these plays in the decade to come. Then perhaps we can begin to think of a definitive work on the development of Canadian drama and theatre.

Patrick B. O'Neill
Mount Saint Vincent University
Halifax

OMR Recording System for Archival Data

Historians face a number of problems in obtaining quantitative data from archival sources. Particularly troublesome are the complex internal accounting systems used by companies which developed business procedures to facilitate exchange with peoples whose economies differed greatly from those of Europe.[1] Records to be used in a study spanning several years may be incomplete. Once these problems are surmounted, the searcher is confronted by the difficulty of transforming the data into machine readable form as quickly as possible. This paper deals with this latter problem by discussing a technique for data recording and storage not widely used by economic historians, one that should save research time and expense. The utility of the method is demonstrated in terms of an ongoing research project involving the use of Hudson's Bay Company records. Rather than presenting research findings, the article concentrates on the problems faced in using the voluminous Company records. We hope that researchers who face similar data acquisition and handling problems may benefit from our experience.

THE PROJECT AND SOURCES

The project involves examining the business records of the Hudson's Bay Company before 1770 in order to understand how the trading system functioned under various economic conditions in Europe and North America. To determine the extent to which market conditions in Europe influenced, and in turn were affected by, circumstances in North America, information regarding the following is being sought:

1. the quantities of each type of fur and hide sent to London annually, as well as the numbers that were sold;
2. the prices received by Indians for the commodities they brought to the posts and the prices the Company received for the same items on the London market; and
3. the prices the Company paid to obtain trade goods in Europe and the tariffs it charged the Indians for these articles at its posts on Hudson Bay.

Data regarding the above issues can be obtained from three sets of Hudson's Bay Company records. Information regarding the shipments of furs from Hudson Bay to London is contained largely in the post account books. For the early years (1670-88), when considerable trade was carried on by ship or from short-lived posts, the post account records must be supplemented with fur shipment data found in the "Grand Jour-

1 An example is the trading post accounting system developed by the Hudson's Bay Company in the seventeenth and eighteenth centuries. The problems that these records pose for fur trade historians were discussed in Arthur J. Ray, "The Early Hudson's Bay Company Account Books as Sources for Historical Research: An Analysis and Assessment," *Archivaria* 1 (Winter 1975-76): 3-38.

nals" which contain records of all the Company's business transactions and include the invoices of furs shipped from Hudson Bay.

The "Grand Journals" also provide information concerning fur sales and the purchase of trade goods from the time of the founding of the Company to 1737. Quantities and prices are usually given. Obtaining average annual prices for key furs and trade goods is very time-consuming because this information was recorded on a transaction-by-transaction basis without any year-end summary. Without the aid of a computer, it would be virtually impossible to extend even this simple form of descriptive analysis over a long period of time.

For the years after 1737, fur sales were recorded on a sale-by-sale basis in the Company's "Fur Auction Books," but again, summaries are not given. For each transaction the number of furs contained in each bale is given, along with the type of fur, the post of origin, the price paid for each pelt and the total value and weight of the bale. The detailed information available from these sections of the Hudson's Bay Company records should make it possible to understand the conditions of supply and demand and prices both at an aggregate level, for example, North America and London, and at the post level, for example, the shipments and sales of furs and trade goods at Fort Albany, Fort Churchill, Eastmain, Moose Factory, Fort Richmond, Fort Severn and York Factory.

HANDLING THE DATA

To date, our efforts have focussed on the information that we have collected from the "Fur Auction Books." Data from this source were originally transcribed in the archives onto ledger sheets, but subsequently a suitable code was devised for keypunching the data onto computer cards.

Two standard options are available for keypunching.

1. Coded information can be transcribed onto keypunching forms, such as a standard IBM Fortran coding form. These forms can then be given to a keypunch operator for processing. For purposes of accuracy, it is generally necessary to have the cards verified, essentially a double punching procedure. All too often, errors still exist after verification and, in order to ensure a high degree of accuracy, it is sometimes necessary to have a duplicate card deck punched and verified by different operators. The two decks can then be cross-checked for keypunching errors.
2. Keypunching can be done directly from the archival notes taken on the ledger sheets. The column format of the ledger sheets can be treated much like a keypunching form. Double checking for errors is still necessary. The computer print-out of the data can also be compared with the original notes taken in the archives.

