24-05-01

[*https://doi.org/10.38750/2303148*](https://doi.org/10.38750/2303148)

# Data Release Report: Missouri Bladderpod Monitoring Data at Wilson’s Creek National Battlefield, 1997-2021

#### James Brown (0009-0004-5867-1904) 1

1 Student Conservation Association, 1201 Oakridge Dr., Suite 150, Fort Collins, Colorado

# Abstract

The data package that this Data Release Report pertains to describes monitoring surveys of the Missouri bladderpod (*Physaria filiformis*) at Wilson’s Creek National Battlefield (WICR), a member of the Heartland Network (HTLN). The datasets within the data package describe the abundance of Missouri bladderpod in select transects, the amount of light reaching the surface of the transects, and the presence and/or size of Eastern red cedars within the transects. These datasets contain data spanning from 1997–2021, and are still being worked on at the time of writing.

# Data Records (required)

## Summary of Datasets Created (required)

This DRR describes the data package *Missouri Bladderpod WICR Data* which contains a metadata file and 4 data files. These data were compiled and processed for dissemination by the National Park Service Inventory and Monitoring Division (IMD) and are available at <https://doi.org/10.57830/2303034> (see Table 1).

**Table 1. Missouri Bladderpod WICR Data: List of data files.**

| **File Name** | **Size** | **Description** |
| --- | --- | --- |
| HTLN\_AccuracyGridA\_INTERNAL.csv | 62.6 KB | HTLN Missouri bladderpod accuracy data for 2006 to 2021 for grid A |
| HTLN\_MoBlad\_CedarData\_INTERNAL.csv | 305.4 KB | Diameter of Eastern red cedars found within bladderpod survey grid A at WICR |
| HTLN\_MoBlad\_CountData1997-2021\_INTERNAL.csv | 1.7 MB | HTLN Missouri bladderpod survey data from 1997 to 2021 for 10 grids |
| HTLN\_MoBlad\_PARData\_INTERNAL.csv | 164.3 KB | HTLN Missouri bladderpod measurement of Photosynthetically Active Radiation exclusively for grid A in 2007, 2008, and 2015 |

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.

The data contained in the four .csv files have not been modified from their original values. The main addition to these files post-cleaning have been the inclusion of standard Darwin Core columns: dwcType and dwcBasisOfRecord. A ParkCode column has also been added, sometimes alongside the ParkName column present in all datasets except HTLN\_MoBlad\_AccuracyGridA\_Cleaned\_INTERNAL.csv. Beyond the added columns and their values, no other values from the original datasets have been modified in processing.

**Required elements for this section**

*Required Table*

**Table 3: Summary of data quality flags for each column [A – Accepted; AE – Accepted but Estimated; P – Provisional; R – Rejected.]**

| **File Name** | **Measure1** | **Number of Reocrds** | **A2** | **AE** | **R** | **P** | **% Accepted3** |
| --- | --- | --- | --- | --- | --- | --- | --- |
| HTLN\_AccuracyGridA\_INTERNAL | NA | NA | NA | NA | NA | NA | NA% |
| HTLN\_MoBlad\_CedarData\_INTERNAL | NA | NA | NA | NA | NA | NA | NA% |
| HTLN\_MoBlad\_CountData1997-2021\_INTERNAL | NA | NA | NA | NA | NA | NA | NA% |
| HTLN\_MoBlad\_PARData\_INTERNAL | NA | NA | NA | NA | NA | NA | NA% |

**Note:** 1The ’\_flag’ suffix has been omitted from column names for brevity. 2All non-missing data in specified unflagged columns are considered accepted. 3% Accepted 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).

*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 (Table 3). Of the data that were evaluated for quality, XX.X% of fields in this data package met data quality standards. The *Missouri Bladderpod WICR Data* is suitable for its intended use as of the date of processing (2024-05-01).

# Usage Notes (required)

As the information contained in this data package directly describes features of the Missouri bladderpod, the access to this data package is limited to internal use only. Any requests to access the data within the package should be made to the Heartland Network’s data manager. At the time of writing (2024-05-01), [Jennifer Haack-Gaynor](mailto:jennifer_haack@nps.gov) serves as the primary point of contact for any inquiries about the data package. Otherwise, inquiries can be directed to [Judd Patterson](mailto:Judd_Patterson@nps.gov).

