[*https://doi.org/10.38750/7654321*](https://doi.org/10.38750/7654321)

# DRR Title

### Data Release Report : get this number from Joe DeVivo

#### Jane Doe,1,2 <https://orcid.org/0000-1111-2222-3333>

1 NPS Inventory and Monitory Division, 1201 Oakridge Dr, Suite 150, Fort Collins, Colorado

2 Managed Business Solutions (MBS), a Sealaska Company, Contractor to the National Park Service, Natural Resource Stewardship and Science Directorate, 1201 Oakridge Dr, Suite 150, Fort Collins, Colorado

#### 24 May, 2023

# Abstract

Abstract Should go here. Multiple Lines are okay; it’ll format correctly. Pay careful attention to non-standard characters, line breaks (), carriage returns, and curly-quotes. You may find it useful to write the abstract in NotePad++ or some other text editor and not a word processor (such as Microsoft Word).

Note that if you need multiple paragraphs or line breaks you can generate them using a combination of backslashes and n’s.

The abstract should succinctly describe the study, the assay(s) performed, the resulting data, and their reuse potential, but should not make any claims regarding new scientific findings. No references are allowed in this section.

# Data Records (required)

## Data Inputs (optional)

If the data package being described was generated based on one or more pre-existing datasets, cite those datasets here.

## Summary of Datasets Created (required)

The Data Records section should be used to explain each data record associated with this work (for instance, a data package), including the DOI indicating where this information is stored, and provide an overview of the data files and their formats. Each external data record should be cited. Below is some sample text:

This DRR describes the data package *Data Package Title* which contains a metadata file and 2 data files. These data were compiled by the National Park Service Biological Resources Division and are available at <https://doi.org/10.57830/12342567> (see Table 1).

**Table 1: Data Package Title: List of data files**

| File Name | Size | Description |
| --- | --- | --- |
| my\_data.csv | 0.8 MB | This is a short description of my\_data.csv (a good guideline is 10 words or less). |
| my\_data2.csv | 10 GB | This is a short description of my\_data2.csv. |

See Appendix for additional notes and examples.

# Data Quality Evaluation (required)

The Data Quality Evaluation section should present any analyses that are needed to support the technical quality of the dataset. This section may be supported by figures and tables, as needed. *This is a required section*; authors must provide information to justify the reliability of their data. Wherever possible & appropriate, data quality evaluation should be presented in the context of data standards and quality control procedures as prescribed in the project’s quality assurance planning documentation.

**Required elements for this section**

*Stock Text to include:*

The data within the data records listed above have been reviewed by staff in the NPS Inventory and Monitoring Division to ensure accuracy, completeness, and consistency with documented data quality standards, as well as for usability and reproducibility. The *Data Package Title* is suitable for its intended use as of the date of processing (2023-05-24).

*Required Table*

**Table 2: Description of data quality flags**

| Flag | Definition | Usage |
| --- | --- | --- |
| A | Accepted | columns ending “\_flag” |
|  |  | unflagged data |
| AE | Accepted, estimated | columns ending “\_flag” |
| P | Provisional | columns ending “\_flag” |
| R | Rejected | columns ending “\_flag” |
| NA | Missing | All data |

**Table 3.1: Summary of data quality flags for the data package**

| A | AE | P | R | % Missing (Mean) | RRU (Mean) | RRU (SD) |
| --- | --- | --- | --- | --- | --- | --- |
| 2968 | 0 | 0 | 32 | 0 | 0.99 | 0.03 |

**Table 3.2: Summary of data quality flags for each data file within the data package**

| File Name1 | A2 | AE | P | R | % Missing (Mean)3 | RRU (Mean)3,4 | RRU (SD)4,5 |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Mini\_BICY\_Veg\_Geography | NA | NA | NA | NA | NA | NA | NA |
| Mini\_BICY\_Veg\_Intercept\_Cleaned | 1469 | 0 | 0 | 31 | 0 | 0.98 | 0.04 |
| Mini\_BICY\_Veg\_Transect\_Cleaned | 1499 | 0 | 0 | 1 | 0 | 1.00 | 0.00 |

**Note:** 1 NAs for a given file indicates no quality controlled data were found. 2 All non-missing data in specified unflagged columns are considered accepted. 3 Means are geometric means. 4 RRU (unweighted relative response) is calculated as the number of accepted (where A and AE are both considered accepted) divided by the total number of observations plus the number of missing observations. 5 SD is standard deviation.

**Table 3.3: Summary of data quality flags for each column**

| File Name1 | Colum2 | A3 | AE | R | P | Total | % Missing | RRU4 |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Mini\_BICY\_Veg\_Geography | NA | NA | NA | NA | NA | NA | NA | NA |
| Mini\_BICY\_Veg\_Intercept\_Cleaned | coordinate | 500 | 0 | 0 | 0 | 500 | 0 | 1.00 |
| Mini\_BICY\_Veg\_Intercept\_Cleaned | taxonomic | 469 | 0 | 31 | 0 | 500 | 0 | 0.94 |
| Mini\_BICY\_Veg\_Intercept\_Cleaned | eventDate | 500 | 0 | 0 | 0 | 500 | 0 | 1.00 |
| Mini\_BICY\_Veg\_Transect\_Cleaned | coordinate | 500 | 0 | 0 | 0 | 500 | 0 | 1.00 |
| Mini\_BICY\_Veg\_Transect\_Cleaned | taxonomic | 499 | 0 | 1 | 0 | 500 | 0 | 1.00 |
| Mini\_BICY\_Veg\_Transect\_Cleaned | eventDate | 500 | 0 | 0 | 0 | 500 | 0 | 1.00 |

**Note:** 1 NAs for a given file indicates no quality controlled data were found. 2 The ’\_flag’ suffix has been omitted from column names for brevity. 3 All non-missing data in specified unflagged columns are considered accepted. 4 RRU (unweighted relative response) is calculated as the number of accepted (where A and AE are both considered accepted) divided by the total number of observations plus the number of missing observations.

Possible content **strongly Suggested to Include**

* Occurrence rates or patterns in data that do not meet established standards or data quality objectives.

