Reports/
Surveys
1957-59
63/9

Sanitized Copy Approved for Release 2009/11/23 : CIA-RDP82T00271R000100120001-9

Copy 1 of 20 Copies

25X1

REPORT

ON

OFFICE OF CENTRAL REFERENCE

25X1

31 July 1959

This document contains information affecting the national defense of the United States, within the meaning of the Espionage Laws, Title 18, U. S. C., Sections 793 and 794, the transmission or revelation of which in any manner to an unauthorized person is prohibited by law.

25X1

I. INTRODUCTION

Approximately 6 months ago representatives [redacted] visited the Office of Central Reference (OCR) and spent some time in examination of its operation. The purposes of this visit were first, to develop an understanding of the operation of OCR with which the developmental activities [redacted] could be more suitably guided, and second, to provide OCR with the benefit of whatever ideas were stimulated by the visit. 25X1 25X1

The following report is a discussion of the thinking which resulted from that visit. In no sense should it be considered an adequate system analysis of the present system and procedures. The examination was entirely too brief and much of the pertinent data must have been missed or misevaluated. Furthermore, the report does not discuss specific equipment or techniques of mechanization as such. In order for any general recommendations concerning mechanization to have concrete realization they must eventually be implemented by specific pieces of equipment. In this respect, [redacted] as a direct result of discussions with OCR, is prepared to propose such equipment and to work with CIA in efficiently mechanizing OCR operations, if this is deemed desirable; however, it was felt that equipment proposals, as such, were outside the purpose of this report. 25X1 25X1

The report consists of essentially three parts. The first is a brief summary of the survey results themselves. The second discusses our reaction to the present status of OCR. The final section is concerned with recommendations concerning mechanization.

The significant conclusions can be summarized as follows: (1) In general, the present system of operation was considered to be reasonably efficient in terms of the existing conditions, i.e., the present work load and prescribed techniques for mechanization; (2) On the other hand, the immediate reaction at the time of the survey was that the existing conditions themselves are completely unsatisfactory and, in particular, that the present work load is shockingly small and the present techniques for mechanization unsuited to the job. Further thought has merely re-enforced the substance of

that initial reaction. In view of this, if OCR expects to continue with the present information retrieval work load, it is recommended that the present system of operation be replaced by a manual one which will be more responsive to these requirements. (4) However, it is very strongly felt that continuation of the present work load is completely undesirable and is inconsistent with the overall aims of CIA. If this feeling on our part is valid, then mechanization of OCR is essential. Since the present mechanization is completely inadequate for any substantial increase in work load, serious study should be made of the requirements for more sophisticated mechanization and the techniques available for their implementation. (5) It is felt that the work load placed on OCR and the problems in mechanization of OCR cannot be considered independently of the overall CIA operation. In particular, increased usage of OCR depends upon its effective relationship with the analysis and production elements in CIA and with associated intelligence agencies. Furthermore, common processing problems and overlapping information requirements must exist throughout CIA; it is essential that compatible data processing techniques also be made to exist. Consequently, the effective mechanization of OCR must and should be considered in relationship to the mechanization of the total CIA operation. (6) This requires, as a minimum first step, a comprehensive study of the entire CIA information processing problem, so that the relationship among various existing CIA information systems, between them and the analysis and production elements of CIA, and between them and the other intelligence agencies is clearly understood.

II. SUMMARY OF SURVEY

A. LIMITATIONS ON THE SURVEY

Our survey of the Office of Central Reference was a short three day examination. It is certain that there were areas which were missed and data which was misinterpreted by us. Furthermore, because of the requirements of security and the lack of contact with the users of OCR it is also certain that our understanding of OCR purpose and operation is at best not precise. Under these conditions--brevity of examination, missed information, ill-defined purpose--it must again be emphasized that we cannot in any way carry out a "system analysis" of the present operation; hence, we are not at all attempting to suggest modifications in procedure, data flow, or other system considerations for the purpose of minor or even major improvement in present operations. In fact, it is doubtful whether such an analysis is necessary since the present equipment seems to be used to its maximum efficiency anyway.

