Data Structure for a Construction Company Management System

A. P. FROEMMING and S. J. FENVES, University of Illinois

The object of this paper is to outline a data structure framework that would make possible the integration of the various data involved in data processing into a management information system for a construction company. The scope of consideration of a construction company system is limited to the four basic operations of accounting, payroll, CPM, and estimating. CPM is envisaged as both an updating and monitoring tool for the actual construction operations, as well as a planning and scheduling tool.

The data structures in the four operations are described with emphasis on the hierarchical arrangement of the data. The interaction between the data structures on a company-wide basis is discussed in terms of the company activities of bidding and pre-project planning, project control, and generation of historical data.

The data structure presented deals only with the format of the data, and does not depend in any way on the contents of the data. The value of the system to the user depends exclusively on his ability to formulate his particular requirements within the context of the structure.

•ONE problem confronting management of a construction company is the ability to integrate easily the various elements of data that constitutes a construction company system. For example, the data resulting from the most commonly used estimating procedures are not convenient to use in keeping cost records on a project. Similarly, the normal accounting systems have no convenient way of associating the various project expenses with individual project cost accounts.

The object of this paper is to outline a data structure framework that would make possible integration of the various data elements and supply the desired management information. The scope of consideration will be limited to the four basic operations of a construction company system: accounting, payroll, CPM, and estimating. Figure 1 is a simplified representation of a construction company data processing and management system, showing the four operations and the existing data paths between the four operations, as well as the data going in and out of the management system. The functions of accounting, payroll, and estimating are straightforward. In this paper, CPM is considered as both an updating and monitoring tool for the actual construction operations, as well as a planning and scheduling tool.

DATA STRUCTURE OF OPERATIONS

In this section, the data structures involved in the four basic operations are described, with emphasis on the hierarchical arrangement of the data and on their interrelationship within each operation.

Paper sponsored by Committee on Highway Equipment and presented at the 48th Annual Meeting.

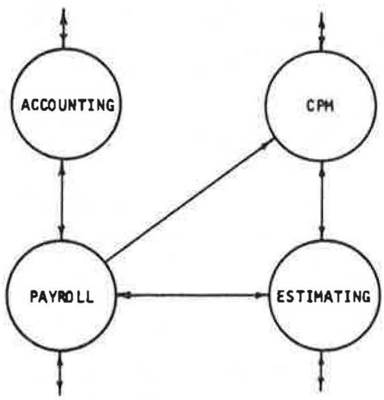

Figure 1. General layout of operations.

Accounting

The hierarchy of the data structure within accounting involves three levels of data: statement entries, open accounts, and account entries. The collection of statement entries forms either the balance sheet or the profit or loss sheet for the company.

The data structure within accounting is shown in Figure 2. Here, as well as in all remaining figures in this section, the data structure is represented in a skeleton form. The columns represent data at various levels of the hierarchy. In each column, one or more typical data items are shown, separated by double lines. Arrows are used to show the interaction between the data at the several levels, either as a cross reference or as an actual transfer of data. The three types of data will be discussed in turn.

1. Statement entries consist of an identification code (ID) for the line of the statement and the line of the statement itself. The following is an example of entries that might be made in a balance sheet:

DESC1	BALANCE SHEET
DESC2	CURRENT ASSETS
100	CASH WEST SIDE BANK
101	ACCOUNTS RECEIVABLE
SUB1	SUBTOTAL CURRENT ASSETS
200	ACCOUNTS PAYABLE
400	JOB MATERIAL EXPENSE

The line ID has two purposes: internal referencing within the statement, i.e., to let the user identify a line that he may wish to change or delete, or following which he may wish to insert a new line; and upward referencing from the open accounts, i.e., to associate the various accounts with the proper line of the financial statement. The line

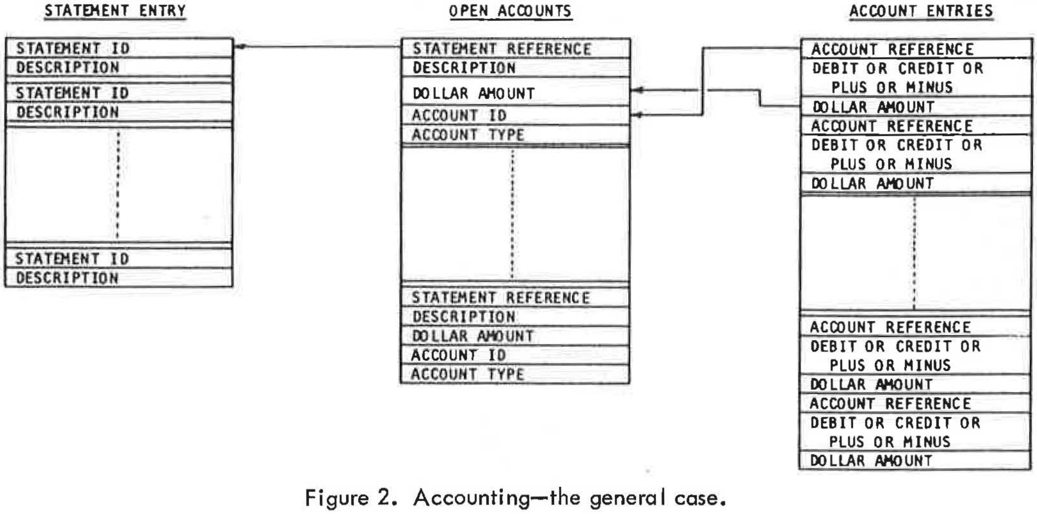

Figure 2. Accounting—the general case.

of the statement itself consists of optional descriptive text, except that the first word of the description may be one of the reserved words "surplus", "profit", "total", or "subtotal". The function of these words is to indicate where the totaling calculations are to be performed and printed.

