Some aspects of recording Graduated National Insurance contributions

By J. Drummond

To control the recording of graduated contributions, the Ministry of Pensions and National Insurance installed an Emidec 2400 system, including 24 magnetic-tape units, two file search units, two Xeronic printers, and a card-to-tape converter.

Introduction

In 1959, when the Government announced proposals for a system of graduated National Insurance contributions to be collected through the Inland Revenue's P.A.Y.E. system, one of the major problems which faced the Ministry was to find the most efficient way of collating the contributions into individual insurance accounts. Time was short, as the collection of contributions was to start in April 1961; therefore, by the time the first flood of returns was received in 1962, the full range of over 26 million individual accounts had to be set up and the recording system fully operative. Because of the time needed to obtain machines, staff and buildings, we had only a few months in which to decide the major issue of whether we were to rely on clerical methods, a punched-card system or a computer system.

Our study of the problem could not be termed a systems study in the normal sense, as the specification of requirements was so scanty at that early date. Nevertheless, after a comparison of possible methods on the basis of known requirements, plus some anticipation of others, an Emidec 2400 system was ordered in August 1959. We were in no doubt as to the risks in setting up a new system involving enormous quantities of data with an untried machine and untried methods, and without an escape route to another system. We have had a very trying time in the last two years, but have not changed our opinion that the computer system provides the best answer to our problems.

In this short paper I propose to outline one or two aspects of the work which were, in our case, inflated to major problems by the volumes of data involved. Some of the methods we use could, however, be usefully employed in smaller installations.

Volumes

The system has to maintain over 26 million accounts in such a way that details of contributions and supporting documents can be produced up to 52 years after receipt—a man's "insurance life" is from age 18 to age 70. Over 40 million additions and amendments have to be entered into the account files each year, but for long periods the intake is at the rate of over $1\frac{1}{2}$ million per week. Around 10% of these items contain errors in identity detail which have to be investigated, and in a further 30,000 or so cases per week exceptional circumstances call for action beyond the computer's discretion. For example, the contributions may appear to have been paid erroneously before age 18 or after age 70. The system has to cope with about 6,000 inquiries a day and has to issue, in the course of about ten weeks at the end of each year, some 11 to 14 million individual statements of contributions paid in the preceding tax year.

Condensing and sorting data

Two of the major problems we had to solve in the initial study were how to confine our standard recording to the minimum essential for dealing with the normal pension claims and inquiries, whilst maintaining access to the supporting documents relating to any individual, and to eliminate the sorting and inter-filing of such documents. Unless we could eliminate this sorting, it would absorb a considerable staff, and employers would have to submit separate returns of income tax and graduated contributions, instead of the combined returns now in use.

The system we evolved not only met these basic requirements, but condensed and simplified the work of the ancillary clerical and punched-card sections to what may best be termed mass-production procedures.

In general, we followed three principles. The first principle was that all information with a low reference activity should be kept off the active computer records, and should be retained in sub-dormant archives. Secondly, there should be no sorting or re-routing of documents, including punched cards, other than to main streams compatible with the major input programs of the computer. The computer record should include a cross reference to the location of the supporting document, and the computer should carry out the sorting of incoming data to the order of its files, and outgoing data to the order most convenient to the recipients. Thirdly, data in all their forms must be condensed as far as practical. The computer records must be very brief and densely packed. The computer programming system had to be designed to eliminate unnecessary repetition in card punching. All archives should be reduced to microfilm.