B. THE FILES

The Office of Central Reference is an information system based on one main file of documentary data and three associated registers: 1) The General File; 2) The Biographic Register; 3) The Industrial Register; and 4) The Photographic Register. The contents of the first three of these are generally intelligence reports derived from CIA and other intelligence sources; a typical document, for example, may be a report by an attache associated with some diplomatic unit. The large mass of the information stored is factual data with some degree of interpretation or evaluation.

The files themselves are each large by almost any standards of comparison. For example, the General File presently contains approximately one million documents. The Biographic Register contains dossiers of various size covering 250,000 individuals and involves an additional one and one-quarter million unfiled items. The Industrial Register covers 275,000 industrial installations with nearly two million supporting documents. The Photographic Register involves a large, but undetermined by the survey, number of individual photographs, prints and negatives, and rolls of motion picture film.

C. THE INPUT

1. General File

The input to the General File involves a standard processing pattern consisting of the following individual steps: First, the incoming documents, averaging some 1,500/day, are screened for their individual relevancy and some 40% are handled in a routine fashion with no further processing; Second, the remaining 60% are each read in some appreciable detail by a group of approximately 28 people for the purpose of specifying dissemination to interested organizations and analysts; Third, following the dissemination, the documents are again read in detail by a group of 18 "coders", this time for the purposes of abstracting, classifying, indexing, and coding the individual documents for later retrieval. Each coder normally handles about 40 documents per day. This operation generally results in approximately six individual codes per document.

Finally, the documents and the subject coding information are processed for insertion in the file. This processing itself involves a standard procedure as follows: the document is photographed on microfilm which is then prepared in the form of a punched card aperture card, identified by document number and some other relevant data. A bibliographic source description and a set of file cards is prepared, each containing the abstract in printed form and the relevant source, subject, or area code on which retrieval may be made.

The aperture cards containing the microfilm document image are stored in document number sequence and, therefore, correspond to an accessions sequenced storage system. The bibliographic source descriptions are stored in sequence by source identification. The individual file cards are sorted into subject code sequence or area code sequence and merged into their respective existing files.

The aperture card file and the bibliographic source description file each contain approximately one card per document and thus each involve one million cards. These files grow at the rate of approximately 1,000 new documents each day. The subject and area index files, on the other hand, together contain approximately 6 million cards (an average of 6 codes per document) and are growing at the rate of over

5,000 cards per day.

2. Biographic Register

The Biographic Register is somewhat closer to an integrated information retrieval-analysis operation than is the General File, since it involves the maintenance of a set of dossiers, the associated analytical work being integrated with the file operation. Thus information is processed by the Biographic Register in a sequence of steps: The first involves the routing of new unfiled information to a "geographic-area desk" where it is stored in strictly date sequence. The set of unfiled information is analyzed by trained experts for its relevancy to the existing 250,000 dossiers. Quick reference to the unfiled information is made available by a punched card tabulated listing and a set of 5x8 file cards. When, as a result of the analysis process, previously unfiled information is found to be relevant to existing dossiers, it is added to them. This results in the addition of approximately 280,000 pieces of new information to old cases. Furthermore, the number of dossiers itself increases by 25,000 to 30,000 new dossiers each year.

The dossiers are stored in sequence by assigned case numbers. Reference to them is provided by three separate indexes; the first are McBee key sort cards also stored in case number sequence; the second, an alphabetical index by name in Soundex sequence; and the third, an alphabetical name index grouped by nationality. There is also a set of IBM punched cards used for preparation of tabular listings. These consist of "who's who" cards arranged by case number within nationality and a set of category cards (averaging four per case number) grouped by professional categories.

25X1

3. Industrial Register

The Industrial Register is very similar in its purpose to the Biographic Register, except that it is concerned with industrial facilities. Its basic file of nearly two million documents is organized into a set of some 275,000 "plant folders" arranged by geo-

graphic area. This file of documents increases at the rate of 20,000 to 25,000 new documents each month, each of which may refer to several installations. Summary information concerning the plants is provided by an industrial card file--a set of 5x8 cards each summarizing the significant information concerning a single plant. Access to the plant folders and the industrial card file is provided by a set of three indices, each of which lists [redacted] separate items. The first list, called the "A" Listing, is an alphabetical listing of geographic location (thus bringing together all plants within a given area). The second listing, called the "B" Listing, is an alphabetical listing of plant name and/or description. The third, called the "C" Listing, is by plant activity. Up until July 1958 a set of punched cards, one for each plant, was maintained with a coding for single products of the plant. This operation has been discontinued. 25X1

Information is collated into the file by an analytical process similar to that involved in the Biographic Register.

