THE JOHNS HOPKINS UNIVERSITY LIBRARY

Benjamin Courtright

The purpose of this paper is to describe two projects relevant to the interests of this conference which are active at the Johns Hopkins University. These are, first, the conversion of the shelf list of the University Library to machine-readable form, and, second, the design and operation of a new type of computer-based circulation system.

These activities were initiated and are being carried out at Johns Hopkins as part of an operations research and systems engineering study of university libraries, under the sponsorship of the National Science Foundation. The leadership in initiating and directing this study has from the beginning been that of Robert H. Roy, Dean of the School of Engineering Science and Chairman of the Department of Operations Research of the University.

The Project team has been fortunate throughout in having a close and cordial working relationship with the staff of the Johns Hopkins University Library. From providing a field for examination and case study of systems problems, to sympathetic and responsive reception of proposals for action, the Library has been friend, teacher, and subject all at once. This fortunate and indeed essential state of affairs has been due above all to the perception, the vision, and the understanding of the University Librarian, John Berthel.

The support of the National Science Foundation has been crucial. The cost of the work is small in terms of the long-range objectives which are at stake here, but without the far-sighted support of NSF these undertakings could not have been attempted at all.

Conversion of the Shelf List to Magnetic Tape

It is hardly necessary, at this date, to dwell on the desirability of having the shelf list of a library in machine-readable form. Suffice

---

Benjamin Courtright is a Research Associate in the Department of Operations Research of the Johns Hopkins University, Baltimore, Maryland.

it to say that Johns Hopkins arrived at this evaluation quite independently, in 1963, after a careful consideration of system problems. The orientation was definitely not that of electronic data processing (EDP), in the first place. In order to apply modern control techniques and optimization procedures to the problems of library organization and operation, we needed, in particular, data on the library's primary resource, its holdings. The shelf list is monumental, a single, unique file, ninety years in the making. For micro-examination, for looking at one or a hundred items in the collection, it was, and is, quite serviceable; but as a basis for control measures for the collection as a whole, for coping with the new generation of library problems, it was simply unusable. No one has ever regarded it as indigestible, only because no one has ever conceived of having to swallow it. Making this file assimilable became an integral, essential part of the solution to problems of inventory and flow in the library.

This account will omit a detailed description of the specific library problems which motivated the shelf list conversion, and the proposed solutions to them, as well as a consideration of the various means available to accomplish the conversion, and the factors affecting the choice to be made among them. Both topics have been covered in published progress reports. The following then is a case report of actual operating experience with the processing involved.

The conversion process decided upon was accomplished by the following steps:

(1) Producing a working copy of the shelf list by microfilming.

(2) Transcribing the desired elements of information for each record in the file and, at the same time, producing optically scannable copy, by typing the elements desired on a typewriter equipped with the special font required for the scanning process.

(3) Automatically converting to magnetic tape on an optical scanning (character recognition) device.

(4) Editing, reformatting, and updating the cumulated record on magnetic tape.

Microfilming—The need for a working copy of the shelf list is abundantly clear to everyone who has had contact with library procedures. The conversion process, including set-up time, took two to three months; to have attempted this from the shelf list itself, even if it had been possible to do it in the immediate vicinity of this unique file, would have produced endless confusion and frustration for everyone concerned. The choice of microfilming for producing the working copy has been a wholly satisfactory one. The cost of making the microfilm copy has turned out to be much less than the cost for any other available method. The completed microfilm copy is very easily

handled and stored; the microfilm replica of the entire shelf list fits into a box about 1-1/2 feet on each side. Above all, the microfilm copy did the job desired—it served as an adequate working basis for the conversion process. Although they had had no prior experience with the use of this medium for any job of this nature, we were delighted to find that the typists worked about as effectively from the microfilm reader as they do from original copy.

Microfilm does have one signal disadvantage: it is impossible to edit it or mark it for any intended use. Acceptance of this limitation forced the staff into the lengthy and laborious process of constructing a very careful and detailed set of decision rules for the typist to follow in reading and abstracting the data. Instead of depending on the editing of an experienced librarian or subject matter specialist the staff was forced into an explicit codification of the treatment for each of the many varieties of form encountered. The end product of such a process cannot be as good as that achieved through pre-editing by an experienced and competent person. On the other hand, the saving in time and in cost is very considerable; the process of forming explicit guidelines is in many ways a salutary one; and we believe that through the exercise of planned safeguards and checks, the final product will stand comparison very well.