Possible content **may include:**

* experiments that support or validate the data-collection procedure(s) (e.g. negative controls, or an analysis of standards to confirm measurement linearity)
* statistical analyses of experimental error and variation
* general discussions of any procedures used to ensure reliable and unbiased data production, such as chain of custody procedures, blinding and randomization, sample tracking systems, etc.
* any other information needed for assessment of technical rigor by reviewers/users

Generally, this **should not include:**

* follow-up experiments aimed at testing or supporting an interpretation of the data
* statistical hypothesis testing (e.g. tests of statistical significance, identifying deferentially expressed genes, trend analysis, etc.)
* exploratory computational analyses like clustering and annotation enrichment (e.g. GO analysis).

# Usage Notes (required)

The Usage Notes should contain brief instructions to assist other researchers with reuse of the data. This may include discussion of software packages (with appropriate citations) that are suitable for analysing the assay data files, suggested downstream processing steps (e.g. normalization, etc.), or tips for integrating or comparing the data records with other datasets. Authors are encouraged to provide code, programs or data-processing workflows if they may help others understand or use the data.

For studies involving privacy or safety controls on public access to the data, this section should describe in detail these controls, including how authors can apply to access the data, what criteria will be used to determine who may access the data, and any limitations on data use.

# Methods

Ideally these methods are identical to the methods listed in the metadata accompanying the data package that the DRR describes.Future versions of this template will pull directly from metadata.

The Methods should cite previous methods under use but also be detailed enough describing data production including experimental design, data acquisition assays, and any computational processing (e.g. normalization, image feature extraction) such that others can understand the methods and processing steps without referring to associated publications. Cite and link to the DataStore reference for the protocol for detailed methods sufficient for reproducing the experiment or observational study. Related methods should be grouped under corresponding subheadings where possible, and methods should be described in enough detail to allow other researchers to interpret the full study.

Specific data inputs and outputs should be explicitly cited in the text and included in the References section below, following the same [Chicago Manual of Style author-date format](https://www.chicagomanualofstyle.org/tools_citationguide/citation-guide-2.html) in text. See the [USGS data citation guidelines](https://www.usgs.gov/data-management/data-citation) for examples of how to cite data in text and in the References section.

Authors are encouraged to consider creating a figure that outlines the experimental workflow(s) used to generate and analyse the data output(s) (Figure 1).



Example general workflow to include in the methods section.

### Data Collection and Sample Processing Methods (optional)

Include a description of field methods and sample processing

### Additional Data Sources (optional)

Provide descriptions (with citations) of other data sources used.

### Data Processing (required if done)

Summarize process and results of any QC processes done that manipulate, change, or qualify data.

### Code Availability (required)

For all studies using custom code in the generation or processing of datasets, a statement must be included indicating whether and how the code can be accessed and any restrictions to access. This section should also include information on the versions of any software used, if relevant, and any specific variables or parameters used to generate, test, or process the current dataset. Actual analytical code should be provided in Appendices.

# References (required)

Provide sufficient information to locate the resource. If the citation has a DOI, include the DOI at the end of the citation, including the <https://doi.org> prefix. If you are citing documents that have unregistered DOIs (such as a data package that you are working on concurrently) still include the DOI. Electronic resources data and data services or web sites should include the date they were accessed.

Include bibliographic information for any works cited (including the data package the DRR is describing) in the above sections, using the standard *NPS NR Publication Series* referencing style.

See the following examples:

### Agency, Company, etc. as Author Examples

Fung Associates Inc. and SWCA Environmental Consultants. 2010. Assessment of natural resources and watershed conditions for Kalaupapa National Historical Park. Natural Resource Report. NPS/NPRC/WRD/NRR—2010/261. National Park Service, Fort Collins, Colorado.

Greater Yellowstone Whitebark Pine Monitoring Working Group. 2014. Monitoring whitebark pine in the Greater Yellowstone Ecosystem: 2013 annual report. Natural Resource Data Series. NPS/GRYN/NRDS—2014/631. National Park Service. Fort Collins, Colorado.

National Park Service (NPS). 2016. State of the park report for Zion National Park. State of the Park Reports. No. 23. National Park Service. Washington, District of Columbia.

U.S. Forest Service (USFS). 1993. ECOMAP. National hierarchical framework of ecological units. U. S. Forest Service, Washington, D.C.

### Traditional Journal Article Examples

Bradbury, J. W., S. L. Vehrencamp, K. E. Clifton, and L. M. Clifton. 1996. The relationship between bite rate and local forage abundance in wild Thompson’s gazelles. Ecology 77:2237–2255. <https://doi.org/10.2307/2265717>

Oakley, K. L., L. P. Thomas, and S. G. Fancy. 2003. Guidelines for long-term monitoring protocols. Wildlife Society Bulletin 31(4):1000–1003.

Sawaya, M. A., T. K. Ruth, S. Creel, J. J. Rotella, J. B. Stetz, H. B. Quigley, and S. T. Kalinowski. 2011. Evaluation of noninvasive genetic sampling methods for cougars in Yellowstone National Park. The Journal of Wildlife Management 75(3):612–622. <https://doi.org/10.1002/jwmg.92>

### Book Example

Harvill, A. M., Jr., T. R. Bradley, C. E. Stevens, T. F. Wieboldt, D. M. E. Ware, D. W. Ogle, and G. W. Ramsey. 1992. Atlas of the Virginia flora, third edition. Virginia Botanical Associates, Farmville, Virginia.

### Book Chapter Examples

McCauly, E. 1984. The estimation of abundance and biomass of zooplankton in samples. Pages 228–265 in J. A. Dowling and F. H. Rigler, editors. A manual on methods for the assessment of secondary productivity in fresh waters. Blackwell Scientific, Oxford, UK.

Watson, P. J. 2004. Of caves and shell mounds in west-central Kentucky. Pages 159–164 in Of caves and shell mounds. The University of Alabama Press, Tuscaloosa, Alabama.

