5

SECRET

CONFIDENTIAL

PHASE II FINAL REPORT

PROJECT CHIVE

Volume V

SYSTEM ORGANIZATION, FUNCTIONS, AND PROCEDURES

CHIVE/R-3-65

1 March 1965

DIRECTORATE OF SCIENCE AND TECHNOLOGY

OFFICE OF COMPUTER SERVICES

CONFIDENTIAL

SECRET

Approved For Release 2000/05/30 : CIA-RDP78-03952A000100050001-7

WARNING

This material contains information affecting the National Defense of the United States within the meaning of the espionage laws, Title 18, USC, Secs. 793 and 794, the transmission or revelation of which in any manner to an unauthorized person is prohibited by law.

CONFIDENTIAL

Project CHIVE:

Phase II Final Report

Volume V

SYSTEM ORGANIZATION, FUNCTIONS AND PROCEDURES

CHIVE/R-3-65

1 March 1965

010365

U.S. Central Intelligence Agency, Office of Computer Services

CONFIDENTIAL

TABLE OF CONTENTS

Page

Page

Page

FIGURES

25X6
25X6

TABLES

25X1A

Chapter 5.1.

INTRODUCTION

5.1.1. GENERAL

This volume of the report is primarily concerned with the non-EDP aspects of the CHIVE system, that is, the organization of personnel required to operate the system and types of personnel needed, the nature and extent of the data base to be exploited, the indexing philosophy and technique, the files which will be identified to the user, system flows and data handling procedures, and the man-machine interactions projected for the computer-centered system. Of course, not all design problems have been resolved. Moreover, even if they had, it would not be possible to describe within the confines of one volume all of the transactions which must be performed in a system as large and complex as this. However, illustrations of representative tasks are included and some concepts of system data flows are presented to demonstrate the impact of hardware and programs upon personnel actions.

The recommendations interspersed in this volume result from Phase II of the design study, including a preliminary evaluation of the CHIVE Indexing Experiment conducted between November 1964 and January 1965, and are supported by material in some of the appendices to this volume as well as in earlier CHIVE documentation. The other appendices present further details on the indexing language and technique and the files to be inherited from the existing central reference repositories. All are recommended reading for recipients of this report who desire more detail on specific aspects of the system, as well as further background on the alternative configurations considered and the steps taken to arrive at the recommended system. A supplementary appendix to this volume will be issued later describing the CHIVE Indexing Experiment in greater detail, and reporting the final conclusions derived therefrom.

5.1.2. SYSTEM OVERVIEW

A simplified graphic view of the CHIVE system can be obtained by referring to Figure 5-1. In this diagram the flow paths within the system are separated for descriptive purposes into three major functional categories--document input processing (flow path 2), document retrieval processing (flow path 1), and information file building and maintenance (flow path 3). The following

25X1B

Approved For Release 2000/05/30 : CIA-RDP78-03952A000100050001-7

paragraphs will summarize briefly the major elements of the system, leaving the more detailed explanations to subsequent chapters in the volume.

In general, the philosophy of the CHIVE system is to combine the required intellectual talents of trained intelligence information analysts with the processing and storage capabilities of the computer. The source documents to be input to the system, the necessary human functions to be performed relative to these documents (i.e., reading, selecting, indexing, querying and reporting) and the outputs to be derived from the system are quite similar to those which characterize one or more elements of the existing central reference operation. Only if the proposed system is compared to an individual register subsystem within the current OCR complex does the contrast appear, and then only with respect to certain features of the existing subsystem.

In terms of file organization, the system follows the approach used in SR/OCR and DD/OCR in maintaining a separation between an index and the document holdings to which it refers. This necessarily has implications in terms of input time which may compare unfavorably with some of the current systems which are oriented toward multiple-filed documents (e.g., BR/OCR), but it also offers certain advantages in such areas as procedural standardization, index integration, number and variety of access points to the files, space requirements, etc.

The information is received primarily in the form of documents; however, index records to maps, photographs, and films will also be included in the system, as will certain machine-language data prepared on contract (but under CHIVE control) by external organizations (e.g., the Library of Congress).

Following preparation of the index record (a function normally performed by humans except where only a limited retrieval capability seems required), the index will be converted to machine storage with the aid of an optical character reader and placed in a random access device, ultimately the IBM/System 360 Data Cell Drive. The information storage capacity of one Data Cell Drive will allow us to accommodate the content of an estimated 600,000 index records (the actual storage capacity is 400 million characters of information), and there is no practical limit on the number of modules that could be provided. The same device would be used to hold what might be called the directory to the index records themselves, i.e., a list of the terms which appear in the index records and, for each term, the record and phrase number(s) containing said term. This would obviate the need to examine every index record in the file to see if it contains the term (or terms) sought. Index entries can be retrieved from the index store at the rate of about two per second depending on the number of terms involved in the search formula.

CHIVE's recommendation is that most textual documents should be converted to microfilm and stored either in the form of 35 mm. aperture cards (containing up to 8 images per aperture) or packed microfiche (sheet microfilm records containing up to 60 letter-size pages on each microfiche). Documents in excess of a certain page limit and those of poor image quality should be kept in hard copy. Maps, films, and photos will continue to be stored in the conventional manner in the physical repositories in which they are now located.

Whether the 35 mm. aperture card or microfiche storage system is chosen, the document images should be filed in motorized card files, but should be retrieved and refiled manually. Assuming 10 million documents were to be stored on site, the estimated floor space required for a packed microfiche system would be an area approximately 30' x 60'; for the 35 mm. system, 40' x 70'. Output from either the hard copy document or microimage files would consist of paper copies. The integrity of the document collection will be maintained such that none of the master microimages, or original documents if filed only in hard copy, will leave the file except for photoduplication or hard copy printing.

Chapter 5.2.

SYSTEM ORGANIZATION

5.2.1. BACKGROUND

The organizational configuration recommended by CHIVE is the product of much thought and discussion extending back into the Phase I study and reflects a variety of views expressed by persons both internal to the CHIVE design team as well as to OCR. One of the most vexing and, at the same time, one of the most important of the CHIVE design problems, it is not anticipated that the organizational plan which has evolved will be attractive to all. Nevertheless, it appears to offer the best hope of achieving the desired system objectives, consistent with the human factor requirements imposed by the environment within which the system must operate.

The search for a revision of the existing central reference organizational structure was largely influenced by the findings of the DD/I survey and the set of system requirements derived therefrom. These findings and inferred goals have been described in the CHIVE Phase I Report, in CHIVE/R-1-63, and (in more abbreviated fashion) in Volume IV, Chapter 2, of this report. The organization study itself may be said to have consisted of three phases:

a. An analysis of the personnel or management requirements imposed on the system by the overall system objectives.

b. A study of various alternative organizational configurations which might be adopted, ranging from a completely decentralized activity to various kinds of centralized operations, including alternative configurations at different hierarchic levels.

c. An evaluation of one organizational concept by the process of subjecting the concept to a practical experiment which simulated to some extent the problems to be encountered in a live environment.

Phases a. and b. are described in some depth in Appendix 5.A. to this volume and are briefly reviewed below. Phase c., which resulted in some revision of the organizational concept, is discussed in Appendix 5.B. and only its conclusions are reflected here.

In considering the managerial problem of how best to organize the input and retrieval functions to be performed, as well as the personnel to carry out these functions, a number of organizational requirements were set forth.

These requirements, or objectives, may be summarized for the purposes of this review as follows:

a. Specialization with minimum processing duplication

b. Minimum customer contact points

c. All-source service from any point

d. Close communication between input and query handlers

e. Close communication between system operators and users

f. Document control as the first priority

g. Operator job satisfaction

h. Personnel flexibility

The next step was to pass various organizational configurations against these objectives to determine which would appear to offer the best hope of accommodating the defined goals. Because of the size of the contemplated activity in terms of the number of personnel needed to operate the system, this required that alternative configurations be considered not only at the first organizational level, but at least at one additional level below that.

For the first cut the following four different organizational concepts were considered:

a. Retention of the existing OCR configuration

b. Development of a single, all-source document retrieval system, with a separate biographic information facility

c. Dispersal of some or all of the information storage and retrieval activity among the research and production components

d. Continuation of the central system, but on an all-source, geographically-organized basis

Where the additional subdivision of personnel would be required because of the size of a particular component, these additional means of grouping the analysts assigned thereto were studied:

a. Organization by document source (Collateral, Comint, etc.)

b. Organization by function (input, retrieval, information file maintenance, etc.)

c. Organization by class of data to be stored and retrieved (biographic, installation, subject/commodity, etc.)

d. Organization by topic (political, scientific, economic, military, etc.)

The study very quickly made clear that none of the alternatives considered resolved all problems that could be anticipated. However, the combination of the geographic approach at the first level, and topical specialization (where required) at the second, seemed to come closest to meeting the organizational objectives outlined above. It remained to be seen, however, whether an information analyst could perform all-topic indexing of all-source documents satisfactorily, and what effect it might have on his morale and attitude if he had to operate in this kind of environment.

The CHIVE Indexing Experiment afforded the opportunity to test the configuration proposed and, as detailed in Appendix 5.B., identified a number of problem areas which suggested that some additional organizational and procedural alternatives might well be considered. Of principal interest from the organizational point of view was the recommendation that the geographic concept be retained but that the coding process *per se* be separated from the function of selecting documents and identifying the subjects or objects to be indexed. Acceptance of this approach meant some compromise of the single-point indexing concept but offered the advantage of increased job satisfaction on the part of the more highly qualified analyst, helped reduce the selection problem, and suggested the possibility of acquiring more personnel for less money to perform the more routine input functions. Since it still permitted achieving all the other organizational objectives, it was selected as the alternative best satisfying the system requirements and is the approach recommended here.

5.2.2. PROPOSED ORGANIZATIONAL CONCEPT

The responsibility for implementing a specific organizational configuration must be left to those who will direct the operation since there are a variety of factors to be considered which are beyond the purview of the system designer. To assist those, however, who will be charged with this activity, it might be useful to summarize the principal CHIVE organizational recommendations in the context of the major functions to be performed within the system, and to give some feel for the interrelationships between these functions since these could have implications for management in terms of communication interface, assignment of physical space, and so forth. This first look will be an abbreviated one since much of the same ground is covered (if from a slightly different point of view) in more detail in other sections of this volume. A set of position descriptions outlining the duties and responsibilities of the various types of personnel within the system concludes the chapter.

5.2.2.1. Input Control and Customer Service

The CHIVE system would be built largely around Information Analysts organized (at the first level) into some four or five geographic components. It is our view that it is difficult to identify any better way of organizing the input and retrieval activity than by grouping the primary individuals involved by geographic area. As stated in earlier documentation, this approach loses the advantage of source specialization in processing and poses the problem of geographic overlap in document analysis and query coordination. At the same time, it contributes to standardization of vocabularies and procedures important in an all-source environment, and is in focus with customer inquiries which normally relate to a particular geographic region of the world. Thus, on balance, while it does not overcome all operational problems that can be envisaged, of all the alternatives considered it seems to come nearest to meeting the system objectives.

Without specific restrictive criteria (which, thus far, seem impossible to obtain) with respect to the content of the documents to be processed, the experienced Information Analyst, operating in close communication with his customers, appears to offer the best hope of resolving the data selection problem. The Information Analyst would, therefore, be responsible for determining not only what documents entered the system files but what data within these documents was captured for retrieval purposes.

