Share this article:

Datum et veritas - Having Faith In Your Data

Simply put, bad information begets bad decisions. Using inaccurate data as an information source puts you at risk for making the wrong decision. The importance of data verification is fundamental to all business and scientific practices, and has been the subject of many conferences and discussion groups. Yes Virginia, paper surveys are still alive and verification practices are an integral part of assuring high quality data capture.

Let's start by defining some terms:

- To "Verify" - verb: to ascertain the truth or correctness of, as by examination, research, or comparison (Dictionary.com)
- "Data verification" is a process wherein the data is checked for accuracy and inconsistencies after data migration is done. (Wikipedia)

When potential clients inquire about DataStar's data entry services, we often find ourselves saying "we are all about the quality of the data." What exactly does that mean? How do we get behind the quality of the data that we are contracted to enter? Clients will often want some reassurances about data quality. Not too long ago we received this email: "I wanted to talk with you about the auditing process for data entry. How do you go back to check that data is entered correctly, etc.?" Our answer: for manual entry of paper surveys, we use a process known as "100% keyverification." Simply translated, this is the act of re-keying or entering a data set a second time, completely, as it was done the first time, but with a tool that "verifies" if the second pass of entry matches the first.

Like many data entry providers, we use professional data entry applications which include verification steps. These tools allow one operator to enter a batch of documents following a screen map previously programmed according to a desired record layout. When the batch has been completely entered, a second operator (the "verifier") pulls up the first operator's file, which is "masked," or hidden from view. The verifier keys the batch again following the screen map. Every time he/she hits a key that does not match what the first operator entered, the program alerts the operator and stops. The verifier must then decide which response is correct by examining the source document. The change is then made and the verifier continues keying the rest of the document. Upon completion, a series of statistics can be run detailing the number of characters keyed per hour, the number of corrections made in the verifying step and more. This information becomes part of ongoing quality control efforts.

It is important to point out that 100% key-verified does NOT mean 100% perfect. It is generally considered to yield an average of 99.96% accuracy. Un-verified data entry yields 96% accuracy (when keyed by professional operators). Not too much difference you say? Let's do the math. If you have 200 surveys with 50 data columns (strokes or characters), this requires a grand total of 10,000 keystrokes to be entered. When the data is 100% verified you might expect to see about four operator errors at 99.96% accuracy. Compare that to un-verified data entry at 96% accuracy, which results in a whopping 400 mistakes!

Buyers beware of data entry providers who claim to do key-verifying or double punching. Sometimes it is not 100%, but rather portions of the data set (i.e. 20%) or just selected fields which are double keyed. Always ask if all of the data will be verified or double keyed.

Keying directly into a program such as Microsoft Excel, unless the operator is extremely careful and efficient, is not going to yield high quality and cost effective results. While it is possible to program in some checks and balances, spreadsheet applications do not have built-in verification features. Some companies may achieve 100% verification by keying two separate files for the data set and then matching the two files. This process shows every discrepancy. It is then up to somebody to pull and review the source documents in order to determine which answer is correct. This is a very time consuming process, with no improvement in accuracy over keyverification.

Regardless of the method used, a solid verification process is essential to ensure a high quality dataset for further analysis. To paraphrase Martha Stewart, "Good data is a good thing