### Published Report Examples

Bass, S., R. E. Gallipeau, Jr., M. Van Stappen, J. Kumer, M. Wessner, S. Petersburg, L. L. Hays, J. Milstone, M. Soukup, M. Fletcher, L. G. Adams, and others. 1988. Highlights of natural resource management 1987. National Park Service, Denver, Colorado.

Holthausen, R. S., M. G. Raphael, K. S. McKelvey, E. D. Forsman, E. E. Starkey, and D. E. Seaman. 1994. The contribution of federal and nonfederal habitats to the persistence of the northern spotted owl on the Olympic Peninsula, Washington. General Technical Report PNW–GTR–352. U.S. Forest Service, Corvallis, Oregon. <https://doi.org/10.2737/PNW-GTR-352>

Jackson, L. L., and L. P. Gough. 1991. Seasonal and spatial biogeochemical trends for chaparral vegetation and soil geochemistry in the Santa Monica Mountains National Recreation Area. U.S. Geological Survey, Denver. Open File Report 91–0005. <https://doi.org/10.3133/ofr915>

### Unpublished Report Examples

Conant, B., and J. I. Hodges. 1995. Western brant population estimates. U.S. Fish and Wildlife Service Unpublished Report, Juneau, Alaska.

Conant, B., and J. F. Voelzer. 2001. Winter waterfowl survey: Mexico west coast and Baja California. U.S. Fish and Wildlife Service Unpublished Report, Juneau, Alaska.

### Thesis/Dissertation Examples

Diong, C. H. 1982. Population and biology of the feral pig (Sus scrofa L) in Kipahulu Valley, Mau’i. Dissertation. University of Hawai’i, Honolulu, Hawai’i.

McTigue, K. M. 1992. Nutrient pulses and herbivory: Integrative control of primary producers in lakes. Thesis. University of Wisconsin, Madison, Wisconsin.

### Conference Proceedings Examples

Gunther, K. A. 1994. Changing problems in bear management: Yellowstone National Park twenty-plus years after the dumps. Ninth International Conference on Bear Research and Management. Missoula, MT, International Association for Bear Research and Management, Bozeman, Montana, February 1992:549–560.

Webb, J. R., and J. N. Galloway. 1991. Potential acidification of streams in Mid-Appalachian Highlands: A problem with generalized assessments. Southern Appalachian Man and Biosphere Conference. Gatlinburg, Tennessee.

### General Internet Examples

Colorado Native Plant Society. 2016. Colorado Native Plant Society website. Available at: <https://conps.org/> (accessed 07 March 2016).

National Park Service (NPS). 2016a. IRMA Portal (Integrated Resource Management Applications) website. Available at: <https://irma.nps.gov> (accessed 07 March 2016),

National Park Service (NPS). 2016b. Natural Resource Publications Management website. Available at: <http://www.nature.nps.gov/publications/nrpm/> (accessed 07 March 2016).

United Sates Fish and Wildlife Service (USFWS). 2016. Endangered Species website. Available at: <http://www.fws.gov/endangered/> (accessed 07 March 2016).

### Online Data Warehouse Sites (sites that allow you see and download data from multiple sources)

National Oceanographic and Atmospheric Association (NOAA). 2016. NOAA National Climatic Data Center website. Available at: <http://www.ncdc.noaa.gov/> (accessed 07 March 2016).

Environmental Protection Agency (EPA). 2016. Storage and Retrieval Data Warehouse website (STORET). Available at: <http://www.epa.gov/storet/> (accessed 07 March 2016).

National Park Service (NPS). 2016c. NPScape Landscape Dynamics Metric Viewer website. Available at: <http://science.nature.nps.gov/im/monitor/npscape/viewer/> (accessed 07 March 2016).

National Park Service (NPS). 2016d. NPSpecies online application. Available at: <https://irma.nps.gov/NPSpecies/> (accessed 07 March 2016).

United States Geologic Survey (USGS). 2016. BioData - Aquatic Bioassessment Data for the Nation. Available at: <https://aquatic.biodata.usgs.gov/> (accessed 07 March 2016).

# Acknowledgements (optional)

The Acknowledgements should contain text acknowledging non-author contributors. Acknowledgements should be brief, and should not include thanks to anonymous referees and editors or effusive comments. Grant or contribution numbers may be acknowledged.

# Appendix A. Code Listing

In most cases, Code listing is not required. If all QA/QC and data manipulations were performed elsewhere, you should cite that code in the methods (and leave the “Listing” code chunk as the default settings: eval=FALSE and echo=FALSE). If you have developed custom scripts, you can add those to DataStore with the reference type “Script” and cite them in the DRR. Some people have developed code to perform QA/QC or data manipulation within the DRR itself. In that case, you must set the “Listing” code chunk to eval=TRUE and echo=TRUE to fully document the QA/QC process.

# The title of your DRR. Should all DRR start with "Data Release Report:"? Should we enforce titles specifically referencing the data package(s) they the report is about?  
title <- "DRR Title"  
  
# Optional and should only be included if publishing to the semi-official DRR series. Contact Joe if you are. If not, leave as NULL  
reportNumber <- ": get this number from Joe DeVivo"  
  
# This should match the Data Store Reference ID for this DRR. Eventually we should be able to pull this directly from the data package metadata.  
DRR\_DSRefID <- 7654321  
  
#Author names and affiliations:  
  
#One way to think of the author information is that you are building a table:  
  
# Author | Affiliation | ORCID  
# Jane | Institute 1 | 0000-1111-2222-3333  
# Jane | Institute 2 | 0000-1111-2222-3333  
# John | Institute 2 | NA  
  
#once the table is built, authors can be associated with the appropriate institute via relevant superscripts and the institutes can be listed only once in the DRR.  
  
# list the authors. If an author has multiple institutional affiliations, you must list the author multiple times. In this example, Jane Doe is listed twice because she has two affiliations.  
authorNames <- c(  
 "Jane Doe",  
 "Jane Doe"#,  
# "John Doe"  
)  
  
# List author affiliations. The order of author affiliations must match the order of the authors in AuthorNames. If an author has multiple affiliations, the author must be listed 2 (or more) times under authorNames (above) and each affiliation should be listed in order. If authors share the same affiliation, the affiliation should be listed once for each author. In this case, Managed Business Solutions (MBS) is listed twice because it is associated with two authors. MBS will only print to the DRR once.  
  
