# SPICE data archive and QC

Assumptions:

* Each site has an archive, where the site data will be stored as collected, together with instrument metadata, information of maintenance and other site logs (power outage, etc);
* Each site is able to transfer data automatically or manually (flash drive, removable hard drive, DVD, etc) to a central location with an Internet connection.
* Not all sites are equipped to perform data review and verification/QC.
* NCAR will store data and perform data QC, as defined by SPICE team.
* Env Canada may be able to offer similar capabilities, but would be a secondary archive facility to NCAR.
* All data will be in a common format as was established by NCAR at the Boulder meeting (see below)
* Will cover all data levels available: (Q: need to derive minute data, if sampling with a higher temporal resolution?)
* Data analysis will be performed using the Level 1 data.

Figure 1, below, is an illustration of the data flow, using Marshall as an example.

The specifics of each site must be added, for completing the picture of the SPICE data flow.

For illustration, Figures 2 and 3, outline the Marshall and CARE internal data flow.

Data QC should include the following levels:

* (Automatic) File presence and integrity:
	+ File exists?
	+ File size
	+ # of records
	+ check for valid data (range of expected values, bad data/error flags, etc.)
* (Manual) File content check:
	+ use web-based quick-view plots
	+ record integrity
	+ site checklists (daily, weekly, monthly, etc.)
* Site events: maintenance, power, failures, etc.
* Flag bad/missing data
* Correct data (if it can be corrected)

FTP Archive structure:

Site(SSS)/Year(yyyy)/ Sensor (SEN)

From sites with good Internet connections

daily /weekly files: SSSyyyymmdd.SEN

From sites with no internet connection: flash drive, by mail, etc.: develop scripts to take data into common looking data

Archiving scripts to search ftp server for today's data (regular intervals may be site specific)

Access to the data will be available via the NCAR SPICE page. Plots and data downloads will be available on demand.

**SPICE data format requirements**

Data shall be delivered as daily files, using the following file naming convention: <site>YYYYMMDD.<Instrument>

where <instrument> is a short description of the data, e.g. 20110621.GEONOR or 20101121.PWD

Files shall contain an ASCII representation of the data. Each line of data shall be delimited by a single 'newline' character, having an ASCII code of 0x0A (i.e. a decimal value of ten).

The top of each file shall contain five or more header lines describing the location where the measurement(s) were made, a short description of the instrument used for measurement, and a description of the data field(s). Header lines shall have a pound sign ('#') as the first character, and shall have the following format:

# <site\_ID> # <site\_description> # Latitude: <lat>, Longitude: <lon>, Altitude: <alt> # <dataset description>

where:

<site\_ID> is a short description of the site, using only alphanumeric characters (i.e. [a-z|A-Z|0-9] and the underscore '\_' character)

<site\_description> is a more verbose description of the site, which may contain spaces and punctuation

<lat> and <lon> are decimal position coordinates (e.g. -108.3754) <alt> is the height of the measurement, in meters above sea level <dataset\_description> is a description of the dataset, e.g. a description of the instrument used

for measurement. This may contain spaces and punctuation followed by one or more data field descriptors using the following format:

# <field\_number>) <data\_ID>, <data\_description>, <data\_units> where:

<field\_number> The field position, starting with field #1. (Does not include date and time fields) <data\_ID> is a short description of the data field, using only alphanumeric

characters (i.e. [a-z|A-Z|0-9] and the underscore '\_' character) <data\_description> is a more verbose description of the data, which may contain

spaces and punctuation OTHER THAN the comma (,) character <data\_units> contains the unit of measurement for the data (e.g. mm, m/s, etc).

All units shall be given in metric.

The header lines shall be followed by lines of data collected at the site on the relevant day, in ascending temporal order, using the following format:

YYYYMMDD, HH:MM:SS, <field1>, <field2>, ...

The time MUST be included for each line of data, and must be specified using UTC time. All data collected at this site at the given time shall be listed on a single line. Data from multiple instruments at a particular site can be uploaded as separate files, using the naming conventions described above.

Each data field shall be either a numeric or ASCII string value, and shall be delimited using the comma (,) character. Numeric values shall include only a combination of the characters 0-9, a negative (-) sign, and a decimal point (.). ASCII strings are permitted, but must be surrounded by double quotation marks ("). If data for a particular field is not available for the given timestamp, the field shall be populated with the string “NULL” (excluding quotation marks).

The following can be used as an example data file, with the filename '20110121.PWD':

# USA\_DIA1 # Denver International Airport Site

#1, Deicing Pad, USA # Latitude: 39.8679, Longitude: -104.6795, Altitude: 1615.0

# Data collected from a Vaisala PWD-22

# 1) Status, Status Message,

# 2) Vis\_One, Visibility 1 minute avg, m

# 3) Vis\_Ten, Visibility 10 minute avg, m

# 4) IPW\_NWS, Instant Present Weather NWS Codes,

# 5) Temp, Ambient Temperature, C

20110121, 00:00:00, "00", 20000, 20000, "C", -4.37

20110121, 00:01:03, "00", 20000, 20000, "C", -4.77

20110121, 00:01:15, "01", 18746, 20000, "-S", NULL

20110121, 00:02:06, "01", 14746, 19375, "S", NULL

20110121, 00:03:26, "01", 12342, 17432, "S+", -5.02



**Figure 1 -** SPICE data flow



Figure 2 - Marshall Internal Data Flow



**Figure 3 –** CARE Internal Data Flow