## Acquiring the Data Package

This data package is available for download from the NPS DataStore at <https://doi.org/10.57830/2303034> and can be directly imported into R data frames using the NPSutils package <https://doi.org/10.57830/2300651>

# Methods

Beginning on the first week of April, the Missouri bladderpod population at grid A is inspected twice weekly to note the phenological status of the plants. Flowering typically begins in early- to mid-April. Monitoring commences during or near peak flowering. At grid A and C, observers mark the corners of 5 m × 5 m cells. During most years at Grid A, only cells in the core area are sampled. Every five years, however, the entire grid is sampled. At the other eight grids, observers navigate in 15 m × 15 m (225 m²) cells using a GNSS (global navigation satellite system) device; corners are not marked. For all sites, Missouri bladderpod density is estimated in grid cells using a density class scale: 0 = no plants, 1 = 1-9 plants, 2 = 10-49 plants, 3 = 50-99 plants, 4 = 100-499 plants, 5 = 500-999 plants, 6 = 1,000-4,999 plants, and 7 = >5,000 plants. Prior to estimating density, observers calibrate their estimates. Observers estimate abundance using the density class scale (0 = no plants, 1 = 1-9 plants, 2 = 10-49 plants, 3 = 50-99 plants, 4 = 100-499 plants, 5 = 500-999 plants, 6 = 1,000-4,999 plants, 7 = >5,000 plants) and then count the plants in the plot at the ground-level. This process is performed over a range of plot densities until estimates match counts. Observers should also compare results until estimates begin to converge. When estimating density, observers should consider counting up to 100 bladderpods in a 5 m × 5 m plot before determining estimates. Observers have the discretion to count as many plants as needed to improve confidence in the density estimate. Such counting may encompass all plants in the cell or only the plants in certain section of the plot (e.g., a particularly dense cluster of plants). If needed, plants may be counted in a portion of the plot and then extrapolated to the 5 m × 5 m plot.

Accuracy Assessment for Density Class Estimation is done on grid A only. After estimating density for each 5 m × 5 m plot at grid A, 60 cells are randomly selected using the rand() function in Excel. Each plot is visited to assess reliability of density class assignments. Additional cells of known density class are selected subjectively for accuracy assessment to bring the total number of cells assessed per density class to 10.

Habitat measurement only takes place on grid A. The percentage of photosynthetically active radiation (PAR) is measured at each plot center using LI-190 Quantum Sensors. PAR is the part of the electromagnetic radiation spectrum (400-700 nm) used in photosynthesis. Field workers also measure dbh of all Eastern red cedar (Juniperus virginiana) stems reaching breast height (1.37 m). Because cedar basal area will not vary greatly in the absence of disturbance, such measurements are only required on a five-year basis. Observers should, however, measure cedar basal area following significant canopy changes (e.g., thinning, fire, ice storms, tornados).

## 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)

# Appendix A. Code Listing

# 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 <- "Data Release Report: Missouri Bladderpod Monitoring Data at Wilson's Creek National Battlefield, 1997-2021"  
  
# 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 <- NULL  
  
# 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 <- 2303148  
  
#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(  
 "James Brown"  
)  
  
# 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(  
 "Student Conservation Association, 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(  
 "(0009-0004-5867-1904)"  
 )  
  
# Replace the text below with your abstract.  
DRRabstract <- "The data package that this Data Release Report pertains to describes monitoring surveys of the Missouri bladderpod (\*Physaria filiformis\*) at Wilson's Creek National Battlefield (WICR), a member of the Heartland Network (HTLN). The datasets within the data package describe the abundance of Missouri bladderpod in select transects, the amount of light reaching the surface of the transects, and the presence and/or size of Eastern red cedars within the transects. These datasets contain data spanning from 1997--2021, and are still being worked on at the time of writing."  
  
# 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(2303034)  
  
# Must match title in DataStore and metadata  
dataPackageTitle <- "Missouri Bladderpod WICR Data"  
  
# Must match descriptions in the data package metadata  
dataPackageDescription <- "The Heartland Inventory and Monitoring Network monitors the Missouri bladderpod (Physaria filiformis, formally Lesquerella filiformis) a federally threatened plant at Wilson’s Creek National Battlefield. The plant is surveyed annually in the spring with the objectives to track changes in the abundance, distribution, and persistence of the Missouri bladderpod through time at Wilson’s Creek National Battlefield. Values for photosynthetically active radiation and Eastern red cedar basal area are also collected to see how that may play a role in plant density amounts over time."  
  