4. Photographic Register

Very little factual information was obtained concerning the Photographic Register. From the short description available, it involves a manually maintained file of individual photographs and negatives. The positives are stored in geographic sequence with color-coded tabs providing a visual index [redacted] 25X1
25X1

D. THE OUTPUT

1. General File

Information can be obtained from the General File either directly from the aperture card document-image section or by means of any of the index files--the bibliographic source description cards, the subject coded cards, or the area coded cards. Thus, requests for specific documents identified by number can be serviced directly from the aperture card file; requests in terms of source are referred to the bibliographic source file; and requests in terms of subject or subject and area combinations are answered through use of the subject-area index files.

No data was available on the number of requests serviced directly from the aperture card file or the bibliographic source file. However, the impression was gained that this represented an appreciable volume of activity, particularly when compared with the activity serviced by the two index files.

Requests serviced by the subject and area index files average approximately 4,000 per year. The procedures in usage of these index files involve the following steps, some of which may not be present in handling particular requests:

a) The requestor may discuss his requirements with a so-called "composite group" consisting of an expert familiar with the operation of the machine system, one familiar with the subject area coding system, and a trained librarian familiar with the problems of research reference. This composite group, working together with the requestor, defines the search problem and establishes the criteria for mechanized decision.

b) Cards are manually selected from sections of the index file considered to be relevant to the search. (This operation frequently results in the complete satisfaction of the request as a result of the manual search and the judgment of the searcher. Sometimes the unavailability of the 100 sorter for the next step forces such a manual decision to be made.)

c) An IBM-101 Statistical Sorter is wired for comparison of the punched information in the manually selected cards with relatively complex combinations of additional subjects and areas. The result of this is a set of file cards satisfying the established criteria of the search.

d) A bibliographic strip is prepared from this set of cards by a photographic reproduction of the abstract printed on each card. This process has in the past been called the Intellofax Output.

e) This bibliographic strip is provided to the requestor for his further screening of the documents on the basis of the abstract data. Specific documents requested are then retrieved from the aperture card file and reproduced for the requestor's usage.

f) Meanwhile, the selected index cards are returned to the file area for eventual remerging in the index file. Several successive generations of intermediate files build up (apparently at times as many as six). This is essential for efficient handling of the file in the remerging process but is unfortunate from the standpoint of subsequent searches which must be made against each of these intermediate files to assure recovery of relevant data.

2. Biographic Register

The Biographic Register services approximately 5,000 requests for information each year. These apparently are primarily handled either by direct reference to the appropriate index and resulting case number or by discussion with the expert at the relevant area desk. There is also generated a moderate volume of summary reports and listings disseminated to using organizations.

3. Industrial Register

The Industrial Register handles approximately 12,000 requests for information each year. These requests can refer to industrial establishments, their cadre of personnel, parameters of production, etc.

4. Photographic Register

No information was available concerning the volume of usage of the Photographic Register.

E. SUMMARY

In summary, the total reference file of the Office of Central Reference involves some three to five million separate documents organized in three main files. Indexed access to individual documents is provided by close to ten million index cards and half a dozen periodically published printed indices, each consisting of between one and two million entries.

This rather large file (certainly one of the large information files in the country) handles less than 25,000 requests a year, i.e., less than 100 a day.

III. SUMMARY OF REACTION

A. PRESENT STATUS

Our first and immediate reaction to the operation of the Office of Central Reference was surprise combined with some degree of dismay at what we feel is the shockingly small usage of its files. This initial feeling has been re-enforced by our later thinking, particularly in the preparation of this report.

This statement certainly requires some degree of explanation. It was based on a number of considerations. Primary among these is the vital importance of the file, in terms of overall CIA and national objectives; to a certain extent this explains the feelings of dismay. The surprise resulted from comparison in our own mind of OCR with other systems of similar size and complexity (but presumably less vital importance) whose usage is greater by factors as large as 100. Among these might be included the files of the Title Insurance and Trust Company of Los Angeles, the files of the Medical Impairment Bureau of the Insurance Industry, the files of the Department of Motor Vehicles of the State of California, any one of several credit bureau files, the files of Time, Inc., the files of the Police Department of Los Angeles, and the files of other intelligence activities. (It must be admitted that the nature of some of these files, but not of all of them, may be somewhat different from those of OCR. As discussed below, however, this should represent a reason for careful examination of the purposes in usage of the file, to determine how more effective usage can be made.)