2. Open accounts consists of (a) a statement of reference, i.e., the ID of a statement line; (b) dollar amount in the account at the time it is opened; (c) ID of the account; and (d) account type.

The account type designates the account as being one of the following: asset, liability, expense, income, account receivable, account payable, or fixed asset. The account type is used in the preparation of the financial statement as well as to check the validity of account entries referencing the account.

The data for an expense account differ from those of other accounts in that, in addition to the actual dollar amount, values of estimated material quantity, actual material quantity used, and budgeted dollar amount may be associated with the account. In addition, a project reference may be included, referring to a particular set of cost accounts in the payroll program. These additional data are of importance both as current information to the user and for the interaction of accounting with payroll and ultimately with the estimating program.

With any expense account it is possible to associate an auxiliary account. More than one expense account may reference the same auxiliary account. The use of the auxiliary account is discussed later. Figure 3 shows the layout of the data in an expense account. The following are examples of data for open accounts:

```
100  $8,000.00 A100 ASSET
101  $1,568.39 AR100 ACCOUNT RECEIVABLE
200  $6,247.00 AP100 ACCOUNT PAYABLE
400  EXP100 EXPENSE BUDGET $35,000.00 QUANTITY 255 JOB-10 AUX10
```

In the last example a new account is being opened. No dollar amount exists in the account; $35,000.00 is the budgeted amount; 255 is the estimated material quantity; JOB-10 references a project ID; and AUX10 is the ID for the auxiliary account.

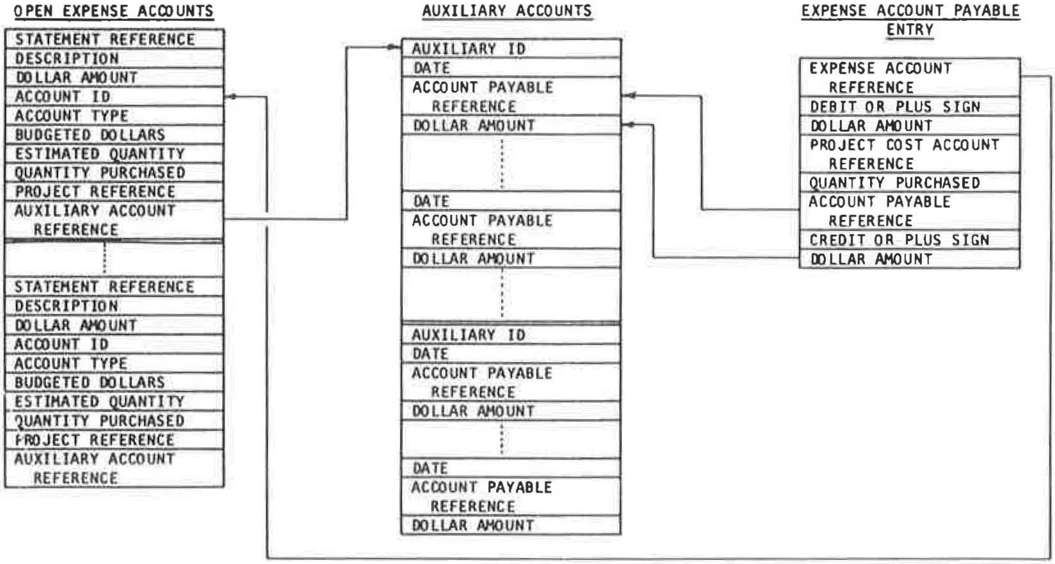

Figure 3. Accounting—the expense and auxiliary account.

3. Account entries consist of a minimum of two sets of data, each with an account reference, the label DR (debit) or CR (credit), and a dollar amount. A plus or minus sign may be used in place of DR or CR. The formula used to determine whether the entry is in balance is Assets = Liabilities + Net Worth + Income - Expense. An example of an account entry is as follows:

A100 DR $1,568.34

AR101 CR $1,568.34

or, using signs,

A100 + $1,568.34

AR101 - $1,568.34

An expense accounts payable entry could look as follows:

EXP100 + $5,000.00 CON684 QUANTITY 2400

AP100 + $4,000.00

In the preceding, CON684 is a project cost account reference and 2400 is the quantity purchased. Since the expense account EXP100 is associated with auxiliary account AUX10 (see last example above for open accounts), the account payable reference and the dollar amount, together with the date of the entry, are stored in the auxiliary account, i.e.,

7/28/68 AP100 $4,000.00

If all the expense accounts for a given project were associated with the same auxiliary account, the balance in that auxiliary account would always represent the total amount of moneys owed on the project. By referencing the data in the auxiliary account, it is possible to write checks to the vendors for a given project and thereby simultaneously reduce cash, accounts payable, and the auxiliary account by the amount of the check.

Payroll

The data used in payroll fall into three main categories: entries made on projects in progress, temporary data for a pay period, and permanent data. The permanent data can be further divided into two main categories: (a) data on the individual man; and (b) data pertaining to the individual cost account on a project.

The data on an individual man (Fig. 4) consists of ID; name and address; social security number; number of dependents; a craft type identifier (i.e., carpenter, laborer, iron worker, etc.); and earning and deduction data, consisting of gross yearly earnings, yearly federal tax, yearly FICA, etc., and quarterly earnings and deductions.

The collection of data pertaining to the men on the company's payroll constitutes the man file. It is to be noted that whereas the ID, name, address, etc., for a man are entered directly into the man file, the earning and deduction data are accumulated as a result of payroll processing.

The data for a typical cost account on a given project are an ID for the account, account description, estimated quantity needed, estimated cost of material, estimated dollars for labor, real quantity purchased, real dollars spent on material, real quantity installed, real labor dollars spent, and number of total hours worked on the account per craft type. The collection of cost accounts constitutes the cost account file. The initial entries (estimates) are entered at the time the project is started, and the remaining entries are accumulated as the project progresses. For report-generating purposes, cost account entries containing the words "subtotal" or "total" in the description may also be defined.

Conceptually, the data processing involved in payroll is merely to accept and check field report data, generate the temporary data (check register) for the pay period, and produce various derived reports. Again referring to Figure 4, the field report data

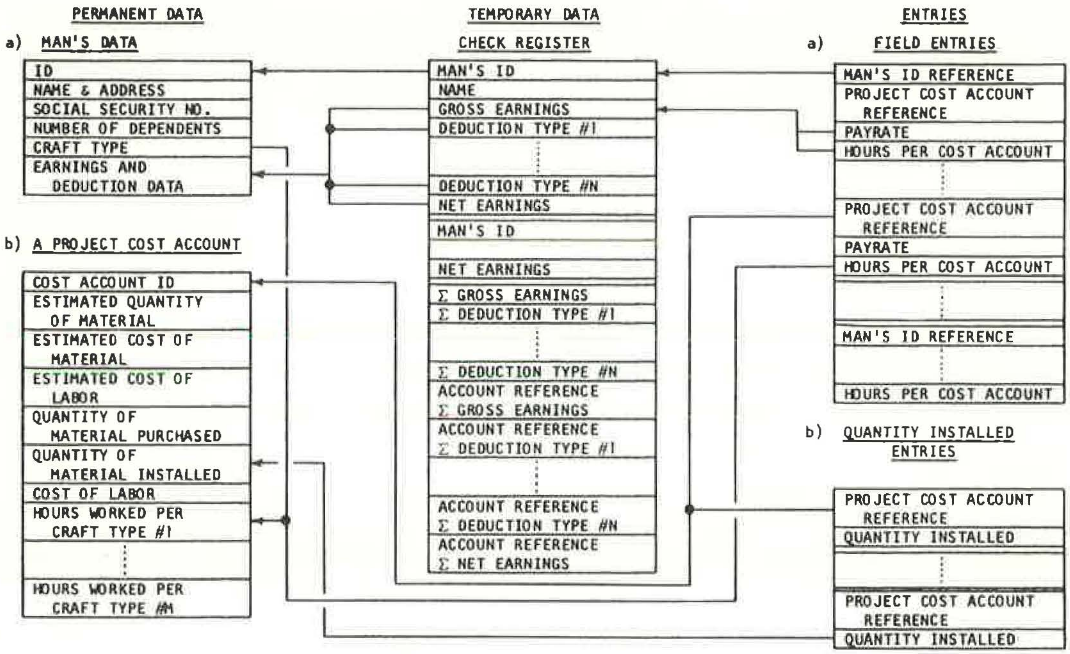

Figure 4. Payroll data structure.

are made up of the man's ID reference, a cost account reference, a pay rate, and number of hours worked on the given account at the stated pay rate. As many sets of account-rate-hours as needed to describe the man's activities may be entered for any given period.

The checks for payroll may be written on the basis of the field data and the man data (Fig. 4), and the man data and cost accounts updated accordingly. The check register, together with the permanent data references, also contains all the information necessary to prepare government reports, union reports, and insurance reports.

The input of quantity installed for each account active during the period completes the data processing necessary to update the cost accounts. By entering the ID's of various pieces of equipment instead of men, this same system can be used to charge a project cost account with equipment rental expenses, using the rental rate of the equipment where the man's pay rate is used. The ID of the equipment would indicate to the system that no check should be written and that the data should not be used in union, government, or other similar reports.

Estimating

The data needed for estimating consists of three levels of hierarchy: the production history, estimate items, and estimate data (Fig. 5). The three types of data will be discussed in turn.

1. Production history consists of two types of data: descriptors and individual records. The intent of the descriptors is to provide the most flexible means possible to retrieve from production history data relevant to a new project being estimated. These data are transferred to the estimate for review by the person pricing the project. For generality, a three-level hierarchy of descriptors is considered.

The first and second levels of description for a new project need be input only once, when the estimating process is started. The first level of descriptors would narrow the search to only those production histories that have the same first level descriptors as the one entered. The second level descriptors would further decrease the area of search.

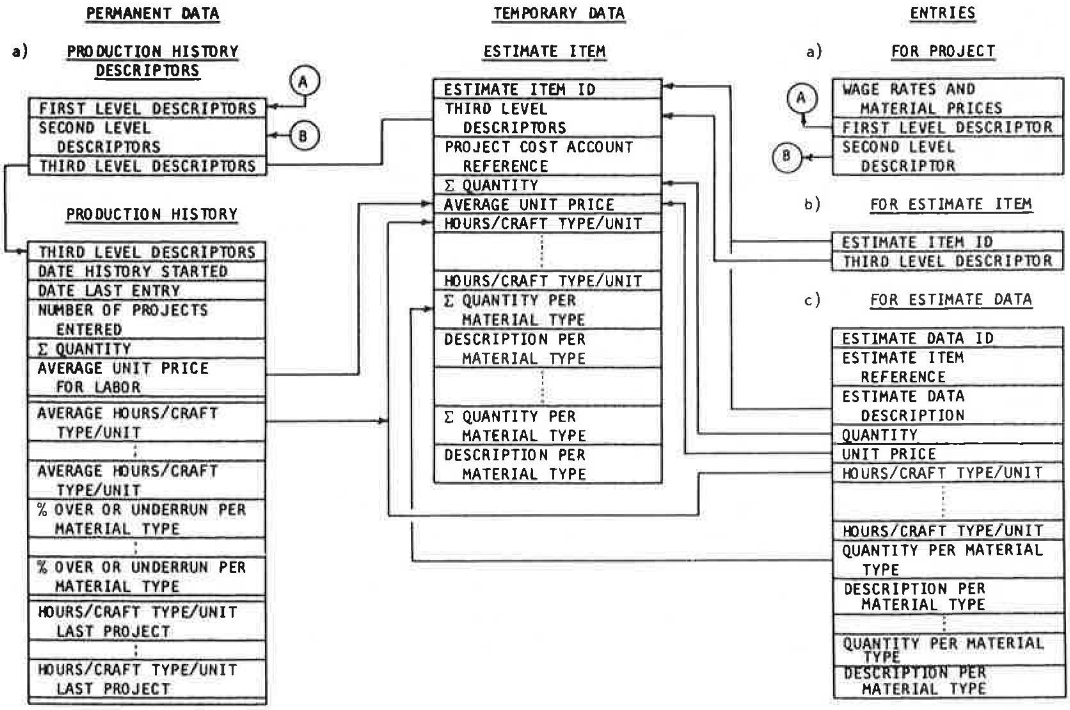

Figure 5. Estimating.

Third level descriptors would be entered for each estimate item on the new project. The third level search is limited to only those histories that have had a match on the first and second level. The third level of description of a production history record must match the descriptors of each estimate item. Examples of possible descriptors would be

Level #1	SOUTHWEST ILL GRADING NORTHEAST WIS GRADE AND PAVE
Level #2	100,000-200,000 CY 20-30% ROCK EXCAV 800,000-1,000,000 CY 10-15% SAND
Level #3	10,000-12,000 CY ROCK RIPPER WASTE 3-4 MI. +3% GR 8,000 CY CLAY 2,000-2,500 FT. HAUL -2% GR. 95% COMP.

The production history record referenced by the descriptors consists of the date the history was started; date the last data were entered into this record; total number of projects making up the data in record; a summation of the quantity of the item installed; average unit price for labor; average hours worked per craft type per unit; and percent overrun or underrun of material estimated as opposed to material purchased.

The data for the last project that was entered into the production history record are also stored separately (Fig. 5).

2. Estimate items comprise a file created as the particular project is being estimated and which remains in existence as long as the project runs. If the company was not awarded the project, the estimate item file for the project would be removed. The estimate item contains information such as estimate ID, third level descriptor, cost account reference (may be filled in later if the company was awarded the contract), summation of estimated quantities, average unit price, hours per craft type per unit, summation of the quantity of various types of material estimated, description of the

material, date of the last job, date the production history was started, and number of jobs that are included in the production history.

If a matching production history record exists, the minimum input necessary is an estimate item reference and the quantities of the various material types. This minimum input is valid only if the summation of quantities is equal to one (i. e., if the production history record refers to a single entity), in which case the data on hours per craft type per unit is transferred to the estimate item, together with the summation of the products of wage rate per craft type times hours per craft type. This summation would appear as the average unit price. The estimated summation quantity per material type would be increased or decreased as indicated by the percentage overrun or underrun from the production history.

The other case where production history would be used is where only quantity and material type are entered into the estimate data along with the estimate item ID. In this case, at the user's option, either the average unit price or the average hours per craft per unit from the production history would be used in the estimate item. The material would be treated as described previously.

Unlike in accounting and payroll, the data on the first level of hierarchy, i. e., the production history, need not exist in order to operate at the second level, i. e., the estimate item. If the production history does exist, it will be accessed by the proper estimate item; if not, a new production history would be created after all the data from a given project were available from the project's cost account file.

3. Estimate data consist of the information that a takeoff man normally assembles. The input (Fig. 5) consists of ID for the estimate data, estimate item reference, estimate data description, quantity of this estimate data, unit price for labor, hours per craft type per unit, quantity of material type No. 1, and description of material type No. 1. There may be several sets of material data for any given estimate data ID. Any information that is supplied to an estimate item by the input of estimate data would override the corresponding data that would normally be retrieved from the production history.

Regardless of the type of data processing employed, the estimating process begins by the takeoff man assigning an ID to each estimate data on the project summary sheet, and then giving the same estimate item ID to all summary sheet estimate data that he wishes to group together.

A meaningful way to utilize the data structure described would be to enter in the estimate item the third-level descriptors described. This would allow for a recall of the relevant information from the production history. The wage rates and material prices would be entered before the estimate items are specified. From these data, plus the production history, the cost of each estimate item would be calculated. The estimator could then review the resulting calculated cost. If the cost was satisfactory to him, he would approve it. If the cost did not meet with his approval, he would make the necessary modifications and reenter the modified estimate item data. After he has satisfied himself concerning each estimate item, the recap sheet and total estimate could be prepared.

A further step in exploiting the available data structure would be to associate cost account references with the estimate item. In this way, it would be possible to obtain from the cost accounts the hours per craft type per unit for the production history; such information could then be retrieved and used in preparing future bids. For example, a third-level descriptor in production history may be:

48" MANHOLE BLOCK 10'-12' 24" CORBEL

Upon completion of a project, data from cost accounts referenced by estimate items having this descriptor would be transferred into the appropriate production history record. When the same descriptor is used for an estimate item on a new project, the user would have to enter only the actual material quantity for the particular manhole. The data returned from production history would be the hours per craft type for that entity. In this fashion, on the assumption that production costs do not vary for similar items where the material quantities vary within certain limits, labor costs on classes of similar items can be treated for estimating purposes as a single entity.

CPM Planning

The data for CPM planning consist of information or individual operations of a project or network and derived quantities obtained from processing the network. The data for a single operation of a CPM network (Fig. 6) consist of operation ID; estimate reference (could be used as the operation ID); cost account reference (could be used as the operation ID); operation description; operation duration; cost of the operation; operations immediately following the given operation; resource types and quantity, e.g., crew size and composition and/or equipment requirements; earliest start target date; earliest finish target date; latest start target date; latest finish target date; number of shifts worked on the operation; and indicator if work will take place on Saturday, Sunday, or holidays.

The minimum information that is required for an operation in order for the network calculations to take place is (a) one of the three possible operation ID's; (b) operation duration; and (c) operations immediately following the given operation. The start date of the project would be entered for calendar dating purposes.

From this information, the CPM calculation can supply the earliest start time, latest start time, earliest finish time, latest finish time, and the total and free float for each operation.

Three files on CPM operations can come into existence as a result of processing a given CPM network (Fig. 6). Two of these files are of a permanent nature, while the third is temporary and is used only for trial-and-error work in an effort to get the best schedule possible. The original network data, generated at the time the project is initiated, are stored in the base network data file. Experimentation or simulation with the network may be performed at any time. If it is desired to save the data entered to modify the network, information on the affected operations is stored in the network update file, but the base network file remains unchanged. The temporary data pertaining to trial network modifications exist only temporarily in the network modification file. In CPM, as in the payroll system, equipment may be treated in the same manner as labor.

INTERACTION OF DATA STRUCTURE

In this section the interaction of the data structures of the individual operations on a company-wide basis is discussed. The interaction is described separately in terms of

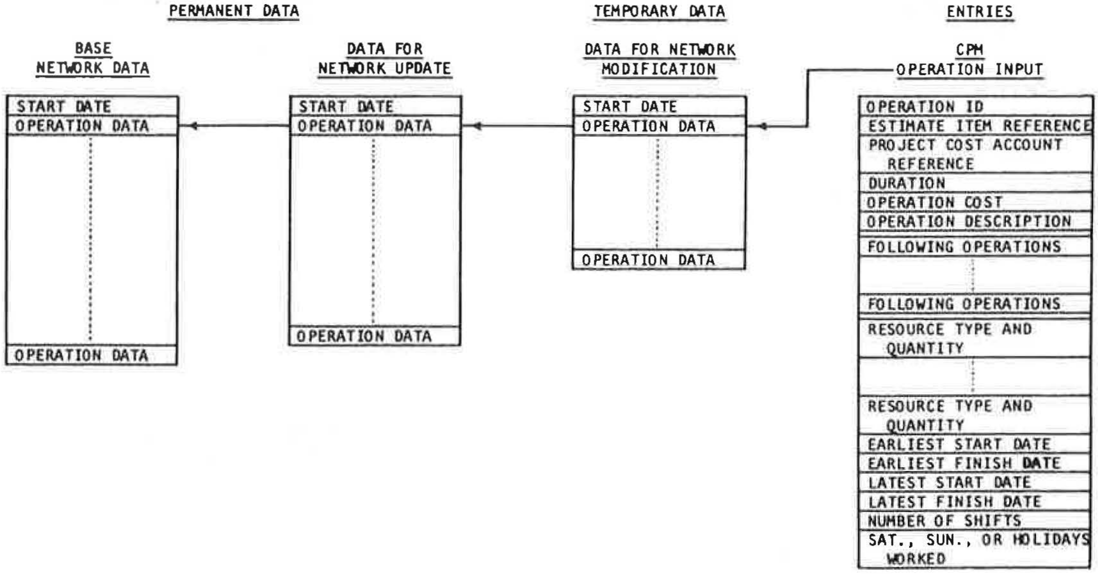

Figure 6. CPM.

the three major activities associated with a given project, namely, bidding and pre-project planning, project control, and generation of historical records.

Bidding and Pre-Project Planning

The initial steps in setting up a project are (a) to prepare an estimate; (b) to set up the project cost accounts; and (c) to obtain an initial CPM network.

The process of estimating was discussed previously. It should be remembered that in preparing an estimate, the pay rate for the various crafts, the material prices, the project overhead cost per day, and the first- and second-level descriptors (referring to the entire project) are initially entered for the entire project.

By associating a project cost account with each estimate item, the output of the estimating process (i.e., estimated material quantity and labor requirement) becomes directly the initial data of the project cost account.

The data for the estimated quantity of material in the project cost account can be transferred directly from the summation of quantity data of the estimate item that references it. The estimated cost of material would be the summation of quantity of the estimate item multiplied by the proper material price. More than one type of material may be involved in a given project cost account.

The estimated cost of labor may be arrived at either from the product of quantity times average unit price or the product of summation quantity times hours/craft type/unit times pay rate/craft type. From the same data, the estimated unit price and cost data are also available to the project cost account.

Figure 7 shows the transfer of information from the estimate and production history data. In this and succeeding figures in this section, only the project cost account data are shown in their entirety. For all other types or data, only the actual items directly interrelated are shown.

Since an operation on the CPM network may reference both a project cost account and an estimate item, the project planner need only supply the operation duration as

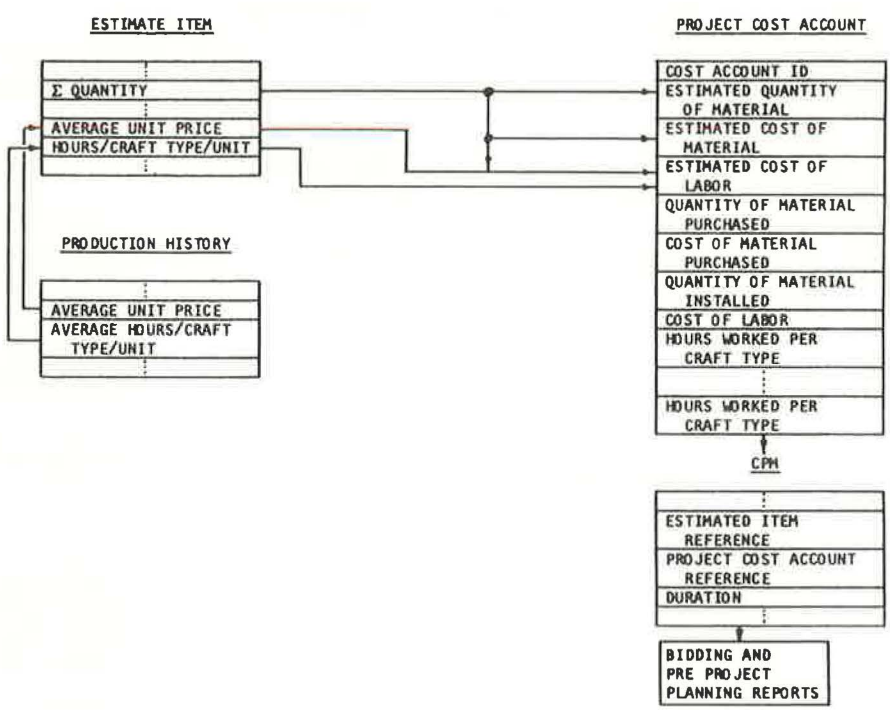

Figure 7. Bidding and project planning.

input for the initial CPM network. The initial contents of the cost account referenced by the operation already contain the labor cost, the material requirement, the material cost, and the manpower resource requirements for the operation from the referenced estimate item. This completes all the information on the operation necessary to run the initial (base) CPM network calculations. In fact, with the manpower resources available, the CPM program can calculate the operation durations directly.

The output of the initial network processing can supply the CPM results described, as well as the total dollar requirement for the project through each operation. The CPM duration multiplied by the project overhead cost per day may be used to calculate total project overhead cost.

In-Progress Project Control

In-progress project control makes simultaneous use of data from accounting, payroll, and CPM (Fig. 8). The data processing for in-progress control is essentially of two types: (a) progress reporting, accomplished by updating of the project cost accounts; and (b) control proper, which involves the preparation of reports directing changes to be made in the conduct of the project.