The first option does not require that the research team do the keypunching, thereby allowing them to devote their attention to less mechanical problems. Since the cards are processed by a keypunch operator, however, they should be verified by another operator. While this approach saves the researcher valuable time, the following disadvantages remain:

1. the researcher must transcribe the data onto keypunch forms, a task almost as tedious as keypunching;
2. the costs of keypunching are significant, amounting to a minimum of $120.00 per thousand cards; and

3. much time is consumed between the original transcription from the archival records and the availability of the data on a computer-compatible input medium.

The time required to collect and transform the data into a format that can be used by a computer can be shortened considerably by recording the initial set of archival notes directly onto keypunching forms. However, this is practicable only when the data source is reasonably consistent internally and has a format allowing the researcher to devise a parallel coding system. The eighteenth-century "Fur Auction Books" and "Post Accounts" of the Hudson's Bay Company are quite satisfactory in this regard.

The second option certainly saves time and reduces expenses since the research team does all of the work, even though a considerable amount of research time is still used punching and verifying cards. In the case of our present research project, transferring the information from the "Fur Auction Books" for the period 1737-70 involved keypunching and checking over 6,500 computer cards. Coding the data was straightforward and the transcription orderly and simple. Notes taken in the archives were arranged in column format on ledger sheets and these sheets were then treated much like keypunching forms.

In short, the first approach to keypunching involves four steps: note-taking in the archives, transcription of notes onto keypunching forms, keypunching by operator number one, and verifying by operator number two. The first two steps are carried out by the research team. One of these steps can be eliminated if the original notes can be taken on Fortran or similar forms that keypunch operators can read. The second method of obtaining a keypunched record involves two steps: transcription of notes in the archives, and keypunching and checking of card decks by the research team. Since the chances of error increase each time the data are handled, any approach which reduces the number of transcription steps not only saves time and money, but also improves accuracy.

AN ALTERNATIVE APPROACH

Keypunching in the archives directly from documentary sources would reduce the amount of tedious recording necessary, thereby greatly expediting research and cutting costs. It is unrealistic, however, to expect such facilities at most archives. Fortunately, an alternative means of processing archival data into a machine-readable format eliminates many of the common problems discussed above without requiring archives to purchase expensive keypunching equipment. This is the Optical Mark Read (OMR) system for transcribing and keypunching data.

The OMR system uses computer cards with an array of "bubbles" in a forty-column by twelve-row matrix that can be pencil marked to indicate either a letter, a number or any other specified sign (see figure 1). It is possible for a researcher to take OMR cards into the archives and directly transcribe information such as that contained in the Hudson's Bay Company "Fur Auction Books" onto the cards in the manner shown. The OMR cards are then used as a computer input medium and the data recorded on them can be transferred directly to disk, tape or another (or the same) set of computer cards using a computer punch programme. In the case of our work with the Hudson's Bay Company documents, this approach could eliminate all of the data transcribing and keypunching steps that we have hitherto had to perform outside the archives. Almost as important, it should save transcription time since a numerical code will be used with the OMR cards that requires considerably less writing than does the original transcription (figure 2).

THE OMR SYSTEM

Initially all IBM mark reading hardware was magnetic rather than optical; now, optical reading facilities available from IBM are in widespread use. Although some magnetic mark reading systems may still be in use, the OMR system has rapidly superseded the magnetic mark system because it is more reliable. Magnetic readers sense electrographic pencil marks along a twelve-row by forty-column matrix. Each card contacts the reading head and the brushes detect the marks present throughout the matrix. This system generally functions quite efficiently, but clogging of the reading head from electrographic pencil dust can cause data cards to be rejected.

In contrast, the OMR card reader utilizes photo-electric cells and a powerful light source to distinguish between light and dark patterns of reflected light. Since there is no contact between the sensors and the cards, the chances of the reading machines becoming fouled are greatly reduced. To date, the main problem with both mark reading systems has been incorrect punching of the recorded data by the computer as a result of smudge marks and accidental pencil marks. Since the machine notes only the absence or presence of pencil marks and has some facility to compensate for slight errors in the positioning of such marks, accidental pencil marks can be interpreted as data. In order to reduce or eliminate this source of error, care and neatness are necessary in the marking and handling of cards. Carefully marked cards and a properly

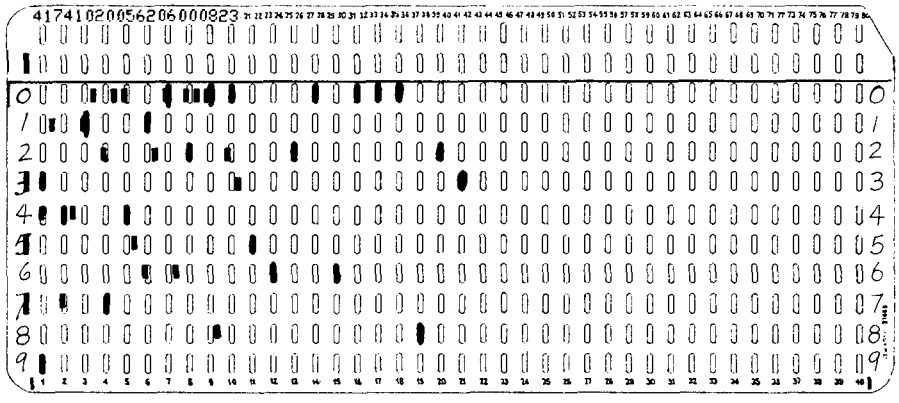

Fig. 1. *An OMR data card marked and punched with fur sale data from the "Fur Auction Books." The punched code columns are indicated across the top and the OMR code columns across the bottom. The data are numerically coded and the punched numerical code is printed across the top over the corresponding punched columns. The punched columns from 1 to 20 correspond with the OMR columns from 2 to 21 (the first OMR column has been used for programming information.) The numerical code used for the OMR marks in the twelve bubble columns is, from the top down, blank, blank, 0, 1, 2, 3, 4, 5, 6, 7, 8, 9. The code in column 1 (punched) indicates the post of origin (e.g., 4 = York Factory); columnns 2 to 5 indicate the year of the sale (eg. 1741); columns 6 to 8 indicate the commodity sold, (eg., 020=parchment beaver); columns 9 to 12 indicate the number of pelts in the bale (eg., 0562* = 562 furs); columns 13 to 16 indicate the value, per fur, paid (eg., columns 13 and 14 indicate shillings [*06 = 6 shillings*] *and columns 15 and 16 indicate pence* [*00= no pence*]; *and columns 17 to 20 indicate the total weight of the bale (eg., 0823 823 pounds).*

functioning OMR reader make the system as reliable a computer input medium as punched computer cards.

TIME AND COST EFFECTIVENESS OF OMR

As mentioned earlier, the use of OMR to transcribe data directly from archival documents eliminates at least one tedious transcription procedure. In the case of the Hudson's Bay Company "Fur Auction Books," the keypunching was performed by two graduate students working ten hours a week during one three-month period at a cost of approximately $2,400.00 for salaries. If the data had been transcribed initially using OMR, analysis could have begun as soon as the research assistants returned from the archives, and the three months could have been devoted to dealing with the issues that the preliminary examination of the data raised.

Information collected on OMR cards should be recorded for storage in keypunch form, since OMR cards may become smudged or soiled, or may even shrink. We suggest that on the first card run, after ensuring that the data have been successfully transferred to either disk or tape, a complete set of cards be punched duplicating the information contained on the OMR cards. This can be done by using a keypunching programme. It is possible to have the computer punch this record directly on the OMR cards so that each card is punched according to the marks that the researcher recorded on it. This procedure preserves the data in the most secure and universally compatible form for computer input and at the same time contains the original transcription on the same card. When there is a question regarding the accuracy of transcription or of the computer-punched record, both can be conveniently compared on this duplicate record.

