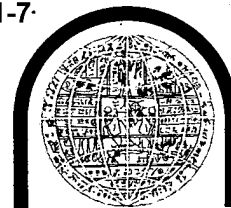


AREA C



*"Record Protection in
an Uncertain World"*

WORLD CONFERENCE ON RECORDS AND GENEALOGICAL SEMINAR

Salt Lake City, Utah, U.S.A.

5-8 August 1969

COMPUTER INPUT AND OUTPUT TOOLS AND TECHNIQUES

Part II

Automatic Document Reading; Optical Scanning, an answer
to the Computer Input Problem?

by Melvin M. Farr, B.A., J.D., LL.M.

Part II
Automatic Document Reading; Optical Scanning, an answer
to the Computer Input Problem?

by Melvin M. Farr, B.A., J.D., LL.M.

INTRODUCTION

Much has been said and written about the population explosion, but an even larger explosion is the "Information Explosion". It is reported that every five years the amount of printed information doubles. More and more of this information is finding its way into the computer, along with myriad of other information; such as information about bank accounts, financial records, traffic violations, etc.

Keeping pace with this information explosion, computer manufacturers continue to announce faster and faster computers. They have developed disc drives that read in one hour, information that would take one keypunch operator 85 years to prepare. This being the case it is easy to see why "Input" is considered to be one of the key bottlenecks in computer systems.

It was the recognition of this bottleneck that caused the development of various optical scanners. Each striving for the ideal in systems design, "the capturing of data at its source in a machine readable form.". Some have been more successful than others in this attempt and it is this fact that makes it difficult for potential users to know

what to do in regards to optical scanning. If there were but one scanning company, with only one machine, reading only one type of recorded media it would be relatively easy for one to determine whether or not he had a scanning application; and once an application was recognized there would be no company selection problem. This, of course, is not the case. There are several companies, some having more than one machine, having different "capabilities", and reading different types of recorded data, making it difficult to know what to do in regards to optical scanning.

The purpose of this paper is to present general information on scanning with the hope that it might aid those who have potential scanner applications to understand some of the factors they should consider in determining if optical scanning can aid them in reducing the "Input" bottleneck.

I. FINDING A SCANNER APPLICATION

The potential for scanner use, obviously, exists anytime there is a system that does tie, or should tie into a computer. The ideal as mentioned before, is where the data can be captured at its source in machine readable form. e.g., the first time the information is recorded on paper have it in a form, mark sense, handprinting, typewritten, etc; so that it can be read by machine and put into computer

readable form. In determining whether there is a scanning application several factors must be considered. Some of them are: The type of scanning equipment, the system itself, the capability of the people working the system, the cost of accomplishing this by scanning vs. other methods.

A. TYPES OF SCANNING EQUIPMENT

1. Bar Code Readers. This type of equipment, reads "Bar Codes" printed mechanically. This is usually accomplished through manually operated imprinters or "Data Recorders."

A well known example of this method, is some oil company credit cards. Account identification information is embossed on a card in bar code form (a series of vertical long and short dashes,) that represent an account number. This card is put in an imprinter at the time a sale is made and variable information, i.e., the amount being charged by the customer is entered into the imprinter key board by the station attendant. The charge slip is put in the imprinter and when operated the imprinter transfers the account number, amount of purchase, plus other information to the charge slips. The record of purchase is put through an optical scanner and the information is converted to cards, paper tape, or magnetic tape and from these entered into the computer for processing.

This is a good use of Bar Code Readers, in that it permits entry of the data into machine readable form at the source where the transaction takes place and the first time it is recorded. In addition, the cost of the imprinter is not great (less than \$100.00) and it is very simple and fast, yet reliable.

2. Mark Readers This equipment reads pencil marks e.g./- , entered by a person on a card or paper. These marks are entered on the forms in predetermined fields (areas on the forms). The presence of a mark in a given area serves to represent a given quantity or meaning, dependent upon the way the system is designed. A very common use for mark readers is in connection with meter reading of gas or power company meters. The "Meter Reader" marks the meter value on a customer account card, e.g. The customer card would probably have four vertical columns of numbers, each column having the digits 0 to 9 down the card. To represent the numbers 2468 the meter reader would put a mark through the 2 in column one (the column farthest to the left). Through the 4 in the second column, through the 6 in column three, and through the 8 in the fourth column (farthest to the right). The card is then passed through an optical mark reader and the information is punched into the card, converted to paper tape, or magnetic tape, so that it can

then be processed by computer. (Some mark readers read optically, while others read it by "sensing" the marks on the paper or card. For the latter, special electrographic pencils must be used.)

