

File Naming / File Structure / Nodes / Metadata for PSWS

Edit History

Rev 0.00 2020-04-28 by JCGibbons / N8OBJ
Initial release
Rev 0.01 2020-05-01 by JCGibbons / N8OBJ
mods per Nathaniel
Rev 0.02 2020-05-05 by JCGibbons / N8OBJ
added content about filename structure , nodes
Rev 0.03 2020-05-05 by JCGibbons / N8OBJ
mods to directory structure, node info
Rev 0.04 2020-05-06 by JCGibbons / N8OBJ
increased verbage wrt project assumptions
removed [minus] from illegal chars in filename spec
modified top descriptive name of file
added metadata definition / file format info

This document defines (in part) the file naming structure, Node numbers, Low Cost Personal Space Weather Station (LC PSWS) file structures, conventions, Metadata formats and header formatting. It was requested that the file naming be informative of its contents by date (in ISO format), Grid Locator, and other info such that rudimentary sorting of the data files (and the information contained within) could be quickly assessed by the simple inspection of the filename itself. This proposal is an effort to define the naming structure that will be used for the PSWS project. It is assumed that the person looking at these files is knowledgeable about this PSWS project and understands what / why / how / when the data was/is being collected.

Note:

This document is intended to be a work in progress as only the entire group knows all aspects of this project and what the needs are for each aspect of the project.

ALL data collection is being done on a one (1) UTC day cycle. All collected data will be time stamped in the filename itself and will be done in UTC time to avoid confusion with time zones. ALL dates are given in the ISO UTC format only. The ISO UTC date will also be included in the first line of the file for data processing purposes (along with other metadata). If data collection is interrupted, the remaining data for that UTC day and type will simply be appended to the already existing data file. For the LC PSWS, each data entry has a ISO formatted UTC time stamp associated with each entry, so nothing else needs to be done to accommodate the data collection interruption.

It is assumed that each node will only collect one data set per submitted data type (TYP defined below), i.e. only one WWV 5MHz file, one magnetometer file, one temperature file, etc. Duplicate data sets for a specific given type measurement will not ALL be stored in the database.

If duplicate data is created by the station, is is up to the station operator to sort this out. A given node should not be collecting multiple identical data sets.

If. for a given node number. multiple same type station data sets are submitted to the database, the last submission will be used and the previous data set for that type (TYP defined below) will be discarded. This will give a mechanism to allow for correcting data collection errors (if they arise).

Node DataBase Contents Definition

The data for each collection node contains the following information:

1. Node number (N00000 – N99999)
2. City and State
3. Latitude, Longitude, Elevation
4. GridSquare (Maiden-Head Locator) AA00aa (full 6 digits)
5. Radio Make and Model
6. Antenna Description (Make and Model if commercial)
7. Magnetometer information
8. Temperature Sensor(s) information
9. Lightning Sensor information
10. TBD Sensor information (?)

Data that needs a higher security level (Contact Info) to view but is included WRT the node

11. Contact Callsign
12. Contact Name
13. Contact Address
14. Contact Cell Phone
15. Misc Info

ALL Info will be all tied back to the node number as the central reference point in the database

Low Cost PSWS Filename Structure

YYYY-MM-DD_N12345_AA00aa_TYP_MOD.csv

Filename will be constructed with the above sequence of info:

1. YYYY-MM-DD - ISO format of UTC day (date) for the collected data in file (10 Chars)
2. N12345 – Node number assigned to this station (6 Chars)
3. AA00aa – 6 position GridSquare locator of station (6 Chars)
4. TYP – 3 character description of data (3 Chars)
 - FRQ = frequency data
 - TMP = Temperature data
 - MAG = Magnetometer data
 - LGT = Lightning strike data
5. MOD – A modifier field to further describe the TYP field. Variable length (0 to ?? Chars)
MOD field starts after TYP field “_” and is terminated by the . for the files extension .csv
MOD identifies data type included in the file:
 - FRQ: WWV2p5, WWV5, WWV10, WWV15, WWV20, WWV25, CHU3, CHU7, CHU14
 - TMP: N/A
 - MAG: N/A
 - LGT: N/A
6. .csv – Descriptor of file format for data storage

LC PSWS data file examples:

