# *Wideband Integrated Bioaerosol Sensor (WIBS) Data Extraction*

**Name: Jeff Peterson**
**Semester: Spring 2020**
**GitHub Portfolio URL: jdp71.github.io**

**Which Domain?**
**What domain is this data going to come from? Please list 10 references (with a brief annotation) to use to make sense of what you're doing with these data**.

The data will come from the Wideband Integrated Bioaerosol Sensor (WIBS) instrument that we currently use at my work. We have multiple WIBS instruments that we use a referee equipment for a variety of tests. The WIBS instrument came with some software that allowed us to view, analyze, and plot the data. Unfortunately, due to security concerns, we are no longer able to install the software on our government computers. This project will involve writing a script that will allow us to extract and process the WIBS instrument data.

Since this is an actual process I am trying to work out, there really aren't many references for this project.

References:

WIBS-5/NEO. (2020). Retrieved from https://www.dropletmeasurement.com/product/wideband-integrated-bioaerosol-sensor/ (This is the website for the WIBS instrument manufacturer)

Takai, T. M. (2011). *Interim Guidance on Networthiness of Information Technology (It) Connected to DoD Networks*. Washington, DC: Department of Defense. (This is the site that has the memorandum dictating the DoD software requirements)

Hulslander, D. (2018, May). Open HDF5 files with Python Sample Code. Retrieved from https://www.neonscience.org/hdf5-intro-python (This is an .h5 file processing site)

Battelle. (n.d.). NEON AOP Hyperspectral Data in HDF5 format with Python - Tiled Data. Retrieved from https://www.neonscience.org/neon-aop-hdf5-tile-py (This is another .h5 processing site)

Baron, P. A. (1986). Calibration and Use of the Aerodynamic Particle Sizer (APS 3300). *Aerosol Science and Technology*, *5*(1), 55–67. doi: 10.1080/02786828608959076 (This is a journal article about calibrating and using the APS instrument)

Chen, B., & Crow, D. (1986). Use of an aerodynamic particle sizer as a real-time monitor in generation of ideal solid aerosols. *Journal of Aerosol Science*, *17*(6), 963–972. doi: 10.1016/0021-8502(86)90022-4 (This is another APS related journal article)

Hogan, J. (2019). *Test Plan For The System X*. Dugway, UT: U.S. Army Combat Capabilities Development Command. (This is the test plan for the System X instrument)

**Which Data?**
**What is the dataset you'll be examining? Please provide a codebook if there is one or a link to the dataset as well as a detailed description.**

Two of the data are files needed are collected directly from the WIBS instrument.

1. Filename: **20191119074400__x2.h5**
   a. This is the data feed of the instrument response during the test. This is the main file we are going to be working with.
2. Filename: **20191119074400_FT__x1.h5**
   a. This is called the Force Trigger (FT) file. This serves as a background measurement during the trial. This file will be used as a correction factor to the WIBS file.

The next data file that is needed is called the Test Log.

3. Filename: **cd_test_log.csv**
   a. This is the test log that is created for each testing day. This file is necessary to know what times the tests start and end.

The final file needed is called the Aerodynamic Particle Sizer (APS) file.

4. Filename:  **20191119_1.txt**
   a. This file is a measure of the detected particle sizes released during the test. The APS instrument is also used as a referee system.
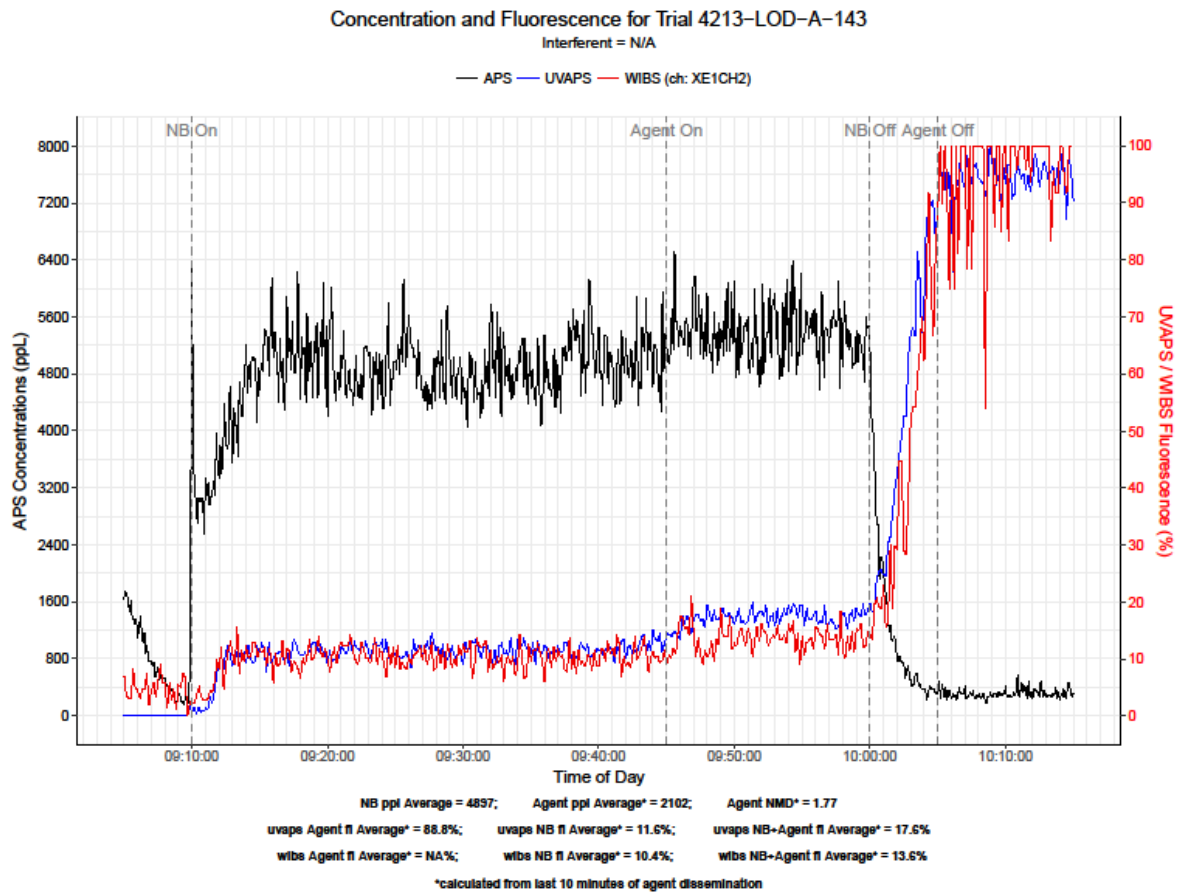
The individual data files have been downloaded to my local laptop where they are stored for use in this code. Due to the nature of the data, I may not be able to post the full datasets to my project portfolio. The user will be able to see how I implemented the data files and what their basic content are, but they will not be able to see the entire dataset.