The Information Analyst operating out of a geographic component would also be solely responsible for the selection and processing of data input to information files required by customers, and would handle all queries levied on the system. By virtue of the fact that he was personally involved in the input process, he would not only be familiar with the current reporting but would know what material had been stored for retrospective searching and how to get at it.

Whether the Information Analyst should also specialize by topic within area or by some class of intelligence data (e.g., biographic, installation, etc.) remains a moot point. CHIVE continues to favor the former in the belief that it would lessen the number of times a document would have to be handled, but additional testing of both concepts is desirable.

5.2.2.2. Index Preparation

The function of physically preparing the index records to documents, including both the header (bibliographic) as well as the content data descriptions, would be assigned to special personnel, known as Header Indexers and Content

Indexers, operating in close communication with the analytical components.

Content Indexers serving one geographic desk, e.g., the Far East, should probably be located together as a unit attached to said component. The Content Indexers, like the Information Analysts, would be subdivided by geographic area and each would normally process the output of his counterpart analyst or analysts.

Content Indexers would each have a set of the dictionaries and other vocabulary control tools pertinent to his area of responsibility. In addition, a master set of other area dictionaries would be located within each content indexing group for reference purposes.

Content Indexers would translate the items of data tagged by Information Analysts into the codes and other descriptors dictated by the vocabulary of the system. To increase their sense of participation in the more intellectual aspects of the input process (and, thereby, reduce turnover), they might be given full responsibility for general subject indexing as distinct from named-object control.

Header Indexers would perform a function similar to content indexing, but on the bibliographic elements of a document. One group of Header Indexers would operate in a centralized mode, serving all geographic components by header indexing, immediately upon receipt, those documents for which CHIVE has a repository responsibility. Other Header Indexers would be assigned to each geographic organization to capture the necessary bibliographic data pertaining to non-repository-type documents which had been reviewed by Information Analysts and selected for retention by the system.

5.2.2.3. Dissemination

The dissemination function, apart from any necessary re-routing of documents within a CHIVE geographic component, is external to the system *per se*. However, it might be advantageous to co-locate dissemination personnel with the centralized header indexing group to shorten the time between document receipt and file availability for repository-type documents.

5.2.2.4. Data Transcription

This function refers to the rather formalized typing operation required if, as planned, optical recognition equipment is to be used to convert index and other records into machine-recognizable form. Header Indexers can type their inputs in a form suitable for processing by a page reader. However, a central pool of typists will also be needed, operating at the system level, to convert the majority of transcript sheets received from Content Indexers as well as search requests from Information Analysts into the graphic quality required. This central pool can be supplemented by typists assigned to the various area desks who, in addition to typing finished reports, memoranda, etc., would also transcribe many of the file maintenance and query transactions for input to the page reader.

5.2.2.5. Image Processing and Document File Maintenance

Image processing is that activity conducted by the so-called "Document Delivery System," i.e., the microfilming and associated operations required to convert incoming documents to microimage form, as well as the reproduction of items retrieved from the document store for delivery to customers. This is a relatively discrete function although, if an aperture card storage system is employed, it requires some support from the machine side of the house. Otherwise, its principal interface is with the document store itself to which materials are passed after microfilming and from which it receives, in turn, items to be reproduced.

During the evolutionary development of the CHIVE system both the new and old system operators will require access to many of the same document collections. If the logistical problems are not too severe, it would seem advisable to co-locate all master document files in one general physical area to lessen the communication problem as well as render file maintenance operations more efficient. This might increase the distance which now obtains between an existing central document collection and a set of users, but over time the majority of users would probably benefit from the establishment of one "Document Center." Similarly, because of the close relationship between the document files themselves and the image processing function, it is recommended that the latter be connected both physically and organizationally to the former.*

5.2.2.6. Machine Functions

The principal machine-related activities and hardware include:

a. EAM personnel and equipment needed to input data to files not yet absorbed into the new system and to retrieve data therefrom. Assuming no conversion to an EDP storage medium, the latter, in particular, will necessitate the retention of an EAM facility for as long as the inherited files have value.

b. EDP hardware needed to operate the new system, including associated I/O devices (e.g., the page reader), and computer operator personnel.

---

* Problems involved in co-locating files are discussed in Volume III.

c. System analysts/programmers (referred to in this report as EDP File Analysts) who will develop and refine the machine operations to be performed, define new files to the system, etc.

Logically, all of these personnel and operations should be centralized in one organizational component whether located within the central reference complex or external to it.

5.2.3. POSITION DESCRIPTIONS

The personnel involved in making up the CHIVE operator complex will include the following: Information Analyst, Content Indexer, Header Indexer, Dictionary Editor, Data Transcriber, Information Control Clerk, Document File Clerk, Reproduction Equipment Operator, EAM Operator, Computer Operator, and EDP File Analyst.

5.2.3.1. Information Analyst

The Information Analyst will be the principal intermediary between the customer and the system. He will be responsible for selecting what goes into the files and will screen all output before it is delivered to a requester. Senior Information Analysts will serve in various supervisory capacities from the sub-Section to the Branch or Division level, directing, coordinating, and reviewing the work performed by their subordinates.

All Information Analysts will hold professional positions, and will specialize in a particular geographic area and (where required by reason of work volume) by topic within area.

Every Information Analyst will be trained in applying the indexing vocabulary to documents by actual involvement in the coding process. He will also be thoroughly familiar with all the CHIVE-built files available within the system as well as the query language used to interrogate or modify said files. In addition, he will know what inherited files were acquired from the existing system and their general content, although not necessarily the vocabulary used in these files.

The duties of an Information Analyst will include:

a. Receiving and reviewing the content of documents, cables, graphics and other incoming data for information worthy of retention by the central reference system.

b. Selectively marking the elements of information to be extracted from the documents for representation in the system's index files and distributing the marked documents to Content and/or Header Indexers.

c. Exploiting the content of document index records for the purpose of building formatted information files pertaining to a specific subject or class of subjects.

d. Preparing file maintenance transcript sheets as the means of adding data to, or changing data within, said information files.

SYSTEM ORGANIZATION
Position Descriptions
5.2.3.1.

e. Receiving requests from customers and preparing the necessary search prescription after consulting the relevant vocabulary control files and other Information Analysts most familiar with the vocabularies of certain inherited files.

f. Requesting copies of documents as well as dossiers and other master records from the central document repository.

g. Reviewing, analyzing, and synthesizing data recovered as a result of the search process and preparing responses in raw or finished form for delivery to the customer.

h. Advising customers about files or persons external to CHIVE that might be worthwhile consulting, and personally contacting same if required.

i. Recording necessary management data relative to requests received, responses furnished, and other system processes.

5.2.3.2. Content Indexer

The Content Indexer will be a semi-professional possessing at least a high school education. His duties will include:

a. Extracting the elements of information in a document identified for him by the Information Analyst.

b. Consulting the relevant dictionaries and other vocabulary control files for the purpose of selecting the appropriate controlled terms to express these items of data.

c. Arranging the data into a form for machine entry using pro-forma content data transcript sheets.

d. Consulting with the appropriate Information Analysts and Dictionary Editors with regard to the application of the index language and possible revisions to the system vocabularies.

e. Initiation of additions or changes to the vocabulary control files through preparation of file maintenance transcript sheets.

f. Reviewing printout of changes and additions to the files including incorrect entries.

5.2.3.3. Header Indexer

The Header Indexer will occupy a clerical position and must be a qualified typist. The duties of the Header Indexer will include:

a. Extracting the standard header (bibliographic) data appropriate to the category of document involved, and expressing this data (where required) in the codes used by the system.

b. Typing the data in the prescribed manner for machine entry using the correct header data transcript sheet.

c. Consulting with the Dictionary Editor for Header Data with regard to the use of the header data codes and format, and recommending changes when required.

5.2.3.4. Dictionary Editor

The Dictionary Editor will be an Information Analyst with primary responsibility for control of one of the system's vocabulary files. Some Dictionary Editors will have system-wide control over the application of terms in their respective subject areas. Others (e.g., an Organization Dictionary Editor) may govern the use of terms only within a given country or other geographic area. The

SYSTEM ORGANIZATION
Position Descriptions
5.2.3.4.

duties of a Dictionary Editor will include:

a. General review of the content and format of sample transcript sheets emanating from Indexers assigned to the area unit of which he is a part.

b. Providing advice and counsel to Indexers on the use of the specific dictionary for which he is responsible.

c. Reviewing all new entries to the dictionary for the purpose of determining whether each was a legitimate entry and whether format and content met established procedures.

d. Personally initiating changes to a dictionary where required.

e. Reviewing printouts of changes and additions and insuring that all revisions to the dictionary are published and disseminated.

f. Consulting with other Information Analysts and customers regarding current requirements and possible improvements to the system's vocabulary control files.

g. Advising Information Analysts preparing request statements on the terms to be used in the query prescription.

5.2.3.5. Data Transcriber

The Data Transcriber includes any person exclusively assigned to operate a key-driven device from copy provided via another system operation. The duties of a Data Transcriber will be as follows:

a. Receive format instructions from Information Analyst, Content Indexer, or other individual for typing, tape perforation, or card punching.

b. Prepare typed copy, punched paper tape, or cards for optical character recognition or other form of computer entry.

c. Check transcribed copy for accuracy and correct if necessary.

d. Operate typewriter, Flexowriter-like device, 026 Key Punch, and 056 Verifier.

5.2.3.6. Information Control Clerk

Information Control Clerks will be assigned to most operational components of the system. Their general duties will include:

a. Receiving material such as hard copy documents, machine listings, document request forms, paper and magnetic tapes, card decks, etc.

b. Accounting for material received and maintaining necessary special-purpose logs of requests and other actions.

c. Intra-office routing and delivery of materials to staff personnel and mailing of system products to customers.

d. Assisting Information Analysts in the routine maintenance of manual files including the insertion of handwritten entries to machine listings and other hard copy records.

5.2.3.7. Document File Clerk

The duties of the Document File Clerk will include:

a. Filing newly-processed documents or refiling old materials into one or more of the following types of document files: personality or installation dossiers, card files, open-shelf document files, and 16 mm. or 35 mm. aperture card collections.

b. Receiving requests for documents or other records to be retrieved and recovering same from the appropriate files on either a routine or priority basis.

c. Maintaining dossier and other special-purpose logs pertaining to transactions affecting the document files.

d. Recording action taken on document request forms, forwarding requests for unrecovered documents to other file repositories for searching, and transmittal of master records to image processing for photographic reproduction.

5.2.3.8. <u>Reproduction Equipment Operator</u>

The duties of the Reproduction Equipment Operator will include:

a. Receiving incoming documents and determining which are photographable and which must be stored in hard copy.

b. Operating the appropriate microfilming equipment required to reduce the documents to a micro-storage medium and reviewing the quality of the photographic record.

c. Receiving documents retrieved from the master files and reproducing same on a variety of image-processing equipment.

d. Servicing and supplying reproducing equipment.

e. Supplying copies of documents to Information Control Clerks for delivery to internal or external requesters.

5.2.3.9. <u>EAM Operator</u>

EAM Operators will be required to process certain [illegible]ch card files inherited from the existing system as well as select new files. In general, the duties of an EAM Operator will include:

a. Operating electrical accounting machines including interpreter, reproducer, tabulator, sorter, and printer units.

b. Performing routine machine operations in accordance with conditions outlined by EDP File Analysts, Information Analysts, and Indexers.

c. Wiring panels in accordance with directions.