Setting up the computer records

We had to set up the computer records by April, 1962. This was, in itself, a major systems study and programming project. Punched cards had to be prepared showing the basic detail of every insured person, including title, surname, initials, sex, date of birth and full postal address. The detail was obtained from the clerical records relating to the flat-rate insurance system. As some cases with long addresses required two cards, over 30 million cards were prepared. The data were then transferred to magnetic tapes. The prototype computer in the E.M.I. factory was used to process the data to produce as an end product two magnetic-tape files—one merely consisting of full names and addresses to be used annually for the issue of statements of account, the other the Contributions File showing essential personal detail which would be in day-to-day use. The programs to be written included a comprehensive editing program, which checked the feasibility of every character as well as every set of characters, and converted the data to the form agreed as best suited to normal operations. Some sets of characters were converted to pure binary—some remained in alpha. The key National Insurance number was recorded in both forms. A point worthy of mention is that we find very strict editing of all new data very time-consuming, but absolutely essential. A number of sort and merge programs were also necessary, as the clerical records were not in straight rising sequence, but in a staggered sequence suited to the method of returning stamped National Insurance cards.

Computer records

The magnetic-tape record of each account on the Contributions File in daily use is reduced to an average of twelve computer words of thirty-six binary digits. In the time available, I cannot detail all we cram into these few words, but one or two aspects are worth mentioning. Whilst the National Insurance number is our main identification key we have to carry part of the name in the account to prevent items with misquoted numbers being incorrectly posted. Analysis showed that three letters of the surname and up to two initials would suffice. No significant benefit was obtained by using the full surname. Other features are that the accounts are of variable length and that, within the account, the sharing of words by a number of items is the rule rather than the exception.

Included in each account is an entry for each item received in the year, showing not only the amount paid, but also the name of the employer and the document's serial number. This detail may seem inconsistent with the principles of condensing data, but the name of the employer is essential to our day-to-day work, and we may need to locate the supporting document. However, as we shall see, the entry of this rather detailed item has involved only a comparatively minor amount of human effort.

Contribution returns are received in employer groups from the Inland Revenue, and have to be returned within ten days. Each group normally consists of a single card per employee showing, among other things, National Insurance number, name, and the amount of contributions paid; preceding the individual cards is a control card showing the total for the group. The cards are serially numbered and microfilmed in the sequence received. The Punched Card Sections then produce for each group a summary card showing the microfilm serial number, employer's name, tax district and total payment. In respect of the employees, the operator punches into cards only a designator, the National Insurance number, three characters of surname, two initials and the amount of payment. She can put four of these items into one 80-column card, thus reducing the time involved in converting the items to magnetic tape. The summary card is fully verified, but of the detail on the employee cards only the cash item is verified. The computer will identify whether the number or name has been incorrectly punched.

The data in cards are written on to magnetic tape on an off-line converter, and the tape is edited by the computer. This program not only checks that the items are feasible and that the amounts paid by the individuals reconcile with the group total, but it also expands the item relating to the individual to include the name of his employer and the precise microfilm frame number of his contribution card.

Subsequent computer processes sort these items to National Insurance number order, and match them against the individual accounts. Many are rejected because the number has been misquoted. These have to be put out for investigation, and read in again when the correct number has been determined. The investigation necessitates reference to the basic documents, and, to ease the work of the Microfilm Library, the items are sorted to Microfilm frame order before being put out in the form of punched cards. These cards are extended in due course by addition of the correct National Insurance number, and returned to the computer. To ensure that these exceptions are eventually brought back into the system, the computer maintains a magnetic-tape file of Outstanding Queries, and issues reminders on items which have been outstanding more than a specified period. The punched cards returning with correct numbers added take the items off the Queries File, amend them, and put them back into the work flow. At the same time they make appropriate additions to a Correction Index which automatically corrects any identical misquotations on future items.

Similar principles are followed in relation to other types of input and output too numerous to mention here, but before leaving the question of condensing data, I would mention one more point. At the end of each year we have to summarize our Contributions File by transferring details of postings during the year to a History Schedule, replacing them by a new figure of accumulated pension rights. In its paper form this schedule would be over 280 miles long and one foot wide, but we have fitted microfilm cameras to our printers and the end product is less than 400 spools of microfilm each 100 feet long.

