

# Data Dictionaries

Data set name: Acoustic files

File name: files.tsv

Data format: tsv

Description: This dataset describes the acoustic recordings included in this dataset.

column_name	description
site	Site name
replicate	An ordinal label indicating the random draw label: 'A,' 'B,' or 'C'
recording_date	Recording date and time formatted as "Year-Month-Day Hour:Minute:Second"
annotated	Categorical assignment describing whether a recording was completely annotated: 'complete,' 'partial,' or 'not annotated'
file	Wav file name
zip_file	The zip file location of the file

Data set name: Acoustic annotations

File name: annotations.tsv

Data format: tsv

Description: This data set lists all annotations from the fully annotated recordings.

column_name	description
file	Wav file name
start	Start time of the 2-second clip in seconds
end	End time of the 2-second clip in seconds
eBird_2021	2021 species identification eBird code
label	Sonotype label concatenates the 2021 eBird taxonomy code and the sound type label

Data set name: Partial annotations

File name: partial\_annotations.tsv

Data format: tsv

Description: This data set lists all annotations from the fully partially annotated recordings.

column_name	description
file	Wav file name
start	Start time of the 2-second clip in seconds
end	End time of the 2-second clip in seconds
clip_complete	Binary indicator for whether the clip was completely labeled
eBird_2021	2021 species identification eBird code
label	Sonotype label concatenates the 2021 eBird taxonomy code and the sound type label

Data set name: Annotation metadata

File name: metadata.tsv

Data format: tsv

Description: This data set describes the focal acoustic sounds included in the recording annotations.

column_name	description
label	Sonotype label comprising a concatenation of the 2021 eBird taxonomy code and the sound type label
eBird_2021	2021 eBird taxonomy species_code
sound	Sound type label
common_name	The common name of the sound source. For avian species, the scientific name follows Clement's taxonomy outlined in the 2021 eBird taxonomy
scientific_name	The scientific name of the biotic sound source. For avian species, the scientific name follows Clement's taxonomy outlined in the 2021 eBird taxonomy
taxonomic_authority	Primary taxonomic authority
description	Biological and phonetic description of the target sound
n_files	Total number of audio files containing at least 1 of the target label
n_annotations	Total number of label-specific annotations in the fully annotated data

Data set name: Environmental characteristics

File name: environmental\_characteristics.tsv

Data format: tsv

Description: This data set lists the environmental characteristics at each recording station. Units of measurements for appropriate covariates are in parentheses.

column_name	description
site	Site name
replicate	An ordinal label indicating whether the row describes a random sample 'A,' 'B,' or 'C'
state	State location of survey site
township_range	Township and range identifier of the survey site. The township was data obtained from three sources: <a href="#">CA</a> , <a href="#">OR</a> , <a href="#">WA</a>
age_dom_2017	Basal area weighted stand age based on dominant and codominant trees (years)
ba_ge_3_2017	Basal area of live trees $\geq 2.5$ cm dbh ( $\text{m}^2/\text{ha}$ )
bac_ge_3_2017	Basal area of live conifers $\geq 2.5$ cm dbh ( $\text{m}^2/\text{ha}$ )
bah_ge_3_2017	Basal area of live hardwoods $\geq 2.5\text{cm}$ dbh ( $\text{m}^2/\text{ha}$ )
bph_ge_3_crm_2017	Component Ratio Method biomass of all live trees $\geq 2.5$ cm ( $\text{kg}/\text{ha}$ )
bphc_ge_3_crm_2017	Component Ratio Method biomass of all live conifers $\geq 2.5$ cm ( $\text{kg}/\text{ha}$ )
bphh_ge_3_crm_2017	Component Ratio Method biomass of all live hardwoods $\geq 2.5$ cm ( $\text{kg}/\text{ha}$ )
cancov_2017	Canopy cover of all live trees (percent)
cancov_con_2017	Canopy cover of all conifers (percent)
cancov_hdw_2017	Canopy cover of all hardwoods (percent)
cancov_layers_2017	Number of tree canopy layers present (number of layers)
conplba_2017	Conifer tree species with the plurality of basal area (raster to alphanumeric lookup table available at source)
covcl_2017	Cover class based on cancov (raster to alphanumeric lookup table available at source)
ddi_2017	Diameter diversity index