A big advantage of this approach is that the "Imprinter" (pencil) is very inexpensive and mobile, permitting the person, responsible for making the entry, to go to many places to gather information, and there make an entry that, at its source, is machine readable. A problem with this method is that in dealing with marks one is dealing with something with which people are not accustomed and it is therefore, more susceptible to people errors.

3. Character Recognition Equipment These machines read hand printed or machine printed characters. The object of this equipment is to permit a person to record data in the form in which humans are accustomed to seeing and using it, and then have the scanning equipment "translate" it to computer language for processing.

HAND PRINTING

Hand printing technology is in an infant stage, and requires that the person printing use great care in the preparation of the data. Reading of handprinted alpha characters is not yet very good. Since this is the case, applications using hand printing are rather limited; though

there are some where department store clerks hand print customer account numbers and purchase amounts on a sales check, which is then optically scanned and converted to computer language for processing. Hand printed character recognition holds out promise since there are so many businesses where the source of entry is made by a person using a pen or pencil. As with mark readers, the input device (pencil) is very inexpensive and mobile. Unlike the mark readers, however, by reading hand printed characters information is read in the form in which people are accustomed to seeing and using it. For this reason it is easier for people to work with, and since it is in "English" a copy of the hand printed entry form can be given to a customer as a receipt.

MACHINE PRINTING

The reading of machine printing is quite advanced and several companies have announced equipment that is capable of reading machine printed characters. Most of the machines are single font machines. These are machines that are capable of reading only one font. A font is a set of characters of the same type, size, and style, such as Underwood Distinctive Elite, IBM Prestige Elite, etc. There are, however, some "Multifont" scanners that are capable of reading two or more fonts; an example of one of these is

the machine being installed at Management Systems in Salt Lake City, Utah. It is capable of reading 19 different typewriters (Those typewriters that have basic pica or elite fonts), 1403 type style, and USASI-1 type style. It reads pica and elite upper and lower case characters along with certain special characters.

Some of the applications using machines to read machine printed information are:

- (1.) To read text material prepared by typewriters for newspapers.
- (2.) To read customer bills being returned with a payment, the bill having been printed on a 1403, mailed to the customer, returned, and the amount of payment encoded on the bill with a bank proof machine.
- (3.) To read airline passenger and audit tickets.
- (4.) For the conversion of large amounts of information to computer readable form. This is accomplished by the use of typewriter and scanners rather than keypunching.

Advantages that exist for the use of scanners for the conversion of large amounts of information over keypunching are as follows:

- (1.) The input device (typewriter) is much cheaper than a keypunch machine, and more mobile.
- (2.) The amount one has to pay for a good typist is less than that paid a good keypunch operator, due to the fact that typing is easier, and a typewriter takes less skill to operate efficiently than a keypunch.
- (3.) It is easier to correct typing errors than keypunching errors.
- (4.) Typing is faster than keypunching.
- (5.) Record length is practically unlimited (not limited to 80 characters as in keypunch).

4. MICR READERS (Magnetic Ink Character Recognition) These are not "Optical Scanning" machines, since they do not read optically; but through the recognition of "magnetic" patterns as the paper is passed under the magnetic head with the head touching the paper. These are used mostly in the Banking Industry and since they are not optical scanners will not be discussed further.

B. The System Itself In considering whether there is a scanner application, the first question, that must be answered is "should the information being considered be processed by computer?" If it should be, then the next step is to "determine where the information originates" Once the origin is found, the question is, "can the system be

designed so that the original entry is in a scanable form?" The greatest possible benefit is derived from scanning if the source record, that has to be prepared as a basic element of the system, can be prepared in scanable form. An excellent example of this is the above mentioned gasoline credit card application. As a part of doing business the gasoline attendant must:

1. Record the purchase as to
 - a. The amount of purchase
 - b. The customer making the purchase
2. Give the customer a receipt of the purchase for his record.

All of these are accomplished, at the source, through the use of an embossed card and an imprinter. In addition, the method used is about as simple as could exist, considering what has to be done.

If the source data cannot be entered in machine readable form, but must be "copied" (keypunch, etc.) for computer use, scanning could still be the best method because of the above mentioned advantages of typing over keypunch.