One such safeguard, for example, was a direction to the typist to enter a special symbol "XXX" in place of the record element, whenever she encountered a form of entry which seemed to her not to be covered adequately by the rules. The last step in the conversion process has as one of its parts the printing of a list of records containing these symbols, by a simple computer program, for checking by a trained technical person. This means that such attention must be given only to the 2 or 3 percent of the records containing the anomalies, instead of to every record in the file, and was the last step in the conversion process.

The microfilming was done on a Recordak Model RM automatic microfilmer. This machine has an automatic feed, with a rated capacity of 375 cards per minute. The effective operating rate was about 200 cards per minute, allowing for the necessary shifting of cards and personnel breaks, and the entire operation was performed over one weekend, when the Technical Services Division was, of course, closed. The costs comprised $480 for the rental of the microfilmer; $229 for twenty-four 200-foot rolls of 16 mm. microfilm, including processing; and $253 for the wages of the personnel employed for the purpose. A major part of the latter cost was for some sorting and re-sorting that had to be done on the file in the process. The total cost, then, was $962.

Transcribing—Whether to convert all of the information on the shelf-list card, or only some part of it, was perhaps the hardest decision

of all. The question was considered from every point of view, over a range of possible alternatives. The final decision was made on the basis of the marginal utility of the various elements of information and on the basis of the specific functions for which the machine-readable record was proposed. A detailed consideration of these functions is set forth in the April 1963 Progress Report of the project. The fields of information entered into the record are as follows:

(1) Call number. The complete call number is taken in every case. This includes the date of publication for modern cataloging.

(2) Main entry. For corporate entries, the entire entry is transcribed. For personal authors, the last name and initials are taken. Authors' dates are omitted.

(3) Title. This is the full title in most cases, a shortened form in the case of very long titles and certain other cases. Subtitles are omitted.

(4) The number of pages, or the number of volumes in the case of a multi-volume set. (The number of pages for each volume is not given in the latter case, in local cataloging.)

(5) The physical size of the book. This is, as usually given, the height in centimeters for normally shaped books, or both the height and the width where given under Library of Congress (LC) cataloging rules for unusually shaped books.

(6) The number of copies of the work in the Library's collection.

These basic elements of information have been transcribed for every item in the Library, with appropriate coding so that any given field of the information can be retrieved at will. The records were transcribed, as indicated above, from microfilm readers onto typewriters equipped with the special font used for the optical character recognition process.

Optical Scanning—In a survey conducted to plan for the optical scanning phase of the process, two organizations in the country having fully operational pieces of equipment with sufficient sophistication to do the job were found. The company chosen was the Control Data Corporation (CDC), whose wholly-owned subsidiary, the Rabinow Engineering Corporation, had developed pioneer models of such devices. Having had no experience with either of the organizations contacted, the choice of CDC was made on quite explicit grounds; the scanning device used by the other company accepted a line of only seventy characters, in contrast to a line of 100 characters for the Rabinow machine; the operating rates of the other company's machine were much slower and the formatting less flexible than that provided by Control Data Corporation's operation. The location of the CDC machine, in Rockville, Maryland, has also proved

a considerable benefit, in view of the frequent coordination required in the planning and execution of the work.

This machine is a one-of-a-kind device, built specifically for in-house developmental testing and service bureau operation. Two different types of agreement were concluded with CDC, both of which have functioned satisfactorily. Under the first of these CDC took the microfilm copy of the shelf list and assumed responsibility for the conversion process. They subcontracted the typing work and assumed responsibility for the implementation of the conversion rules referred to above, subject of course to continuous coordination with Johns Hopkins in their application. Following the optical scanning process, they reprogrammed the raw output of the scanner, converting it into the bit code appropriate to the computers. The guaranteed accuracy for the whole operation, specified in the contract, is as follows:

(1) For the call number. Since this serves as an identifying code for the machine record, as well as for the many functions it has in its own right, the maximum obtainable accuracy was specified here. The terms of the guarantee are expressed as "virtually 100 per cent accuracy"; the procedure used was to make a visual check from the print-out of the final product against the original microfilm for this field, in addition to the error detection routines incorporated for the record as a whole, with both CDC and the typing subcontractor assuming responsibility for the checking.