# 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 <- list.files("../EML\_processing/Data/", pattern = "\*.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("62.6 KB", "305.4 KB", "1.7 MB", "164.3 KB")  
  
# list a short, one-line description of each data file. Descriptions must be in the same order as the filenames.  
dataPackage\_fileDescript <- c(  
 "HTLN Missouri bladderpod accuracy data for 2006 to 2021 for grid A",  
 "Diameter of Eastern red cedars found within bladderpod survey grid A at WICR",  
 "HTLN Missouri bladderpod survey data from 1997 to 2021 for 10 grids",  
 "HTLN Missouri bladderpod measurement of Photosynthetically Active Radiation exclusively for grid A in 2007, 2008, and 2015")  
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] "widgetframe" "htmlwidgets" "QCkit" "NMFSReports" "lubridate"   
## [6] "forcats" "stringr" "dplyr" "purrr" "readr"   
## [11] "tidyr" "tibble" "ggplot2" "tidyverse" "devtools"   
## [16] "usethis" "kableExtra" "yaml" "knitr" "pander"   
## [21] "rmarkdown" "markdown" "stats" "graphics" "grDevices"   
## [26] "utils" "datasets" "methods" "base"   
##   
## [[2]]  
## [1] "widgetframe" "htmlwidgets" "QCkit" "NMFSReports" "lubridate"   
## [6] "forcats" "stringr" "dplyr" "purrr" "readr"   
## [11] "tidyr" "tibble" "ggplot2" "tidyverse" "devtools"   
## [16] "usethis" "kableExtra" "yaml" "knitr" "pander"   
## [21] "rmarkdown" "markdown" "stats" "graphics" "grDevices"   
## [26] "utils" "datasets" "methods" "base"   
##   
## [[3]]  
## [1] "widgetframe" "htmlwidgets" "QCkit" "NMFSReports" "lubridate"   
## [6] "forcats" "stringr" "dplyr" "purrr" "readr"   
## [11] "tidyr" "tibble" "ggplot2" "tidyverse" "devtools"   
## [16] "usethis" "kableExtra" "yaml" "knitr" "pander"   
## [21] "rmarkdown" "markdown" "stats" "graphics" "grDevices"   
## [26] "utils" "datasets" "methods" "base"   
##   
## [[4]]  
## [1] "widgetframe" "htmlwidgets" "QCkit" "NMFSReports" "lubridate"   
## [6] "forcats" "stringr" "dplyr" "purrr" "readr"   
## [11] "tidyr" "tibble" "ggplot2" "tidyverse" "devtools"   
## [16] "usethis" "kableExtra" "yaml" "knitr" "pander"   
## [21] "rmarkdown" "markdown" "stats" "graphics" "grDevices"   
## [26] "utils" "datasets" "methods" "base"   
##   
## [[5]]  
## [1] "widgetframe" "htmlwidgets" "QCkit" "NMFSReports" "lubridate"   
## [6] "forcats" "stringr" "dplyr" "purrr" "readr"   
## [11] "tidyr" "tibble" "ggplot2" "tidyverse" "devtools"   
## [16] "usethis" "kableExtra" "yaml" "knitr" "pander"   
## [21] "rmarkdown" "markdown" "stats" "graphics" "grDevices"   
## [26] "utils" "datasets" "methods" "base"   
##   
## [[6]]  
## [1] "widgetframe" "htmlwidgets" "QCkit" "NMFSReports" "lubridate"   
## [6] "forcats" "stringr" "dplyr" "purrr" "readr"   
## [11] "tidyr" "tibble" "ggplot2" "tidyverse" "devtools"   
## [16] "usethis" "kableExtra" "yaml" "knitr" "pander"   
## [21] "rmarkdown" "markdown" "stats" "graphics" "grDevices"   
## [26] "utils" "datasets" "methods" "base"   
##   
## [[7]]  
## [1] "widgetframe" "htmlwidgets" "QCkit" "NMFSReports" "lubridate"   
## [6] "forcats" "stringr" "dplyr" "purrr" "readr"   
## [11] "tidyr" "tibble" "ggplot2" "tidyverse" "devtools"   
## [16] "usethis" "kableExtra" "yaml" "knitr" "pander"   
## [21] "rmarkdown" "markdown" "stats" "graphics" "grDevices"   
## [26] "utils" "datasets" "methods" "base"   
##   
## [[8]]  
## [1] "widgetframe" "htmlwidgets" "QCkit" "NMFSReports" "lubridate"   
## [6] "forcats" "stringr" "dplyr" "purrr" "readr"   
## [11] "tidyr" "tibble" "ggplot2" "tidyverse" "devtools"   
## [16] "usethis" "kableExtra" "yaml" "knitr" "pander"   
## [21] "rmarkdown" "markdown" "stats" "graphics" "grDevices"   
## [26] "utils" "datasets" "methods" "base"