C. CAPABILITY OF THE PEOPLE WORKING THE SYSTEM

A system should not be built around personalities or individuals, but one must consider the capabilities of the type of person that is to perform the job, and the

environment in which the job will be performed. For example, in the gasoline credit card application, factors to consider and test are:

1. Can the station attendant, who probably has only a high school education or less, accurately operate the imprinter.
2. Will his hands be clean enough so that he will not get grease on the charge slip, that might interfere with the scanner's reading of the slip?

If this application required that the attendant type the information, it would probably not work, since relatively few attendants know how to type accurately and it would also be too slow.

Tests should be conducted to determine whether the typical employee working the system has the capability to prepare the entry in machine readable form.

D. COST OF SCANNING VS. OTHER METHODS

Once it has been determined that a potential scanning application exists, the question of cost comes up. Although it has been said that typing has advantages over keypunch; if the volume of the input data is sufficient, say, for only one keypunch machine, it would be difficult, if not impossible to cost justify a scanner. Therefore, one should conduct a study to determine the number of transactions

needing entry to the computer, along with the average number of characters per transaction. Then using this information, determine the comparable cost for various methods of input.

Some factors to be considered are:

1. Can the source entry be prepared in a scannable form, as a by product of the normal system, thereby eliminating the need to "copy" it (keypunch, etc.)
2. What will the error and reject rate be on the application if a scanner is used? In scanning an error (substitution) is where the scanner mis-reads a character, e.g.. a B as an 8 an E as an F, etc. A reject is where the scanner cannot read a character and therefore outputs a reject symbol, such as a question mark, in place of the character.
3. Under the system as planned, what is the method (and its cost) of finding and correcting errors?

Some possible methods are as follows:

- a. As the material is read on the scanner a print-out of what was read by the scanner is prepared on a high speed printer. This print-out is visually compared against the transactions and errors and rejects are noted. Corrections are then typed and re-entered into the system.

- b. The scanner output tape is edited on the computer for obvious errors.
 - (1.) Incorrect account numbers (found through use of check digits.)
 - (2.) Incorrect amounts. (found through batch balancing procedures)
 - (3.) Numeric information in an alpha field, etc.
 - c. A combination of the above two methods
4. What method will be used for entering reject characters? Some possible methods are:
- a. Entry in connection with error correction as discussed above.
 - b. Through the use of a CRT terminal, the transactions are displayed and the reject characters are entered through the terminal.
- NOTE: Consideration should be given to the effect of an error or reject on the system if it is not corrected. It might be found that errors and rejects in some fields will not be a problem. If this is the case those fields would not be checked for errors or rejects.

5. The number of keypunches and verifiers (operators and equipment) that will be displaced if scanning is used.
6. The cost of the scanner rental. In figuring this, recognize that in an eight-hour day the scanner will not be running eight hours due to forms handling, paper jams, etc.
7. Scanner operator costs, including overhead; such as sick leave, vacations, holiday, and supervision.
8. Space cost for the equipment, keypunch, scanner, etc., along with facility costs, such as air conditioning, power, etc.

II SELECTING THE OPTICAL SCANNER

Once a scanning application has been found, and it is known what type of scanner is needed, e.g. bar code reader, mark reader, etc. the next step to be taken is to select the specific scanner to be used.

Some of the steps to be taken in this phase are the following:

- A. Select from the many companies, three or four that have a scanner to fit the system. There are many scanning companies manufacturing many different types of scanners, at different prices. Some of the companies are:

ADDRESSOGRAPH MULTIGRAPH CORP.

1200 Babbitt Road
Cleveland, Ohio 44117

COGNETRONICS CORP.

333 North Bedford Road
Mount Kisco, New York 10549

CONTROL DATA CORPORATION

8100 34th Ave. 5
Minneapolis, Minn. 55440

CUMMING-CHICAGO CORP.

4740 N. Ravenswood Ave.
Chicago, Illinois 60640

FARRINGTON MANUFACTURING CO.

5881 Leesburg Pike
Falls Church, Va. 22041

GENERAL ELECTRIC CO.

3500 North Central Ave.
Phoenix, Ariz. 85012

HONEYWELL EDP

60 Walnut Street
Wellesley Hills, Mass. 02181

IBM CORP.

112 East Post Road
White Plains, N. Y. 10601

INTERNATIONAL COMPUTERS LTD

839 Stewart Ave.

Garden City, Long Island, N. Y. 11533

MEASUREMENT RESEARCH CENTER

P.O.B. 30

Iowa City, Iowa 52240

THE NATIONAL CASH REGISTER CO.