Finally, a rather cursory examination of the costs and expenses involved in input, output, filing, and (perhaps most important) collections indicated that the cost of a search in OCR is perhaps 10 to 100 times greater than searches in other systems of varying size and complexity. This extreme cost is a direct result of the low usage.

In summary, the immediate explanation for our reaction that the usage of OCR is shockingly small is the fact that it represents an enormous file which is vitally important and yet is almost completely unexploited.

B. WHY IS THE USAGE SO SMALL?

After the initial reaction of surprise the next reaction was one of attempting to explain this anomaly. There seem to be at least five relevant factors, each of them somewhat related to the others so that the explanation is not a completely clean-cut one. Considering each of these factors in turn:

1) First is the extremely effective dissemination provided by OCR. This is obviously an important function because of the need to provide analysts with relevant information at the earliest possible moment. The result of this effective dissemination is that the individual analyst presumably sees all information relevant to his problem. There is a resulting reduction therefore in the tendency to use the files.

2) Resulting from the fact that the dissemination presumably provides the analyst with all current information relevant to his interests is the natural tendency by the analyst to depend upon his memory and his personal files. Thus, he not only reads the daily "take" in terms of its relevancy to an immediate situation or research problem, but also remembers it and, in fact, probably files the document itself in some readily accessible personal file. Now, certainly, each of these effects is desirable. It is only by the picture which the analyst builds up in his own mind that he is able to interpret the significance of new data; it is this comprehensive picture stored in the analyst's memory which makes him the expert and capable of quickly providing an evaluated answer to requests for information. Furthermore, there is little question concerning the value to the analyst of having immediately available on his desk highly significant and current documentary information to re-enforce his memory and support his conclusions.

However, desirable though these two factors of the analyst's reliance upon his memory and his personal files may be, it is evident that these are usually carried to extremes. Attempting to remember every relevant fact, most of them usually remote from immediate interests, imposes a memory burden beyond the analyst's capacity. In establishing personal files the analyst is further burdening his memory with the necessity of remembering

not only the information from the documents he has read but also where he put those documents in case he requires them for supporting purposes. This additional burden on the memory, combined with the human tendency to forget, means that not only the information but also the retrieval of documents become more and more uncertain.

3) The third factor causing the low usage of the OCR files is the virtual restriction of present operation to answering research type questions. This fact is evidenced by the logical complexity of the questions and by the necessity for the composite group to act as an intermediary between the researcher and the file. Such usage by its very nature is not likely to be large, since research is time consuming and the file is used primarily to provide the initial starting point. In part, this restriction to a very limited type of usage is a historical situation and probably the result of OCR position in the CIA organization. In effect, it has served as the supporting library for the research function alone.

4) The first three causes discussed are to a large extent the result of the very nature of the work carried out by analysts. The fourth cause, however, is quite different. The OCR file has, unfortunately, developed a reputation among users for providing a slow and cumbersome response. There is thus a resulting lack of confidence and a feeling that the information will arrive too late to be useful anyway. The obvious effect of such a lack of confidence is the avoidance of using the file.

5) This unfortunate situation has been enormously complicated by the historical compartmentalization of intelligence work. The effects of security are obvious by its nature. However, in addition, the effects of both intra-agency and inter-agency politics have probably also been operative. To a certain extent this politics may be based on a valid consideration: namely, the necessity of adequately supporting the specialized needs of the individual component. Unfortunately, there is a resulting tendency to carry this to extremes and to perpetuate the compartmentalization in the usage of the files. Thus, OCR has not been effectively used as a central information service for the intelligence community.