devtools::install\_github("EmilyMarkowitz-NOAA/NMFSReports")

## Using GitHub PAT from the git credential store.

## Skipping install of 'NMFSReports' from a github remote, the SHA1 (fbd67654) has not changed since last install.  
## Use `force = TRUE` to force installation

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

## Using GitHub PAT from the git credential store.

## Downloading GitHub repo nationalparkservice/QCkit@HEAD

## Error in utils::download.file(url, path, method = method, quiet = quiet, :   
## cannot open URL 'https://api.github.com/repos/nationalparkservice/QCkit/tarball/HEAD'

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

## # Data Release Report: Missouri Bladderpod Monitoring Data at Wilson's Creek National Battlefield, 1997-2021

if (!is.null(reportNumber)) {  
 subtitle <- paste0("Data Release Report ", reportNumber)  
 cat("###", subtitle)  
}  
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$authorNames, sep="")  
 if (is.na(curr$authorORCID)) {  
 }  
 if (!is.na(curr$authorORCID)) {  
 orc <- paste0(" ", curr$authorORCID, "")  
 cat({{ orc }})  
 }  
 cat(" ^",script,"^", " ", " ", sep="")  
   
 #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="")  
}

## #### James Brown (0009-0004-5867-1904) ^1^

#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, " ", sep="")  
 if (is.na(curr$authorORCID)) {  
 }  
 if (!is.na(curr$authorORCID)) {  
 orc <- paste0(" ", curr$authorORCID, " ")  
 cat({{ orc }})  
 }  
 cat( " ^", script, "^", ", ", " ", sep = "")  
 }  
   
 # 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, sep="")  
 if (is.na(curr$authorORCID)) {  
 }  
 if (!is.na(curr$authorORCID)) {  
 orc <- paste0(" ", curr$authorORCID, " ")  
 cat({{ orc }})  
 }  
 cat(" ^",script,"^", ", ", sep="")  
 cat("and ", sep="")  
 }  
   
 #If only 2 authors, omit comma before "and":  
 if(length(seq\_along(unique\_authors$authorNames)) == 2){  
 cat("#### ", curr$authorNames, sep="")  
 if (is.na(curr$authorORCID)) {  
 }  
 if (!is.na(curr$authorORCID)) {  
 orc <- paste0(" ", curr$authorORCID, " ")  
 cat({{ orc }})  
 }  
 cat(" ^",script,"^ ", sep = "")  
 cat("and ", sep="")  
 }  
 }  
   
 # if IS the Last author :  
 if(i == length(seq\_along(unique\_authors$authorNames))){  
 cat(curr$authorNames, sep="")  
 if (is.na(curr$authorORCID)) {  
 }  
 if (!is.na(curr$authorORCID)) {  
 orc <- paste0(" ", curr$authorORCID, " ")  
 cat({{ orc }})  
 }  
 cat(" ^", script, "^", sep = "")  
 }  
 }  
}  
cat("\n\n")

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

## ^1^ Student Conservation Association, 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. Missouri Bladderpod WICR Data: List of data files.**

| **File Name** | **Size** | **Description** |
| --- | --- | --- |
| HTLN\_AccuracyGridA\_INTERNAL.csv | 62.6 KB | HTLN Missouri bladderpod accuracy data for 2006 to 2021 for grid A |
| HTLN\_MoBlad\_CedarData\_INTERNAL.csv | 305.4 KB | Diameter of Eastern red cedars found within bladderpod survey grid A at WICR |
| HTLN\_MoBlad\_CountData1997-2021\_INTERNAL.csv | 1.7 MB | HTLN Missouri bladderpod survey data from 1997 to 2021 for 10 grids |
| HTLN\_MoBlad\_PARData\_INTERNAL.csv | 164.3 KB | HTLN Missouri bladderpod measurement of Photosynthetically Active Radiation exclusively for grid A in 2007, 2008, and 2015 |

# To turn off, set eval=FALSE.  
# Generates a table of acceptance criteria for each of the data quality fields in your data package. Mitigations taken when data did not meet the acceptance criteria should be described textually in the Data Quality Evaluation section.  
  