#Note that the entirety of each affiliation is enclosed in quotations. Do not worry about indentation or word wrapping.  
authorAffiliations <- c(  
 "NPS Inventory and Monitory Division, 1201 Oakridge Dr, Suite 150, Fort Collins, Colorado",   
 "Managed Business Solutions (MBS), a Sealaska Company, Contractor to the National Park Service, Natural Resource Stewardship and Science Directorate, 1201 Oakridge Dr, Suite 150, Fort Collins, Colorado"#,  
# "Managed Business Solutions (MBS), a Sealaska Company, Contractor to the National Park Service, Natural Resource Stewardship and Science Directorate, 1201 Oakridge Dr, Suite 150, Fort Collins, Colorado"  
)  
  
# List the ORCID iDs for each author in the format "xxxx-xxxx-xxxx-xxxx". If an author does not have an ORCID iD, specify NA (no quotes). If an author is listed more than once (for instance because they have multiple institutional affiliations), the ORCID iD must also be listed more than once. For more information on ORCID iDs and to register an ORCID iD, see https://www.orcid.org.   
  
#The order of the ORCID iDs must match the order of authors in AuthorNames.In this example, Jane Doe has an ORCID iD but John Doe does not. Jane's ORCID iD is listed twice because she her name is listed twice in authorNames(because she has two authorAffiliations).  
authorORCID <- c(  
 "0000-1111-2222-3333", "0000-1111-2222-3333"#, NA)  
 )  
# Replace the text below with your abstract.  
DRRabstract <- "Abstract Should go here. Multiple Lines are okay; it'll format correctly. Pay careful attention to non-standard characters, line breaks (<br>), carriage returns, and curly-quotes. You may find it useful to write the abstract in NotePad++ or some other text editor and not a word processor (such as Microsoft Word).\n\n  
  
Note that if you need multiple paragraphs or line breaks you can generate them using a combination of backslashes and n's. \n\n  
  
The abstract should succinctly describe the study, the assay(s) performed, the resulting data, and their reuse potential, but should not make any claims regarding new scientific findings. No references are allowed in this section."  
  
# DataStore reference ID for the data package associated with this report. You must have at least one data package.Eventually, we will automate importing much of this information from metadata.  
dataPackageRefID <- c(12342567)  
  
# Must match title in DataStore and metadata  
dataPackageTitle <- "Data Package Title"  
  
# Must match descriptions in the data package metadata  
dataPackageDescription <- "Short title for data package1"  
  
# generates your data package DOI based on the data package DataStore reference ID. This is different from the DRR DOI! No need to edit this.  
dataPackageDOI <- paste0("https://doi.org/10.57830/", dataPackageRefID)  
  
# list the file names in your data package. Do NOT include metadata files.  
dataPackage\_fileNames <- c(  
 "my\_data.csv",  
 "my\_data2.csv"  
)  
  
# list the approximate size of each data file. Make sure the order corresponds to the order of of the file names in dataPackage\_fileNames  
dataPackage\_fileSizes <- c("0.8 MB", "10 GB")  
  
# list a short, one-line description of each data file. Descriptions must be in the same order as the filenames.  
dataPackage\_fileDescript <- c(  
 "This is a short description of my\_data.csv (a good guideline is 10 words or less).",  
 "This is a short description of my\_data2.csv.")  
RRpackages <- c("markdown",  
 "rmarkdown",  
 "pander",  
 "knitr",  
 "yaml",  
 "kableExtra",  
 "devtools",  
 "tidyverse")  
  
inst <- RRpackages %in% installed.packages()  
if (length(RRpackages[!inst]) > 0) {  
 install.packages(RRpackages[!inst], dep = TRUE, repos = "https://cloud.r-project.org")  
}  
lapply(RRpackages, library, character.only = TRUE)

## [[1]]  
## [1] "QCkit" "lubridate" "forcats" "stringr" "dplyr"   
## [6] "purrr" "readr" "tidyr" "tibble" "ggplot2"   
## [11] "tidyverse" "devtools" "usethis" "kableExtra" "yaml"   
## [16] "knitr" "pander" "rmarkdown" "markdown" "stats"   
## [21] "graphics" "grDevices" "utils" "datasets" "methods"   
## [26] "base"   
##   
## [[2]]  
## [1] "QCkit" "lubridate" "forcats" "stringr" "dplyr"   
## [6] "purrr" "readr" "tidyr" "tibble" "ggplot2"   
## [11] "tidyverse" "devtools" "usethis" "kableExtra" "yaml"   
## [16] "knitr" "pander" "rmarkdown" "markdown" "stats"   
## [21] "graphics" "grDevices" "utils" "datasets" "methods"   
## [26] "base"   
##   
## [[3]]  
## [1] "QCkit" "lubridate" "forcats" "stringr" "dplyr"   
## [6] "purrr" "readr" "tidyr" "tibble" "ggplot2"   
## [11] "tidyverse" "devtools" "usethis" "kableExtra" "yaml"   
## [16] "knitr" "pander" "rmarkdown" "markdown" "stats"   
## [21] "graphics" "grDevices" "utils" "datasets" "methods"   
## [26] "base"   
##   
## [[4]]  
## [1] "QCkit" "lubridate" "forcats" "stringr" "dplyr"   
## [6] "purrr" "readr" "tidyr" "tibble" "ggplot2"   
## [11] "tidyverse" "devtools" "usethis" "kableExtra" "yaml"   
## [16] "knitr" "pander" "rmarkdown" "markdown" "stats"   
## [21] "graphics" "grDevices" "utils" "datasets" "methods"   
## [26] "base"   
##   
## [[5]]  
## [1] "QCkit" "lubridate" "forcats" "stringr" "dplyr"   
## [6] "purrr" "readr" "tidyr" "tibble" "ggplot2"   
## [11] "tidyverse" "devtools" "usethis" "kableExtra" "yaml"   
## [16] "knitr" "pander" "rmarkdown" "markdown" "stats"   
## [21] "graphics" "grDevices" "utils" "datasets" "methods"   
## [26] "base"   
##   
## [[6]]  
## [1] "QCkit" "lubridate" "forcats" "stringr" "dplyr"   
## [6] "purrr" "readr" "tidyr" "tibble" "ggplot2"   
## [11] "tidyverse" "devtools" "usethis" "kableExtra" "yaml"   
## [16] "knitr" "pander" "rmarkdown" "markdown" "stats"   
## [21] "graphics" "grDevices" "utils" "datasets" "methods"   
## [26] "base"   
##   
## [[7]]  
## [1] "QCkit" "lubridate" "forcats" "stringr" "dplyr"   
## [6] "purrr" "readr" "tidyr" "tibble" "ggplot2"   
## [11] "tidyverse" "devtools" "usethis" "kableExtra" "yaml"   
## [16] "knitr" "pander" "rmarkdown" "markdown" "stats"   
## [21] "graphics" "grDevices" "utils" "datasets" "methods"   
## [26] "base"   
##   
## [[8]]  
## [1] "QCkit" "lubridate" "forcats" "stringr" "dplyr"   
## [6] "purrr" "readr" "tidyr" "tibble" "ggplot2"   
## [11] "tidyverse" "devtools" "usethis" "kableExtra" "yaml"   
## [16] "knitr" "pander" "rmarkdown" "markdown" "stats"   
## [21] "graphics" "grDevices" "utils" "datasets" "methods"   
## [26] "base"

