

## Cost Comparison of Four Data Input Methods

DONNA M. MENDENHALL

Uniroyal Chemical, Division of Uniroyal, Inc., Naugatuck, Connecticut 06770

Received April 30, 1974

**A study of four methods of input of Wiswesser Line Notations is presented in which costs for each method (keypunch, magnetic card Selectric typewriter, cathode ray tube, and optical character recognition) have been determined. Equipment associated with optical character recognition was found to be the most efficient, versatile, and acceptable as well as least expensive.**

The purpose of this study was to consider relative costs of input by different methods of WLN notations of company-prepared compounds. Wiswesser Line Notation (WLN)<sup>1</sup> was the notation system chosen because of its minimal keyboard requirements.

A limitation of five years was placed on the study of the literature because of changes in the technology and the costs involved. This study revealed input methods for chemical compound information which included the chemical typewriter,<sup>2</sup> a modified teletype,<sup>3</sup> a magnetic tape Selectric typewriter,<sup>3</sup> an IBM communications terminal,<sup>4</sup> a paper tape typewriter,<sup>4</sup> and the keypunch.

To determine which available equipment to adopt for our purposes, four types were studied: keypunch, IBM magnetic card Selectric typewriter, cathode ray tube display station, and optical character recognition equipment.<sup>5,6</sup> Data for each type were compiled in four categories: cost, efficiency in use, versatility, and acceptability by the input operators.

Cost was determined as the sum of the time spent in coding, proofing, transcribing, keyboarding, and verification leading to a computer-readable product, multiplied by an average wage rate for the region. For the sake of simplicity in costing, the keyboarding for each unit was held constant and well below what can be expected by professional typists or keypunch operators. In this facility, input by typewriter or keypunch was found to be in the order of 75/hr for both methods. However, a study exists which shows typing to be twice as fast as keypunching.<sup>7</sup> In the following, each input method is discussed separately. The costs are calculated for the machine use alone, including either lease or purchase arrangements. All other costs of the notation process are summarized in the later section titled Labor Costs.

Efficiency in use was considered to include the elements of fatigue on the part of the operator, the proximity of the equipment, and the ease of installation of the equipment. Speed in adapting to the keyboard was also weighed.

Versatility was defined as including the possibility of multiple usage and the ease of installation (again) of the equipment. The feasibility of entering with the same type of equipment all screening data generated at different locations within the company was also included as an element of versatility.

Acceptability was judged as a subjective analysis by users, current and prospective. This would reflect the combination of factors described under versatility and efficiency. The small size of the Information Center-Library precluded training assistants for the clerical portion of the notation input. Thus the coding and keying operations and subsequent editing of the notations, serving as repeated

checks, were done by one person. This is schematically illustrated in Figure 1. The first step in the process of input of data consists of reproducing the original documents borrowed from a commodity group (at a rate of 500/hr). This is followed by a rearrangement of the documents according to WLN type, which is done at a rate of 500/hr. Notations are then assigned at a rate of 100/hr, followed by a visual review (200/hr). This error checking loop is also noted in the flowchart.

After these preliminary steps, Figure 1 depicts the various steps necessary for the four different input methods: (1) keypunch, (2) optical character recognition (OCR), (3) cathode ray tube (CRT), and (4) magnetic card Selectric typewriter (MCST). Where applicable, the rates at which work progressed are included in the flowchart.

### CATHODE RAY TUBE (CRT) TERMINAL

This involves on-line data entry by communication with devices peripheral to the central processing units of the company computers, requiring installation of telephone lines and modem (data set). For purposes of calculation of on-line charges, I have used a hypothetical facility preparing approximately 2500 compounds per year in a steady flow. Each individual can calculate his own usage of on-line time with these hypothetical figures as a base.

IBM 3275 CRT lease	\$140/mo.
Phone line lease	20/mo.
"On-line" charges (\$/hr, 4 hr/mo.)	8/mo.
Total cost	<hr/> \$168/mo.

The efficiency and versatility of the CRT were judged good. The fatigue factor inherent in long-term visual observation of the CRT left it only barely acceptable overall. It has overstrike correction-playback features as does any terminal typewriter. It would seem that this high-glamour instrument has drawbacks in this application.