The updating of project cost accounts is accomplished as follows. The quantity of material purchased for a project cost account and the cost of that material are obtained from expense account entries of the accounting system. The information for material installed is a result of the quantity report input data. The labor cost for a project cost account is calculated by multiplying the hours worked per cost account from field report entries times the appropriate craft pay rate. Hours worked per craft type per unit for a project cost account come directly from the payroll system field report entries. An additional interaction of payroll and accounting in the in-progress phase is the transfer of the payroll cost information to the proper accounts in the accounting system, i.e., to reduce the cash account by the value of the net payroll or increase an account payable by the value of a deduction type.

For control purposes the system has available, through the CPM operations referencing project cost accounts, the data needed to update all CPM operations with regard

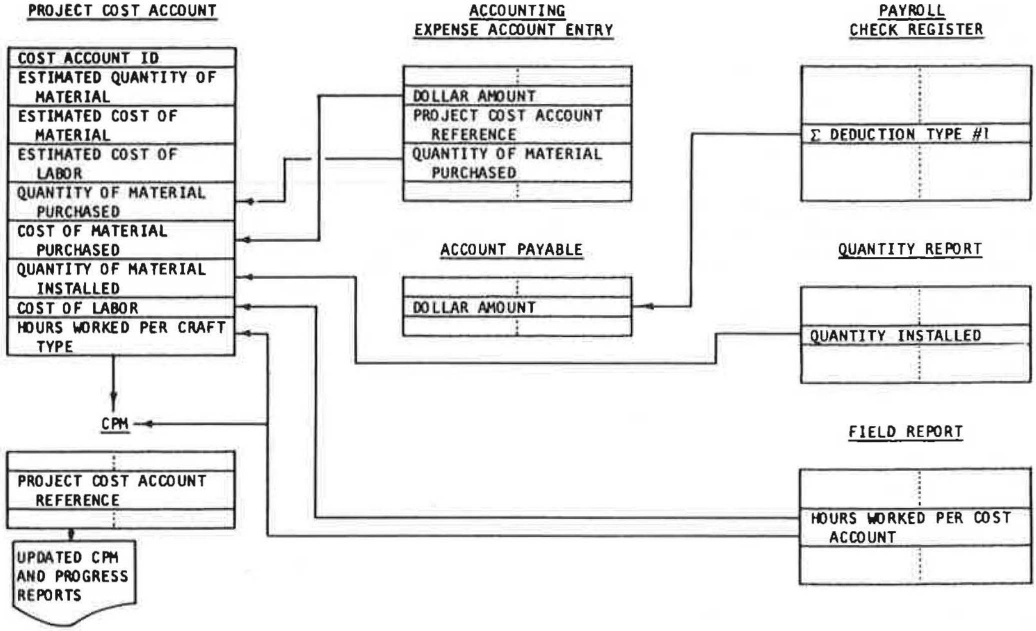

Figure 8. In progress project control.

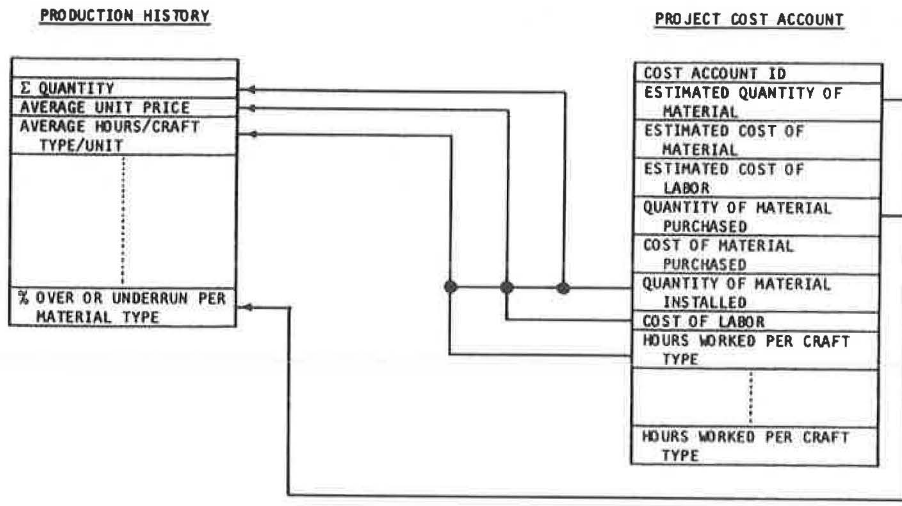

Figure 9. Generation of historical record.

to duration and cost. These data include (a) material yet to be installed; (b) rates at which installation has been occurring; and (c) size of the crew presently working (obtained directly from the field report entries).

The data on operations to be modified would be stored in the network modification file. The user can inspect this file or ask to run a new CPM network calculation. Should the automatic update data meet with the user's approval, it can be saved by transferring the data to the network update file. If not, the user can make any additional modifications necessary before accepting the results and transferring the data to the network update file.

Generation of Historical Record

At the successful completion of a project, information on the project may be transferred into the historical record. The generation of historical records involves interaction of only two parts of the system, payroll and estimating.

As described, a project cost account can receive its initial data directly from an estimate item that, in turn, received its data from production history. In the generation of historical records the data flow is reversed and goes from the project cost account into the production history (Fig. 9).