devtools::install\_github("nationalparkservice/QCkit")

## Skipping install of 'QCkit' from a github remote, the SHA1 (23478752) has not changed since last install.  
## Use `force = TRUE` to force installation

library(QCkit)  
date <- format(Sys.time(), "%d %B, %Y")  
cat("#", title, "\n")

## # DRR Title

if (!is.null(reportNumber)) {  
 subtitle <- paste0("Data Release Report ", reportNumber)  
 cat("###", subtitle)  
}

## ### Data Release Report : get this number from Joe DeVivo

author\_list <- data.frame(authorNames, authorAffiliations, authorORCID)  
unique\_authors <- author\_list %>% distinct(authorNames,  
 .keep\_all = TRUE)  
unique\_affiliation <- author\_list %>% distinct(authorAffiliations,  
 .keep\_all = TRUE)  
  
#single author documents:  
if(length(seq\_along(unique\_authors$authorNames)) == 1){  
 for (i in seq\_along(unique\_authors$authorNames)) {  
 curr <- unique\_authors[i, ]  
   
 #find all author affiliations  
 aff <- author\_list[which(authorNames == curr$authorNames),]  
 aff <- aff$authorAffiliations   
   
 #identify order of affiliation(s) in a unique list of affiliations  
 #build the superscripts for author affiliations  
 super\_script <- unique\_affiliation$authorAffiliations %in% aff   
 super <- which(super\_script == TRUE)  
 script <- super  
   
 if(length(seq\_along(super)) > 1){  
 script <- NULL  
 j <- 1  
 while(j < length(seq\_along(super))){  
 script <- append(script, paste0(super[j],","))  
 j <- j+1  
 }  
 if(j == length(seq\_along(super))){  
 script <- append(script, super[j])  
 }  
 }  
 }  
 cat("#### ", curr$authorNames, ",", "^",script,"^", " ", " ", sep="")  
 if (is.na(curr$authorORCID)) {  
 }  
 if (!is.na(curr$authorORCID)) {  
 orc <- paste0("https://orcid.org/", curr$authorORCID, " ")  
 cat({{ orc }})  
 }  
   
 #cat("#### ", unique\_authors$authorNames, "^1^", sep="")  
 #if(!is.na(authorORCID)){  
 # orc <- paste0(" https://orcid.org/", unique\_authors$authorORCID)  
 # cat({{ orc }}, "\n")  
 #}  
 #cat("#### ", unique\_authors$authorAffiliations, sep="")  
}

## #### Jane Doe,^1,2^ https://orcid.org/0000-1111-2222-3333

#multi author documents:  
if(length(seq\_along(unique\_authors$authorNames)) > 1){  
 for (i in seq\_along(unique\_authors$authorNames)) {  
 curr <- unique\_authors[i, ]  
   
 #find all author affiliations  
 aff <- author\_list[which(authorNames == curr$authorNames),]  
 aff <- aff$authorAffiliations   
   
 #identify order of affiliation(s) in a unique list of affiliations  
 #build the superscripts for author affiliations  
 super\_script <- unique\_affiliation$authorAffiliations %in% aff   
 super <- which(super\_script == TRUE)  
 script <- super  
   
 if(length(seq\_along(super)) > 1){  
 script <- NULL  
 j <- 1  
 while(j < length(seq\_along(super))){  
 script <- append(script, paste0(super[j],","))  
 j <- j+1  
 }  
 if(j == length(seq\_along(super))){  
 script <- append(script, super[j])  
 }  
 }  
   
 # if NOT the second-to-last author:  
 if(i < (length(seq\_along(unique\_authors$authorNames)) - 1)){  
 cat("#### ", curr$authorNames, ",", "^",script,"^", " ", " ", sep="")  
 if (is.na(curr$authorORCID)) {  
 }  
 if (!is.na(curr$authorORCID)) {  
 orc <- paste0("https://orcid.org/", curr$authorORCID, " ")  
 cat({{ orc }})  
 }  
 }  
   
 # if IS the second-to-last author  
 if(i == (length(seq\_along(unique\_authors$authorNames)) - 1)){   
   
 #if 3 or more authors, include a comma before the "and":  
 if(length(seq\_along(unique\_authors$authorNames)) > 2){  
 cat(curr$authorNames, ",","^",script,"^ ", sep="")  
 if (is.na(curr$authorORCID)) {  
 }  
 if (!is.na(curr$authorORCID)) {  
 orc <- paste0("https://orcid.org/", curr$authorORCID, " ")  
 cat({{ orc }})  
 }  
 cat("and ", sep="")  
 }  
   