5.2.3.10. <u>Computer Operator</u>

In general, the duties of the Computer Operator will include:

a. Maintaining a schedule and operating log of the components of the computer complex.

b. Loading and unloading Tape Units.

c. Loading and operating stored programs.

d. Tracing and correcting program errors.

e. Correcting failures in card, paper tape, or optical character reading equipment.

f. Wiring and/or selecting control panels for use in card reading machines.

5.2.3.11. <u>EDP File Analyst</u>

The duties of the EDP File Analyst will include:

a. Determining from Information Analysts requirements for new system files and developing the record structures, file formats, and output products needed to establish and maintain such files.

b. Preparing general and special-purpose programs either for the purpose of converting extant machine-language files or to provide new data processing capabilities.

c. Testing newly designed programs utilizing the computer and necessary input/output units.

d. Conducting studies of system data flow for the development and refinement of programs.

e. Determining utilization requirements for input/output devices including displays, and designing programs to permit exploitation of input/output capabilities.

f. Designing quantitive techniques and statistical devices for special program applications.

g. Preparing procedures descriptions including coding formats and flow charts for operator task guidance.

Chapter 5.3.

DATA BASE

5.3.1. THE SELECTION PROBLEM

The selection problem has been with OCR since its inception. No coordinated study of selection as an entire OCR problem has ever been made. Individual registers have established selection criteria, some more formalized than others. An attempt to summarize these criteria for compatibility, or to establish common criteria to be used by all registers was not deemed necessary heretofore. Since each register has been more or less independent, _ipso facto_, its criteria have for the most part been unrelated to those of any other register. This condition has led to non-uniform levels of coverage and, in some cases, duplicative processing of the same subject matter. Regardless of CHIVE, if OCR adopts a geographical organization posture, uniform criteria for document series and depth of subject indexing become mandatory within geographic component.

5.3.2. BASIC SELECTION CRITERIA

Selection criteria will depend on several factors:

(a) the documents used and information needed by the analytic offices; (b) the all-source concept and organizational configuration thereof. These two factors have to be balanced against the manpower and resultant capability available for the operation.

There seems to be a consensus of opinion that several levels of indexing should be applied to the various categories of documents:

- Entire series to be indexed in depth.
- Entire series to be rejected for depth indexing, but to receive header or bibliographic control.
- Entire series to be rejected completely.
- Specific documents within a series to be indexed in depth.

Selection of an indexing level for a particular document category is contingent upon customer reaction and acceptance, which determination requires discussion of interest in series not covered now and re-examination of series presently covered. Customer participation in determining selection criteria can mean the success or failure of the system in terms of usage. Once the level of indexing is agreed upon, document priorities will need to be established for implementing the CHIVE system since all categories cannot be implemented within the initial system simultaneously.

DATA BASE
Basic Selection Criteria
5.3.2.

5.3.3. SOURCES TO BE EXPLOITED

The following major document series are planned for CHIVE control:

Raw Intelligence Reports (Collateral)

State Airgrams

Military Attache Reports

CIA Reports--OO,CS, etc.

Military Command Reports

Selected Other Governmental--AID, USIA, etc.

25X6 International Organizations--NATO, etc.

[redacted]

Cables (Collateral)

CIA-TDCS

Non-CIA

Finished Intelligence

25X6 U.S. [redacted]

25X1A [redacted]

Open Publications and Translations

FDD

JPRS

DATA BASE
Sources to be Exploited
5.3.3.

COMINT

Messages

Reports

Photo Interpretation Reports (T/KH)

Maps, Films, and Ground Photos

Miscellaneous

25X1A

Select Contractual-[redacted], etc.

State Biographic Cards

Unclassified Selected Periodicals, e.g., for China: Peking Review, Survey of China Mainland Press, etc.

Criteria for the depth of coverage will be developed by the CHIVE information analyst working in concert with the research offices. He will direct the indexer as to coverage and depth, i.e., which personalities, which organizations, and/or which subjects should be indexed. The CHIVE Indexing Experiment has shown the need for title coverage of most documents regardless of the level of indexing unless the document or series is completely rejected. This includes title preparation for those types to be selectively indexed which have no titles, e.g., non-CIA cables.

5.3.4. LEVEL OF COVERAGE

5.3.4.1. Raw Intelligence Reports

Since the information content of IR's supports a variety of intelligence interests, all IR's will be considered for some level of indexing. Duplicative information which frequently occurs between sources will be eliminated wherever possible, based on the information analyst's recall capability supplemented by data contained in dictionaries and identifier lists.

5.3.4.2. <u>Cables</u>

CIA cables (TDCS's) have always been handled as Information Reports and should be continued as such. As for non-CIA cables, the very fragmentary and highly perishable nature of these cables and the frequent duplication by follow-up reporting would indicate that only a small percentage of these cables are worthy of storage for retrospective search purposes. Only those cables containing positive foreign intelligence information will be indexed for header control as well as content. All others will be rejected completely--the Cable Secretariat continuing to retain repository responsibility for same.

5.3.4.3. <u>Finished Intelligence</u>

The Intelligence Publications Index (IPI) and Special Register's Job 3 are published by OCR to provide current awareness, and, to a lesser extent, retrospective subject and area searching for finished intelligence. In addition, some finished intelligence is incorporated into the files of BR and FIB.

Since the Agency has a repository responsibility for finished intelligence, bibliographic control over such material will be established in the CHIVE system for document retrieval purposes. Furthermore, some named-object indexing of finished intelligence documents will be performed similar to the control currently maintained by BR and FIB.

During the evolution of the CHIVE system, the bibliographic and named-object control achieved by the [redacted] 25X1A
25X1A [redacted] and subsequent branches will in part duplicate the contents of the IPI and Job 3. This duplication seems unavoidable since the issuance of these publications should continue in order to serve the current awareness needs of analysts, and it does not seem feasible during the implementation period to split the preparation of the publications between CHIVE and the existing activities. When implementation has been completed, however, it will be desirable to investigate the feasibility of producing a permuted title index to finished intelligence from the machine-stored data base as a replacement for both the IPI and Job 3.

DATA BASE
Level of Coverage
5.3.4.3.

STATSPEC

[redacted]

5.3.4.5. Foreign Translations

FDD and JPRS translations can be considered as one type. Heretofore in CIA no subject indexing scheme has incorporated both of these open literature sources. The manpower needed to cope with this large volume (see Table 5-1) is of significant concern. However, broad customer interest dictates in-depth subject and named-object control.

25X1A

Table 5-1

CHIVE INPUTS

Series	Approximate Annual Volume	Repository Responsibility	Bibliographic (B) and/or Content (C) Control Majority	Remainder
Raw Intelligence (including TDCS's)	253,500	X	B and C	B only
Cables	192,000	-	Not processed	B and C
Finished Intelligence	7,800	X	B only	B and C
[redacted]				
Translations				
FDD	78,000 items	X	B and C	B only
JPRS	44,300 items	X	B and C	B only
COMINT	109,050	X	B and C	B only
Photo Interpretation Reports	7,900	X	B and C	B only
Maps	6,000	-	B and C	B only
Films and Ground Photos	87,000	-	B and C	B only
	903,035			

25X6

[redacted]	7,200	-	Not processed	B and C
	36	-	B and C	Not processed
	625 items	-	B and C	Not processed
	1,560 items	-	B and C	Not processed
	3,400 items	-	B and C	Not processed
	104 items	-	B and C	Not processed
	12,925			

5.3.4.6. COMINT

All hard-copy SI material with the possible exception of military order-of-battle data will be considered for indexing in depth. Teletypes are excluded in their entirety pending the design of an automatic processing capability which will take advantage of the fact that the data is available in machine-language. An information analyst knowledgeable in both collateral and SI may be able to spot duplicative information if such exists. One large series of SI material which, in the present OCR/SR system, is given cursory control, will be studied to determine whether it should receive any title or subject control whatsoever. A few items in this series were processed during the experiment, but the titles were so general as to be practically worthless for retrieval.

5.3.4.7. Photo Interpretation Reports

The unquestioned value of this category requires that all published reports receive in-depth content and header indexing.

5.3.4.8. Maps, Films, and Photos

These categories of receipts will be excluded from CHIVE processing control because of the specialized knowledge needed for their analysis and input, the difficulty of separating the indexing function from the acquisition activity, etc. It has been agreed, however, to have these materials indexed by GR and the Map Library according to the CHIVE indexing scheme and the index records will be incorporated into the CHIVE data base.

5.3.4.9. Other

A number of miscellaneous classes of documents will also be processed by CHIVE. Most (e.g., press reviews and surveys) will receive named-object indexing primarily. The large volume of State biographic cards needs rigid selection and weeding not only to determine names of interest but also to eliminate repetitive information.

25X1A

Like the [redacted] there is little high-grade ore contained therein in relation to volume.

Chapter 5.4.

CHIVE INDEXING TECHNIQUE

5.4.1. INTRODUCTION

The most critical design element of the proposed system is the indexing system to be applied to input documents; the performance of the system is no better than the data which it is supplied. The transformation of textual material to the system language is an expensive process - one which has been given more attention than any other in the Phase II effort.

5.4.2. CONCEPTS

5.4.2.1. Document/Information Retrieval

The system will provide combined information retrieval and document retrieval capability. Documents themselves will be at the heart of the system, with their index records providing access to them through content control. The index records will also be the base from which information files will be built. That is, in the process of indexing documents, facts about named things of intelligence interest will be extracted and stored. The approach will be to extract information about specific named objects, keep this information in the context of the document for document retrieval, and manipulate this information out of context for information retrieval. It is not proposed to create non-redundant summary records from index records at input time either through human or machine collation. Summary records will be formed and maintained on select high-interest personalities, installations, and other finite subjects, but the creation of these records will be an analytic activity requiring the synthesis of index records and documentary information.

In addition to the index records, the indexer working aids will themselves be a source of answers to questions. For example, the Organization Identifier List will contain names of organizations, their locations, type of activity, etc.

5.4.2.2. Manual Indexing

An investigation of the state-of-the-art of automatic indexing reveals that it is still largely experimental and is not sufficiently precise to meet most of the Agency's retrieval requirements. Automatic indexing techniques usually involve word frequency counts, assigning weights to high-frequency words, and storing these words as index terms. Other techniques include syntactic analysis, sometimes in conjunction with the above statistical process. It is obvious that these techniques could not be applied to an intelligence storage and retrieval system requiring a high relevance/recall rate, since much intelligence information is inferential and interpretive and requires analysis for high-quality indexing.

Human indexing, therefore, with its recognized faults is still superior to automatic techniques and is the only feasible system for CHIVE. However, some documents will require only title indexing and in these cases automatic title-indexing techniques can be applied. The most notable title-indexing system is the Key-Word-In-Context (KWIC) method. In this system, the key words in titles are permuted so that each word appears in its alphabetic file position along with the other significant surrounding words from the title. The permuted titles can be machine stored for searching on demand, or printed listings can be generated for manual perusal.

5.4.2.3. Depth--Subjects vs. Named-Objects

It need hardly be argued that intelligence interests are catholic in nature, and that if an information storage and retrieval system arbitrarily decides to limit its coverage to personalities, installations, or conceptual-type subjects, it automatically limits its ability to satisfy its total customer population.