2020-05-03_N00001_AA00aa_FRQ_WWV2p5.csv
2020-05-03_N00001_AA00aa_FRQ_WWV5.csv
2020-05-03_N00001_AA00aa_FRQ_WWV10.csv
2020-05-03_N00001_AA00aa_FRQ_WWV15.csv
2020-05-03_N00001_AA00aa_FRQ_WWV20.csv
2020-05-03_N00001_AA00aa_FRQ_WWV25.csv
2020-05-03_N00001_AA00aa_FRQ_CHU3.csv
2020-05-03_N00001_AA00aa_FRQ_CHU7.csv
2020-05-03_N00001_AA00aa_FRQ_CHU14.csv
2020-05-03_N00001_AA00aa_TMP.csv
2020-05-03_N00001_AA00aa_MAG.csv
2020-05-03_N00001_AA00aa_LGT.csv

LC PSWS 1st Line Metadata Format

First Row

The first row of the stored *.csv file will contain metadata information about the node collecting the data. It is from these fields that the information presented on the daily data plot will be taken from. This data is obtained every day from the *.txt files defined below for this station's info.

Structure is as follows (each field is separated by a comma so as to be compatible with the CSV format) Refinement of structure for fields follows in the Metadata Files and Contents section below.

- COL 1 – ISO format DATE Format: YYYY-MM-DD (10 chars)
- COL 2 – Node Number for station Format: N00000 (6 chars)
- COL 3 – GridSquare for Station Format: AA00aa (6 chars)
- COL 4 – Latitude for Station Format: sNN.NNNN (s=sign*, MIN of 4 decimal places)
- COL 5 – Longitude For Station Format: sNNN.NNNN (s=sign*, MIN of 4 decimal places)
- COL 6 – Elevation for Station Format: NNNNN (in meters, rounded to nearest meter)
- COL 7 – City and State for Station Format: City State (separated by 1 space, NO commas)
- COL 8 – Radio Format: Make Model (separated by 1 space, NO commas)
- COL 9 – Antenna Format: Make Model or brief description (spaces but NO commas)
- COL 10 – Freq Standard Used Format: type or Make Model (spaces but NO commas)

* - sign indicates:

+ [plus] for Northern Latitude, - [minus] for Southern Latitude, Range = +90 to -90

+ [plus] for Eastern Longitude, - [minus] for Western Longitude, Range = +180 to -180

An example of the 1st line [row] of days collection file:

2020-05-03,N00001,EN91fh,41.3219273, -81.5047731,284,Macedonia Ohio, LC PSWS Gen 1, 135 foot OCF dipole up 30 feet, GPSDO LB mini GPS

Second... Row(s)

The 2nd row can be a metadata file whose contents have each line starting with the 2 characters

#,

specifically the characters <hashtag><comma>

This allows the parser to skip over them for the end of day parsing task. This information is obtained from the file Metadata.txt. The file can have as many lines as wanted as long as each line starts with #,

It is also valid to have a zero length Metadata.txt file which will result in nothing being added to the header info of the days data file contents.

Operational Header Info

This row of header info defines the columns of stored data for this file. It is defined by the data engine storing the data every second (or specified time period) to the file. It always starts with the ISO time stamp as the first column.

It immediately follows the Metadata field(s) [if any exist]

- COL 1 – ISO Time stamp HH:MM:SS followed by a comma
- COL 2 – Column 2 Header Info followed by a comma
- COL 3 – Column 3 Header Info followed by a comma
- COL 4 – Column 4 Header Info followed by a comma
- ...
- COL N – Column N added as needed

For the different file headers, they would look like:

FRQ Example: UTC,Freq,Freq Err,Vpk,dBV(Vpk)

MAG Example: UTC, X(uT), Y(uT), Z(uT)

TMP Example: UTC, In Tmp (C), Out Tmp (C)

LGT Example: UTC, Dist (Mtrs)

After the operational header, the collected data set starts. Each second specified has its own row by itself. Rows do not wrap around to the next row.