(2) For the remainder of the record. The guarantee specifies that the number of errors shall not exceed 2 per 100 records.

These guarantees have seemed eminently satisfactory, probably exceeding the accuracy of most original copy. We are not as yet prepared to make a statement concerning our own sampling for verification of these levels, but there is no reason to believe, on the basis of experience, that CDC would fail to take full responsibility in the event that our checking shows these rates not to be maintained for some part of the completed record. Their responsibility would take the form either of cash rebates for our cost in instituting accurate correction procedures, or of their redoing the work in question.

The total cost for this processing is at the rate of 4.8 cents per record.

Under the other form of agreement, the typing is done by us, and the typescript is shipped directly to the scanner for automatic conversion. The output in this case is the raw output from the scanner written on magnetic tape; here we do the subsequent conversion processes and reformatting on our computer. The cost for this part of the operation is at the rate of 2 cents per line scanned. The scanning device accommodates 100 characters per line; our records exceed this length in 8 to 10 percent of the total cases.

The second form of agreement mentioned above was instituted in order to provide a solution to a specific operating problem, but the procedure has become interesting in its own right. The problem referred to is the conversion of items from a separate serials catalog. In the preliminary survey and planning, because the items in the serials catalog were sufficiently more complex in their bibliographic record than the monographs, it was preferable not to follow the first procedure outlined above in this case. After much exploring, a very intelligent and competent person with library experience was hired, to work solely on the serial records for a period of time. A suitable typewriter was obtained, an IBM Selectric, fitted with a typing element having the special scannable font of characters, and set up in a place immediately adjacent to the serial librarian's office for ease of reference. This was the first time that such a procedure had been tried, but after some initial difficulties it was found to work very well. Moreover, it was found that the overall cost, per record converted, is very nearly the same as under the other arrangement. Having no experience for a guide, it was expected that the unit cost might be significantly higher (although the total cost would in any case be small compared with the monographic conversion). The knowledge of this as an additional resource at about the same cost level seems highly propitious.

Total Cost—The total cost for the conversion of all records, including the microfilming, payments to CDC for their conversion processes, clerical wages, the purchase of magnetic tapes for the storage of the final product, an allowance for wages and computer time for the editing and reformatting processes, plus an allowance for overhead to the University, amounts to $18,170. On a per character basis, this is at a rate of $.0038 per word of 5 characters, or $.00076 for each character converted.

New Circulation Control Procedure

The circulation control system outlined below was first proposed in a Progress Report issued in June 1964. Testing and development of the requisite means to put it into practice continued for several months after that; and in April 1965, shortly after the opening of the new central library building on the Johns Hopkins campus, the new system was put into operation for the Library as a whole.

Books brought to the circulation desk to be charged out are not opened. They are placed together on the counter, as many as are brought by a single borrower, and oriented so that the book labels

(bearing the call numbers) all face the same way. The borrower's card is set on a small stand facing in the same direction and the books and card together are then recorded on microfilm by a suitable camera positioned a few feet away (see Figure 1). Ten to twelve books can be photographed together in this manner, so that, with few exceptions, all of the books borrowed by one person are recorded at the same time. The charging process is completed by affixing a removable pressure sensitive label, imprinted with the due date, on the outside back cover of each book (see Figure 2). The idea of the labels was originated in connection with the system in order to obviate altogether the need for opening books, so as to speed up the charge-out process, and to provide a more easily noted reminder of the due date.

At convenient intervals the film is developed in a processing unit and then placed on a microfilm reader situated alongside a key-punch machine. The reader currently in use here projects the microfilm image at a magnification which is slightly greater than the reduction ratio of the camera lens, so that the image of the books is somewhat larger than life-size. The call number only is read from the book label and key-punched, together with the borrower's card number by the operator (see Figure 3). Experience has shown that if the call number can be read on the book itself, it can be read from the projected microfilm image. Since this number has had to be legible in order for the book to be in place on the shelf, it is generally read without difficulty; in the occasional case of a torn or mutilated label, however, identification can also be made from author-title information appearing on the cover of the book, the call number then being obtained from the Library catalog. This does not occur frequently.

