have been detected during quality assurance reviews.

In collaboration with the IMD Central Office, SCPN is in the process of establishing

specific data certification processes that will be implemented in the near future. As recently announced revisions to the IMD data certification guidance are completed, a more detailed Quality Assurance Plan will be developed for this protocol to include: data quality objectives, required QA/QC procedures, and how QA/QC data will be used to improve overall data quality.

1.4 Data modification

Any changes to the data that result in discrepancies between the values in the database and the data sheet are tracked in tbl_DBChanges and are identified by the table level unique ID. In addition, the master database is under version control, so that any changes made after certification are tracked. This allows any previous changes to be reversed or the state of the data at any point in time to be recreated.

2 Geospatial data

The project lead and GIS specialist work together to review the surveyed coordinates and other geospatial data for accuracy. The purpose of this joint review is to make sure that geospatial data are complete and reasonably accurate, and also to determine which coordinates will be used for subsequent mapping and field work. A certification procedure is being developed for the geospatial data as well.

3 Revision history

Provide the following information for each revision made to SOP #16.

New version:

SCPN_WATER_QUALITY_SOP16_YYYYMMDD

Date:

Author:

Sections affected:

Changes & reasons: