Woman C.P.A.

Volume 38 | Issue 2 Article 5

4-1976

Electronic Data Processing: Controlling the Accuracy of Computerized Data

Elise G. Jancura

Electronic Data Processing

Controlling the Accuracy of Computerized Data

Dr. Elise G. Jancura, CPA
The Cleveland State University
Cleveland, Ohio

Accounting is the process of identifying, evaluating, recording, and communicating financial information about an economic entity. Accounting controls may be characterized as a system of procedures designed to facilitate the proper discharge of accounting responsibilities for the protection of company assets and creation of a satisfactory audit trail. These control procedures should document all transactions and further insure that only correct and authorized data enter the accounting records. Care must be taken that any operations performed on the data are done accurately and in a time frame that will produce useful results. The fundamental importance of accurate data in any information system is succinctly described by the oft-repeated phrase GIGO — garbage in, garbage out.

In a data processing environment the tasks of verifying and controlling data generally involve the activities of capturing source data, accurately transcribing nonmachine-readable data, and guarding against any subsequent loss or distortion as the data is subjected to the many different manipulations, summaries, and transformations performed upon it during the course of normal processing operations. A variety of techniques, some manual and some computerized, are employed to insure the accuracy of the data used.

Machine-Readable Data

Before a computer can successfully process data, these data must exist in a form intelligible to the equipment. Machine-readable data can take many physical forms. It can appear as holes in a punched card, as magnetic spots on a strip of tape, or as magnetic spots along a track in a disk file. These machine readable forms are usually different from those directly intelligible or convenient for human use. Thus the information about transactions must be converted into one of these storage forms before it can be processed by the equipment. This conversion or transcription process in which human-readable data is converted into machine-readable data is one of the key procedures in any data-processing environment.

The transcription process has a role of great importance in computerized record-keeping. First, it is the basic source of data for subsequent processing and, incidentally, the point at which many of the data controls are established. Therefore, much time and attention is devoted to developing techniques for verifying the accuracy of the machine-readable records produced. Second, the process is time-consuming and expensive. The key punch is still the predominant method of original data entry, representing 80 percent of the keyboards involved in computer data preparation.

Keypunching is a manual key-driven process similar to that used in typing that effectively copies desired data on punched cards. A trend that seems to be increasing is the replacement of keypunching with key-to-tape or key-to-disk operations. This is functionally similar to key-punching — that is, it is a manual data-recording process — but it produces magnetically recorded data that can be used by higher-speed input devices, constitutes a more compact form of storage than punched cards, and is frequently subject to fewer intermediate processing steps than punched card data.

Many of the key-to-disk systems are trending toward an on-line mode of operation in which the computer can perform certain editing functions as the data are stored on disk storage for subsequent processing. This is quite similar to the input process in real-time systems where an operator keys in a transaction at a terminal for immediate processing, although the typical key-to-disk operation is a data collection activity for subsequent batch-processing.

Much of the development work in the data-recording or data-entry area has been aimed at speeding the manual transcription process or eliminating it where possible. Thus preprinting magnetic characters on checks or developing optical scanning equipment that can read printed or even handwritten documents would eliminate the requirement of manual transcription which is time-consuming and a source of many errors.

The need to convert data into a machine-usable format introduces additional considerations into the control process. Procedures must be developed to insure that the conversion process takes place without error, so that the machine-readable records reflect accurate and com-

plete data. The fact that these records are usually not intelligible to a human being without interpretation from the data-processing equipment raises the possibility that errors in the data will go undetected until a fairly long period of time has elapsed or until the error is reflected in the final reports. A delay in detection of an error can be costly, particularly if these records are being used to control or otherwise affect the actual operation. In addition, correction becomes more difficult when undetected errors are allowed to mingle with large volumes of other transactions.

Responsibility for Data Controls

The data-processing department does not create information, nor does it in fact act as the end user. It is only a processing facility that makes it possible for the various user departments to minimize the cost of recording information generated by those departments and to maximize the use or benefit derived from this information. Since the user departments generate the information, they are the departments that have responsibility for proper authorization of data and in many instances for verifying the accuracy of the source data.

The data-processing group has as its responsibility insuring that this source information is accurately converted into machine-readable or otherwise usable form. Certainly the data-processing department is not excused from the responsibility of insuring full capture of information wherever possible. But the ultimate responsibility for recognizing sources of information and for insuring that all data is properly generated and all transactions properly initiated falls with the user department.