C. MECHANIZATION UNDER PRESENT STATUS

In view of this over-riding fact of the small usage of OCR files, the mechanization of OCR operation becomes highly questionable. Indeed, it was our impression that, because of this marginal character of mechanization, previous investigators have recommended that further mechanization not be considered, and that the present mechanized system should be eliminated and replaced by some effective manual library procedures. The basis for this recommendation is the low usage combined with the unsuitability of the present system for the research type questions with which OCR has been predominately concerned--on the one hand, it is too involved a system for easy human access and utilization of human judgment and, on the other hand, it is not sufficiently involved for adequate machine logical processing.

Our conclusions, assuming that the present conditions of usage will continue, essentially agree with those implied by previous investigations. In other words, if the present low rate of usage is expected to continue in the future, mechanization should not be considered.

However, we must emphasize that we firmly believe the present status of usage should not and, it is hoped, will not continue. Under increased conditions of usage entirely different conclusions concerning mechanization are then valid. The remaining section of this report is concerned with precisely this question: namely, why should there be increased usage and what should be the corresponding effects upon mechanization?

IV. THE ROLE OF THE OCR FILE AND ITS MECHANIZATION

A. THE GENERAL ROLE OF AN INFORMATION SYSTEM

Information systems can be used for any one, or several, of a number of distinct purposes. Specifically, they can be used for simple accumulation or archival storage; they can be used for interrogation concerning specific facts; they can be used as a research tool with "judgment" scanning as a critical element; finally, they can be used as a basis for file-generated analysis and reporting of information. Each of these purposes is to a certain extent present in the operation of any information system, although generally one or two are of a predominant importance. In view of this, two questions are immediately relevant to the purposes of this report: the first relates to the role of the OCR files as the information system for national central intelligence and therefore how it should be used; the second relates to the role of mechanization with respect to implementing the OCR functions.

B. ROLE IN CENTRAL INTELLIGENCE

In the past there has been an almost traditional division of the national intelligence activity and its supporting functions. This division is a result of several factors. These include the need by line activities for direct intelligence support closely attuned to the individual line requirements, the highly significant requirements of security, and, apparently, the psychological makeup of intelligence workers themselves. In part, the present volume and type of usage of OCR files is a direct result of the historical fact of its position as the supporting library for a single division of a single agency.

Despite this historical state of affairs, however, there has become evident the need for centralized intelligence activity -- not as a replacement for the traditional line-responsive intelligence groups but as an additional intelligence capability in support of overall national strategic needs. Such a centralized intelligence activity provides a single point responsibility and a basis for common understanding and communication among the individual line groups. It is only in this way that a consolidated national intelligence picture can be derived from the frequently conflicting reports of the line groups.

Once the necessity for a centralized national intelligence responsibility has been recognized, the concommittant need for a centralized supporting information system also becomes immediately evident. It is only by the consolidation of all relevant information at a single point that its reconciliation and analysis in terms of overall requirements can be accomplished. Once such a centralized file exists, the economics of operation require that it act as a supporting information system for the entire intelligence community. In this way, although the individual line groups may require specialized information systems to meet their specialized requirements, they will draw on the centralized information system for their basic data; in turn, they will provide the input to the centralized system of the results of their own analysis.

C. THE NEED FOR RESPONSIVENESS

The great difficulty with a centralized information system, however, is its responsiveness -- in terms of both speed and nature of response -- to the requirements of the user. If the traditional reaction against centralization is to be overcome, the centralized information system must provide a quality of service at least equivalent to that previously obtained.

This responsiveness, as indicated, has two aspects which are closely related. In the following, three levels of response are discussed: the first (called "research response") involves complex, sometimes badly formulated, requirements but with a relatively relaxed time requirement; the second (called "indications response") requires an almost instantaneous reaction in terms of an intuitively understood total situation; and the third (called "information response") involves the answering of specific, well-defined, fairly simple requests on a reasonably rapid time scale.

In the past the major emphasis of the OCR library has been in terms of response of the research type. (As indicated above, this is probably a direct result of its organizational position as the supporting library for a single division of CIA). By its very nature, it is unlikely that such activity will ever be of large volume. Similarly, important though the indications response is in the overall intelligence picture, it again is unlikely that an information system of any presently available type will be involved to a great extent, because of the critical role played in indications response by human judgment. On the other hand, it is our feeling that information response requirements will

involve enormous volume, if the capability is present to provide the service necessary. Information response requirements have not previously been a significant factor in OCR operations; yet they represent a critically important intelligence capability and usually constitute the major load in other large size information systems.