 #If only 2 authors, omit comma before "and":  
 if(length(seq\_along(unique\_authors$authorNames)) == 2){  
 cat("#### ", curr$authorNames, "^",script,"^ ", sep="")  
 if (is.na(curr$authorORCID)) {  
 }  
 if (!is.na(curr$authorORCID)) {  
 orc <- paste0("https://orcid.org/", curr$authorORCID, " ")  
 cat({{ orc }})  
 }  
 cat("and ", sep="")  
 }  
 }  
   
 # if IS the Last author :  
 if(i == length(seq\_along(unique\_authors$authorNames))){  
 cat(curr$authorNames, "^", script, "^", sep="")  
 if (is.na(curr$authorORCID)) {  
 }  
 if (!is.na(curr$authorORCID)) {  
 orc <- paste0(" https://orcid.org/", curr$authorORCID, " ")  
 cat({{ orc }})  
 }  
 }  
 }  
}  
cat("\n\n")

for(i in 1:nrow(unique\_affiliation)){  
 cat("^",i,"^ ", unique\_affiliation[i,2], "\n\n", sep="")  
 }

## ^1^ NPS Inventory and Monitory Division, 1201 Oakridge Dr, Suite 150, Fort Collins, Colorado  
##   
## ^2^ Managed Business Solutions (MBS), a Sealaska Company, Contractor to the National Park Service, Natural Resource Stewardship and Science Directorate, 1201 Oakridge Dr, Suite 150, Fort Collins, Colorado

filelist <- data.frame(dataPackage\_fileNames, dataPackage\_fileSizes, dataPackage\_fileDescript)  
  
knitr::kable(filelist, caption = paste0("\*\*Table 1: ", dataPackageTitle, ": List of data files\*\*"), col.names = c("File Name", "Size", "Description"), format = "pandoc")

**Table 1: Data Package Title: List of data files**

| File Name | Size | Description |
| --- | --- | --- |
| my\_data.csv | 0.8 MB | This is a short description of my\_data.csv (a good guideline is 10 words or less). |
| my\_data2.csv | 10 GB | This is a short description of my\_data2.csv. |

#Generates a table with definitions for the data flags, A, AE, P, or R. Honestly not super useful and can probably be turned off (set eval=FALSE) if these are defined elsewhere in text.  
flags<-c("A"," ", "AE", "P", "R", "NA")  
def<-c("Accepted"," ", "Accepted, estimated", "Provisional", "Rejected", "Missing")  
app<-c("columns ending \"\_flag\"", "unflagged data",   
 rep("columns ending \"\_flag\"", 3), "All data")  
data\_flags<-data.frame(flags, def, app)  
  
knitr::kable(data\_flags, caption = "\*\*Table 2: Description of data quality flags\*\*", col.names=c("Flag", "Definition", "Usage"), format="pandoc")

**Table 2: Description of data quality flags**

| Flag | Definition | Usage |
| --- | --- | --- |
| A | Accepted | columns ending “\_flag” |
|  |  | unflagged data |
| AE | Accepted, estimated | columns ending “\_flag” |
| P | Provisional | columns ending “\_flag” |
| R | Rejected | columns ending “\_flag” |
| NA | Missing | All data |

# To turn off, set eval=FALSE.  
# Generates a table with a single row summarizing data quality across all flagged columns of the entire data package. To add additional non-flagged columns, specify them with column names: cols=("my\_unflagged\_data1", "my\_unflagged\_data2)" or column numbers: cols=c(1:4). All non-missing data in unflagged columns is assumed accepted.   
  
#if your package is not in your working directory, you need to specify the directory:  
path<-"../../DataPackageWorkflow/dataPackages/BICY\_Veg\_Data\_Package\_Example"  
dp\_flags <- QCkit::get\_custom\_flags(directory = path, output="package")  
  
#generate table:  
kableExtra::kbl(dp\_flags, caption = '\*\*Table 3.1: Summary of data quality flags for the data package\*\*', col.names = c("A", "AE", "P", "R", "% Missing (Mean)", "RRU (Mean)", "RRU (SD)"), format = "pandoc", digits=2)

**Table 3.1: Summary of data quality flags for the data package**

| A | AE | P | R | % Missing (Mean) | RRU (Mean) | RRU (SD) |
| --- | --- | --- | --- | --- | --- | --- |
| 2968 | 0 | 0 | 32 | 0 | 0.99 | 0.03 |

# To turn off, set eval=FALSE.  
# Generates a table summarizing data quality across all flagged columns of each data file. To add additional non-flagged columns, specify them with column names: cols=("my\_unflagged\_data1", "my\_unflagged\_data2)" or numbers: cols=c(1:4). All non-missing data in unflagged columns is assumed accepted. If a file has no flagged columns and no specified custom columns, all values for that data file will be listed as "NA".  
  
#set directory to the location of your data package  
path<-"../../DataPackageWorkflow/dataPackages/BICY\_Veg\_Data\_Package\_Example"  
df\_flags <- QCkit::get\_custom\_flags(directory = path, output="files")  
df\_flags$filename <- gsub(".csv", "", df\_flags$filename)  
colnames(df\_flags) <- c("File Name", "A", "AE", "P", "R", "% Missing (Mean)", "RRU (Mean)", "RRU (SD)")  
colnames(df\_flags)[1] <- paste0("File Name", "^1^")  
colnames(df\_flags)[2] <- paste0("A", "^2^")  
colnames(df\_flags)[6] <- paste0("% Missing (Mean)", "^3^")  
colnames(df\_flags)[7] <- paste0("RRU (Mean)", "^3,4^")  
colnames(df\_flags)[8] <- paste0("RRU (SD)", "^4,5^")  
  
#Generate the table  
kableExtra::kable(df\_flags, caption = '\*\*Table 3.2: Summary of data quality flags for each data file within the data package\*\*', format = "pandoc", digits=2) %>%  
add\_footnote(c("NAs for a given file indicates no quality controlled data were found.",  
 "All non-missing data in specified unflagged columns are considered accepted.",  
 "Means are geometric means.",  
 "RRU (unweighted relative response) is calculated as the number of accepted (where A and AE are both considered accepted) divided by the total number of observations plus the number of missing observations.",   
 "SD is standard deviation."), notation = "number")