The main advantages of the OMR system are:

1. the original data are transcribed quickly and easily;
2. the time and costs for data analysis and presentation of research results are reduced; and

Fur Sale Books, Hudson's Bay Co. PAC HBC A 48/1, pp. 58-66 | x Per Skin ▲ Per bale

	Parchment Beaver (020)				Damaged Beaver (021)				Cub Beaver (030)				Marten (060)		
	Qty.▲	Post	Val.×	lbs.▲	Qty.▲	Post	Val.×	lbs.▲	Qty.▲	Post×	Val.	lbs.▲	Qty.▲	Post	Val×
1	562	F	6/2	823	622	F	5/3	913	380	F	5/10	241	271	AR	8/-
2	281	"	5/10	411	312		5/4	459	"	"	6/1	244	276	MR	5/7
3	"	"	6/-	406	"		"	458	1520	"	"	1,015	282	F	5/8
4	"	"	5/11	395	"		"	459	380	"	"	240	271	AR	7/6
5	"	"	6/4	392	"		"	459					"	"	7/8

Fig. 2. *A transcribed page based on "Fur Auction Books." The data have been organized in a column format by commodity and the quantity of pelts, post of origin, price per pelt in shillings and pence. The total weight of each bale is recorded. Cross-outs, check marks and code numbers are a result of the keypunching procedure indicating verification and the correct codes. The first four entries on the left have been illustrated in final punched form in figure 1.*

3. automated keypunching of the original OMR transcription provides a compact, accurate and convenient method of storing data, including the original archival transcription.

The coding systems used with OMR records, like coding systems used with punched cards, can be tailored to the needs of individual research projects. Furthermore, IBM will print OMR cards with a format or code system to suit the needs of the user making the original transcription easier by eliminating a separate record of data codes and formats and by providing space for additional notes.

SUMMARY AND CONCLUSIONS

For the historian to deal adequately with a problem he must use all of the documentary information that is at his disposal, a significant portion of which often exists in quantitative form. To analyze these data satisfactorily, it is frequently necessary to use a computer since the latter enables the researcher to handle large quantities of data that might otherwise be unmanageable. In our present project, given the volume of data to handle, even sampling would be difficult if we did not use a computer. Yet, while the computer makes it possible for the historian to analyze his data properly, it also presents him with a problem. In a time of dwindling research budgets, it is increasingly necessary to find ways of transforming documentary material into machine readable form quickly and cheaply so that computer analysis can begin with a minimum of delay. The OMR system offers one solution to this problem and, for this reason, warrants the attention of historians.

The adoption of the OMR system or other means of expediting the transfer of archival data into machine readable form has implications for archivists. The increased ease of access to data through computer analysis will mean that historians of the future will want to work with records that are as complete as possible in order to enable them to utilize a greater array of statistical techniques. It will be necessary to re-examine the current policy of storing only samples of voluminous government records, even though this would result in greater storage problems.

As historians make use of archival quantitative data sources and base their conclusions increasingly on statistical confirmation, it would serve the interests of scholarship to have researchers provide archives with copies of data sets they have used in their research. The acquisition of this kind of record is within the mandate of the Machine Readable Archives Division of the Public Archives of Canada, and such collecting of researchers' data is already extensively practised at the University of Michigan's Inter-University Consortium for Political and Social Research.

In this way, other scholars choosing to challenge or re-examine an inference or those wanting to examine other dimensions of a problem could do so without having to repeat the time-consuming transcription of archival information onto a compatible input medium. If such services will be made available, many archives would have to develop facilities suitable for storing computer cards, disks and tapes as well as for providing the necessary hardware in the form of card readers and computer terminals to permit users access to these data. If the archives provide these facilities and scholars co-operate by making transcribed data available, then quantitative historical research can proceed with a minimum of duplication of initial data collection.

Arthur Roberts and Arthur J. Ray
York University