Intelligence analysts have found that "named-objects"--e.g. installations, personalities, organizations--most often provide the clues to resolving research problems. OCR request experience is an accurate reflection of this interest. We recommend, therefore, that these subjects receive the greatest emphasis; and, in view of OCR experience relating to the kinds of things users are interested in concerning named-objects, we recommend that an increased number of attributes of named-objects be brought under control. The latter are the elements of information which identify a named object, e.g., a person's address, organizational affiliation, etc. In-depth indexing of named-object attributes does not necessarily have to mean an equivalent increase in the volume of data indexed or in indexing time since common attributes, such as addresses, types of organizations, and products of an installation, will be stored in indexer identifier lists (see Section 5.4.2.5. below), and it will not be necessary to re-index this data when it is reported repetitively in documents.

We recommend that subject indexing, that is, the kind of indexing performed by the Intellofax system and the Subject/Commodity Section of the Special Register be continued at least to the present level, but on a broader data base to include important document series (e.g., foreign translations) which are excepted today.

5.4.2.4. Index Language; Linkage

The CHIVE indexing language consists of controlled entries taken from identifier lists and code schedules, as well as words and phrases extracted directly from documents.

5.4.2.4.1. Identifier Lists and Code Schedules

In the case of certain kinds of named-objects, identifier lists are required to ensure that the same organization, place, etc., is always entered in the same manner so that information is not missed during retrieval because of incorrect or synonomous entries. In the subject indexing area, a subject authority list or code scheme is required to control the depth of indexing, synonyms, and homographs.

In some cases, the authorized entry form will be identical to the way the entry will frequently appear in documents. In other instances, the entry will be converted to a code to either express the hierarchic structure built into the identifier list--e.g., the hierarchic arrangement of organizations in a Communist country--or to compress a long entry into more abbreviated form to conserve storage space.

5.4.2.4.2. Extracted Words and Phrases

Words or phrases extracted from documents are used (a) to index certain kinds of named-objects which will not receive identifier list control, (b) to give greater specificity to subject indexing, and (c) to provide information retrieval via the index record.

In the first instance, it is felt that identifier list control of all named-objects is impractical and impossible. Where the volume of reporting is reasonably restricted, or where one can predict fairly well which named-objects will be the subject of customer queries, it makes sense to control input through identifier lists. Such is the case, for example, for place names and priority organizations and installations. Personalities, however, are neither few in number nor can one readily anticipate which names will be requested. Similarly, in the case of lower-level installations, it would not pay to exercise a high degree of input control when it is probable that the referenced information will be retrieved infrequently, if at all. For both these categories, therefore, we recommend that the burden of overcoming the synonym problem be transferred to the output end of the system.

Key words taken from documents are added to subject index categories to provide greater retrieval specificity without complicating the subject schedule. The subject indexing vocabulary provides a medium-depth, generic searching capability. Key words added to the subject schedule provide a specific search capability, e.g., equipment nomenclatures, types of research, new concepts, etc.

The third application for entering key words from documents is to provide a level of information retrieval.

CHIVE INDEXING TECHNIQUE
Concepts
5.4.2.4.2.

In this case, the entry is uncontrolled, but the class of entry is searchable. For example, one of the personality attributes in the CHIVE system is "Reason for Travel." [redacted] information would be provided by the index record which would aid in selecting documents or in some cases obviate the need to refer to documents.

25X6
25X6

5.4.2.4.3. Linkage

Index entries which are related (e.g., an organization and its address) will be linked together in the index record so that the relationship can be interrogated at the index record level, thus negating the need to refer to documents to determine ties among elements of information. This is necessary because intelligence documents typically include many people, organizations, areas, subjects, and their interrelationships. If there were no way to determine the contextual relationship between these subjects, the system would be overburdened with false retrieval matches (false drops) requiring reference to many irrelevant documents.

Linkage can be accomplished through the use of formatted input, as is typical in punch card systems (i.e., all entries in one defined record are by definition linked), or by appending a linkage symbol to each index entry, as is typical in systems utilizing unformatted input. Formatted input records are not practical for CHIVE because of the long record lengths and large number of variable elements of information included. Experimentation with appending the linkage symbol to each entry has worked very successfully and will be adopted.

5.4.2.5. Requirements for Identifier Lists and Thesauri

The use of identifier lists is recommended for the following reasons:

(a) There is little consistency in the way named-objects are reported, e.g., the Institute of Physics of Moscow University may be referred to as the Moscow Institute of Physics, or the Moscow Physics Institute, or the Physics Institute of Moscow University, or the Nuclear Physics Institute, etc. Even place names are translated and transliterated in a variety of ways. Therefore, if named-objects were entered as reported, it would be a very difficult retrieval problem to determine the right synonyms to use in order to find the variant entries. An identifier list includes variants but allows only one correct entry format.

(b) An identifier list (e.g., for organizations) contains not only the name of the organization, but also a number of identifying attributes of the organization, including address, commodities produced, etc. This capsule summary aids the indexer in identifying and discriminating among organizations and improves the quality of the indexing.

(c) As was pointed out earlier, an identifier list helps decrease redundant indexing because the common attributes of a named-object do not have to be repetitively indexed when they are listed in the identifier list.

(d) Identifier lists are of value for answering queries of a non-complex nature such as the correct spelling of an organization or place, the precise location of a facility, etc.

Identifier lists will be required for installations and organizations, place names, significant national and international meetings and conferences, and personalities on whom physical or logical dossiers are maintained. The initial identifier lists will be constructed from the machine language data which exists in OCR, and will be issued to indexers in machine-listing form organized geographically in the various sort orders as required.

Key words will be appended to hierarchic classification terms to reflect the terminology of documents and to provide greater search specificity. The initial concept is that these words will be entered as written in documents and will not be subject to thesaurus control. The key words may be printed out, however, in answers to queries on the hierarchic subject codes to which they are appended, and should aid in determining which documents are relevant. For example, if a requester is searching for a particular aluminum alloy and three of the index records retrieved refer to alloys in which he is not interested, the requester can screen out these references from further consideration.

If in the future it is determined that dictionary control over key word entries will raise the quality of the indexing and retrieval, key word thesauri can be created by obtaining listouts of the key words which have been applied to the individual hierarchic codes. These key word lists would be turned over to dictionary editors who would resolve synonym and homograph problems and weed out undesirable terms. It is felt that this method of building a thesaurus, i.e., building it from the actual terminology used in documents, is both superior to and cheaper than trying to adapt an established dictionary to the Agency's indexing problem. In addition, one can take advantage of the uncontrolled key word indexing prior to the building of the thesauri.

5.4.2.6. <u>Header Data Indexing</u>

The foregoing discussion dealt with CHIVE concepts related to indexing the subject content of documents. Another important aspect of document indexing relates to the so-called header (or bibliographic) elements of the document such as title, author, control number, etc.

Header data indexing is required for the following reasons:

(a) To obtain bibliographic control of documents over which the Agency has a repository responsibility.

(b) As searching parameters in conjunction with subject or named-object searches.

(c) To provide minimum index control over documents which are not indexed in depth.

In the first instance above, header data control would perform a service comparable to that performed by the source card file maintained by the CIA Library. The machine-stored header data record will be used to verify the receipt of documents in the Agency and to recover specific documents whose control numbers are unknown. In the second instance, header data control will be used most often to limit searches (e.g., searches can be restricted to certain document series or dates), or a subject request can specify that information is required only when authored by a particular scientist. In the third instance, header data will provide minimum, but important, search keys at very little input cost. Permuted title indexes can be published for certain series (e.g., finished intelligence) in lieu of in-depth indexing. Similarly, searches can be made for all reports issued by a particular post during a specific time period when an important event occurred. In this latter case, all documents can be retrieved whether they were subject indexed or not.

Whereas the selection of documents for content indexing will be subject to well-defined criteria and therefore limited, it is anticipated that most documents can be brought under header data control. This possibility is rendered more likely by the fact that header data indexing (with the exception of title expansion) can be performed by clerical personnel, as borne out by the recent CHIVE Indexing Experiment.

5.4.3. SYSTEM DESCRIPTION

What follows is a summary description of the indexing technique. A detailed description is given in Appendix 5.C.

5.4.3.1. Elements of Information and Indexing Tools

As stated above, the CHIVE indexing concept includes "named-objects" and "subjects." Named-objects refer to people, places, organizations/facilities, and conferences/meetings. Subjects include commodities, concepts, research activities, military activities, and all other topics and events which do not fall under the above-defined named-objects.

5.4.3.1.1. Personalities

Personality names will be entered more or less as they appear on documents. Only those misspellings will be corrected which it is possible to recognize without reference to identifier lists or other support files.

25X1A

The use of name search tools such as the [redacted] Name Tables and printouts of unique personal name/surname combinations entered into the system will be investigated as substitutes for controlling names during input processing. When a specific name is searched, and all of the records relating to that personality have been identified, this identification will be retained so that subsequent searches for the same personality will have to address only those records which have been entered since the previous search.

A detailed list of the attributes of personalities which will be indexed is included in Appendix 5.C. Most of these attributes will be entered in a prescribed manner and thus will be available for direct searching in term files. For example, all locations will be entered from approved gazetteers, dates will be formatted, organization affiliations will be entered from organization identifier lists, etc. This will provide the capability to make information retrieval type queries from the index

25X6 [redacted]

5.4.3.1.2. Organizations/Installations

Two levels of control will be applied to organizations and facilities. Priority organizations will be included in identifier lists. These lists will also include significant attributes of the organization, e.g., addresses, synonymous names, function code, products, etc. The lists will be built from the machine language data which exists in SR, BR, and FIB. The organization identifier lists will be issued on a country basis in several arrangements, i.e., by name of organization, by function, and by place name location.

For organizations on the list, the indexer will enter an identifying number in lieu of the organization's name, thus ensuring that all indexed information relating to a specific organization can be retrieved exclusive of other organizations with the same or similar names. Attributes of organizations included in the identifier lists will not be re-indexed when the same information is repetitively reported.

Low-level installations and organizations will not be identifier list controlled. They will not be indexed by name but rather by location and a function code. It may be desirable later to produce listings of these facilities for mapping and aerial photographic customers. Once these listings are established, it is unlikely that any further indexing of these facilities would be required unless the status of the facility changed.

5.4.3.1.3. Area/Locations

For indexing large geographic areas, e.g., blocs, countries, and provinces within countries, the ISC area code has proven a satisfactory tool and it is recommended that it, or a similar country code, be adopted. For place names within countries, there are a number of gazetteers available, e.g., OCR generated gazetteers, the NIS gazetteer, etc. The NIS gazetteer has recognized faults, but it is generally conceded to be the most authoritative tool available and it is recommended that it be used as the authority for entering place names.

The basic gazetteer will be updated with new place name entries encountered in documents and will be issued on a country-by-country basis. Place names will be entered in clear text as they are spelled in the gazetteer, appended to the appropriate country code. Geographic coordinates will be entered in index records only when they are not associated with a place name. Coordinate searches will be accomplished by a machine search of the gazetteer to locate the appropriate place names having the desired coordinates, followed by a search of the place name term file plus a search for those coordinates that were disassociated with a place name.

5.4.3.1.4. Meetings/Conferences

Significant national and international meetings and conferences will be controlled in identifier lists.

Earlier comments on the use of identifier lists for organization control apply to this category also. Less significant conferences will not be indexed by name, but will be subject indexed with appropriate ISC subject codes.

5.4.3.1.5. Subjects

The Intelligence Subject Code has been used throughout the Intelligence Community for a number of years for subject indexing, and it is generally recognized as the best general indexing tool for intelligence documents. For these reasons, CHIVE has recommended that it be used as the basic subject indexing tool in a revised OCR system. However, during the CHIVE Indexing Experiment, several weaknesses were noted which should be corrected prior to its adoption in a going system.