# Specify which columns in your data package are data quality fields in the dataQualityFields variable. In the example below, data quality fields/columns in the data package are listed in the format [FieldName]\_flag. These data quality fields relate to the respective temporal, taxonomic, and geographic data.  
  
dataQualityFields <- c(  
 "eventDate\_flag",   
 "scientificName\_flag",   
 "coordinate\_flag"  
 )  
  
# Brief description of the acceptance criteria for each respective data quality field. The order of the acceptance criteria must match the order of the data quality fields.   
  
dataQualityAcceptanceCriteria <- c(  
 "Sampling event date within the start and end dates of the project",  
 "Taxon exists within Integrated Taxonomic Information System and GBIF",  
 "Sampling location is within the park unit boundaries"  
 )  
  
data\_criteria<-data.frame(dataQualityFields = str\_remove(dataQualityFields, "\_flag"), dataQualityAcceptanceCriteria)  
  
data\_criteria %>%   
 NMFSReports::format\_cells(1:3, 1, "bold") %>%   
 knitr::kable(caption = "\*\*Table 2. Acceptance criteria for data evaluated.\*\*", col.names=c("\*\*Field\*\*", "\*\*Acceptance Criteria\*\*"), format="pandoc", align = 'c')

**Table 2. Acceptance criteria for data evaluated.**

| **Field** | **Acceptance Criteria** |
| --- | --- |
| **eventDate** | Sampling event date within the start and end dates of the project |
| **scientificName** | Taxon exists within Integrated Taxonomic Information System and GBIF |
| **coordinate** | Sampling location is within the park unit boundaries |

# 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 <- "../EML\_processing/Data/"  
dc\_flags <- QCkit::get\_custom\_flags(directory = path, output="columns")  
dc\_flags$`File Name` <- gsub(".csv", "", dc\_flags$`File Name`)  
  
  
colnames(dc\_flags)[2]<-paste0("Measure", "^1^")  
colnames(dc\_flags)[4]<-paste0("A", "^2^")  
colnames(dc\_flags)[8]<-paste0("% Accepted", "^3^")  
  
#Generate the table:  
dc\_flags %>%   
 knitr::kable(  
 caption = '\*\*Table 3: Summary of data quality flags for each column [A – Accepted; AE – Accepted but Estimated; P – Provisional; R – Rejected.]\*\*',   
 format = "pandoc",   
 digits = 2,   
 align = 'c',   
 col.names = c("\*\*File Name\*\*", "\*\*Measure^1^\*\*", "\*\*Number of Reocrds\*\*", "\*\*A^2^\*\*", "\*\*AE\*\*", "\*\*R\*\*", "\*\*P\*\*", "\*\*% Accepted^3^\*\*")) %>%  
kableExtra::add\_footnote(  
 c("The '\_flag' suffix has been omitted from column names for brevity.",  
 "All non-missing data in specified unflagged columns are considered accepted.",  
 "% Accepted 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: Summary of data quality flags for each column [A – Accepted; AE – Accepted but Estimated; P – Provisional; R – Rejected.]**

| **File Name** | **Measure1** | **Number of Reocrds** | **A2** | **AE** | **R** | **P** | **% Accepted3** |
| --- | --- | --- | --- | --- | --- | --- | --- |
| HTLN\_AccuracyGridA\_INTERNAL | NA | NA | NA | NA | NA | NA | NA% |
| HTLN\_MoBlad\_CedarData\_INTERNAL | NA | NA | NA | NA | NA | NA | NA% |
| HTLN\_MoBlad\_CountData1997-2021\_INTERNAL | NA | NA | NA | NA | NA | NA | NA% |
| HTLN\_MoBlad\_PARData\_INTERNAL | NA | NA | NA | NA | NA | NA | NA% |

**Note:** 1The ’\_flag’ suffix has been omitted from column names for brevity. 2All non-missing data in specified unflagged columns are considered accepted. 3% Accepted 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.