fortypba_2017	Forest type, which describes the dominant tree species of current vegetation (raster to alphanumeric lookup table available at source)
hdwplba_2017	Hardwood tree species with the plurality of basal area (raster to alphanumeric lookup table available at source)
mndbhba_2017	Basal-area weighted mean diameter of all live trees (cm)
mndbhba_con_2017	Basal-area weighted mean diameter of all live conifers (cm)
mndbhba_hdw_2017	Basal-area weighted mean diameter of all live hardwoods (cm)
qmd_dom_2017	The quadratic mean diameter of all dominant and codominant trees (cm)
qmd_ht25_2017	The quadratic mean diameter in inches of trees whose heights are in the top 25% of all tree heights (cm)
qmdc_dom_2017	The quadratic mean diameter of all dominant and codominant conifers (cm)
qmdh_dom_2017	The quadratic mean diameter of all dominant and codominant hardwoods (cm)
sbph_ge_25_2017	Biomass of snags $\geq 25$ cm dbh and $\geq 2$ m tall (lb)
sdi_reineke_2017	Reineke's stand density index
sizecl_2017	Size class, based on QMD_DOM and CANCOV (raster to alphanumeric lookup table available at source)
stndhgt_2017	Stand height, computed as the average height of all dominant and codominant trees (m)
stph_ge_25_2017	Density of snags $\geq 25$ cm dbh and $\geq 2$ m tall (trees/ha)
strucond_2017	Structural condition (raster to alphanumeric lookup table available at source)
svph_ge_25_2017	Volume of snags $\geq 25$ cm dbh and $\geq 2$ m tall ( $\text{m}^3/\text{ha}$ )
tph_ge_3_2017	The density of live trees $\geq 2.5$ cm dbh (trees/ha)
tphc_ge_3_2017	The density of live conifers $\geq 2.5$ cm dbh (trees/ha)
tphh_ge_3_2017	The density of live hardwoods $\geq 2.5$ cm dbh (trees/ha)
treeplba_2017	Tree species with the plurality of basal area (raster to alphanumeric lookup table available at source)

vegclass_2017	Vegetation class based on CANCOV, BAH_PROP, QMD_DOM (raster to alphanumeric lookup table available at source)
vph_ge_3_2017	The volume of live trees $\geq 2.5$ cm dbh ( $\text{m}^3/\text{ha}$ )
vphc_ge_3_2017	The volume of live conifers $\geq 2.5$ cm dbh ( $\text{m}^3/\text{ha}$ )
vphh_ge_3_2017	The volume of live hardwoods $\geq 2.5$ cm dbh ( $\text{m}^3/\text{ha}$ )
dem_30m	Digital elevation model at 30 $\text{m}^2$ resolution (m)
northness_30m	A cosine transformation of aspect to demonstrate the orientation of land relative to a north-facing land derived from dem_30m
slope_30m	Estimate of land slope at 30 $\text{m}^2$ resolution derived from dem_30m
tpi5x5_30m	Mean difference of the central point to a focal square of the surrounding $5 \times 5$ grid cells derived from dem_30m
vrn_30m	Variation in slope and aspect derived from dem_30m
an_precip_1km	Average precipitation at a 1 $\text{km}^2$ resolution averaged from 1970-2000 (mm)
minT_1km	Average minimum temperature at a 1 $\text{km}^2$ resolution averaged from 1970-2000 (degrees Celcius)
maxT_1km	Average maximum temperature at a 1 $\text{km}^2$ resolution averaged from 1970-2000 (degrees Celcius)

Data set name: Environmental characteristics metadata

File name: environmental\_characteristics\_metadata.tsv

Data format: tsv

Description: This data set describes the environmental characteristics included in environmental\_characteristics.tsv.

column_name	description
covariate	Covariate name
type	Value type of variable
range	The range of values extracted across our survey sites. The values in this cell represent the value minimum to the value maximum
description	A description of the variable, including a brief discussion of the methods used to create the variable
source	Variable source citation