5.4.3.1.5.1. ISC Structure

The 1960 revision of the ISC did much to simplify its structure. However, experience in using this edition points to several areas where further simplification is desirable.

(a) Expanded Use of Modifiers: The ISC subject modifiers are a faceting device which can be combined with certain subjects to specify actions or states which affect those subjects. For example, the modifier "049 Production" can be combined with any commodity to indicate production of the commodity. The 1960 revision greatly expanded the use of these modifiers over previous editions, but further expansion is desirable in two ways:

(1) The 1960 revision limited the use of the modifiers, i.e., each modifier could only be used with specific chapters or sections of the ISC. As a result, in sections where a modifier cannot be applied, it has been necessary to set up a subject code in lieu of the modifier. For example, modifier "069 Government Policies, Laws, Legislation, etc." can only be applied to the commodity chapter of the ISC. As a result, a subject code for government policy has had to be set up in various non-commodity sections of the ISC. If the modifiers were freed and the redundant subject codes deleted, it would increase the efficient application of the ISC. During the CHIVE Indexing Experiment, the modifiers were freely applied, and no particular difficulties ensued.

(2) In a subject classification system, the same subject is often repeated in several different sections because each section gives a different meaning or emphasis to the subject. For example, in most classification systems, guided missile subjects would be found under engineering, production activities, and military activities. This repetition is logical, but it complicates the structure of the system and makes it hard to apply. A generalist indexer often

finds it difficult to determine whether the information he is reading is oriented toward engineering or production. If he mistakenly puts engineering information under production, it may be lost in a later retrieval run. With the addition of some new subject modifiers, much of this repetition could be eliminated, i.e., the various subject facets could be shown through the use of modifiers to distinguish production from military activities, etc. This would also considerably reduce the size of the ISC.

(b) Expanded Use of Clear Text: Some of the detailed subject breakdowns in the ISC could be eliminated with a more liberal use of clear text. During the experiment's indexing consistency test, it was found that there was a low-level of consistency in applying the ISC. This can be attributed to the depth of subject detail in the ISC, i.e., one indexer will use "621.349 Uranium" and another indexer will index the same subject matter using "621.351 Natural Uranium."

If some of this subject detail were further reduced so that there was only one subject code for uranium, the consistent application of the ISC would rise measurably. Moreover, indexing specificity (e.g., natural vs. enriched uranium) could still be achieved by using controlled clear text as an extension of the subject code. The advantage of this approach is that with more consistent application of the ISC there is less likelihood of losing information. This may often put a burden on the searcher in that with fewer subject categories, more material will initially be retrieved, but this is preferable to losing information and the free use of clear text can help alleviate the problem. Thus, if the clear text is uncontrolled, it can be used as a screening device to get rid of unwanted references, or if it is controlled, it can be used as a searching device to restrict the volume retrieved.

CHIVE INDEXING TECHNIQUES
System Description
5.4.3.1.5.1.

5.4.3.1.5.2. Subject Schedules for Occupations and Installations

A subject schedule or code is required for occupations and installation types in order to respond to queries on such subjects as all [redacted]

25X6

25X6

During the recent Indexing Experiment, specified subject codes in the ISC were designated for this purpose. Since the ISC was not constructed with this aim in mind, the designated codes proved quite inadequate. Problems were caused by the previously alluded to duplication of subjects (e.g., an atomic installation could be indexed in several different places), and by the multiplicity of subjects in the ISC (i.e., a rather simple code schedule was required, and the ISC was too detailed for the required need). In addition, the ISC did not have appropriate subjects for some occupation and installation categories.

---

*This need for a generalized subject schedule for occupations and installation types is to be distinguished from the requirement to retrieve by specific activity. The latter capability will be provided either through the ISC code itself or, where necessary, through ISC plus key word.

In view of the above problems, it is recommended that the ISC not be modified to perform this function, but that separate subject schedules be developed based on the experience available in the Foreign Installations Branch and Biographic Register.

5.4.3.1.5.3. Area Rules

The present ISC area rules proved inadequate on several counts during the recent experiment.

(a) The terminology is sometimes confusing--e.g., some rules read "nationality is primary area." Since the CHIVE indexing procedures provide for area tags for nationality and primary country, the terminology is subject to ambiguous interpretation.

(b) Some of the rules are illogical--e.g., there are two subject codes in Chapter VII which can be used for foreign military training and the area rule for one of the codes is the opposite of the other.

(c) The CHIVE indexing technique allows more flexibility in area relationships than is allowed in the ISC as used in the Intellofax system. Consequently, there are many subjects which do not have area rules which should have them appended for CHIVE purposes.

All these rules should be re-examined and modified before the CHIVE system goes operational.

5.4.3.1.5.4. Subject Gaps

Prior to the recent experiment, it was felt that a number of subjects occurred in Codeword materials which did not appear in Collateral documents and that, therefore, the ISC would not be adequate for indexing these materials. For this reason, sections of the Special Register code manual were utilized as a supplement to the ISC during the experiment. As it turned out, the SR supplement was not used a great deal because the ISC had subject categories which were almost comparable. However, there are a limited number of special-purpose subjects which should be added to the ISC to make it fully suitable for all-source indexing.

5.4.3.2. Tags

Each entry in the CHIVE indexing system is preceded by a tag. A tag is a three-digit mnemonic symbol which identifies the entry which follows. Tags are used to:

(a) Distinguish between homographs, e.g., Washington a person vs. Washington a city or street.

(b) Organize machine files, i.e., separate people's names from organizations and subjects and thereby facilitate searching.

CHIVE INDEXING TECHNIQUES
System Description
5.4.3.2.

The CHIVE tags were made mnemonic as a memory aid. The first character of a tag represents a major subject category, e.g., "P" = Personality, "O" = Organization, etc. The second and third characters further specify the element of information being indexed, e.g., "PVN" = Personality Name Variant, "POH" = Personality Organization Head, etc. (See Appendix 5.C. for a detailed list of the CHIVE elements of information and their associated tags.)

5.4.3.3. Phrasing

The requirements for linkage were discussed earlier. In the CHIVE system, linkage is accomplished through a system defined as phrasing. A phrase is simply a group of tags and terms which the indexer relates together with a unique number which is assigned to each tag and value in the group. On retrieval, queries can specify that the input linkage must be present for the query to

25X6 be satisfied--i.e., a query may specify all information on a person [redacted]

25X6

Without phrasing (linkage), all documents which contained

CHIVE INDEXING TECHNIQUES
System Description
5.4.3.3.

these two terms would be retrieved and in some cases the relationship would be accidental. On retrieval, the phrase linkage can be reconstituted by testing for those terms which have the same document accession number and phrase number.

The rule for phrasing is very simple. All terms which are logically related can be combined in a phrase. Thus, if a <u>person</u> is affiliated with an <u>organization</u> in

25X6 [redacted] these three elements of information can be combined together in a phrase. However, if additional information were given that this individual also traveled

25X6 to [redacted] an additional phrase would have to be constructed otherwise it might be interpreted that the organization, if it appeared in the same phrase, was located in both

25X6 [redacted]

Phrases can be very simple or complex. The simplest phrase contains only a place name or an area and one subject. A complex phrase may contain a number of index terms which constitute a rather detailed biographic sketch of an individual. Further details on phrasing with examples are contained in Appendix 5.C.

CHIVE INDEXER TECHNIQUES
System Description
5.4.3.3.

5.4.3.4. Header Data Indexing

Header data indexing will be performed by clerical personnel who will type the information on formatted transcript sheets. During the recent experiment, as illustrated in the header section of Appendix 5.C., a single transcript sheet was used for all documents. This does not appear to be as efficient as developing unique transcript sheets for different series.

The elements of information comprising header data will be taken from the document or the information from the document will be translated into code to achieve conciseness and uniformity of entry. A formatted transcript sheet can be used since the header data elements are fixed in number for each document series, and the length of entries is either fixed or a maximum field length can be determined. The use of a formatted sheet obviates the need for tags and the only linkage required is to the document control number. The latter will be appended automatically to each header data term. A detailed list of the elements of information comprising header data is included in Appendix 5.C.

Chapter 5.5.

SYSTEM FILES

5.5.1. INTRODUCTION

This chapter classifies and describes the logical files and sub-files which will be available in the CHIVE system. These are the files which are identified to the user--i.e., the CHIVE information analyst and, perhaps ultimately, the research analyst. They are the files he must be familiar with, if he is to take full advantage of the resources of the system and exploit it intelligently.

The total number of individual system files, including old as well as new, might easily exceed a hundred. However, it is possible to classify all the various files into no more than nine types, each with very distinctive functions and properties. These nine categories are as follows:

Document Index Files: Files containing all the raw document index records in the system, including not only the complete index records themselves but the access mechanism to these records. The documents referenced by these records may include any form of information carrier --e.g., maps, photos, films, or other, and need not necessarily be readily accessible to the system.

Vocabulary Control Files: Files required to insure consistent entry of index terms (tag and value) into the Document Index Files and other system files. The principal function of these files is to reduce the synonym problem at search time. They include "identifier files" for named objects (which, like scope notes in a code schedule, help to distinguish one specific subject from another), code books, dictionaries, thesauri, and other authority lists.

Unsynthesized Information Files: Files consisting of select phrases or terms extracted from document index records or directly from the raw documents themselves. Such files would be built to facilitate retrieval where a substantial number of requests for the pertinent data can be anticipated on a continuing basis. Unlike Summary Information Files (see below), records in these files would often contain duplicative and/or contradictory information. Periodically, however, information in such files might be reviewed and added to the appropriate Summary Information Files.

Summary Information Files: Files built either from records (or portions of records) in the Document Index Files, from records in Unsynthesized Information Files, or from the raw documents themselves during or after input processing. The distinguishing feature of these files is the fact that they will ordinarily contain evaluated, non-redundant data about named objects or events associated with named objects. Named-object identifier files could be placed in this file category, the only apparent difference being the limited amount of historical data ordinarily found in such files.

<u>Special Project Files</u>: The unique features of these files are as follows: (a) the inputs to the files originate outside CHIVE; (b) CHIVE actually acquires the files and not simply "profiles" thereof; (c) additions or modifications to the files can be anticipated; (d) the files do not use the elements of information and/or vocabulary controlled in CHIVE. Special Project Files may otherwise have the properties of any of the file classes named above. These files will be processed by CHIVE but maintained by CIA or other agency analysts. The degree of CHIVE involvement in such files remains to be determined since the responsibility for such files is currently assigned to the Applications Division of OCS.

<u>Referral Service Files</u>: These files differ from Special Project Files in that they are not substantive data files but rather descriptions or profiles of files located outside the CHIVE system. Referral Service Files will consist both of profiles of analysts' special fields of competence as well as files maintained by analysts and/or information repositories external to CHIVE. CHIVE will not maintain, or retrieve data from, the substantive files themselves. It will simply inform customers of those files potentially relevant to a given query.

Document Image Files: Files of documents stored by the CHIVE system. From a functional point-of-view they include "aspect" systems (where the index is stored separately from the documents) as well as self-indexed document files. Both existing OCR document collections as well as CHIVE-originated document repositories are encompassed by this category. The storage media for such files will include hard copy, various types of microimages, and even digital storage in some instances. Similarly, the categories of documents involved will differ widely in size, shape, classification, and point of origin.