# 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<-"../EML\_processing/Data/"  
dp\_flags <- get\_custom\_flags(directory = path, output="files")  
  
#generate table:  
dp\_flags %>%   
 kableExtra::kbl(caption = '\*\*Table 4: Summary of data quality flags for the data package [A – Accepted; AE – Accepted but Estimated; P – Provisional; R – Rejected.]\*\*',  
 format = "pandoc",   
 col.names = c("\*\*File Name\*\*", "\*\*A^1^\*\*", "\*\*AE\*\*", "\*\*R\*\*", "\*\*P\*\*", "\*\*% Accepted^2^\*\*"),  
 digits=2,   
 align = 'c') %>%   
 kableExtra::add\_footnote(c("All non-missing data in specified unflagged columns are considered accepted.",  
 "% Accepted 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 4: Summary of data quality flags for the data package [A – Accepted; AE – Accepted but Estimated; P – Provisional; R – Rejected.]**

| **File Name** | **A1** | **AE** | **R** | **P** | **% Accepted2** |
| --- | --- | --- | --- | --- | --- |
| HTLN\_AccuracyGridA\_INTERNAL.csv | NA | NA | NA | NA | NA% |
| HTLN\_MoBlad\_CedarData\_INTERNAL.csv | NA | NA | NA | NA | NA% |
| HTLN\_MoBlad\_CountData1997-2021\_INTERNAL.csv | NA | NA | NA | NA | NA% |
| HTLN\_MoBlad\_PARData\_INTERNAL.csv | NA | NA | NA | NA | NA% |

**Note:** 1All non-missing data in specified unflagged columns are considered accepted. 2% Accepted 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.

sessionInfo()

## R version 4.3.3 (2024-02-29 ucrt)  
## Platform: x86\_64-w64-mingw32/x64 (64-bit)  
## Running under: Windows 11 x64 (build 22631)  
##   
## 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   
##   
## time zone: America/Denver  
## tzcode source: internal  
##   
## attached base packages:  
## [1] stats graphics grDevices utils datasets methods base   
##   
## other attached packages:  
## [1] widgetframe\_0.3.1 htmlwidgets\_1.6.4 QCkit\_0.1.7   
## [4] NMFSReports\_0.0.1.3 lubridate\_1.9.3 forcats\_1.0.0   
## [7] stringr\_1.5.1 dplyr\_1.1.4 purrr\_1.0.2   
## [10] readr\_2.1.5 tidyr\_1.3.1 tibble\_3.2.1   
## [13] ggplot2\_3.5.0 tidyverse\_2.0.0 devtools\_2.4.5   
## [16] usethis\_2.2.3 kableExtra\_1.4.0 yaml\_2.3.8   
## [19] knitr\_1.46 pander\_0.6.5 rmarkdown\_2.26   
## [22] markdown\_1.12   
##   
## loaded via a namespace (and not attached):  
## [1] gtable\_0.3.4 xfun\_0.43 remotes\_2.5.0 tzdb\_0.4.0   
## [5] vctrs\_0.6.5 tools\_4.3.3 generics\_0.1.3 parallel\_4.3.3   
## [9] curl\_5.2.1 fansi\_1.0.6 pkgconfig\_2.0.3 uuid\_1.2-0   
## [13] lifecycle\_1.0.4 compiler\_4.3.3 textshaping\_0.3.7 munsell\_0.5.1   
## [17] httpuv\_1.6.15 htmltools\_0.5.8.1 crayon\_1.5.2 later\_1.3.2   
## [21] pillar\_1.9.0 urlchecker\_1.0.1 ellipsis\_0.3.2 openssl\_2.1.2   
## [25] cachem\_1.0.8 sessioninfo\_1.2.2 mime\_0.12 zip\_2.3.1   
## [29] tidyselect\_1.2.1 digest\_0.6.35 stringi\_1.8.3 fastmap\_1.1.1   
## [33] grid\_4.3.3 colorspace\_2.1-0 cli\_3.6.2 magrittr\_2.0.3   
## [37] pkgbuild\_1.4.4 utf8\_1.2.4 withr\_3.0.0 scales\_1.3.0   
## [41] promises\_1.3.0 bit64\_4.0.5 timechange\_0.3.0 officer\_0.6.5   
## [45] bit\_4.0.5 ragg\_1.3.0 askpass\_1.2.0 hms\_1.1.3   
## [49] memoise\_2.0.1 shiny\_1.8.1.1 evaluate\_0.23 miniUI\_0.1.1.1   
## [53] viridisLite\_0.4.2 profvis\_0.3.8 rlang\_1.1.3 Rcpp\_1.0.12   
## [57] xtable\_1.8-4 glue\_1.7.0 xml2\_1.3.6 pkgload\_1.3.4   
## [61] vroom\_1.6.5 svglite\_2.1.3 rstudioapi\_0.16.0 R6\_2.5.1   
## [65] systemfonts\_1.0.6 fs\_1.6.3

