

Intro-Create-Read-HDF5

Goals / Objectives

After completing this activity, you will:

1. Understand how HDF5 files can be created and structured in R using the rhdf5 libraries.
2. Understand the 3 key HDF5 elements: the HDF5 file itself and groups and datasets.
3. Understand how to add and read attributes from an HDF5 file.

First, let's get R setup. We will use the RHDF5 library.

```
# To access HDF5 files in R, we will use the rhdf5 library which is part of the Bioconductor suite of R  
  
#Install the R HDF5 Library  
# source("http://bioconductor.org/biocLite.R")  
# biocLite("rhdf5")  
  
#Call the R HDF5 Library  
library("rhdf5")
```

We'll start by creating a new HDF5 file. Let's call it vegData.

```
# Create hdf5 file  
h5createFile("vegData.h5")
```

```
## file 'C:\Users\kthibault\Documents\GitHub\NEON-Data-Skills\code\R\vegData.h5' already exists.
```

```
## [1] FALSE
```

```
#create a group called aNEONSite within the H5 file  
h5createGroup("vegData.h5", "aNEONSite")
```

```
## Can not create group. Object with name 'aNEONSite' already exists.
```

```
## [1] FALSE
```

```
#view the structure of the h5 we've created  
h5ls("vegData.h5")
```

```
##      group      name      otype dclass    dim  
## 0      /    aNEONSite  H5I_GROUP  
## 1 /aNEONSite temperature H5I_DATASET  FLOAT 20 x 2
```

```

# create some sample, numeric data
a <- rnorm(n=40, m=1, sd=1)
someData <- matrix(a,nrow=20,ncol=2)

# add some sample data to the H5 file located in the aNEONSite group
# we'll call the dataset "temperature"
h5write(someData, file = "vegData.h5", name="aNEONSite/temperature")

# let's check out the H5 structure again
h5ls("vegData.h5")

```

```

##          group          name          otype dclass      dim
## 0          /    aNEONSite    H5I_GROUP
## 1 /aNEONSite temperature H5I_DATASET  FLOAT 20 x 2

```

```

# we can look at everything too
# but be cautious using this command!
h5dump("vegData.h5")

```

```

## $aNEONSite
## $aNEONSite$temperature
##           [,1]      [,2]
## [1,]  0.59320932  2.0558270
## [2,] -1.51228352  0.4675319
## [3,]  0.88255032  2.2556916
## [4,]  1.93612342  1.0770521
## [5,]  1.54227034  0.5046503
## [6,]  0.72985530  1.8625616
## [7,] -0.76820976  2.1599439
## [8,]  2.91871759  0.8175413
## [9,]  1.03258012  1.1870309
## [10,] 2.43477588  1.4855226
## [11,] 0.85395930  1.1949661
## [12,] 1.03953360  0.5397565
## [13,] 0.05501753  0.4336031
## [14,] 1.78427902  0.5011792
## [15,] 3.23141869  2.4551404
## [16,] 1.82001528  0.5404874
## [17,] 0.68681068  1.1834248
## [18,] 0.30706192  1.0759600
## [19,] 2.04645660 -1.4181957
## [20,] 0.36826337  1.4578805

```

```

#Close the file.
H5close()

```

Next let's add some Metadata (attributes) to our data

```

#open the file, create a class
fid <- H5Fopen("vegData.h5")

```

```

#open up the dataset to add attributes to, as a class
did <- H5Dopen(fid, "aNEONSite/temperature")

# Provide the NAME and the ATTR (what the attribute says)
# for the attribute.
h5writeAttribute(did, attr="Here is a description of the data",
                 name="Description")
h5writeAttribute(did, attr="Meters",
                 name="Units")

#let's add some attributes to the group
did2 <- H5Gopen(fid, "aNEONSite/")
h5writeAttribute(did2, attr="San Joaquin Experimental Range",
                 name="SiteName")
h5writeAttribute(did2, attr="Southern California",
                 name="Location")

#close the files, groups and the dataset when you're done writing to them!
H5Dclose(did)
H5Gclose(did2)
H5Fclose(fid)

```

```

#look at the attributes of the precip_data dataset
h5readAttributes(file = "vegData.h5",
                 name = "aNEONSite/temperature")

```

```

## $Description
## [1] "Here is a description of the data"
##
## $Units
## [1] "Meters"

```

```

#look at the attributes of the aNEONsite group
h5readAttributes(file = "vegData.h5",
                 name = "aNEONSite")

```

```

## $Location
## [1] "Southern California"
##
## $SiteName
## [1] "San Joaquin Experimental Range"

```

```

# let's grab some data from the H5 file
testSubset <- h5read(file = "vegData.h5",
                    name = "aNEONSite/temperature")

testSubset2 <- h5read(file = "vegData.h5",
                     name = "aNEONSite/temperature",
                     index=list(NULL,1))

H5close()

```

```
#create a quick plot of the data  
hist(testSubset2)
```



Challenge –

1. Open up the D17_2013_vegStr.csv in R.
2. Create a new H5 file called vegStructure.
3. Add a group in your h5 file called SJER. Add the veg structure data to that folder!
4. Add some attributes the data. Do the same with the SOAP csv. NOTE: R is not good at subsetting compound datasets.

```
#options(stringsAsFactors = FALSE)  
#newData <- read.csv("D17_2013_vegStr.csv")
```