Management Data Files: Files of data collected on the activity of the CHIVE system to (a) enable operational management to evaluate the cost/performance ratio of the system and (b) to guide system designers in improving hardware and software support. From the point-of-view of what data is collected, most of the Management Data Files will have to do with either system processing times or processing volumes.

System Processing Files: Files used to support the system in processing data. Most such files will be organized in table form enabling values to be obtained from arguments. Examples would include a file of legal tags and other error correction files, decode dictionaries which would convert codes into clear text for display to a reader, intermediate files which exist only temporarily during the processing of a transaction, working storage files, etc. Since these files are largely internal to the CHIVE EDP System and the information analyst need not interact with them in any direct way--only know what functions the system is capable of performing--they will not be covered further in this volume but rather in Volume VII of the report.

For each of the file categories listed above a second-level categorization may be required, i.e., one which classifies CHIVE files from the point-of-view of the origin of the files. These classes are three in number:

Chive-Built Files: Files built by and for the CHIVE system either from new inputs or through the conversion of existing OCR files to the format and vocabulary of CHIVE. These files will be continually updated as part of the regular processing cycle.

Inherited Files: Files originally established by the various OCR systems which it was not found possible to integrate with new CHIVE files. Such files will include records in hard copy as well as machine language. In some instances these files may be transferred to another storage medium (e.g., magnetic tape) if querying and output can thereby be improved. Similarly, some existing machine-readable files may be restructured and interrogated in the vocabulary of a single CHIVE language. Neither of these changes, however, implies true conversion to the CHIVE system. Another significant difference between these files and Chive-Built Files is that while both will be used by the CHIVE information analyst, no additions will be made to the Inherited Files once the CHIVE system is fully operational.

Supplemental Files: Files not built or maintained by CHIVE, nor inherited from OCR, but which contain data functionally useful to CHIVE as a secondary source of information. All Special Project Files (see above) fit this category, as do reference aids of various kinds (e.g., Who's Who compilations, gazetteers, commercially published indexes, etc.) obtained from external sources and left essentially in the form in which they were received.

In the broadest sense the CHIVE system must necessarily include not only the new files it creates but the files it inherits from the existing system. The discussion of these separate but related subjects, however, has been divided in the pages to follow to lessen the possibility of losing the reader in the file forest. In the main body of this chapter, we will focus primarily on the CHIVE-Built Files, providing a summary description of their functions, data content, maintenance criteria, and other characteristics. Appendix 5.D to this volume describes the principal Inherited Files which must be accommodated by the system, with primary attention given to those which fall in the categories of Document Index and Document Image files, as defined above.

It should be emphasized that the basic objective in this chapter is to communicate a more or less static image of the files in order to simplify understanding of the structural framework (or file philosophy) of the system. In Chapter 5.6. we will examine the more dynamic aspects of file activity within the system, i.e., the transactions which will affect the files, interactions which might take place between files, etc.

5.5.2. DOCUMENT INDEX FILES

5.5.2.1. <u>Master Index File (MIF)</u>

The Master Index File of the CHIVE system will contain the index entries for all the documents available in the CHIVE system as well as certain classes of documents located in repositories not under CHIVE management. Examples of the latter include maps, the storage responsibility for which will be retained by the Map Library Division of ORR, and select open-source books and periodicals which may be accessible only at the Library of Congress or at some other holding agency. Conceivably, certain documents indexed by

CHIVE may not even be available at all--for example, select Soviet periodicals never received in this country, but which were described in secondary sources that were accessioned. In all cases, however, whether the original source document is readily available or not, the preparation of index records for the CHIVE Master Index File will be under CHIVE format and vocabulary control no matter where the records are physically prepared.

All index records will be stored in such a manner that a search, based on certain criteria, will produce all the records in the system or, at the customer's option, phrases and/or terms within records which may apply to the search criteria. The index records will contain sufficient information to enable the requester to determine if the document referred to in the index entry should be requested for detailed study. In the case of named-object associated information, the entries will have sufficient information-bearing content to permit summary data files to be built and responses given to certain queries directly from the index records themselves without referral to the source documents.

Records entering the Master Index File will originate from the following sources:

- CHIVE information analysts processing incoming documents in the CHIVE geographic divisions.
- Graphic analysts indexing photos and films in the Graphics Register (GR).

- Map catalogers processing maps in the Map Library Division (ML), ORR.

- Miscellaneous additional organizations (either under contract to CHIVE or agreeing to follow CHIVE input procedures) exploiting primarily foreign language documents. Examples of such organizations might be the Library of Congress, FDD, etc.

- Documents received by CHIVE in machine language (e.g., Comint teletype) on which a limited form of automatic indexing is to be performed.

- Machine-converted document index files from existing central repositories.

With regard to input selection criteria, assuming continuation of present practices, CHIVE will have the responsibility to serve as the Agency's repository for community-published positive intelligence materials (with the exception of cables and maps), and to provide reference service on "active" documents. In addition, it will presumably assume OCR's obligation to serve as the office of record for archival storage of certain CIA document series [redacted]

25X1A

In order to fulfill these responsibilities, CHIVE will be obliged to index at least the header (or bibliographic) data for every "intelligence" document received. By "intelligence" documents we mean all categories of textual materials generally considered to be in the mainstream of intelligence reporting. These include Comint (messages, reports, and teletype), T/KH reports, USIB-produced IR's USIB-produced finished intelligence, the FBIS, photo enclosures to IR's, and USIB-produced translations of foreign documents. By agreement with the Map Library it will also store map index records generated by ML.

The preparation of index records on other categories of materials, e.g., cables, non-USIB-produced reports and studies, films, and original open-source literature, will depend on the substantive content therein.

The content of document index records can include any legal term type permitted by the vocabulary of the CHIVE indexing system. (For a list of all permissible term types see Appendix 5.C.) No single record will, of course, contain all possible term types since some terms will be unique to certain kinds of documents.

Outputs from the Master Index File will consist of both scheduled and ad hoc products. The principal items to be provided within each category are briefly described below.

Scheduled Products

- KWIC listing of titles or expanded titles of all documents which have not been content indexed, as well as the FBIS Daily Reports. The permuted portion of the listing will be the title and expanded title, while the reference portion will include basic header data for the document including document control number. Separate, as well as combined, listings will probably be run for the different categories of documents involved, e.g., SI Teletype, FBIS, Finished Intelligence, and Raw Intelligence Reports.

- Map catalog cards in 3" x 5" form containing in clear text on each card the entire index record for a map. This record would include accession number, area code, subject, scale, classification, map title, date of publication and name of publisher. The cards would be outputted in the sequence of the Map Library Card Catalog to facilitate interfiling at the Map Library.

- Output similar to the map cards, but reflecting index records on films stored in the Master Index. In this instance, the records will probably be of tab card size to conform with the size of the existing file. The sequence will also conform with the existing Intellofax reference card file on the film collection.

- Accessions lists comprised of clear-text index records on maps, ground photos, and perhaps select additional document receipts (e.g., tables of contents of foreign scientific periodicals) processed by CHIVE.

Ad Hoc (Query) Products

- Listings in natural language of document index records or subsets thereof (i.e., phrases within records) containing the search terms specified in the query. Subject or concept-oriented queries will normally require output of complete document index records including header as well as content data. Named-object-oriented queries will ordinarily result in the output of select phrases only which match the search criteria, together with a limited amount of header data (e.g., document classification, document type, and appropriate document control numbers).

- Listings of control numbers only for documents whose index records match the search parameters.

- Listings containing simply a computed figure of the number of index records matching the search parameters. This kind of intermediate output will enable the customer to broaden or narrow the search prescription depending on the volume of the anticipated output.

- Listings of index records containing terms which match standing customer queries or analyst interest profiles. Whenever hits occur because new information is received on a subject or person of interest to a particular research analyst, the customer would be notified through the transmittal of the listing containing the pertinent record.

5.5.3. DOCUMENT IMAGE FILES

The Document Image File (DIF) is the central repository for active textual intelligence documents. Maps and graphics are to be retained within the respective organizations currently responsible for their retention, although all of these will be used in conjunction with the computer-based Master Index File described in the previous section. The Document Image File shall consist of those textual intelligence documents for which the Agency has repository responsibility as well as other documents which are judged to contain information of potential value within the intelligence community. As a central repository it is to be all-source, containing

USIB finished and unfinished intelligence reports, FBIS, Open-Source literature (including JPRS and FDD translations), COMINT (messages, reports, and teletypes) and selected cables. The system will be inclusive of inherited document image files (see Appendix 5.D.) as well as newly-accessioned, CHIVE-processed documents. To effect this, files currently maintained in various locations (SR, LY/Circ, etc.) would be moved to a single physical area within the headquarters building along with the CHIVE document system, thus offering the user a single point of entry for his reference needs. The discussion of a proposed approach to implementing this central repository is contained in section 5.7.3.

The primary purpose of the Document Image File is to serve as a central reference point from which identified documents may be retrieved and copied for distribution. The identification of the documents (by unique identification number) may be accomplished via a computer search of the Master Index File, or it may be known by some other means by the requester. The file must be responsive to either type of demand. Documents are not to be circulated outside of the file area; and requests are to be serviced by producing a durable, hard-copy replica of the document master for distribution to the requesting user. The design goals for the volume and turn-around times in responding to these file demands are outlined in sections 6.2.1. and 6.5.6.

Aside from its primary purpose of providing a repository for retrospective reference, a number of secondary purposes must be served by the system. First, provision must be made for a backup file capability. This duplicate file must be produced as a by-product of the input procedure, and must be suitable both as an alternate reference point in the event of loss or destruction of items in the main file, as well as a means of reconstructing the main file in the event of catastrophic destruction. Provision for selective protection of vital records is also within the scope of the document image subsystem although no special design consideration has been devoted to this requirement in this study. An additional implicit requirement of the document system is the need to provide archival quality records for those items requiring prolonged retention.

5.5.4. VOCABULARY CONTROL FILES

5.5.4.1. Personality Identifier Files

5.5.4.1.1. Master Dossier File (MDF)

The functions of the Master Dossier File are:

- To identify hard copy folder files maintained by CHIVE on select personalities.*

25X1B

[redacted]

- To reduce search time on requests for select personalities by virtue of the fact that the information analyst determined in advance of the request which incoming name references pertained to these personalities. An analogy could be drawn here to the difference between searching a tightly controlled classification

---

*The reason for maintaining hard copy folder files, in addition to storing documents in the Master Image File in microimage form, would be the anticipated high request activity on these select documents which would increase reproduction costs significantly. An alternative to maintaining hard-copy personality files would be to maintain lists of documents referring to select individuals, and, when one of these individuals' file is requested, to reproduce all the documents referred to in the list. This approach of building a dossier on a "demand" basis would make sense if experience proves that redundancy in name searches is minimal. Pending further study, however, of the redundancy factor during Phase III, we have assumed a requirement for some hard-copy personality files and made provision for same in the system design described here.

system and an uncontrolled keyword index. A listing of the Dossier File, wherein is contained a unique file number and set of attributes for each personality cited, is directly analogous to a listing of a classified schedule which contains both a unique code and often scope notes defining each term in the listing.

- To provide by means of a printout of the identifying information on each dossier personality a summary-type information record which can be used to answer requests, serve as a reference aid for research analysts, facilitate screening of dossier files without the necessity for examining the files themselves, etc.

The initial CHIVE Master Dossier File will be derived from the BR dossier system with possibly some deletions to the latter file. The subsequent creation of new dossier records will occur largely as a result of name searches, taking advantage of the fact that document index records and their related documents have been analyzed in the course of answering the customer's inquiry.