### MAGNETIC CARD SELECTRIC TYPEWRITER (MCST)

This typewriter was rated as the most versatile of all modes considered. It can serve as a terminal as well as a standard office typewriter. Playback capacity *via* the modem, which is a function of all terminals, permits off-line corrections and provides for reformatted or corrected data entry. The MCST was also rated as excellent in the categories of efficiency and acceptability. The cost of the magnetic cards, \$1.25 each, has not been included as they

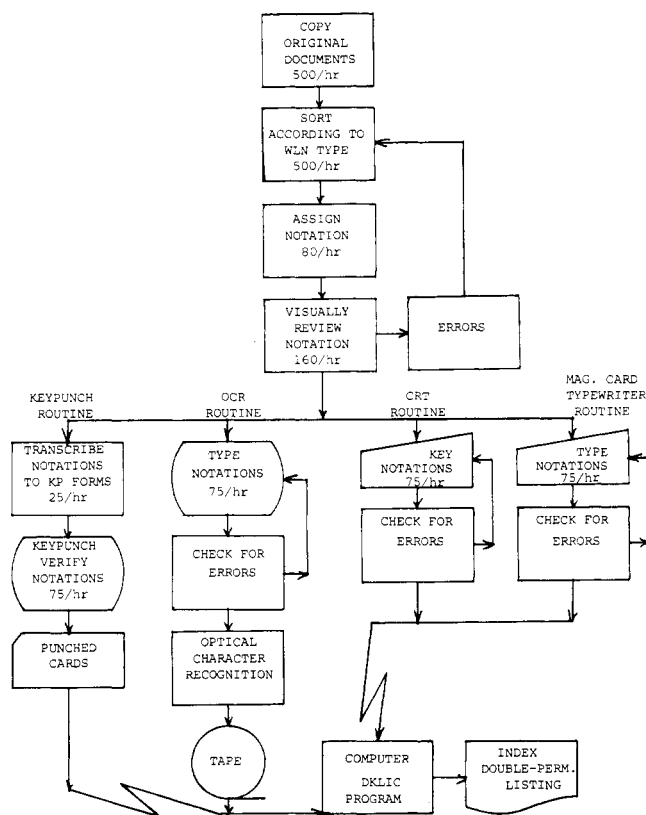


Figure 1. Flowchart for steps in input of WLN data.

are reusable for any other applicable usage. Its cost factors include

Lease	\$245/mo.
Phone line	20/mo.
"On-line" charges	8/mo.
<b>Total cost</b>	<b>\$273/mo.</b>

On-line charges have been estimated here as one-third of the costs of on-line charges used for the CRT calculation with the same hypothetical facility being considered. The time required for batch-type entry of one or more magnetic cards containing the recorded data is a function of the terminal speed, 10-30 CPS. This is in contrast to on-line methods where the limiting factors are the speed of the typist and the computer software.

#### OPTICAL CHARACTER RECOGNITION (OCR)

This type of equipment, now beginning to come into its own, comprises a computer that can "read" the alphabet, numerals, and special characters and translate them into standard configurations on magnetic tape. Leasing begins around \$3000/mo., but one can type the data with any typewriter with an American Standards Association OCR-A typefont, which can be leased (\$22/mo.), and have the pages scanned-to-tape on a service bureau basis for \$23/1000 lines and a nominal one-time programming charge of \$100.

Figure 2 is a reproduction of a page ready to be scanned-to-tape, illustrating the field separator (the vertical bar), and one correction overstrike feature (the black rectangle or "blob"). The latter obliterates an error, following which is the correct symbol; a string of "blobs" is acceptable to the OCR equipment. Thus the entire record need not be re-typed as must be done with a punched card, saving staff time. The length of each line is 80 characters, 10 to the inch.

```

12345|TSMV DN CHJ CMVZ EO & AL2 Q4 G
12346|ZR D & C..N-03*2
12347|QV1XQR DQ&R DQ
12348|GR DNUNO&R D&G
12349|LVMR CZ BVR& EV01
12350|LY&LN1&2X&&10P0&OR CG DX&&&0LY&LN1&1X

```

Vertical bar

Figure 2. Illustration of WLN typed for OCR.