The summation of quantity in the production history record is a direct transfer from quantity of material installed. The average unit price for production history is arrived at by dividing the labor cost by the quantity installed, both taken from the project cost account, and then averaging it into the data of the production history record. The hours worked per craft type involve a direct transfer and computation of the average. The last item of production history, percent overrun or underrun per material type, is obtained as the quantity purchased divided by the quantity estimated and then averaged into the production history for each type of material.

In preparing the production history, the user would specify whether he wanted to operate with unit prices or if the production history is to maintain the summation of quantity equal to one, i.e., deal with separate entities. In the latter case, quantity of material installed from the cost account is used to calculate only the percent overrun or underrun per material type in the history record.

In the case where summation of quantity is kept in the production history, it would refer to the first material type listed in the project cost account. There are cases where this may be a dummy entry, such as square feet of forms. The remaining material types could then be wall ties, plywood, 2 × 4's, etc. In this way, the average

unit labor cost, as well as the average hours worked per craft per unit, would be available from production history in terms of one type of material, i.e., forms in the example given.

SUMMARY AND CONCLUSIONS

A construction company data processing and management system consists of three main data sources:

1. Accounting—which is made up of two parts, the financial statements and the supporting accounts.
2. Payroll—where two data files are utilized, one for the data on the employees and one for the project cost accounts.
3. Estimating—where one part of the data, the production history files, can be used in aiding in preparation of the estimate. The other part of the data, the detailed project estimate, is used for pricing and buying the job.

Interaction of data takes place within each of the three parts. However, a much more important aspect is the interaction that occurs between the three parts and makes up the total construction management system.

The CPM network representation of a project, in conjunction with the project cost accounts, crosses the lines of all of the three main parts of the system, with the project cost accounts acting as the central data repository and CPM taking on the role of an interpreter of this data. The result of this combined action of CPM with the project cost report provides management with a simulation tool by enabling them to modify the network and interpret the effects of the modifications on the project under consideration. This combination also gives management a project control tool by reporting on current conditions of the company's projects and predicting the outcome of these projects. The automatic update of the CPM also can suggest changes in operating procedure that management may wish to implement.

The last task performed on a project is the storing away of the pertinent data for future reference. This information comes from the project cost accounts of completed projects and feeds back into the estimate of new projects.

It can be concluded that with the data structure described, a management system that would be of considerable help to a construction company could be implemented. It is to be emphasized that the data structure presented in this study deals only with the format of the data, and does not depend in any way on the contents of the data. The assignment of contents, i.e., the identification of cost accounts, descriptors, estimate items, accounts, etc., is entirely up to the user and can in fact resemble quite closely present practice. It also follows, however, that the value of the system to the user depends exclusively on his ability to formulate his particular requirements within the context of the structure presented.

A common complaint about large-scale integrated systems is that they require a considerable amount of additional information to identify the interrelationships involved. It is believed that in the structure presented, these additional data are an absolute minimum, in that only one level of referencing is required at any point. Furthermore, in some cases even this can be omitted, e.g., by assigning identical ID's to estimate items, CPM operations, and project cost accounts.

It is obvious that much work remains to be done before the total system outlined in this study could be brought into existence. With the idea of the total system operating in an on-line conversational mode, careful consideration will have to be given to the physical structure of the data files involved. However, the existence at the present of on-line payroll (1), accounting (2), and CPM (3, 4) programs incorporating many of the concepts described, even though operating independently, attests to the fact that the implementation of the system is feasible.

Several extensions of the system presented are possible. In particular, the usefulness of the production history can be increased in at least two ways. First, by storing the variance as well as the averages of all the data listed in the production history record, these data could be used for PERT-type calculations on project durations, man-

power requirements, and costs. Second, the production history may contain CPM subnetworks associated with third-level descriptors. Whenever a CPM operation references an estimate item that fits an existing third-level descriptor, the corresponding CPM subnetwork could be retrieved and automatically linked into the project network.

ACKNOWLEDGMENTS

The motivation for this study, as well as many of the concepts involved, resulted from the authors' work on the research/service program, "Building Computer Technology" conducted by the Civil Engineering Systems Laboratory of the Department of Civil Engineering, University of Illinois, under sponsorship of the State Technical Services Act and participating professional firms. The project is under the direction of Drs. L. R. Shaffer and S. J. Fenves.

Professor L. R. Shaffer, Dr. L. T. Boyer, and Mr. A. C. Brooks were extremely helpful by contributing of their time to long discussions relative to the subject matter of this study. Many of their ideas are incorporated herein, for which the authors are grateful.

REFERENCES

1. CESL/STSA Remote Terminal Payroll User's Manual. Civil Engineering Systems Laboratory, Univ. of Illinois, Urbana, March 1968.
2. CESL/STSA Remote Terminal Accounting User's Manual. Civil Engineering Systems Laboratory, Univ. of Illinois, Urbana, July 1968.
3. Brooks, A. C., Leahy, J. P., and Shaffer, L. R. A Man-Machine CPM System for Decision-Making in the Construction Industry. Construction Research Series No. 9, Department of Civil Engineering, Univ. of Illinois, Urbana, June 1967.
4. CESL/STSA Remote Terminal CPM User's Manual. Civil Engineering Systems Laboratory, Univ. of Illinois, Urbana, July 1968.