**Table 3.2: Summary of data quality flags for each data file within the data package**

| File Name1 | A2 | AE | P | R | % Missing (Mean)3 | RRU (Mean)3,4 | RRU (SD)4,5 |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Mini\_BICY\_Veg\_Geography | NA | NA | NA | NA | NA | NA | NA |
| Mini\_BICY\_Veg\_Intercept\_Cleaned | 1469 | 0 | 0 | 31 | 0 | 0.98 | 0.04 |
| Mini\_BICY\_Veg\_Transect\_Cleaned | 1499 | 0 | 0 | 1 | 0 | 1.00 | 0.00 |

**Note:** 1 NAs for a given file indicates no quality controlled data were found. 2 All non-missing data in specified unflagged columns are considered accepted. 3 Means are geometric means. 4 RRU (unweighted relative response) is calculated as the number of accepted (where A and AE are both considered accepted) divided by the total number of observations plus the number of missing observations. 5 SD is standard deviation.

# To turn off, set eval=FALSE.  
# Generates a table summarizing QC at the column level within each file. All flagged columns are included. To add additional non-flagged columns, specify them with column names: cols=("my\_unflagged\_data1", "my\_unflagged\_data2)" or numbers: cols=c(1:4). All non-missing data in unflagged columns is assumed accepted. If a file has no flagged columns and no specified custom columns, all values for that data file will be listed as "NA".  
  
#set directory to the location of your data package:  
path <- "../../DataPackageWorkflow/dataPackages/BICY\_Veg\_Data\_Package\_Example"  
dc\_flags <- QCkit::get\_custom\_flags(directory = path, output="columns")  
dc\_flags$filename <- gsub(".csv", "", dc\_flags$filename)  
dc\_flags$column <- gsub("\_flag", "", dc\_flags$column)  
colnames(dc\_flags)<-c("File Name", "Column", "A", "AE", "R", "P", "Total", "% Missing", "RRU")  
colnames(dc\_flags)[1]<-paste0("File Name", "^1^")  
colnames(dc\_flags)[2]<-paste0("Colum", "^2^")  
colnames(dc\_flags)[3]<-paste0("A", "^3^")  
colnames(dc\_flags)[9]<-paste0("RRU", "^4^")  
  
#Generate the table:  
kableExtra::kable(dc\_flags, caption = '\*\*Table 3.3: Summary of data quality flags for each column\*\*', col.names = , format = "pandoc", digits=2) %>%  
add\_footnote(c("NAs for a given file indicates no quality controlled data were found.",  
 "The '\_flag' suffix has been omitted from column names for brevity.",  
 "All non-missing data in specified unflagged columns are considered accepted.",  
 "RRU (unweighted relative response) is calculated as the number of accepted (where A and AE are both considered accepted) divided by the total number of observations plus the number of missing observations."), notation = "number")

**Table 3.3: Summary of data quality flags for each column**

| File Name1 | Colum2 | A3 | AE | R | P | Total | % Missing | RRU4 |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Mini\_BICY\_Veg\_Geography | NA | NA | NA | NA | NA | NA | NA | NA |
| Mini\_BICY\_Veg\_Intercept\_Cleaned | coordinate | 500 | 0 | 0 | 0 | 500 | 0 | 1.00 |
| Mini\_BICY\_Veg\_Intercept\_Cleaned | taxonomic | 469 | 0 | 31 | 0 | 500 | 0 | 0.94 |
| Mini\_BICY\_Veg\_Intercept\_Cleaned | eventDate | 500 | 0 | 0 | 0 | 500 | 0 | 1.00 |
| Mini\_BICY\_Veg\_Transect\_Cleaned | coordinate | 500 | 0 | 0 | 0 | 500 | 0 | 1.00 |
| Mini\_BICY\_Veg\_Transect\_Cleaned | taxonomic | 499 | 0 | 1 | 0 | 500 | 0 | 1.00 |
| Mini\_BICY\_Veg\_Transect\_Cleaned | eventDate | 500 | 0 | 0 | 0 | 500 | 0 | 1.00 |

**Note:** 1 NAs for a given file indicates no quality controlled data were found. 2 The ’\_flag’ suffix has been omitted from column names for brevity. 3 All non-missing data in specified unflagged columns are considered accepted. 4 RRU (unweighted relative response) is calculated as the number of accepted (where A and AE are both considered accepted) divided by the total number of observations plus the number of missing observations.

include\_graphics("vignettes/common/ProcessingWorkflow.png")



sessionInfo()

