

Project Data Certification Form

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; and 3) that they are appropriately documented and in a condition for archiving, posting and distribution as appropriate. 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.

1) Certification date: July 11, 2016

2) Certified by: Regina Rochefort, Natalya Antonova, and John Boetsch
Title: Science Advisor/Protocol Lead, GIS Specialist, and Data Manager/Ecologist
Affiliation: National Park Service (NOCA/NCCN)

3) Project code: VCa03
Project title: NCCN Prairie Vegetation Monitoring Project

4) Range of dates for certified data: 04/30/15-07/31/15

5) Description and scope of data being certified: Data collected during the 2007-2009, and 2012-2015 field seasons, including annual and alternating panels sampled to date across years.

6) List the parks covered in the certified data set, and provide any park-specific details about this certification.

Park	Details
San Juan Island	Only park for which this protocol applies

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

☒ Database file(s): VCa03 Prairie Veg BE.mdb
☐ Spatial data theme(s): _____
☐ Geodatabase file(s): _____
☐ Other (specify): _____

☐ Certified data are already in the master version of a park, NCCN or NPS database.
Please indicate the database system(s): _____

8) 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 and quality assurance measures. (Note: These can be cut and pasted from appropriate sections of the protocol.)

See the attached certification quality assurance report for a list of the automated validation checks that were performed on the data prior to certification. After the queries were run, the results were reviewed and compared to the original data source. Particular note was taken to the records classified as 'critical' and 'warning'. Inconsistencies in the database were fixed (e.g., sites that were classified as 'proposed' but had data associated with them were changed to 'active' and errors were corrected (e.g., duplicate points). However, not all errors and inconsistencies could be fixed, in which case a description of the resulting

errors and why edits were not made was then documented and included in the metadata and certification report. Once all data checks were run and errors were fixed, none of the queries returned records.

The queries are named and numbered hierarchically so that high-order data—for example from tables on the parent side of a parent-child relationship such as sample locations—are fixed before low-order data (for example, individual species observations). The rationale for this is that one change in a high-order table affects many downstream records, and so proceeding in this fashion is the most efficient way to isolate and treat errors.

One of the first steps taken during the quality assurance reviews included reviewing GPS data as they were processed to confirm that the proper transect was sampled, that segment coordinates were not excessively distant from the transect, and to evaluate for any gaps or missing segments. These checks were done during the season much of the time.

Other steps included reviewing segment sequencing for errors, and to group successive repeats of segments having the same categories. In a few cases, treed sections of the transects were not sampled on the ground but instead had events created post-hoc using aerial imagery to digitize vegetation boundaries.

In addition to these automated checks, the person performing the quality review remained vigilant for errors or omissions that may not have been caught by the automated queries. Another task that cannot be automated is the process of ensuring that all data for the current season are entered into the database.

10) Results and summary of quality assurance reviews, including details on steps taken to rectify problems encountered during data processing and quality reviews.

- Locations sampled out of sequence with their panel - Two transects were unintentionally sampled in 2009 due to a mixup in the field. Both were retained in the data set as they are complete transect events.
- Locations not visited as scheduled - Four annual panel transects were not sampled in 2014: AC.1-16, AC.1-12, AC.1-15, and AC.1-18.
- Events excluded from summaries and analysis - There were five transects in 2014 for which only the north ends of transects were surveyed in order to confirm ground conditions and transect end points, and to determine if these should be surveyed in the field each year or could be digitized from aerial photos. These five events were excluded because only parts of transects were surveyed.
- Multi-day sampling events - Multi-day events for individual transects occurred in several instances across all years. Initially, each visit to a transect on a given date was treated as a separate sampling event. However, to facilitate data grouping for summarization and analysis, a decision to support only a single event per transect per year was made circa 2009, after which an end date was added to the field form. For prior years where there were multiple return visits to a transect in a given year, the event records and their observation data were merged into a single, multi-day event.
- Transect start and end point changes over time – Original GRTS transects were clipped to the park boundary circa 2007. Survey start and end points varied every survey year based on tide conditions and safety considerations. In 2014, after consultation with Shay Howlin (statistician with West Inc.), it was decided that transect lengths should be the same every year; as a consequence, transects were truncated to high tide level for beach end points, or a safe stopping point in cases where the transect traversed a cliff or steep slope. All affected transect start and end points had new GPS positions collected in 2014 and 2015, and start and end segments for previous years were either truncated or extended in GIS to the new start and end points.

- Digitized segments and transect events – During initial surveys, it was discovered that several transects on the west side of the park were entirely under tree cover, resulting in only one segment along their length. We decided that future field surveys of treed portions of such transects would only be conducted if the local Resource Specialist was aware of a storm or wind event in that area that would have resulted in tree fall (such local knowledge being considered a realistic expectation in a small park like SAJH since several hiking trails do cross the transect area). These were not surveyed in the field in the following years, since there were no known disturbance events, but were instead digitized manually in GIS using transect start and end points: AC.1-1 (all years except 2012), AC.3-3 (2008, 2015), AC.4-4 (2009), AC.5-5 (2012). Aerial photos were examined and park staff interviewed to determine if there was any disturbance along these transects in the previous year(s) that would warrant a field survey. Similarly, north ends of other transects that were entirely in trees were not visited after initial surveys and instead digitized to the end of transect in GIS, including: AC.1-2, AC.1-4, AC.1-6, AC.1-9, AC.1-12, AC.1-13, AC.1-15, AC.1-18, AC.2-2, AC.2-5, AC.4-1, AC.5-2, AC.6-1, and AC.6-3.
- Other data gaps and errors –
 - Observers were not recorded for 2007, but were estimated using available information.
 - Phenology observations were not recorded consistently across years (phenology was added to the SOP and field form in 2012).
 - Segment numbering may not always represent the sequence of transect segments because one or more segments may have been added or merged/deleted after the fact, or in some cases the numbering in the field may have intentionally or unintentionally been out of sequence.
 - GPS offsets – If the field crew could not reach a segment start point in the field because of difficult vegetation cover (e.g., shrub thicket) or safety reasons, an offset to the desired location was determined using a range finder and a compass. Offsets were not recorded on the field forms until 2012, but were instead used to modify the recorded position manually. Starting in 2012, offset values were either entered in the GPS while position was being collected or applied in GIS after the fact using field notes. Field coordinates recorded in the field for segments with offsets represent the location where the field crew was prior to offset having been applied.
 - GPS errors - There are several potential sources of GPS error. From 2007 to 2009, an older model of the GPS unit was used that did not have as high horizontal accuracy as newer models. The horizontal error for these units was under 2 meters in most cases, but could be up to 5 meters. Horizontal error for newer units was less than 1 meter in most cases, but could be as large as 2 meters. Proximity to trees produced larger horizontal errors as GPS signal bounced off the tree cover before reaching the GPS receiver. This also resulted in bigger discrepancies between field and processed GPS coordinates. Finally, especially in early years, operator error resulted in survey locations being off the transect because crew often did not zoom the display in far enough while navigating to make sure they were on the transect line.
 - Hours spent was added to the field form in 2012, but this was not consistently recorded across years after that point. Missing were estimated using available information about start/stop times, including that derived from GPS files.
- Protocol changes
 - The SOP was refined in 2012 and again in 2014 to clarify data categories and how to fill in missing or invalid combinations. Most notably, Tree and Shrub segments no longer had a cover class recorded for them (instead this was changed to optional only, with the stipulation that a rationale be given if a cover is noted); during the quality review, cover values entered for these types were moved to the Office_notes field.