The discharging of returned books follows the same flow described above. The borrower drops books off at the circulation desk in the usual manner, and the attendant simply photographs them with a card marked "Discharge," in place of a borrower's card (see Figure 4). The key-punch operator converts the call number as above, putting a one-character symbol denoting "discharge" in the place of a borrower's number. Renewals are handled in an analogous manner.

Capabilities of the System—Given the charge and discharge records on punch cards, all of the remaining functions of circulation control are accomplished by a computer and computer program. The program was originated by Dr. Willis Gore, Associate Professor of Electrical Engineering of this University. It performs the common tasks which are always associated with circulation, and in addition has the ability to do a number of things which seem highly desirable but which have rarely if ever been feasible in a large library.

In the category of common tasks, the computer, guided by the program, keeps track of all outstanding charge-outs and deletes the record for items returned; it determines when an item is overdue and prints out notices bearing the proper address, ready for mailing; it calculates fines, where appropriate, for items returned late; and it provides a readily accessible record of the whereabouts of all materials in circulation, or in use internally within the Library.

The capabilities of the system which are new in some sense do not form a well-defined list, but they include:

(1) The ability to compile circulation statistics for each item in the collection and, equally important, to present this information in concise, usable form, for any selected set of material.

(2) The ability to characterize circulation reliably in terms of various categories of borrowers, the types and subject areas of materials borrowed, periodic changes in rates of circulation and long-term trends, or any combination of the foregoing.

(3) The ability to print out at any time a listing of items charged out by individual borrowers.

None of the capabilities listed above is new in the sense that its attainment is theoretically impossible under older systems of circulation control, but in general the work involved would be tedious and time-consuming. Under a manual card file system or any of its logical derivatives, the file can, in the nature of the case, be kept in only one order. Conventionally this is an arrangement by call number. The file however has many other attributes, or dimensions, of potential interest: borrower, category of borrower, type of material, main entry, time and date of borrowing, etc. To retrieve information for any of these other dimensions, the librarian must either make a painstaking search through the whole file, employ a system of mechanically tagging or reordering throughout the file, or generate and maintain a duplicate file for each different approach.

In a computer-based system, on the other hand, the file may be regarded for practical purposes as multi-dimensional; roughly speaking, using it in any dimension does not affect its integrity in any other. There is, of course, a cost associated with the computer's operations, but the cost is small relative to the value which will be obtained for the capabilities listed above and, with the exceedingly rapid growth in the technology of computers, it seems likely that this will be true for other functions in the not too distant future. In addition, the "real time" taken by the computer operation is negligible, in marked contrast to the length of time required for such purposes under manual systems.

Means of Implementation—Encouraged by the Library staff, we embarked upon the conversion of all circulation control processes in the

new Johns Hopkins Library building, then in the final stages of construction, to the proposed design. Our lack of trepidation was matched only by our lack of knowledge of the problems to be encountered. The microfilming of the shelf list was a considerable task but that was a straightforward application using a standard piece of equipment, quite a different affair from finding, modifying, and fitting together pieces of equipment to perform new functions. We have since learned a small amount about lenses and prisms, microfilming equipment, films and film processing, pressure-sensitive label materials, etc., but in the main it has been the humanity and forebearance of our associates and the Library staff which have made the development possible. Our boldest claim is that, if the system works with what we have put together here, it should work anywhere.

The equipment installed at the circulation counter is in a somewhat different configuration from that originally planned. The original design contemplated a camera set in an overhead position, shooting down on the books and borrower's card, with the books placed spine upward in the usual case. The architectural design for the circulation area in the new Library building had, however, already been drawn up at that time, and it turned out that a configuration such as that proposed would be artistically and architecturally unacceptable. Accordingly, the photographic recording process was adapted in such a way that the books and the borrower's card would be photographed on the horizontal axis. The books, therefore, are set on the counter on their ends, in the position they normally occupy on a shelf, with a suitable movable block to hold them in this position, while the transaction is recorded.

In order to avoid having the camera on top, however, thus cluttering up the new circulation counter, it was placed just to the side and below the counter top with a reflecting surface set at a 45-degree angle to direct the image down into the lens (see Figure 5). At first a front surfaced mirror was used for this purpose, but a total-internal-reflection prism does the optical job at least as well and is a great deal easier to support in a proper position and to keep clean.

In operation the camera is completely enclosed in a housing constructed for the purpose, with a removable side to allow for the insertion and removal of film. The housing serves three purposes: it keeps the camera anchored firmly in position, since the housing is bolted securely to the supporting shelf and the camera is fitted to its interior; it provides physical protection to the camera; and it is lined with soundproofing material, to reduce the slight noise of the relays in the camera during operation. The camera itself is an automatic microfilm unit removed from its normal supports and slightly adapted for the purpose. It is actuated by a toggle switch which has been connected to the operating relays and located remotely, convenient to the location where books are placed for charging. The toggle switch is

spring loaded, and simply pushing it down takes the exposure and automatically advances the film ready for the next transaction. It is an exceedingly reliable piece of equipment; in addition to a highly developed engineering design, it has a comprehensive alarm system which gives a signal whenever the camera is not in condition to take the exposure properly—if the supply reel of film is nearly exhausted, if the film is not properly seated in its channels when reloaded, or in almost any other anomalous situation which might occur.

The film and pressure-sensitive labels now in use are still being explored. The materials in use do reasonably well, but work is still very much in progress to improve their efficiency and performance. The labels are secured with a background design preprinted on them; the due date is imprinted on them by a small automatic label imprinter secured for the purpose, and they are available at the circulation counter by an automatic dispenser which feeds out a new label each time a label is removed. The cost is less than a tenth of a cent per label at the present time and should be reduced with further development.

All computer operations have been carried out on an IBM 1401. Although there are two IBM 7094's in the University's main computing facility, the smaller machine has seemed preferable for this work. It is located right in the new Library building, so that access has been excellent; and it is more suited in basic design to the type of processing being done. This may be subject to drastic change soon; the University has filed a "Letter of Intent" for an IBM 360 Model 50, to replace the IBM 1401. The decision is not yet firm, but the prospect is an exciting one; if it is found that the circulation control operation can afford an input-output console at the circulation desk, and if a suitable time-sharing package can be included in the computer's operating system, a true on-line circulation system would seem to be a possibility.

The IBM 1401 in use has 8000 positions of core storage, four 729 magnetic tape drives (no disk storage), and a fairly complete array of special features on the central processing unit. The 8K storage limitation is a decided nuisance, but so far it has been possible if somewhat awkward to program around it. The daily circulation control update program, for example, has nine overlays.

Programming has been done entirely in the Autocoder language. In addition to the basic circulation control package, there has been the task of assembling data for and issuing library user cards; prior to this, library cards were not issued at the University at all. Identification numbers for circulation purposes are generated in the computer, and the cards are prepared on preprinted continuous forms on the computer's printer, for the several classes of users of the Library.

The amount of time required for making indicated modifications and changes in the programs has been rather surprising. The original circulation control updating program was carefully made, and yet, as experience has been gained, the need for adaptations and the opportunities to make improvements have been virtually continuous. By far the largest part of the author's time for the past several months has been devoted to writing programs, debugging, and operating the computer. On the basis of this experience, it seems that the improvement of means for man-machine communication remains a most important objective.