Dossier identifier records may refer either to physical or logical dossiers. In those cases where it seems desirable to establish a hard-copy file on a personality, a digital record containing select elements of identifying information (see below) will be prepared and added to the Master Dossier File. In addition, all

SYSTEM FILES
Vocabulary Control Files
5.5.4.1.1.

documents containing information on the individual will be reproduced and stored in a folder on the personality involved. Logical dossiers will also be represented by identifier records in the digitalized MDF, but the documents relevant thereto will be accessible only in the Master Image File.

Maintenance of both the hard copy dossiers as well as the digitalized Master Dossier File--although left to the discretion of the analyst--would also ordinarily be performed as a corollary function to name searches and not at the time incoming documents are indexed. This means that a given hard-copy folder file will not necessarily contain all the available documents on an individual which may be held in the Master Image File except immediately subsequent to a request having been answered on said personality. Similarly, the digital record on a dossier personality will only be current as of the time of the last request.

The contents of a digital record in the Master Dossier File will be:

SYSTEM FILES
Vocabulary Control Files
5.5.4.1.1.

- Personality Name
- Variant Name
- Telegraphic Code
- Dossier Number
- Birth Date
- Citizenship
- Date of Death
- General Occupation
- Organization Affiliation
- Position Title
- Organization Affiliation Date (year only)
- Date Record was Last Updated
- Document Reference Numbers

Whenever new dossier records are added to this file or changes made to existing records as a result of name searches, the following actions will take place:

(a) Documents not previously filed in physical dossiers will be reproduced for same.

(b) Dossier identifier records will be created or updated.

(c) The list of document control numbers attached to each identifier record will be compiled or updated.

The effect of (c) will be to establish the identity of the individual mentioned, thus capturing the results of the analysis. Persons searching the same name at a later date will have to employ standard search strategy techniques only to recover those records from the Master Index File which might have entered the system subsequent to the previous search. Earlier references will be available either via the hard copy dossier itself or, in the case of logical dossiers, through an "absolute" search on the document numbers known to be relevant to the individual concerned. The latter can be a semi-automatic process in which the information analyst need only specify the dossier number involved--i.e., the computer will find the document numbers pertinent to the dossier, and either print them out or use these numbers to locate and output the corresponding index records.

In the initial CHIVE system, scheduled outputs from Master Dossier File will consist of:

- Master listings in natural language of the personality identifier records arranged by name within citizenship.
- Cumulative supplemental listings in the same arrangement as the master listings.

SYSTEM FILES
Vocabulary Control Files
5.5.4.1.1.

Demand (ad hoc) products of the file will include natural language printouts of the records on a variety of media (e.g., cards, listings, etc.) in any sort order desired by customers.

5.5.4.1.2. Name Group Tables

The function of name group tables is essentially that of any dictionary of synonym and "see also" references. Such tables properly belong in a list of "vocabulary control files" since, like any term dictionary, they serve to relate the several ways in which a term (in this case personality name) can be spelled to a standard code.

In the CHIVE system, it is proposed to experiment with the two kinds of name group tables developed by

25X1A

and for [redacted], the Surname Table and Given Name Table. Each of these tables contains a list of all the surnames or given names, as the case may be, which have occurred within the system. Listed with each name is a reference to the name group to which it has been assigned.

The functions of these name tables are: (a) to determine if a specifically spelled surname or given

SYSTEM FILES
Vocabulary Control Files
5.5.4.1.2.

name is contained within the system, and (b) to associate a group number to the name. A new name entry in a document index record, before it can be filed, must match a name in the name table. If it finds no match, the machine will print out a notice to the information analyst to this effect. The latter will then consult

25X1A

his tables or the [redacted]/CHIVE expert concerned (procedure to be determined), assign a group number to the name, and re-enter the record into the machine file.

Ideally, the name group table concept reduces the intellectual problem for the name searcher by providing for a guided search of potentially relevant, alternative name spellings. This capability will not, however, preclude the searcher from bypassing the name grouping feature if he wishes the machine to yield only those records which exactly match the spelling(s) in his request.

Scheduled products of the Name Group Tables will include listings of both the surname and given name tables arranged in both name and group number order. Query products may include the variant names searched within a given name group as well as "see also" references to names in related groups.

5.5.4.2. Organization/Facility Identifier Files

The Master Organization/Facility Identifier File (MOFIF), like other vocabulary control files, is required to insure consistent indexing of items of information derived from documents--in this case organizations or facilities (installations). In the sense used here, organizations and facilities are defined in the broadest possible terms. They include political, economic, military, cultural and scientific bodies, as well as physical installations which are relatively fixed in terms of geographic location (e.g., a weather station).

Like the Master Dossier File, the functions of the MOFIF will be:

- To identify hard-copy folder files maintained by the system on select organizations where high request activity is anticipated.
- To provide, via printouts from the file, identifying information about organizations which an information analyst can browse in order to determine (a) whether he has previously assigned a code or unique identifying number to an organization and/or (b) whether there is a hard-copy dossier available on an organization.
- To reduce search time on requests for "controlled" organizations.

SYSTEM FILES
Vocabulary Control Files
5.5.4.2.

- To provide a summary-type information record which can be used to answer requests, serve as a published reference aid, etc.

Not all organization and facility names will be placed under MOFIF control. Furthermore, not all organization references encountered in documents will be indexed by the specific name and/or identifying number of the organization mentioned. Some organizations will be indexed only by type using the OTF tag. Still others (e.g., a laboratory or committee) will be indexed by their parent organization's code, but will not be assigned a unique identifying number of their own, consequently they too will not appear in the MOFIF.

The initial CHIVE Master Organization/Facility Identifier File, which will be composed [redacted] will be built during Phase III of the CHIVE project from existing organizational dictionaries developed by FIB, SR, and BR. Each organization record resulting from this process of analysis and synthesis will include, in addition to the CHIVE-assigned identifying number and name, the code number or numbers (if any) by which the organization was previously

25X6

25X6

SYSTEM FILES
Vocabulary Control Files
5.5.4.2.

identified in the OCR register(s) from which it was derived. These "cross reference" numbers will help indicate to the searcher whether there is information stored on an organization in one or more of the files inherited from OCR (e.g., SR Detail Index, FIB card and plant folder files, BR dossier, etc.). The absence of such cross references in a record would mean either that there was no inherited information available on the organization, or that the organization was so loosely controlled in the earlier system that a subject search or some other method for accessing the files would be required to uncover the pertinent data.

As in the case of personalities, it is planned that certain organizations would have hard-copy dossier files where a high request rate is anticipated (but see footnote in section 5.5.4.1.1. regarding request redundancy study). If this plan is implemented, incoming documents containing information on dossier-controlled organizations would probably be added to the dossiers as a part of the initial processing activity rather than at the conclusion of search operations on said organizations (as was proposed in the case of personality dossier maintenance). The reason for this is that the MOFIF would ordinarily have to be consulted when indexing an organization in order to obtain the correct CHIVE identification number for the organization. Such being the case, little additional effort would be required to determine from the MOFIF listing whether a dossier was being maintained on the organization and, if so, to direct that a copy of the document be deposited in the dossier concerned.

The contents of a digital record in the Master Organization/Facility Identifier File will be as follows:

- Translated Name and/or Number
- Functional (Assigned Name)
- Foreign Language or Transliterated Name
- Variant Name(s)
- Previous Name(s)
- Name Abbreviation
- Telegraphic Code
- CHIVE-Assigned O/F Number
- Dossier Indicator
- Cross-Reference Numbers
    - FIB
    - SR

- COMOR
- BR
- BE
- NPIC
- TDI

- Address
  - Country
  - Political Subdivision
  - Place Name
  - Coordinates
  - Street Address
  - Cable Address
  - Post Box Number
- Parent Organization
- Type O/F
- Source Citations
- Remarks

Not all the elements of information listed above will appear in every organization/facility record in the MOFIF. Not only will the type of organization have an effect on the elements of information that will customarily

SYSTEM FILES
Vocabulary Control Files
5.5.4.2.

appear in its identifier record (e.g., a political body will not have a COMOR or BE number, nor perhaps a specific address), specific elements of information will be unavailable on many organizations and facilities.

Source citations, where desirable, for items of data carried in the MOFIF can be included in the identifier records by referencing the control number of the document which provided the information. One source reference for each element of information in an MOFIF record would probably be sufficient. If additional supporting evidence for a given fact was required, the index records in the Master Index File could be searched.

The "Remarks" field of an MOFIF record is intended for use in recording historical facts about changes in organizational nomenclature, hierarchic relationship to other organizations, etc. This information will not be directly accessible, but may be displayed on printout to enable searchers to determine how to formulate a request which will insure recovery of all pertinent data about an organization despite organizational changes which might have taken place over the years.

SYSTEM FILES
Vocabulary Control Files
5.5.4.2.

In the initial operational system, current thinking is that human interface with this file, as well as all other vocabulary control files, will be through the medium of the printed listing. It is recognized that this is not a wholly satisfactory solution (although a familiar one), particularly in view of the probable increase in the number and size of vocabulary control files which the CHIVE indexer must routinely consult. For this reason, an in-depth study of the matter is planned during Phase III of the CHIVE project.

Scheduled outputs of the Master Organization/ Facility Identifier File will consist of:

(a) Master listings in natural language of the organization identifier records pertaining to a single country. The types of master listings required are:

(1) A permuted title listing of all organization names (including official, assigned, variants, etc.). The reference portion of the listing will be ordered by O/F number and contain the complete identifier record for each organization referenced.

(2) A listing of MOFIF records, without organization name permutation, ordered on place name.

(3) A listing identical to (2) but ordered on type of organization/facility.

(4) A listing identical to (2) but ordered on geographic coordinates.

(b) Cumulative supplements to the master listings issued in the same arrangements as the master listings but bound together. Alternatively, supplementary listings may be issued in the form of pages to be inserted in master listings.

Demand (ad hoc) products of the file will include natural language printouts of the records on a variety of media (e.g., cards, listings, etc.) in any sort order desired by customers.

5.5.4.3. Meeting/Conference Identifier Files

In the planned central retrieval system a requirement exists for retrospective searching of documents dealing with certain meetings and conferences. The conditions which dictate whether a given conference should be indexed by name or identifying number cannot be stated with complete precision at this time. Nevertheless, the fact that some conferences (possibly international scientific meetings attended by USSR nationals) must be controlled dictates that the capability be provided in the CHIVE system to retrieve the pertinent data, whatever the input criteria might ultimately turn out to be.

SYSTEM FILES
Vocabulary Control Files
5.5.4.3.

The function, therefore, of the Master Conference Identifier File (MCIF) is to relate the several ways in which the name of a conference or meeting may be spelled to a standard code and, in addition, to supply other identifying information which would facilitate distinguishing meetings having similar names.

The initial data base for the MCIF may be derived from the [redacted] BR's International Conference and Travel File. In this instance, a requirement does not exist to merge separate OCR system vocabularies since BR is the only organization which maintains a conference authority file. Informal consultation between the CHIVE conference dictionary editor and BR's authority in this area during the evolution of the CHIVE system should enable standardization to be achieved between the two systems on the identification of international meetings which both systems index.