## R version 4.2.3 (2023-03-15 ucrt)  
## Platform: x86\_64-w64-mingw32/x64 (64-bit)  
## Running under: Windows 10 x64 (build 19045)  
##   
## Matrix products: default  
##   
## locale:  
## [1] LC\_COLLATE=English\_United States.utf8   
## [2] LC\_CTYPE=English\_United States.utf8   
## [3] LC\_MONETARY=English\_United States.utf8  
## [4] LC\_NUMERIC=C   
## [5] LC\_TIME=English\_United States.utf8   
##   
## attached base packages:  
## [1] stats graphics grDevices utils datasets methods base   
##   
## other attached packages:  
## [1] QCkit\_0.1.0.4 lubridate\_1.9.2 forcats\_1.0.0 stringr\_1.5.0   
## [5] dplyr\_1.1.0 purrr\_1.0.1 readr\_2.1.4 tidyr\_1.3.0   
## [9] tibble\_3.1.8 ggplot2\_3.4.1 tidyverse\_2.0.0 devtools\_2.4.5   
## [13] usethis\_2.1.6 kableExtra\_1.3.4 yaml\_2.3.7 knitr\_1.42   
## [17] pander\_0.6.5 rmarkdown\_2.20 markdown\_1.2   
##   
## loaded via a namespace (and not attached):  
## [1] httr\_1.4.4 pkgload\_1.3.1 bit64\_4.0.5 vroom\_1.6.1   
## [5] viridisLite\_0.4.1 shiny\_1.7.2 highr\_0.10 remotes\_2.4.2   
## [9] sessioninfo\_1.2.2 pillar\_1.8.1 glue\_1.6.2 digest\_0.6.29   
## [13] promises\_1.2.0.1 rvest\_1.0.3 colorspace\_2.0-3 plyr\_1.8.8   
## [17] htmltools\_0.5.4 httpuv\_1.6.6 pkgconfig\_2.0.3 xtable\_1.8-4   
## [21] scales\_1.2.1 webshot\_0.5.4 processx\_3.7.0 svglite\_2.1.0   
## [25] later\_1.3.0 tzdb\_0.3.0 timechange\_0.2.0 generics\_0.1.3   
## [29] ellipsis\_0.3.2 cachem\_1.0.6 withr\_2.5.0 cli\_3.4.1   
## [33] magrittr\_2.0.3 crayon\_1.5.2 mime\_0.12 memoise\_2.0.1   
## [37] evaluate\_0.20 ps\_1.7.1 fs\_1.5.2 fansi\_1.0.3   
## [41] xml2\_1.3.3 pkgbuild\_1.4.0 profvis\_0.3.7 tools\_4.2.3   
## [45] prettyunits\_1.1.1 hms\_1.1.2 lifecycle\_1.0.3 munsell\_0.5.0   
## [49] callr\_3.7.2 compiler\_4.2.3 systemfonts\_1.0.4 rlang\_1.0.6   
## [53] grid\_4.2.3 rstudioapi\_0.14 htmlwidgets\_1.5.4 miniUI\_0.1.1.1   
## [57] gtable\_0.3.1 curl\_4.3.3 R6\_2.5.1 bit\_4.0.5   
## [61] fastmap\_1.1.0 utf8\_1.2.2 stringi\_1.7.8 parallel\_4.2.3   
## [65] Rcpp\_1.0.9 png\_0.1-7 vctrs\_0.5.2 tidyselect\_1.2.0   
## [69] xfun\_0.36 urlchecker\_1.0.1

Sys.time()

## [1] "2023-05-24 10:56:31 MDT"

# Appendix B. Session and Version Information

In most cases you do not need to report session info (leave the “session-info” code chunk parameters in their default state: eval=FALSE). Session and version information is only necessary if you have set the “Listing” code chunk to eval=TRUE in appendix A. In that case, change the “session-info” code chunk parameters to eval=TRUE.

## R version 4.2.3 (2023-03-15 ucrt)  
## Platform: x86\_64-w64-mingw32/x64 (64-bit)  
## Running under: Windows 10 x64 (build 19045)  
##   
## Matrix products: default  
##   
## locale:  
## [1] LC\_COLLATE=English\_United States.utf8   
## [2] LC\_CTYPE=English\_United States.utf8   
## [3] LC\_MONETARY=English\_United States.utf8  
## [4] LC\_NUMERIC=C   
## [5] LC\_TIME=English\_United States.utf8   
##   
## attached base packages:  
## [1] stats graphics grDevices utils datasets methods base   
##   
## other attached packages:  
## [1] QCkit\_0.1.0.4 lubridate\_1.9.2 forcats\_1.0.0 stringr\_1.5.0   
## [5] dplyr\_1.1.0 purrr\_1.0.1 readr\_2.1.4 tidyr\_1.3.0   
## [9] tibble\_3.1.8 ggplot2\_3.4.1 tidyverse\_2.0.0 devtools\_2.4.5   
## [13] usethis\_2.1.6 kableExtra\_1.3.4 yaml\_2.3.7 knitr\_1.42   
## [17] pander\_0.6.5 rmarkdown\_2.20 markdown\_1.2   
##   
## loaded via a namespace (and not attached):  
## [1] httr\_1.4.4 pkgload\_1.3.1 bit64\_4.0.5 vroom\_1.6.1   
## [5] viridisLite\_0.4.1 shiny\_1.7.2 highr\_0.10 remotes\_2.4.2   
## [9] sessioninfo\_1.2.2 pillar\_1.8.1 glue\_1.6.2 digest\_0.6.29   
## [13] promises\_1.2.0.1 rvest\_1.0.3 colorspace\_2.0-3 plyr\_1.8.8   
## [17] htmltools\_0.5.4 httpuv\_1.6.6 pkgconfig\_2.0.3 xtable\_1.8-4   
## [21] scales\_1.2.1 webshot\_0.5.4 processx\_3.7.0 svglite\_2.1.0   
## [25] later\_1.3.0 tzdb\_0.3.0 timechange\_0.2.0 generics\_0.1.3   
## [29] ellipsis\_0.3.2 cachem\_1.0.6 withr\_2.5.0 cli\_3.4.1   
## [33] magrittr\_2.0.3 crayon\_1.5.2 mime\_0.12 memoise\_2.0.1   
## [37] evaluate\_0.20 ps\_1.7.1 fs\_1.5.2 fansi\_1.0.3   
## [41] xml2\_1.3.3 pkgbuild\_1.4.0 profvis\_0.3.7 tools\_4.2.3   
## [45] prettyunits\_1.1.1 hms\_1.1.2 lifecycle\_1.0.3 munsell\_0.5.0   
## [49] callr\_3.7.2 compiler\_4.2.3 systemfonts\_1.0.4 rlang\_1.0.6   
## [53] grid\_4.2.3 rstudioapi\_0.14 htmlwidgets\_1.5.4 miniUI\_0.1.1.1   
## [57] gtable\_0.3.1 curl\_4.3.3 R6\_2.5.1 bit\_4.0.5   
## [61] fastmap\_1.1.0 utf8\_1.2.2 stringi\_1.7.8 parallel\_4.2.3   
## [65] Rcpp\_1.0.9 png\_0.1-7 vctrs\_0.5.2 tidyselect\_1.2.0   
## [69] xfun\_0.36 urlchecker\_1.0.1

## [1] "2023-05-24 10:56:31 MDT"