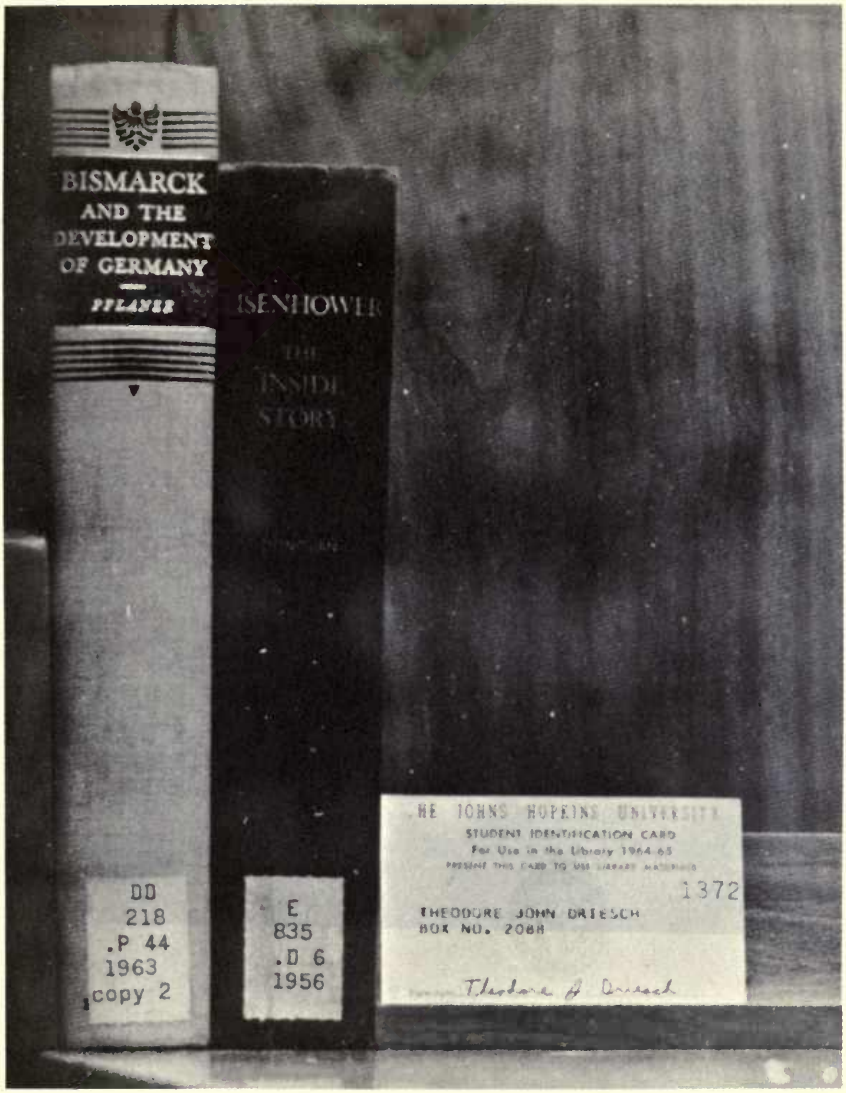

Figure 1
A Frame of the Microfilm Circulation Control Record

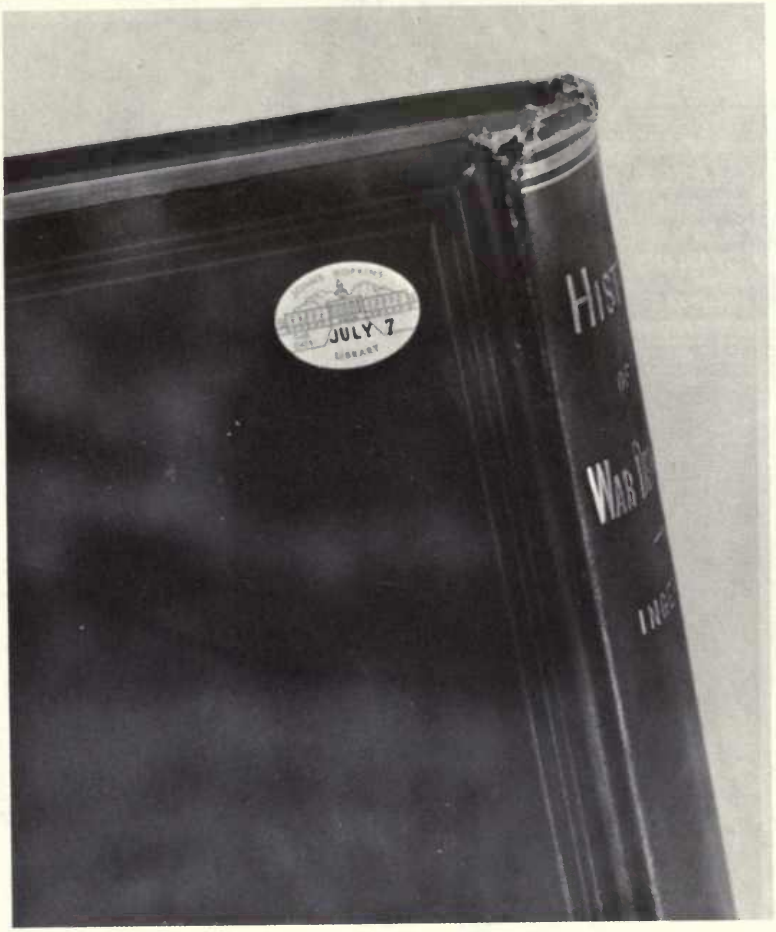

Figure 2
Pressure-Sensitive Label with Due Date