Controls for the Data-Recording Process

The data control techniques employed by computer installations can be divided into three basic groups: those techniques concerned with the accurate transcription of information into machine-readable records, those techniques concerned with testing the accuracy of the data itself, and those techniques concerned with preventing distortion of the data through subsequent processing procedures. The techniques employed to accomplish these three goals represent a mixture of manual and programmed activities.

Most machine-readable data is created by a manual process, usually a keying operation, in which the operator copies the information from some previously recorded document and produces as output of this operation a machine-readable version of the document. The key-driven device used for this "copying" or transcription process may be physically located in the data processing department or it may be located in the user department. These devices vary in nature from traditional keypunches, key-to-tape, and key-to-disk devices to various manually operated terminal devices such as teletype (and other typewriter-like units) and CRT units, which are usually located at the individual user sites. In the latter case, although data is captured at the source by the user, the requirements for accurate transcription and verification of the data remain.

Assuming that the original document is correctly or accurately recorded, the control techniques for the transcription process center around comparing the contents of the machine-readable record with the contents of the source record. The simplest way to accomplish this is to check the machine-readable record visually. This may be accomplished by looking at the records and comparing the list to original source documents. Visual checking is rarely a practical approach, however, for if it is done accurately it is a slow, time-consuming technique and is subject to the fallibility of the visual scanning.

Visual verification is useful in those instances where the data being recorded is to be used immediately, as is the case in real-time systems. When an operator is recording information on a terminal for immediate use in updating an installation's files, it is imperative that that information be checked for accuracy at the point of recording rather than through some subsequent more automated technique. In these circumstances the terminal should be equipped with a facility to provide automatic feedback for immediate visual checking. If the terminal has a cathode-ray tube device (CRT), the information being recorded by the operator is displayed on the tube. The operator can then visually check the information recorded before using it. A similar approach is the attachment of a hard copy facility to a terminal. This is even more desirable, since it provides an audit trail record as well as an opportunity for the operator visually to check the accuracy of the data as it is being recorded.

Visual checking of data usually should be limited to those instances where information must be used immediately. It is time-consuming because it delays the operator's ability to record subsequent information, and in a high-volume situation it is subject to the same fallibility that affects all human activities. A much more common technique for verification of the original keying procedure is one which involves a duplication of the original keying procedure in a process called key verifying. In key verifying the original recording operation in which the operator produces the machine-readable record is followed by a second operation in which the operator rekeys the data from the source document while the device (usually a key verifier) is concurrently reading the previously produced record. The device then compares the information keyed by the operator with the information recorded in the newly created machine-readable record. If they agree, the record is passed through as having been properly produced. If they do not agree, the operator is notified.

The failure of the machine-readable record to compare with the source document may indicate a keying error by the verifier, or it may indicate an error in the recording of the original machine record. Which it is is determined by repeating the keyed verification process. If on the repetition of the verification attempt the source document is found to agree with the machine-readable record, it may be assumed that the original verification was in error but that the machine-readable record is correct. If the disagreement persists during the second verification procedure, the assumption is that the machine-readable record produced is in error, and it is corrected. Key verification provides a fairly accurate way of checking the transcription process in high-volume applications, but it represents a fairly expensive technique. Since key verification represents a complete duplication of the effort and time involved in the original data recording, key verification of all data represents a cost to the installation that is equivalent to the actual data-recording cost. Because of this cost consideration key verification is frequently employed on only key fields within a record rather than on the entire record.

There are a number of different devices for manual key-driven recording of machine-readable data. The most common is the keypunch. This is a device in which the operator, through a keying operation, produces punched card records. These keypunches still represent about 80% of key-driven data-recording activity. Other key-driven data recording devices are those that produce standard magnetic tape or tape cassettes and those that produce various forms of disk storage. The functional advantage of the key-to-tape and key-to-disk devices lies

in the fact that they produce a form of machine-readable record that represents a higher speed facility than that represented by punched cards. Frequently when punched cards are the initial form of machine-readable records, a procedure will be executed to copy those cards to some form of magnetic storage. If the transcription process produces tape or disk directly, it eliminates the need for this separate operation.

The introduction of a buffer facility to any of these key-driven devices makes error correction procedures somewhat easier. When a key-driven device has a buffer facility, the keying operation causes storage of the data in the buffer. The actual machine record is not produced until the operator depresses a key causing transfer of the information in the buffer to the recording medium. The advantage of this approach is that the operator can then key the entire record before having it transcribed into machine format. Recognition of an error during the keying process provides the operator with the opportunity to rekey the stroke immediately before the machine-readable record has been produced. The process is somewhat analagous to that involved in typing. Frequently an operator recognizes that an error has been made in the process of keying a stroke and can correct it immediately. However, it should be recognized that key-driven data-recording operators, like typists, can make keying errors that they do not recognize and therefore that the introduction of a buffer facility to any of these key-driven recording devices does not eliminate the need for subsequent verification.

The Use of Check Digits

Use of a check digit (or self-checking number) is another technique employed to verify the accuracy of certain fields. A self-checking number is one that has a precalculated digit, called the check digit, appended to the basic information for the purpose of catching transmission or recording errors. For example, a five-digit code with a self-check digit would appear as a six-digit or six-position field. When such a field is read, the self-check digit would be used to verify the correctness of the field by recalculating the check digit and comparing the calculated check digit with the digit recorded as part of the field. When key-driven recording devices are equipped with a check digit device, the recorder automatically calculates the check digit as the operator keys the field. If the operator keys the same check digit that the device automatically calculates, this is an indication that the field as recorded by the keying operation is correct.

The check digit does more than merely verify the transcription operation, for it checks on the validity of the source data itself. Thus if the source data is incorrectly generated, the check digit approach identifies the error, whereas key verification identifies only transcription errors, not errors in the original input data. Check digits can be used to validate fields such as part number, customer number, or employee number, where the content of the field is not subject to change by the processing involved and where the value in the field is predictable. A check digit cannot be used to validate a quantity field or an amount field where the value is not a preestablished identification code.

The check digit technique catches approximately 97% of the transposition and substitution errors that are the most common type of keypunching and clerical errors. The possibility does exist, however, that a transposition error will not be detected by this technique. Furthermore, a check digit will not identify those instances in which a valid but incorrect code is used. For example, a payroll record may contain a valid employee identification number as one of its fields, but this valid employee identification number may be identifying the wrong employee.

A self-checking number can be used in a number of ways. As discussed above, it can be part of the transcription-checking process when source data is converted to machine-readable form. Keypunches can be equipped with a hardware feature to handle check digits. The technique can also be easily implemented through programmed control in the computer. Any time a data field that is appropriate as a self-checking number is inputted, the computer program can calculate and check for the check digit. This can be done whenever a data record is accessed by the computer and is especially useful when original source data is being entered directly — that is, in real-time operations where an operator is keying in source data through a remote terminal. A self-checking number is also useful when data is being entered through optical scanning techniques that read original documents. As mentioned earlier, self-checking numbers verify more than the accuracy of the transcription process — they also provide a technique for validating the contents of the fields themselves. This is especially valuable in real-time applications where the data can be used to effect master files as soon as it is recorded.

Batch Controls

Much processing is done in batch mode. Where instantaneous reaction to a transaction is not required, it is usually more efficient to collect a group of transactions (and make any necessary corrections) before using the data for processing. When transactions can be held for batching, control totals for a group of transactions can be calculated from the original source data. When the machine-readable records for this group of transactions are created and subsequently read by data processing equipment, control totals from the machine-readable records can be calculated. If these control totals balance to the control totals accompanying the original source data, the assumption can be made that the transcription process has recorded the data accurately. These control totals can be taken on a number of different fields. Control totals can be calculated in quantitative fields or on identification fields. Frequently, when the control total is calculated on a nonquantitative field (for example, part number), the total is called a hash total rather than a control total. This type of total is taken solely for checking purposes, since the value calculated has no quantitative significance.

A batch may be made up of a group of records having a common identity such as a department, or it may be made up of a specified number of records. The initial balancing operation is designed to determine whether the group of machine-readable records has been properly recorded in the fields checked under the control total; after that takes place several batches may be combined into one larger batch. It is usually desirable to keep the initial batches fairly small so that, should an error be detected by an out-of balance condition, it would be simpler to locate the individual record in error. The larger the batch, the longer the search process to locate such an error record. Once, however, the initial batches are balanced, indicating the machine-readable records have been properly transcribed, accumulation of the small batches into larger, more efficient groupings is desirable. Usually a number of totals are calculated for each batch. If, for example, there are two or three quantity fields as well as an identification field, it would not be unusual to calculate a control total for each of the quantity fields as well as for the identification field.

The balancing operations can be completely automated and executed by com-

puter program. This can be accomplished by having a control card punched that contains the batch totals for each batch and then read along with the machine-readable records produced by the transcribing procedure. The computer program can calculate the batch controls for each group of records and compare them against the control card or control record read in with the original balancing totals.

The use of control totals and record counts is primarily useful in a batch-processing environment, and in fact these control totals are frequently referred to as batch totals. The principles of control totals or batch totals can also be applied in a real-time environment, although the use of these control totals will differ in a real-time environment from that in a batch-processing environment. The purpose of various processing controls and data controls is, first, to prevent erroneous data from producing incorrect results and to prevent loss and distortion of data through processing errors, and, second, to facilitate correction of errors when errors are discovered. Because the batch totals can be calculated and used before the data is actually used to update master records, the batch controls in a batch-processing environment allow accomplishment of both goals. However, this is not possible in a real-time environment because, by definition, in a real-time environment transactions are processed immediately as they occur. It is possible, however, to make good use of control totals in a real-time environment to accomplish the second objective of error control — this is, to facilitate correction of errors when they occur and, furthermore, to facilitate recognition of these errors.

Other Techniques

An installation will usually implement some combination of the verification and balancing techniques to verify the accuracy of its data-recording procedures. For example, it may key-verify or use a check digit on the identification or control fields of each record, and it may use control totals for the quantity fields. It is possible that descriptive information recorded within each record will not be verified at all. One-hundred-percent accuracy is better guaranteed by complete use of all techniques on all fields within every record, but it must be remembered that each of these verification techniques represents a cost to the installation that must be balanced against the importance of accuracy for each individual field. Further, it has to be recognized that these techniques are oriented toward verifying the transcription or recording procedures of the installation. In general, they do not guarantee the accuracy of the data but rather guarantee that the source information provided to the data-processing department has been faithfully transcribed.

Because the transcription process itself can introduce errors into the data, and because it represents an additional cost to the installation, much attention is given to techniques in which the source document itself can be made machine-readable. One such approach is to use turn-around documents in which some of the identifying information is recorded on the document in machine-readable form. An example of this technique would be the use of customer bills in the form of punched cards where customer identification information is prepunched in the card. This reduces the amount of information that has to be manually recorded and thus verified. Another example of this approach is the use of prerecorded magnetic ink characters on the bottom of checks and deposit slips. The only information that must then be manually recorded on the document is the amount of the check, done by a keying operation on an encoding machine.

Another approach is to use a printed document itself as an input form. This can be accomplished by the use of optical character readers that read documents printed with certain special typefonts. These typefonts can be installed on printers and typewriters so that the hard copy they produce can also be used subsequently as input data. The accuracy of this printed information is essential, and since there is no separate verification of a keying process, applications that use this type of data make greater use of other checking procedures such as check digits, batch controls, and other program tests to detect reading error.

One technique for minimizing the cost of transcribing source data to machine-readable form is the production of machine records as a by-product of other operations. Cash registers can produce paper tape or magnetic cartridges or can even transmit to a central recording facility (which might be a computer) as the cashier rings up sales. Some retail stores are experimenting with magnetic encoding on price tags that can be read by a sensing device at the check-out counter as part of a point-of-sale (POS) data entry system.

Checking on the Data Itself

Although the primary responsibility for insuring the integrity of data rests with user/initiator departments, there are some techniques that the data-processing department itself can employ to check on the accuracy of data. The responsibilities of the data-processing group should extend wherever possible beyond mere verification of the transcription and processing procedures and attempt to establish the accuracy of the data itself before or as it is recorded in machine-readable form and used for processing.

As discussed previously, the employment of a check digit or self-checking number is an example of a technique that concerns itself with the accuracy of the data itself. A check digit can be used to insure that certain fields within machine records contain valid information. These check digits can be employed during the recording process or during subsequent processing techniques to insure that records being used to update master files are valid.