To illustrate the importance of information response, consider even the individual research analyst. He is concerned with three types of situations each corresponding to one of the response patterns described above. Thus he is a research specialist in some area. As such it is his responsibility to develop a total picture of that area, to define and research unresolved questions, and to convey background information to others in a form adequate for their usage. On the other hand, being the expert in an area, he is in the position to evaluate the significance of specific events and, if they are critical, to quickly define the strategic situation. He thus has the responsibility of providing immediate indications response to critical situations at the earliest possible moment -- presumably on the basis of information disseminated to him. Finally, however, the individual analyst, being an expert in an area, is also in the best position to provide intelligent evaluation of the significance of information which, although not necessarily critical from his standpoint, is certainly involved with the inter-relationship between his area and others'.

It is this third situation wherein the need for information response is significant. To illustrate its importance, consider the following hypothetical situation. A specific analyst "A" is responsible for Iraq. During the past several weeks he has been principally concerned with two problems. The first is some basic research on the nature of the Communist activities in Iraq (the structure of the local party, its front activities, its agents in governmental positions, etc.). He has also been concerned with the possibilities of nationalist revolution against the existing rule as a particularly critical problem area. Documents disseminated to him have been very naturally scanned primarily for their relevancy to these two problems, although careful watch has been kept for other critical situations. Suddenly, the Israeli invasion of Egypt arises as a critical situation. The Iraqi expert should be able to provide quick and reliable information response on the significance of this invasion on relationships between Iraq and Egypt. Similarly, the experts on other areas, such as Syria, Algeria, Turkey, etc., should be able to provide the same quick and reliable information

response to the same critical situation.

This example should illustrate a number of significant points:

1) The response required is certainly not of the same nature as a research activity. The aim is primarily to provide information, although certainly screened by the evaluation of the expert analyst involved.

2) On the other hand, it is certainly not an indications response wherein the recognition of a critical situation is the significant element.

3) It is unlikely that most analysts would be in a position to provide the information required out of their own memories or even out of their own files. The information required in most cases will not have been uppermost in their minds and thus will have been at best vaguely remembered.

4) Furthermore, if the realities of personnel turnover are considered, all too frequently the analyst, if newly arrived, will not even have seen the relevant information.

5) If the capability is present for providing to each analyst the information required for him to evaluate and answer such precisely defined situations, this capability certainly will be utilized. As confidence is gained, the number of situations to which this type of response will be made will be increased and the overall effectiveness of our strategic knowledge will be improved.

Some measure of the effect upon present OCR operation of the requirements for information response in such a critical situation can be gained by considering the volume of activity following the outbreak of the Israeli-Egypt crisis. At that time the number of daily requests serviced by the OCR General File increased from twelve to fifty, and the service requested was provided only by going into multiple shift operation. It seems clear that the fourfold increase in number of requests was a direct result of an information response requirement. It further is apparent that such a usage of the library was essentially extraordinary rather than usual.

The importance of information response has been illustrated in terms of the individual research analyst. Yet clearly the need for information must exist throughout the operation of CIA and the intelligence community in general. The other

divisions of CIA, the field operations, the other intelligence agencies - each must require a maximum of specific information for effectiveness in its operation. But, from personal experience, it is evident to us that they frequently do not have easy access to information which must have been stored in OCR files. If rapid information response were available from OCR and were utilized by those requiring it, our total intelligence position would immeasurably improve.

A number of conclusions can be made:

1) OCR files are not being used for information response situations except under extraordinary circumstances.

2) The potential of information response as a tool in CIA analytical work is un-utilized.

3) The operation of OCR, its mechanization, and its relationship to the analyst should be modified to provide information response service. This capability, not only of the OCR file, but also of the analyst, should be exploited to the utmost.

4) If this situation is developed, the usage of OCR in terms of number of daily requests from CIA analysts very probably will increase from the present 100 per day to about one thousand per day.

5) As confidence increases in the capability of OCR to provide rapid information response, its usage not only by the research analysts within CIA but by the entire intelligence community will correspondingly increase. The resulting usage may add another one thousand to two thousand requests per day.

6) In this way, the OCR files can become a core library system for the entire intelligence community with only specialized peripheral libraries being necessary outside OCR.

