

"If I'd Only Known the Smart Way to Estimate My Data Conversion Costs"

by James L. Miller III

Beth Fizzle (not her real name) had mountains of text—depositions, pleadings, statutes, etc.—she needed keyed or scanned into an ASCII data base. She contacted a vendor, who gave her an estimate based on "cost per thousand keystrokes," plus setup, programming, and delivery costs. There were costs added for the medium she wanted to have the data delivered in (her choices were floppy disk, CD-ROM, the almost-obsolete 9-track tape, or having the data downloaded directly to her hard drive). Beth went ahead and had the work done, paid the vendor's bill . . . and unknowingly wasted a lot of money. One day one of the firm's partners came into her office and said, "There are too many errors in this data. What is the accuracy level? What was your cost per thousand characters?" Beth shrugged her shoulders and responded, "Accuracy level? What's that? The vendor only talked about keystrokes. Is there some difference between characters and keystrokes?" The next day, while cleaning out her desk, Beth pondered the partner's questions again, still wondering what her former employer was talking about.

Here are some of the answers:

Definitions

Character(s). Any letter, special character (*,&,%,\$,#,@,?, et al.), punctuation mark, tab, or carriage return. Simply put, any "finger move."

Keystrokes. The same as "characters" for a single-pass keying operation. But a single pass does not include verification of the data. Higher levels of accuracy require multiple passes (and multiples of keystrokes). Using keystrokes as a measure of output at higher levels of accuracy instead of characters usually results in a distortion of the output and may produce an overcharge of the client (you)! You, the client, can count the characters that you receive, but how do you count keystrokes?

Accuracy

A grade of **99%** is a terrific grade in calculus, or chemistry, or almost any academic endeavor. But, it's a **lousy grade** if you are describing the accuracy of keyed data in an ASCII database. Examine the following table, based on

10,000 pages of double-spaced data, at 1000 characters (not keystrokes) per page:

Accuracy (%)	Errors per 10MM chars.
95.00	500,000
99.00	100,000
99.50	50,000
99.90	10,000
99.95	5,000
99.99	1,000
99.995	500

It is unfortunate, but true, that most of us have no idea of the accuracy of the work that we do, or of the data which we send out. But, it is critical to the reliability of our database (do you want to be sued because of an omission, misspelling, or misplacing a dollar sign, comma, or digit?). In the example above, at 95% accuracy, your 10,000 pages of data would have 50 errors per page! That's awful! Do you want your name associated with such poor work? Even worse, do you want to pay for such work? You probably have many times . . . and were not aware of it until complaints surfaced about the unusual number of errors. But, wait a minute! Is 99.00% that much better? Or 99.5%? Or 99.9%? **Even at 99.5% you may have five errors per page, and 50,000 errors in the entire data set!**

I'm sure that you are getting the point, which is that reliable databases require an accuracy level of 99.95% or higher. Most online legal databases (e.g., WESTLAW) are even higher than 99.95%. Further, most databases with extremely high accuracy have been keyed, not scanned. If they have been scanned, they have also required substantial "cleanup" from an editor who keys in the corrections. Scanning and editing can be an expensive, time-consuming process at the higher accuracies.

Defining the Task

It is not particularly easy to make an accurate estimate of the hours of work required to key your conversion project. But, it is straightforward! To begin, you need the following:

- Type(s) of document
 - Full text
 - Depositions, pleadings
 - Statutes

Quality of source document

- Original
- First copy
- Second, third, fourth copy
- Printed, typed, or typeset

Density of printed material

- Single-spaced
- Double-spaced
- Multiple columns
- Size of print

Turnaround

- One week
- Two weeks
- Six weeks
- Three months
- Needed it yesterday

Accuracy

To return to estimating the task, one begins by counting (or estimating) the number of pages to be keyed. Then, use the following **rules of thumb:**

- 1200 characters per page for double spaced pages, with 10–12 point type. Twice that for single spaced pages. For two columns of single spaced print, use 4500 characters. Or, you may count the characters on a "typical" page.
- Use 5000 characters per hour of work as the production rate for keying at 99.95% accuracy.
- Use 3300 characters per hour for 99.99% accuracy.

So, the basic estimate for keying 5000 pages of depositions at 99.95% accuracy is as follows:

5000 pages x 1200 characters per page = 6,000,000 characters.

6MM characters divided by 5000 characters/hour = 1200 hours.

1200 hours/40 hours per week = 30 weeks to complete the task.

(Note: The hour estimate does not include any costs for supervision, editing, quality control, etc.)

Tips from a vendor on the practicalities and pitfalls of data conversion

Reality

You can now take the estimated number of hours for the task and multiply it by an estimated (or actual) hour cost for the local area (e.g. \$15/hour, \$20/hour) to arrive at your estimated cost, which in this case is \$30,000.00 or \$40,000! That's a lot of money for only 6000 pages of documentation! Too much, you say. Well, add in a factor of 25% for a two-week turnaround and another factor of 30% for poor quality source documents and your cost has become \$46,500 and \$62,000, respectively. You are up to \$10 per page! Waw! You can't afford this, but you have to get the job done, don't you?

Many keying and scanning businesses do not guarantee an accuracy level. Most, in fact, don't quote prices that include a guaranteed level of accuracy. They use keystrokes, turnaround, number of records or fields, and document quality (ease of keying). But, accuracy should be essential in discussing delivered keying or scanning jobs!

What alternatives do you have regarding the high costs of keying? Well, there are many keying operations around the world,

referred to as "offshore." Some of them have developed excellent reputations over many years for high quality, fast turnaround (depending on the volume), and prices which are 1/3 to 1/2 of those in the USA. You should contact one of these top-of-the-line offshore installations whenever you are considering a conversion project which has significant volume, and therefore, significant costs. The savings makes the project affordable, and the savings go directly to your firm's profitability!

I often tell my clients: "Price, accuracy, turnaround. Pick two!" The point here is that to have your job done quickly, at a high level of accuracy, and at a bargain price usually doesn't happen.

Final words of caution:

- Be very careful in dealing with a vendor who will not discuss accuracy with you, and will not guarantee the level of accuracy which you require.
- It is usually better, particularly with full legal text, to estimate your task in terms of "per thousand

characters." You can extrapolate to create a price/page if it is needed. Using records or fields does not work well when dealing with full text work such as is the case with most legal documentation.

- Shop around for a keying vendor. Ask as many detailed questions as you can think of. And, be suspicious when a vendor starts adding cost elements into your estimate. Make the vendor price your task in terms of a "per thousand character" price which includes all of the elements in his pricing scheme. A good vendor will give you a composite price which will include everything, and on which you can depend.

James L. (Jim) Miller III

(sundata@sundcs.com) is President of SUN Data Conversion Services, Ltd., a Dallas-based offshore company. He has also been a university professor (accounting, finance, economics, ethics). He can be reached at 800/626-9058.