Frequently the logical consistency of individual fields within a record can be checked by a processing program designed to edit the data. This editing program can be combined with a balancing operation or in batch-mode processing it may be executed independently as a separate program. In those instances where the computer itself captures the original data, the computer can be programmed to perform the editing function as it is recording the transaction rather than waiting to collect a batch of transactions to be edited. Further, some of the editing functions may be incorporated in subsequent processing programs. It is frequently possible to predict for various applications the validity of certain classes of data. For example, if customer numbers in the accounts-receivable file do not contain alphabetic data, when transactions affecting that customer file are being edited it is plausible to test the customer identification field in these new transactions for any alphabetic characters. The existence of an alphabetic character signals something wrong with that record. It is possible that a transaction from a different kind of activity has become mixed in with the file, or it is also possible that the customer number for this particular record was incorrectly recorded and somehow escaped the verification procedures associated with the transcription process.

The field containing an action code has certain predefined acceptable actions for a given application. Thus that field can be edited to make sure that it contains only codes that are acceptable or recognizable by subsequent processing programs. Other fields that contain blanks, negative amounts, alphabetic data, or information that exceeds certain values can also be recognized as being inconsistent with data acceptable for a particular application. Frequently it is not possible to predict what a particular value should be in a field, but it is possible to predict what the maximum and minimum values should be in that field. For example, when editing the data records that reflect hours worked in a payroll application, it is quite possible to recognize potential errors in the hours-worked field if the number of hours recorded exceeds a logical maximum for a week or if an employee receiving sick pay receives credit for hours worked during the same time period.

When using a given transaction to effect an updating operation against the master file, it is possible to use certain information recorded in the master file to verify the transaction further before using it for updating purposes. For example, the employee master records that contain year-to-date earnings may also contain a skills inventory indicating those jobs for which an employee is qualified. Thus, before using the transaction record indicating the number of hours the employee has worked, it is possible to verify the job code in the transaction record against the skills inventory recorded for that employee. A transaction record directing payment for a job code or compensation rate for which the employee is unqualified can thus be identified and printed out on an exception report for special review by the payroll department.

Sometimes none of the verification procedures in effect will detect an error in a transaction record. It is quite possible to record a valid customer number that happens to be the number for another customer and is incorrect for this particular transaction. But some other tests of logical consistency can be used — for example, if a transaction indicates receipt of cash from a customer whose balance due is zero, this can be an indication that the transaction somehow recorded the wrong customer number or that, in fact, there could be an error by the customer and an overpayment. In any event, such a condition can be printed out on an exception report and brought to the attention of the accounts-payable department. It is much more desirable to verify that a customer has overpaid than to neglect to record a receipt from one customer by giving credit to another.

The longer an error is allowed to exist, the more difficult it is to correct it or the more pervasive is its effect. Data that is incorrectly transcribed or that is in error at the time it enters the data processing department and is allowed to affect records through several subsequent operations will create incorrect results for each of those operations. As a consequence of this, many validation and checking techniques are concentrated at the beginning of a processing application. Much of the checking takes place immediately after the transcription or recording operation. Usually data is edited before being used to update master files.

The attempt to preserve data integrity involves execution of a combination of techniques, not all of which can be carried out at the same time. Where possible, correction should be made immediately, and the corrected data reentered into the system during the same processing cycle.

(Continued on page 28)

authors have listed practices that seem to them to account for the excellent management control they have found to exist in some nonprofit organizations.

This book could be very helpful for anyone involved in the operation of a nonprofit organization. Accountants are often asked to serve as board members for such organizations.

An active and concerned board of directors is a good point of beginning to improve the management control system within a nonprofit organization. An organization with a good control system can be assured that its goals are being reached in an efficient manner and that its service is of use to those it intends to serve.

Management Control in Nonprofit Organizations should be required reading for all members of government, particularly those involved with such mammoths as the Department of Health, Education, and Welfare, and the Department of Defense.

Bert Scott, Jr., CPA
Graduate Student
Mississippi State University

"THE MARKETING OF ACCOUNTING SERVICES," D. Larry Crumbley, Ray Barnhardt, and Robert J. Boewadt, THE CPA JOURNAL, Vol. XLV, No. 5, May 1975.

The main thrust expressed by the authors is that CPA's must become "client oriented." The first step is a determination of client's needs by a program of marketing research.