Dayton, Ohio 45409

NATIONAL COMPUTER SYSTEMS

1015 South 6 Street

Minneapolis, Minn. 55415

OPTICAL SCANNING CORP.

Newton, Pa. 18940

PHILCO-FORD CORPORATION

3900 Welsh Road

Willow Grove, Pa. 19090

R.C.A.

Cherry Hill, New Jersey

RECOGNITION EQUIPMENT INC.

1500 West Mockingbird Lane

Dallas, Texas

REPUBLIC ADVANCED TECH. SYSTEM GROUP

9754 Deering Ave.

Chatsworth, Cal. 91311

SCAN-DATA CORP.

800 East Main Street

Norristown, Pa. 19401

UNIVAC DIVISION OF SPERRY RAND

Roseville, Pennsylvania

A letter could be written to the above companies, asking for information about the particular scanner, bar code reader, mark reader, etc. needed for the application. Information about the application should be passed on to the companies with this letter. In response to this letter it should be possible to narrow down the number of companies that have scanners that interest you. Factors to consider are:

1. If your application is a bar code application, does the company have a bar code reader?
2. What is the speed of the scanner? It must be fast enough to handle your daily transactions but not so fast that there is too much excess capacity.
3. What is the cost of the scanner? Does it fit in with the cost estimates you have made.

B. Run Tests With the Scanning Companies.

Once the selection has been narrowed down to a few companies, tests should be run with these companies to satisfy yourself that their "promises" in fact, can be met.

Before running a test it is important to find out from the companies:

1. The Brand of paper to be used (should it be bond or sulphite?). Most paper companies put out a scanning paper and the scanning manufacturer should know which is the best for his particular scanner.
2. The color of inks that are "blind" to the scanner. The form should be printed in this color.
3. The minimum and maximum form size.
4. The Location of the "read" areas on the forms. (Margin requirements, etc.)
5. If a machine print reader, what type of ribbon should be used (carbon or fabric ribbon?) What is the best brand?

C. Visit users of the particular scanner you are interested in. Upon narrowing your choice down to a particular scanner, to help you finalize your selection; it is well to make arrangements to talk with people who are using the scanner you have selected. This can be done through correspondence, but is better accomplished via telephone, and at least one visit to a user location is recommended. In this manner much valuable information can be gained from those who have "been down the road" and can point out pitfalls to watch out for.

III SIGNING A CONTRACT WITH THE SCANNING COMPANY.

When the particular scanner has been selected a contract should be signed with the scanning company. Elements that should be present in the contract are:

1. Whether it is for a lease or purchase (if for a lease, a purchase option should be provided for).
2. Maintenance support to be provided. What and at whose expense?
3. What customer training and support is to be provided by the scanning company.
4. Delivery date.
5. What minimum facilities must be provided for the scanner, and at whose expense.
6. Cost of the scanner.
7. Who is responsible for delivery costs (the purchaser usually bears the expense).
8. Who is responsible for providing insurance on the equipment?
9. An understanding as to cost for running second and third shifts including maintenance costs.
10. Guaranteed performance level for your particular application.
 - a. Minimum numbers of documents to be processed in a given time.

- b. Maximum number of errors and rejects per given number of characters (or maximum number of documents with rejects out of a given number of documents).
 - c. Mean time between failure
 - d. Mean time to repair the equipment.
 - e. Size of documents handled.
 - f. Paper weight, etc. handled.
 - g. Jam rate of the paper handler.
 - h. Multiple feed rate of the paper handler.
 - i. Other criteria relied upon by you in selecting the particular scanner.
11. An understanding as to an acceptance test, that will measure the scanner's performance as agreed upon.
12. Other items as agreed upon.

IV. ACCEPTANCE TEST

When a contract has been signed for the purchase or lease of a scanner, arrangements should be made for an acceptance test that thoroughly tests the performance of the machine, to make certain that it meets all the agreed upon performance levels. If at all possible the test should be

conducted, using "live" data, of a large enough sample to cover all conditions that might arise. If live data cannot be used, then data should be used that imitates as nearly as possible "live" conditions. In planning an acceptance test it would be well to formalize it by putting the plan in writing. The plan should clearly indicate how the acceptance test will be conducted including a clear definition of each one's responsibility (the scanner vendor and the customer).

CONCLUSION

Optical scanning is a very good solution to the input bottleneck, when it is properly applied. As with any other good business tool it has its limitations and it is therefore important that careful study and thought be given to any potential scanning application, to make certain that the right equipment for the right job are brought together.