Programming

Circumstances did not allow us to follow the normal procedure of carrying out a detailed systems study and laying down a firm job specification before we started programming. Time was short and we had to find programmers, train them, and start them working on the design of the system before all requirements were known. Discussions within the Ministry, and with other Departments and outside interests, were to continue over the succeeding two years whilst the programmers were building up the system from the skeleton created by the initial study group. It says much for the patience, as well as the ingenuity, of the programmers that, though they had no previous experience, they have coped with the difficulties of expanding and amending programs to meet changes in requirements, in addition to the burden of putting a job of this magnitude into a new type of machine.

One of the major difficulties was that the system consists of complex series of programs integrated in what might well be described as a "clock process." Some are carried out daily, others weekly, others fortnightly and some annually, but all are geared together. With the exception of the annual processes—for which a little more time could be allowed—all programs had to be built up concurrently.

The general line of approach we adopted was to divide the skeleton flow chart designed by the initial study group into Programming Areas and make a small team, usually one Senior Programmer and two to three programmers, responsible for detailed systems study and programming of each Area. A detailed specification of the content of the main contributions file was drawn up and used as the corner stone of the system. Internal design of programs was left very much to the discretion of each team, subject to the format of inputs and outputs being agreed with teams responsible for adjacent sectors of work. This was by no means an ideal method, but thanks to the team spirit which seems to stem from a job of this nature, no real difficulties arose.

What did cause a great deal of extra work was that, as all programs were being built up concurrently, all teams, other than those dealing with basic input, had to design special programs to generate data which would simulate that normally passed on from preceding programs.

Comprehensive control routines

We had, of course, specified that all the teams were to employ certain common housekeeping routines, but we soon realized that the normal type of housekeeping was not adequate for this type of "continuous flow" involving a complex integration of programs. We had to progress to the stage where the computer master-minds the whole system instead of merely safeguarding the program on which it is engaged at any given time. To explain this at all adequately would require a separate paper and, on this occasion, I can do little more than outline what the control routines cover.

Program interrupt system

The control routines must cater for every possible type of program interrupt signal. The Program Interrupt signalling built into the Emidec 2400 system is extensive. In general terms the signals can be divided into four groups. One group signals failures in the computing unit and the associated four buffers, including failure to obey instructions, and irrecoverable loss of parity (the logic of the tape units and buffers covers restoration of single errors in parity). A second group covers the various routine signals of section and sub-section markers from magnetic tapes and paper-tape units. A third group covers the non-routine signals from tape units and the slow input and output equipment. A fourth group caters for a variety of interruptions by the console operator. These include routines to remedy the situation where one of the off-line units is in a condition where time sharing with the computer is necessary. A simple illustration is that where a Xeronic printer signals that it cannot print a form, either because of a misread of the tape or a buffer fault, the console operator injects an interrupt instruction which makes the computer jump to a subroutine to take the Xeronic magnetic tape under control, set it back one block and reconnect to the Xeronic before returning to main program.

Tape control

We maintain over 140 groups of magnetic tapes and follow the three-generation principle. In total over 4,000 tapes are held. Every tape is labelled, both externally and internally, with a heading block consisting of a permanent stock number followed by a code showing the type of data currently held on the tape, a program generation number denoting the age of the data, and a spool sequence number within that generation number.

In assembling tapes for any program, the Tape Library staff check from the external labels that the correct input tapes are sent to appropriate tape units in the right sequence, and the loaders make further checks on sequence. The Library staff should release to the output units only tapes containing redundant data. However, they have the normal human failings and make mistakes, so it is necessary to have computer checks.

To design subroutines to check that the input tapes are correct was not difficult. What did present a problem was the system for checking that the output tapes are redundant. It was necessary to build into the system a list of tape groups showing, in respect of each group, which generations are still active, and arrange for the heading blocks of all output tapes to be compared against the list before allowing the current program to re-head and use them. The "active list" has to be kept up-to-date, and, ideally, the running of any program should amend the list, but the amendment is far from simple. The use of multiple inputs and outputs, and the running of connecting programs in various time cycles, creates the position where the running of one group of tapes releases a number of other groups written at varying times. In practice, the list is only brought up-to-date at certain points in the chain of programs. A teleprinter output informs the Librarian of changes. It could almost be said that the computer dictates to, and checks on, the Librarian, but we have to provide an escape route to meet situations where the Librarian has to break the standard rules—for instance, if he is short of free tapes. This can be done by injecting on paper tape an order which overrides normal control.