While informal client research may yield information as to a client's needs and wants, market research should provide data involving what client's want, what they do not like, and why they have changed accountants. Long-term client relations will benefit from an analysis of data of this nature.

In some cases clients may not realize a firm can perform the service it wants. Exchanging views with lawyers, bankers, and other accountants provides useful information as to what these services might be and whether or not they should be added.

Once an accountant has located in a community, a marketing orientation toward clientele should be adopted. The accountant should attempt to create a favorable image among both present and possible future clients. This may be accomplished by accepting speaking engagements before both accounting and nonaccounting groups; writing accounting articles aimed at the public-at-large or specific business groups; and engaging in activities of civic, religious, or political organizations.

To enhance the marketing of accounting services, the accounting profession should continue to prepare films and pamphlets to inform the public about the services performed by CPA's.

It is the opinion of this reviewer that a marketing orientation toward clientele by the individual accountant backed by group public information and publicity programs by the accounting profession will result in not only a cultivation of new clients, but more longstanding relationships with present clients.

John F. Dockery
Graduate Student
Mississippi State University

Accounting and Matrix
(Continued from page 9)

Conclusions

The applications presented here are by no means singular. There is virtually no limit to the number of possibilities for application, except, of course, if one limits the number of processes in accounting practice. Multi-reciprocal consolidations, secondary overhead allocations, period budgeting, responsibility accounting, and variance analysis for costs are other suggested areas for matrix uses.

The advantages of the matrix models are basically two: (1) their compact and simple format, and (2) the ease of application. As long as there can be identified input and formulated transformations, any output report can be constructed through the use of matrices.

Notes

[1]John W. Buckley, Paul Kircher, and Russel L. Mathews, "Methodology in Accounting Theory," *Accounting Review* XLIII (April, 1968), 281.

[2]A. Wayne Corcoran, "Applied Mathematics and Accounting," *Management Accounting* LI (August, 1969), 29.

[3]*Ibid.*

[4]*Ibid.*, pp. 29-30.

[5]Carl T. Devine, "Some Conceptual Problems in Accounting Measurement," *Research in Accounting Measurement: American Accounting Association Collected Papers*, Edited by Robert K. Jaedicke, Yuji Ijiri, and Oswald Nielson (Menasha, Wisconsin: American Accounting Association, 1966), p. 20.

[6]A. Wayne Corcoran, *Mathematical Applications in Accounting* (New York: Harcourt, Brace & World, Inc., 1968), p. 133.

[7]*Ibid.*, p. 162.

[8]*Ibid.*, pp. 160-165.

[9]Allen B. Richards, "Input-Output Accounting for Business," *Accounting Review* XXXV (July, 1960), 429-437.

[10]*Ibid.*, p. 435.

EDP
(Continued from page 18)

Frequently, however, the correction process involves a certain amount of manual checking, discussion with the operational department involved, or reconstruction of the data. In these instances processing the correct items is usually allowed to continue without waiting for correction of the error item. When the error is corrected, the corrected transaction can be introduced into a subsequent or the next processing cycle.

Failure to exercise control over the correction procedure can be a serious weakness in a program for data control. There should be formal, carefully prescribed procedures to insure that once errors in data are discovered corrections are effected and the corrected data is properly reintroduced into the processing activities. One of the observations the auditor should make when reviewing a client's system of controls is the attention paid to the error correction program and to the assignment of responsibility for its implementation.

Procedures aimed at determining the accuracy of data are equally applicable to batch systems and real-time systems. In both cases data relationships such as the internal consistency of fields within a transaction record, the consistency of the transaction record to the master record against which it is applied, and the existence or absence of required fields for particular types of transactions can be used to check validity, whether one is handling one transaction in a real-time environment or a batch of transactions in a batch-processing environment. The primary distinction between correction in the two environments occurs in the time frame during which errors are recognized and acted upon.

In determining how extensive the checking facilities should be in a given installation, its management should make a conscious effort to measure the expense to the installation of instituting certain controls as well as the expense to the installation if the errors that these controls are designed to prevent should occur. The optimum situation from an installation management point of view would be to provide controls to the point where the cost of the controls would equal the cost of failure that could have been prevented by those controls. This optimum control level may be further affected by legal requirements, but it does not excuse the installation management from recognizing the relative advantages of alternative operations.