For a typical given UTC day, UTC time will go from 00:00:00 to 23:59:59 for a total of 86,400 data points

Metadata Files and Contents

To assist the user in maintaining this information, the primary data collection directory ~/WWVdata/ (or usually /home/pi/WWVdata/) will contain the following files (and their defined formats):

- **CityState.txt** – Contains the City and State of the station separated by a space and NO comma
Example: Macedonia Ohio
- **CallSign.txt** – Callsign of the operator or station
Example: N8OBJ
- **LatLonAlt.txt** – Latitude, Longitude, Elevation of the station as defined above
A Comma separates the Latitude and Longitude as well as the Longitude and Elevation but NO comma after the Elevation
Example: 41.3219273, -81.5047731,284
Example:
- **FreqStd.txt** – Defines the frequency standard being used to make the measurements
Options are:
 - XTAL – standard quartz crystal oscillator
 - TCXO – Temperature Compensated Crystal Oscillator
 - OCXO – Oven Compensated (temperature controlled) Crystal Oscillator
 - Rubidium – Rubidium Frequency Standard (Make Model with spaces but NO commas)
 - GPSDO – GPS Disciplined Oscillator (Make Model with spaces but NO commas)
 - OTHER – custom Frequency standard (brief description with spaces but NO commas)Example: GPSDO Leo Bodnar mini GPS
- **NodeNum.txt** – Node number that has been assigned to this station (N12345) (6 chars)
Example: N00001
- **GridSqr.txt** – Maiden-Head Gridsquare Locator for station – 6 Chars in AA00aa format
Example: EN91fh
- **Radio.txt** – Radio Make and Model (spaces but NO commas)
Example: LC PSWS Gen 1
Example: ICOM IC-7610
- **Antenna.txt** – Make and Model of Antenna used or brief description (with spaces NO commas)
Example: 135 foot OCF Dipole 30 feet up
Example: Gap Titan

Initial Assignment of Node Numbers (and what they mean)

- N00000 – Test Node - data contained is for testing Low Cost PSWS - not to be used in real DB
- N00001 – N00049 – Initial development/testing teams nodes for low cost PSWS. Data is valid
- N00050 – Test Node – data contained for testing of Tangerine PSWS - not to be used in real DB
- N00051 – N00099 – Initial development/testing teams nodes for Tangerine PSWS. Data is valid
- N01000 – N04999 – Normal user installed nodes for low cost PSWS installations
- N05000 – N09999 – Normal user installed nodes for High Performance PSWS installations

Special significance will be given to N00000 and N00050 as test nodes for coding and testing development by both development teams.

The database should be able to sort and view all aspects of these database fields from a database engine designed to do these inquiries. Security to see contact info needs to be controlled based on who is using the query engine.

File Structure used on Low Cost PSWS

This is intended to be running on a Raspberry Pi running the Raspbian Linux Kernel.

This is the file structure currently in use on the Low Cost PSWS's presently (since July, 2019)

~ = user home directory for filesystem – which for user pi would translate to /home/pi

Since the PSWS network could contain Windows 7/10 systems, the use of a . [period] or other WIN7/10 incompatible characters in the filename are not allowed.

From here:

- ~/WWVdata/ - directory where all PSWS day's data files for data processing are stored
- for user pi this becomes = /home/pi/WWVdata/

Subdirectories of ~/WWVdata are as follows

- /WWVPython - where all python code is located for data processing
- /temp - daily cron status file (and temp location for data processing)
- /WWV2p5 - processed data and plots for 2.5MHz WWV data
- /WWV5 - processed data and plots for 5MHz WWV data
- /WWV10 - processed data and plots for 10 MHz WWV data
- /WWV15 - processed data and plots for 15 MHz WWV data
- /WWV20 - processed data and plots for 20 MHz WWV data
- /WWV25 - processed data and plots for 25 MHz WWV data
- /CHU3 - processed data and plots for 3.330 MHz CHU data
- /CHU7 - processed data and plots for 7.850 MHz CHU data
- /CHU14 - processed data and plots for 14.670 MHz CHU data
- /MAG - processed data and plots for Magnetometer data
- /TMP - processed data and plots for Temperature Data
- /LGT - processed data and plots for Lightning Data

This directory ~/WWVdata/ also contains the following files (defined above):

- CityState.txt
- CallSign.txt
- LatLonAlt.txt
- FreqStd.txt
- NodeNum.txt
- GridSqr.txt
- Radio.txt
- Antenna.txt

File Structure used on Tangerine PSWS

tbd...