There is not a codebook for these datasets and most of the data columns that are present will not be used. The columns of interest were determined after personal communication with the scientist running the test and with the WIBS manufacturer.

**Research Questions? Benefits? Why analyze these data?**
**How are you proposing to analyze this dataset? This is about your approach. Here, you'll be proposing your research questions as well as justifications for why you'd offer these data in this way.**

Here is an output plot that the old software produced.

Concentration and Fluorescence for Trial 4213-LOD-A-143

Interferent = N/A

— APS — UVAPS — WIBS (ch: XE1CH2)

NB On                    Agent On                    NB Off  Agent Off

APS Concentrations (ppL) — UVAPS / WIBS Fluorescence (%)

Time of Day

NB ppl Average = 4897;    Agent ppl Average* = 2102;    Agent NMD* = 1.77

uvaps Agent fl Average* = 88.8%;    uvaps NB fl Average* = 11.6%;    uvaps NB+Agent fl Average* = 17.6%

wibs Agent fl Average* = NA%;    wibs NB fl Average* = 10.4%;    wibs NB+Agent fl Average* = 13.6%

*calculated from last 10 minutes of agent dissemination

The software that made this plot was the only way we had to analyze the test data. Once that option was taken away from us, the WIBS instrument was rendered useless and all testing was forced to stop. The manufacturer said that we could pay them to analyze our data for us if we would just send them the output file. We originally had to do this because we had no other option.

My goal with this project is to be able to extract the necessary data from the WIBS instrument, combine it with the APS and the Test Log files, and be able to produce a plot similar to what is shown above.

## What Method?
### What methods will you be using? What will those methods provide in terms of analysis? How is this useful?

I will use Jupyter Notebook so that I can develop a step-by-step process to making this plot. From what I have been told by the manufacturer, this is not a straightforward process. I'm guessing part of the reason they told me this is so that we will keep paying them.

I am going to have to do a lot of joining of datasets, time manipulations, and plots. I will also need to make additional columns in some of the dataframes to support various calculations that are required.

## Potential Issues?
**What challenges do you anticipate having? What could cause this project to go off schedule?**

I'm not sure at this point which plotting package I am going to use. I wanted to use ggplot2, but from what I have read, it doesn't really support a dual y-axis very well. Also, all of the supplemental text that is required on the plot might be problematic.

Also, the WIBS and WIBS FT file formats are .h5 formats and I am unfamiliar with these. I found a couple of websites that offer some guidance, but I haven't gone though them yet, so I don't know if they will apply to my analysis methods.

The biggest thing that will cause this to go off schedule is simply not being able to get the files to import properly. The other issue might be that I simply can't get it done in the few weeks allotted for this project. If that is the case, I will just turn in what I have and continue on later because one way or another, I need to find a solution for this problem we are facing.

## Concluding Remarks
**Tie it all together. Think of this section as your final report's abstract.**

Dugway Proving Ground is currently testing a lightweight, battery operable system designed to rapidly detect, sample, and identify airborne biological warfare agents (BWA) that pose a threat to the soldiers. This instrument (called System X) is a man-portable system that can be carried by one warfighter over short distances.

A special test chamber has been constructed that will house the various instruments during this test. The instruments will include three System X units, one APS, and one WIBS. The APS unit will be located with an associated System X to provide a concentration for that particular location. The WIBS will provide a percent fluorescent in the chamber.

The current state of this test is on hold due to the inability to extract and process the data from the WIBS instrument. The WIBS is a key referee system in these tests, and as such, we are unable to accurately determine the efficacy of System X measurements.

I am really excited about this project because it is something that is actually needed at work and something that is going to save us a lot of time and money.