Figure 3
Microfilm Reader and Card Punch

Figure 4
Returned Items in Position for Discharge

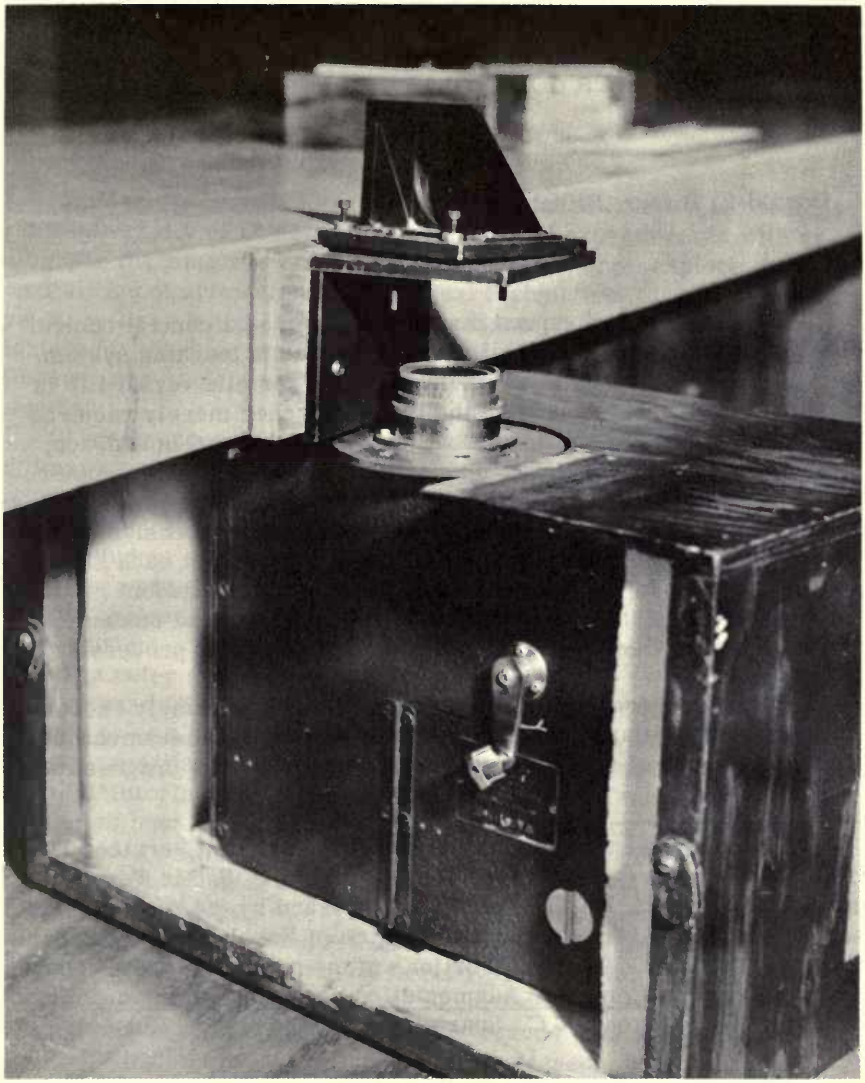

Figure 5
Close-up of Camera