#### KEYPUNCHING (KP)

The keypunch method for data input was judged barely acceptable in all three categories: efficiency, versatility, and acceptability. The careful hand lettering of notations onto keypunch forms often necessary in keypunching data, when it or other keyboard operations must be performed outside, contrasts with the ability to assign Wiswesser Line Notations at a rate of 100/hr.

Versatility was judged low principally because the equipment might be installed in a location to which documents would need to be moved and also because the keypunch can serve only one function. Cost of the KP method was moderate at a total rental of \$77/mo.

#### LABOR COSTS

Figure 1 is sufficiently detailed so that the reader can insert his own staff costs. The fact that a professional handled all processing of the notations, where other information centers could make less expensive arrangements, should be borne in mind. The hourly estimates of work done are conservative since each step was carefully and deliberately done so as to minimize input errors. The typing step also served as an additional notation check. Table I presents the basic labor costs for preparation of one notation for input.

Table I. Basic Labor Costs

Preparation of One Notation for Input		
Assign WLN	100/hr @ 6.00/hr	6¢
Review WLN	200/hr @ 6.00/hr	3¢
Key WLN	75/hr @ 6.00/hr	8¢
Transcribe WLN to KP forms (optional)		12¢
Total costs without transcription		17¢
Total costs with transcription		29¢

tion for input. It can be seen that the assignment and review of a single WLN involves a total cost of 9¢. Keying of the resultant WLN from the original documents by a keyboard service might be considered, if the confidentiality of the material and/or location of the keypunch posed no problem. Two local keypunching firms contacted offered rates of 4-5¢ per notation. If transcription of the WLN to keyboard forms is required, an additional 12¢ per notation is added. Labor costs are therefore 17¢ per notation without transcription, 29¢ if it is included. The transcription from the original document is seen to be an expensive procedure.

#### CONCLUSIONS

Table II, Overhead Costs, summarizes all the costs associated with notation input. The cost of transcribing the notations to keyboard forms was omitted. It is clearly evident that the input costs per notation do not vary widely.

# COST COMPARISON OF FOUR DATA INPUT METHODS

**Table II.** Summary of Costs for Input of WLN

	KP	OCR	CRT	MCST
Basic costs per notation	\$0.17	\$0.17	\$0.17	\$0.17
Scan-to-tape per notation (Service Bureau)		0.023		
"On-line" charges per notation (@ \$2.00/hr)			0.026	0.009
Total cost per notation	\$0.17	0.195	\$0.196	\$0.179
Overhead, cost per month	\$77.00	\$22.00	\$168.00	\$273.00

**Table III.** Rating of Input Modes for WLN Data<sup>a</sup>

Mode	Cost	Efficiency	Versatility	Acceptability
Keypunch	\$77/mo.	D	D	D
OCR	22/mo.	A	A	A
CRT	560/mo.	B	B	C
MCST	305/mo.	A	A	A

<sup>a</sup> A, excellent; B, good; C, fair; D, barely acceptable.

In this study the basic presumption is that the actual key-boarding operation for each type of input device is similar.

The wide range in monthly out-of-budget charges reflected in the last line of Table II makes the typewriter and the optical character recognition on a service bureau basis the easiest method to present to management. The OCR method is also easily the method of choice if cost is coupled with the other factors of efficiency, versatility, and acceptability, as shown in Table III.

## LITERATURE CITED

- (1) High, J., and Granito, C. G., "Wiswesser Line Notation Workbook," C. G. Associates, Cherry Hill, N. J., in press.
- (2) Jacobus, D. P., *et al.*, "Experience with the Mechanized Chemical and Biological Information Retrieval System," *J. Chem. Doc.*, **10**, 135 (1970).
- (3) Feldman, A., "A Chemical Teletype," *J. Chem. Doc.*, **13**, 53 (1973).
- (4) Culkowski, A. H., *et al.*, "A Study of Four Methods of Entering Text into Computers," AEC, Report ORNL-TM3600.
- (5) Scan-Optics, Inc., 22 Prestige Park, East Hartford, Conn. 06108.
- (6) Stiles, H. J., and Maier, J. M., "Automated Cataloging of Technical Reports Via Optical Scanning," *Spec. Libraries*, **63**, 576 (1972).
- (7) Skolnik, H., and Snyder, J. C., presented in part at the "Large Data Bases" conference, Washington, D. C., May 1974.