

Share this article:   

Quality vs. Quantity - A Look At The Options For Data Capture (July 2010)

We are often asked about the pros and cons of scanning data as opposed to traditional data entry. If the quantity of survey forms and the schedule demand it, using an advanced data capture system can make sense, but in the majority of situations we prefer to employ manual data capture methods.

A previous StarTips article ("[Datum et veritas](#)") explained and promoted the use of 100% keyverification for manual data entry to achieve optimal data quality. The most commonly cited reasons to use scanning systems are speed and lower cost. These benefits are not mutually exclusive however.

There exist multiple types of scanning systems. The main techniques used for survey scanning are OMR (Optical Mark Recognition - reading check boxes), OCR (Optical Character Recognition - dealing with machine print) and ICR (Intelligent Character Recognition - automatically translating handwritten text into machine-readable characters).

Contrary to popular belief, scanning alone does not yield data of acceptable quality. It is still subject to a verification process. How does one verify scanned data? After the scanning and recognition passes, data is reviewed by an operator. This is usually called "verification," "reject repair" or "character correction." A verification screen displays both the recognition results and the original page or document image, so an operator can decide if the recognition was accurate. The problem responses, characters or fields are highlighted and operators can display the entire document image or zoom in for contextual information. Verification is a tedious process and in cases where there are numerous fields to be verified, the speed benefit of scanning can be lost.

The accuracy rate of properly scanned and verified data from survey forms is generally found to lie somewhere between 95% and 99%. Compare this to manual data entry which, when 100% verified, yields accuracy rates of 99.96%. The scanning process is subject to a variety of impediments and challenges, including improper marking, printing issues, or misfeeds due to torn/damaged documents, staples and paper clips. Any of these impediments may slow down the process or result in capture errors. Scanned paper surveys are by no means foolproof.

The relative cost of scanned vs. conventional data capture is largely a factor of volume. Scanning and machine recognition alone is a fast and inexpensive process. Employing verification techniques, which foster improved quality as described above, is more labor intensive and thus costly. Scanning typically has much higher fixed costs. The printing, programming and testing required of a scannable document generally make low volumes less cost effective. For very large quantities (i.e., >10,000), savings in processing will offset the fixed costs.

The same goes for speed - turnaround time for scan data capture may be faster, but generally only for very high volume needs. The bottom line is, never assume that a scan solution will be faster or less expensive. Traditional data entry remains a viable option for all but the most demanding schedule requirements.

Offshoring, or the export of data entry or scanning operations to regions where labor is cheap, is also a consideration, but is not limited to either method. We advocate transparency in offshoring - always be aware of where a data capture supplier conducts its operation. Without transparency, a low price is sometimes the only indicator of this practice. Keep in mind the added value of verification for both manual and scanned data capture - this step is too important to rely on the lowest rung of the global labor force.

Call us old fashioned, but it is still our opinion that many, if not most, conventional paper surveys are best handled using traditional data entry methods. When the quantity is small, manual entry is highly accurate and turnaround times are fast. The added human element remains an advantage in terms of decision-making ability vs. machine-readable data capture. However, with advanced planning for the design of the survey tool and an expectation of a high response rate, scanning may be a viable option for capturing data. Regardless of the method used, a solid verification process is essential to ensure a high quality dataset needed for further analyses.