5.9 Tip Sheet: How do I load data using batch processing?

❖ Batch processing is used for loading electronic data files, generally from laboratories, into the database.

❖ The NWIS software accepts two batch file formats—the "1 and * card" format and the "tab-delimited" format.

❖ The "1 and * card" format uses a file called qwcards (or qacards for QC data), and the "tab-delimited" format uses files called qwsample and qwresult (or qasample and qaresult for QC data).

❖ Descriptions of the batch file formats are in *Appendix F*. A document that describes the requirements for external laboratories that may want to produce the batch format is located at *http://wwwok.cr.usgs.gov/nawqa/phoenix/*.

❖ To begin batch processing, choose option 8 – *Batch Processing* from the main menu. See *Section 3.8* for more details about this option.

❖ If you are loading data from the National Water Quality Laboratory (NWQL), you must first retrieve the data by choosing option 1 – *Retrieve National Water-Quality Laboratory Data* from the batch-processing menu. This will retrieve the available NWQL data to the current directory in the qwcards or qwsample and qwresult batch files needed by the batch-processing program.

❖ If you are loading data from a source other than NWQL, the program will look for the qwcards or qwsample and qwresult batch files in the directory where the program is initiated.

❖ The batch process can be run in one of two ways:

1. The batch process updates sample records that are already in the database (option 2 – *Process batch file for all logged in samples*). The data is matched based on agency code, site id, begin date and time, end date and time, and medium code which are contained in the 1-card (qwcards/qacards) or the sample-level file (qwsample/qasample). The result-level data (* cards in qwcards/qacards or qwresult/qaresult records) are appended to the existing sample.

2. The batch process enters and updates samples in the database (option 3 – *Process batch file for all samples*). The data are checked based on agency code, site id, begin date and time, end date and time, and medium code, which are contained in the 1-card or the sample-level file (qwsample/qasample). If a match is found, the result-level data (* cards in qwcards/qacards or qwresult/qaresult records) are appended to the existing sample. If a match is not found, a new sample is created and the result-level data are processed.

❖ During batch processing, the batch file(s) (qwcards or qwsample and qwresult) are separated into a file(s) to be processed in the environmental database (qwcards or qwsample and qwresult) and a file(s) to be processed in the QC database (qacards or qasample and qaresult). The separation is based on medium code. Medium codes 0-9 and A-P are used for environmental samples. Medium codes Q-Z are used for quality-control samples. A complete description of the valid medium codes can be found in *Appendix A*.

❖ If you are dealing only with QC data, you can run the batch files directly into the database using menu options 4 – *Process batch file for all logged in QA samples* or 5 – *Process batch file for all QA samples*.

❖ A record of the actions taken when a batch file is processed is written to files named *watlist.yymmdd.hhmm* **and** *watlist.qa.yymmdd.hhmm* for option 2 or 4 or *watlist.qwenter.yymmdd.hhmm* and *watlist.qaenter.yymmdd.hhmm* for option 3 or 5.

❖ The parameter codes for the parameters listed in the *watlist* are preceded by a one-letter code (N, U, C) to indicate whether the value is new, updated, or computed, respectively. The *watlist* also contains a cation/anion balance and messages from validation checks. The *watlist* should be reviewed for chemical accuracy and preserved as part of the official record for the sample as it is considered original data.

❖ Data for samples that are not successfully processed with the batch program are output to files for clean up (*badqw.yymmdd.hhmm* and *badqa.yymmdd.hhmm* if using 1 and * card format; and *rejected.sample.yyyymmdd.hhmmss*, *rejected.result.yyyymmdd.hhmmss*, *rejected.sample.qa.yyyymmdd.hhmmss* and *rejected.result.qa.yyyymmdd.hhmmss* if using the tab-delimited format).

❖ Samples that are not successfully processed have one or more of the following problems:
(a) invalid formats (columns have been shifted, etc),
(b) invalid site IDs, dates, times, or medium codes,
(c) results for samples that do not already exist in the water-quality file (except for menu options 3 and 5), and
(d) samples that contain results that are overwrite protected with a DQI value.
All results are rejected for a sample if a rejection error is found for one or more results.

❖ There is a batch-file editor available through the batch menu, option 7 – *Review/Edit batch files*, for fixing errors in files that are in the tab-delimited format. A text editor can also be used to edit the tab-delimited files, but the user should be extremely careful not to delete or add tabs used to separate the columns of data. Any text editor can be used for fixing data in the "1 and * card" format.

❖ To help verify the format of tab-delimited files, the program *qwcat* can be used. For more information about this program refer to *Section 3.9.*

❖ After the necessary steps are taken to correct any problems, the corrected samples can be entered by renaming the files to the appropriate batch filename (qwcards or qwsample and qwresult) and initiating one of the batch input programs (options 2 or 3).

❖ If data are rejected during the batch loading because of DQI protection but you have reviewed the new data and accepted it, you can reload the data from the batch file and override the DQI protection using option 9 – *Reload QW data from batch file, overriding DQI* or option 10 – *Reload QA data from batch file, overriding DQI*.