Program control

The control routines cater for adjustment of programs to suit the current availability of tape units and buffers, the amendment or replacement of normal Program Interrupt routines to suit particular programs, and provision of management statistics.

The control routines, which must always be available and up-to-date, are held in 2,000 words of the store. At the commencement of any program, the contents of these 2,000 words are copied out on to a magnetic tape, so that they may be restored to their original value if through any mischance the program fails.

At the end of the programs which include revision of key locations, the state of the locations is dumped onto magnetic tape. This tape is retained until the input/output tapes involved in the relevant group of programs are redundant, so that if a reprocess of any program becomes necessary some time later, the key locations can be restored to the appropriate values.

This is merely an outline of the processes—the detailed instructions relating to them take a large manual—but this outline is perhaps sufficient to indicate the difference in controls required for an integrated system of programs and the type required where programs are not so tightly interlocked.

Running times and change-over routines

Our experience indicates that programs should, where possible, be arranged so that a computer "run" is around 30–45 minutes, and small-volume programs should be grouped to produce a run of this dimension. Time needed to change over from one run to another is about five minutes. Conversely, the risk of re-run or re-process being necessary dictates that, where the volume of data is too great to run through in $\frac{3}{4}$ hour or so, either it should be divided to give runs of this dimension, or re-start points with dump routines should be written into a single run.

General

The present position is that the system is running under the full load. Operational faults and delays were rather high at the outset, but are substantially reduced now the staff are gaining experience of the system. Faults in programs have been remarkably few and have usually been caused by unexpected varieties of data fault of the type which would normally come to light in parallel running with an existing system. The full loading has shown some weaknesses in the tape units and these are now being remedied. We are taking nearly three shifts to do the work we expected to clear in two shifts, but we are confident we can achieve that target in a reasonable time.

Reference

Polley, D. W. (1960). "A Progress Report on the Introduction of A.D.P. for Recording Contributions Paid under the new Graduated Pensions Scheme," *The Computer Journal*, Vol. 3, p. 117.

Summary of Discussion at Cardiff Conference

Mr. M. R. Mills (*Honeywell Controls Ltd.*): Why are records on magnetic tape expensively and slowly copied on to microfilm at the end of a year rather than stored on magnetic tape, and why are certain exceptions stored both on magnetic and paper tape? In other words, does the speaker lack confidence in filing long-term on magnetic tape?

Mr. J. Drummond: The detail copied on to microfilm at the end of the year could be stored on magnetic tape rather than microfilm, but the degree of future reference would be low—probably less than one reference per reel of tape or microfilm per day. The overriding consideration is cost. Long-term storage on magnetic tape would tie up a lot of capital—about £30 per reel. Furthermore, reference at low activity would be very expensive. Details of the inquiry would have to be punched on to cards, and the data transferred to tape, then the inquiries put against the archival tapes on the computer or on File Search Units, and the reply would have to be printed. The Computer or a File Search Unit would take nearly five minutes to scan a reel of tape and in this time the most junior clerk could take the inquiry in its basic form, find the appropriate item in a reel of microfilm, using a comparatively cheap microfilm viewer, and fill in a form relating to the inquiry.

The dumping of some control detail on to both magnetic tape and paper tape is also dictated by cost considerations. The magnetic-tape version is used for any restoration of detail during, but not beyond, the relevant program cycle.

If it became necessary to recreate "current" conditions at any time after the end of the cycle, the paper-tape version would be used. The cost of thousands of pieces of paper tape was not significant compared with the cost of freezing thousands of magnetic tapes.