Sys.time()

## [1] "2024-05-01 12:20:24 MDT"

# Appendix B. Session and Version Information

## R version 4.3.3 (2024-02-29 ucrt)  
## Platform: x86\_64-w64-mingw32/x64 (64-bit)  
## Running under: Windows 11 x64 (build 22631)  
##   
## 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   
##   
## time zone: America/Denver  
## tzcode source: internal  
##   
## attached base packages:  
## [1] stats graphics grDevices utils datasets methods base   
##   
## other attached packages:  
## [1] widgetframe\_0.3.1 htmlwidgets\_1.6.4 QCkit\_0.1.7   
## [4] NMFSReports\_0.0.1.3 lubridate\_1.9.3 forcats\_1.0.0   
## [7] stringr\_1.5.1 dplyr\_1.1.4 purrr\_1.0.2   
## [10] readr\_2.1.5 tidyr\_1.3.1 tibble\_3.2.1   
## [13] ggplot2\_3.5.0 tidyverse\_2.0.0 devtools\_2.4.5   
## [16] usethis\_2.2.3 kableExtra\_1.4.0 yaml\_2.3.8   
## [19] knitr\_1.46 pander\_0.6.5 rmarkdown\_2.26   
## [22] markdown\_1.12   
##   
## loaded via a namespace (and not attached):  
## [1] gtable\_0.3.4 xfun\_0.43 remotes\_2.5.0 tzdb\_0.4.0   
## [5] vctrs\_0.6.5 tools\_4.3.3 generics\_0.1.3 parallel\_4.3.3   
## [9] curl\_5.2.1 fansi\_1.0.6 pkgconfig\_2.0.3 uuid\_1.2-0   
## [13] lifecycle\_1.0.4 compiler\_4.3.3 textshaping\_0.3.7 munsell\_0.5.1   
## [17] httpuv\_1.6.15 htmltools\_0.5.8.1 crayon\_1.5.2 later\_1.3.2   
## [21] pillar\_1.9.0 urlchecker\_1.0.1 ellipsis\_0.3.2 openssl\_2.1.2   
## [25] cachem\_1.0.8 sessioninfo\_1.2.2 mime\_0.12 zip\_2.3.1   
## [29] tidyselect\_1.2.1 digest\_0.6.35 stringi\_1.8.3 fastmap\_1.1.1   
## [33] grid\_4.3.3 colorspace\_2.1-0 cli\_3.6.2 magrittr\_2.0.3   
## [37] pkgbuild\_1.4.4 utf8\_1.2.4 withr\_3.0.0 scales\_1.3.0   
## [41] promises\_1.3.0 bit64\_4.0.5 timechange\_0.3.0 officer\_0.6.5   
## [45] bit\_4.0.5 ragg\_1.3.0 askpass\_1.2.0 hms\_1.1.3   
## [49] memoise\_2.0.1 shiny\_1.8.1.1 evaluate\_0.23 miniUI\_0.1.1.1   
## [53] viridisLite\_0.4.2 profvis\_0.3.8 rlang\_1.1.3 Rcpp\_1.0.12   
## [57] xtable\_1.8-4 glue\_1.7.0 xml2\_1.3.6 pkgload\_1.3.4   
## [61] vroom\_1.6.5 svglite\_2.1.3 rstudioapi\_0.16.0 R6\_2.5.1   
## [65] systemfonts\_1.0.6 fs\_1.6.3

## [1] "2024-05-01 12:20:24 MDT"