D. THE ROLE OF MECHANIZATION

An increased activity of the type described (nam ly, from the present 100 per day to 2,000 to 3,000 per day) can certainly be handled by manual systems. In fact, many of the systems previously mentioned are entirely manual and are handling volumes at least this great. However, the economics of mechanization becomes extremely favorable as larger volumes are involved. This fact has led each of the large information systems presently involving manual operation to consider techniques for mechanization. In doing so, the existence of an adequate manual system has usually represented a severe stumbling block because of the difficulties in conversion.

OCR, on the other hand, is in the position of not having a system adequate for this type of service. It, therefore, can very easily consider mechanization since, in any event, a conversion of operation will be necessary if the recommended service is to be provided. Furthermore, the conversion itself is very much facilitated by the present semi-mechanized state of file index records.

The questions arise, however, as to what role mechanization can fill in the operation of OCR. First, as has been repeatedly emphasized, large volume is an essential element for a mechanized system. It presumably has the capacity to handle the volume at a lower total operating cost than that of an equivalent manual system. On the other hand, it needs the volume to justify the expense (mechanized installations generally have a high unit cost and a high installation and conversion cost in comparison with manual systems).

Second, however, mechanized systems must operate in a well-defined environment. Economic operation depends upon organization for efficiency and maximum usage. In addition, machinery by its very nature must be programmed with all operations finely detailed. In particular, all judgment and decision functions must be reduced to programmed operations. This means that undefined judgment and, in general, human contact with the system can become remote and difficult.

Third, the availability of mechanized information provides the possibility for exploitation of this data through file-generated activity. This means that many requests can be pre-answered and that the file can be continually re-organized for efficiency in response. Furthermore, such file-generated activity is usually an essential factor in

economic operation. If effectively used, it anticipates costly non-routine activity and replaces it with routine file activity. In this way, file-generated activity can even provide service prior to need.

It is clear that the present conditions of operation completely fail to meet these three criteria for effective mechanization. In fact, OCR is concerned with a small volume of requests each requiring a large element of human judgment, and there is virtually no file-generated activity.

On the other hand, if the recommended service of information response is implemented, there will be a many-fold increase in volume. Furthermore, requests for information are usually well-defined and thus fit easily into the patterns of machine operation. Finally, the importance of file-generated activity will immediately increase. In particular, file-generated activity, as mentioned above, is a necessity if quick response is to be provided to the large volume of information requests. This implies a need for the file to be responsive to new environment, both of requests and of file input. It also implies an awareness by the file operation of its own status and activity, through routine statistical and simple analytical work suitable to mechanized operation.

To implement the preceding, we can define a set of gross specifications on the requirements for mechanization of OCR:

1) Capacity for large volume documentary and information storage combined with capacity for handling large volume usage is essential.

2) Provision must be made for capability of close human contact. (Few mechanized systems provide this capability; yet the need for maintaining at least the present level of service to research type requests is an essential element of continued OCR operation. Close human contact is required to develop confidence in the file, understand its contents, and easily apply human judgment where it may be essential.)

3) To implement the need for information response, the mechanization must provide rapid, inexpensive access to well-defined information.

4) Processing capability must provide awareness of the changing environment of input and requests and must support the research function with mechanized access to patterns.

5) Rapid file handling capability for reporting, file organization and reorganization, and other file-generated activity must be inexpensive and easy.

V. RELATIONSHIP BETWEEN OCR MECHANIZATION AND THE OVERALL MECHANIZATION OF CIA

Essentially CIA itself represents an enormous information processing system -- collection, analysis, storage and retrieval, translation, processing, statistical analysis -- all being involved. We do not know the status of mechanization of overall CIA operation. It is clear, however, that such a complex information processing system must eventually be aided in large part by mechanization. Considerationn of economics, speed of response, and necessary processing capability all combine to force it upon CIA. It is essential however, that mechanization not be developed in a piece-meal fashion. The very close tie-in between the analytical process, the input and translation process, and the information storage and retrieval process is merely one evidence of this need. There is thus a necessity for an integrated picture and a definition of common elements throughout the operation.

In this report we have been able to discuss only those considerations relevant to OCR operation. The conclusions must therefore be limited in their awareness of overall aims and needs. We recommend that a completely integrated study of total CIA requirements be implemented.