25X6

The contents of the MCIF will be as follows:

- Name of Meeting/Conference
- Assigned Code Number
- Location
    - Country
    - City

- Date of Conference
- Type of Meeting (Subject)
- Sponsor Organization

Scheduled outputs from the file in the initial CHIVE system will consist of the usual master listings arranged in this instance by name of conference, location, and sponsoring organization. The name listing will be a permuted title arrangement with the reference portion of the listing ordered by code number and containing the complete identifier record for each conference referenced. Supplementary listings will also be provided either in cumulative form or (as indicated earlier) as pages to be inserted into master listings. Demand products will, of course, be issued in any sort order desired.

5.5.4.4. Location Identifier Files

The function served by the Master Location Dictionary (MLD) is to specify the approved entry form for certain classes of locational-type information listed below. (Detailed address information, e.g., cable address or street name, will not be under vocabulary control and, consequently, will not appear in this file.) Additional uses of the file will be to confirm file coverage by location, to show hierarchical and synonomous relationships between place names and political/administrative regions of the world, and to support requests by location defined in terms of country, political subdivision, or place name.

Location identifier (authority) files will ultimately be maintained on all countries, but in the initial system primary concentration will be placed on the [redacted] element of the Master Location Dictionary. Specific senior content indexers will serve as dictionary editors for certain geographic portions of the file, rejecting or approving all new entries generated as a byproduct of the document input process. 25X6

The initial Master Location Dictionary will be constructed on the base of the NIS Gazetteer. The ISC 4-digit classification system will be used to identify country and political subdivision. Place name entries will be carried in full text.

Map catalogers at the Map Library, according to the terms of a tentative agreement arranged between CHIVE and

SYSTEM FILES
Vocabulary Control Files
5.5.4.4.

the Map Library, will employ a modification of the ISC area code which appends the ML provincial codes to the ISC area code. This expanded code will permit ML to continue to index provinces and other political subdivisions where this degree of index specificity may not be required for document retrieval.

During the preparation of the map index transcript sheets at the Map Library, the map cataloger will either (a) enter the modified ISC area code in addition to the ML area code on the transcript form, or (b) enter the ML area code only. If the former procedure is followed, both codes will be converted to machine language but only the modified ISC area code will be stored in the CHIVE Master Index File. The map catalog cards returned to ML, however, will carry the ML area code since the ML card catalog employs this system and the use of another area code would upset the existing file arrangement. Alternatively, a conversion table may be built which would permit the computer to convert the ML area code appearing in the index records (option [b] above) to the modified ISC area code, thus obviating the need for both codes to be entered on the transcript forms by the cataloger.

SYSTEM FILES
Vocabulary Control Files
5.5.4.4.

The Master Location Dictionary records will contain the following elements of information:

- ISC 4-Digit Numeric Notation
- Major Area, Subordinate Geographic Region, Country, or other Political Subdivision (including cross references)
- Remarks (scope notes and comments on historical changes)
- Place Name (including cross references)
- Remarks (place name scope notes and comments on historical changes)
- Geographic Coordinates

Scheduled master and supplemental listings of the Master Location Dictionary will include:

(a) A listing arranged hierarchically by ISC area code and containing the major area and political subdivision names together with any "Remarks" pertaining thereto.

(b) A listing identical to (a) but ordered alphabetically on area and political subdivision name.

(c) A listing arranged hierarchically by ISC area code with the minor sort alphabetical by place name. This listing will also include the "Remarks" field pertaining to place names, as well as geographic coordinates.

(d) A listing identical to (c) but ordered on geographic coordinates.

Ad hoc (demand) products of the file will include a geo-coordinate computation capability. This program,

SYSTEM FILES
Vocabulary Control Files
5.5.4.4.

which uses a mathematical technique based on the overlap of two convex polygons, will allow an information analyst to retrieve all references to place names falling within any regular (or irregular) shaped area whose vertices are known.

5.5.4.5. Subject/Commodity Authority Files

5.5.4.5.1. Intelligence Subject Code (ISC)

A modified form of the ISC classified schedule will be used in CHIVE as the dictionary authority for entry of all terms of a descriptive, semi-abstract nature whether they modify named objects or stand alone. The file is designed to perform three main functions: (a) to display relationships among these descriptive terms, (b) to define these terms when required, and (c) to serve as a code book for input to the computer. The relationships displayed include synonyms and alternate spellings as well as class inclusion and class membership. The file also serves an important mechanical role by requiring that every ISC code in a new index or query be present in the file before the transaction of file maintenance or searching is processed, hence controlling input errors. The file will be maintained manually, that is, all relationships and original entries will be externally controlled with changes made only by a change sheet following approval by the ISC dictionary editor.

To increase the specificity of the ISC, it will be augmented by the addition of key words in clear text. Initially, the information analyst will be permitted to append these key words to any ISC code without reference to a controlled list. After a suitable length of time, however, a key word dictionary separate from the ISC classified schedule may be developed to provide guidance for consistent entry.

The ISC is generally recognized as a satisfactory mechanism for indexing intelligence documents to a medium level of specificity. It is detailed enough to organize a document collection into manageable categories, but not so detailed that it is difficult to learn or apply with reasonable uniformity. The ISC is particularly strong for indexing political and socio-economic concepts.

Key word indexing will be used to supplement the ISC in those areas where it is weakest, and to obtain more specificity in commodity indexing. Heavy emphasis will

SYSTEM FILES
Vocabulary Control Files
5.5.4.5.1.

be placed on key word indexing of equipment nomenclatures and model types. Key words will also be used to index scientific processes and concepts, as well as military strategy and tactics. In other fields, e.g., politics, there will be less need for key word enhancement of ISC codes.

Specific revisions required of the ISC to accommodate it to an all-source document base include the following:

- Reduction of the depth of subject coverage in selected areas to simplify its application.
- Development of separate schedules for the classification of such subjects as organization types and personality occupational categories. (This would require the deletion of organization types which are currently scattered throughout the ISC.)
- Expansion of the list of coded modifiers.
- Expansion of the ISC to provide for special subject requirements unique to certain sources, e.g., photos and SI documents.

The contents of a digital record in the ISC file will consist of:

- ISC 6-Digit Numeric Notation
- Clear-Text Term Definition
- Scope Notes

Outputs from the file will include:

- Master listings of the complete ISC dictionary arranged hierarchically by ISC code with appropriate indentations for each lower-level category. In addition to codes and term definitions, the listing will contain pertinent scope notes.
- Master listings of the subject index to the ISC arranged alphabetically by index term, including "see references."

Both of these types of listings must be classified as demand products since the frequency and number of changes to the ISC vocabulary will dictate the periodicity of master re-runs. Indeed, it is likely that this vocabulary control file more than any other will be updated by page inserts rather than by re-issuance of the complete master schedule.

5.5.4.5.2. Header Data Dictionary

This file encompasses a variety of separate and distinct system dictionaries which control the entry of data pertaining primarily to the header portion of document index records. It includes such specialized system tables or dictionaries as the following:

- Document Category File
- Report Producing Component File

- Series/Periodical Name File
- Classification File
- Codeword Control Stamps File
- Dissemination Controls File
- Photo Type File

None of these files are of such magnitude that their size alone or access requirements would justify their storage in a digital medium. Nevertheless, an important goal of the system is to present information to the external customer in a language with which he is familiar. This must be accomplished even though the information is carried within the system in a different form. For this reason, wherever a convention of codes has been established for a certain type of information, and this information must be displayed to a user on output, the file must be available in digital storage and a conversion routine provided to substitute clear text for codes on output.

The data content of records in all the separate authority files making up the Header Data Dictionary is identical, i.e., code (whether numeric, alphameric, or alpha) and applicable term. No machine-generated products

SYSTEM FILES
Vocabulary Control Files
5.5.4.5.2.

either on a scheduled or demand basis, are presently envisaged from the file as such, although, as indicated above, the file will be machine searched to serve other system functions.

5.5.5. UNSYNTHESIZED INFORMATION FILES (UIF)

No attempt will be made in this section to specify the particular Unsynthesized Information Files which will be built by information analysts in the CHIVE system. It is assumed there will be a continuing requirement for certain of the analogous information files currently being maintained (e.g., BR's International Conference and Travel File), but this is a decision best left to the information analysts within the system, working in concert with their external customers. This section merely outlines the characteristics of Unsynthesized Information Files (UIF), explains the rationale underlying their establishment, and describes some of the methods by which such files will be constructed and maintained.

As indicated in the introduction to this chapter, Unsynthesized Information Files consist of select elements of information about a given subject whether it be a personality, an installation, or some class of activity or event. They are to be distinguished from Special Project Files (discussed in section 5.5.7.), some of which may be information rather than document reference type files, in that they reflect only the elements of information contained in CHIVE document index records. Similarly, they are distinguishable from Summary Information Files (see section 5.5.6.) which are evaluated, concise statements of fact about similar topics.

Most Unsynthesized Information Files being maintained in OCR today are the products of specialized input activity which is separate and distinct from other input processing. The principal reason for this situation is that the regular processing system (or systems) cannot readily be modified to accommodate these specialized indexing requirements, primarily because of the limitations of the supporting EAM equipment. A good example is the BR travel index which, since its inception, has been functionally and physically separate from the dossier processing system.

In the CHIVE concept of document processing, however, wherein all the data of significance in the document is captured in one pass, so to speak, by the person originally assigned to index the document, the resultant product will feed both the document reference system (i.e., the Master Index File) as well as such Unsynthesized Information Files as the information analyst has decided to build. This means, obviously, that the UIF contain data no different from that stored in the Master Index, only select subsets of the same records or phrases.

The reader might well ask at this point why have Unsynthesized Information Files at all if their content is identical with elements of information stored in the Master Index File? Why not simply query the Master Index when a specific set of data is desired?

This brings us to the criteria for establishment of a UIF:

- The customer's information requirements must be capable of definition in terms of logical data units which have specified characteristics--i.e., that there is a logical separation of data elements into related files so that any one file contains data relative to a given subject or function.

- A sufficient number of requests can be anticipated on a continuing basis for the particular set of data elements contained in an information file to justify establishment of the file.

SYSTEM FILES
Unsynthesized Information Files
5.5.5.

Where neither of the above conditions obtain, the data would remain in the Master Index File, and requests for the retrieval of specific elements of information would be handled like any other ad hoc queries levied on the system. On the other hand, if these requirements are met, it is generally agreed that it is useful to group the data elements involved into files, organized on a functional basis, since they can then be handled as logical elements in the system for maintenance, retrieval, and system output.

The basis of organization is affected not only by the type of information to be processed, but also by the relative activity of data within a file and the user's control of the information stored in the file. Thus, in establishing an information file system, the user will probably want to functionally group his data (e.g., personality travel, leader appearances, missile site order-of-battle, directories of government officials, etc.).

For particular applications the user may desire to have his stored information combined on a different basis.

Ordinarily, he would exercise this option only on the Summary Information Files discussed in the next section, since the resultant output would be an evaluated, higher-quality product. However, the system proposed will provide for a multi-file output capability from any file stored within the system. This capability will be achieved by allowing the user to query several files and selectively assemble on an output work tape the resultant information. The data will then be presented to the user in the format he specifies.

For example, if the user has three files containing [redacted] 25X1B

With regard to the means by which inputs to the UIF are to be obtained, it is evident that a separate indexing and transcription process is not required if the CHIVE one-time indexing concept is implemented. In other words, the plan is to exploit the document retrieval system in order to build information files. On the other hand, once the data has been put into machine readable form, the information file inputs might be derived either by (a) automatic duplication of portions of the index records during their input processing into the Master Index, or (b) by periodically querying the Master Index. Figure 5-2 illustrates these alternatives graphically.

CHIVE proposes to follow the latter path for the following reasons:

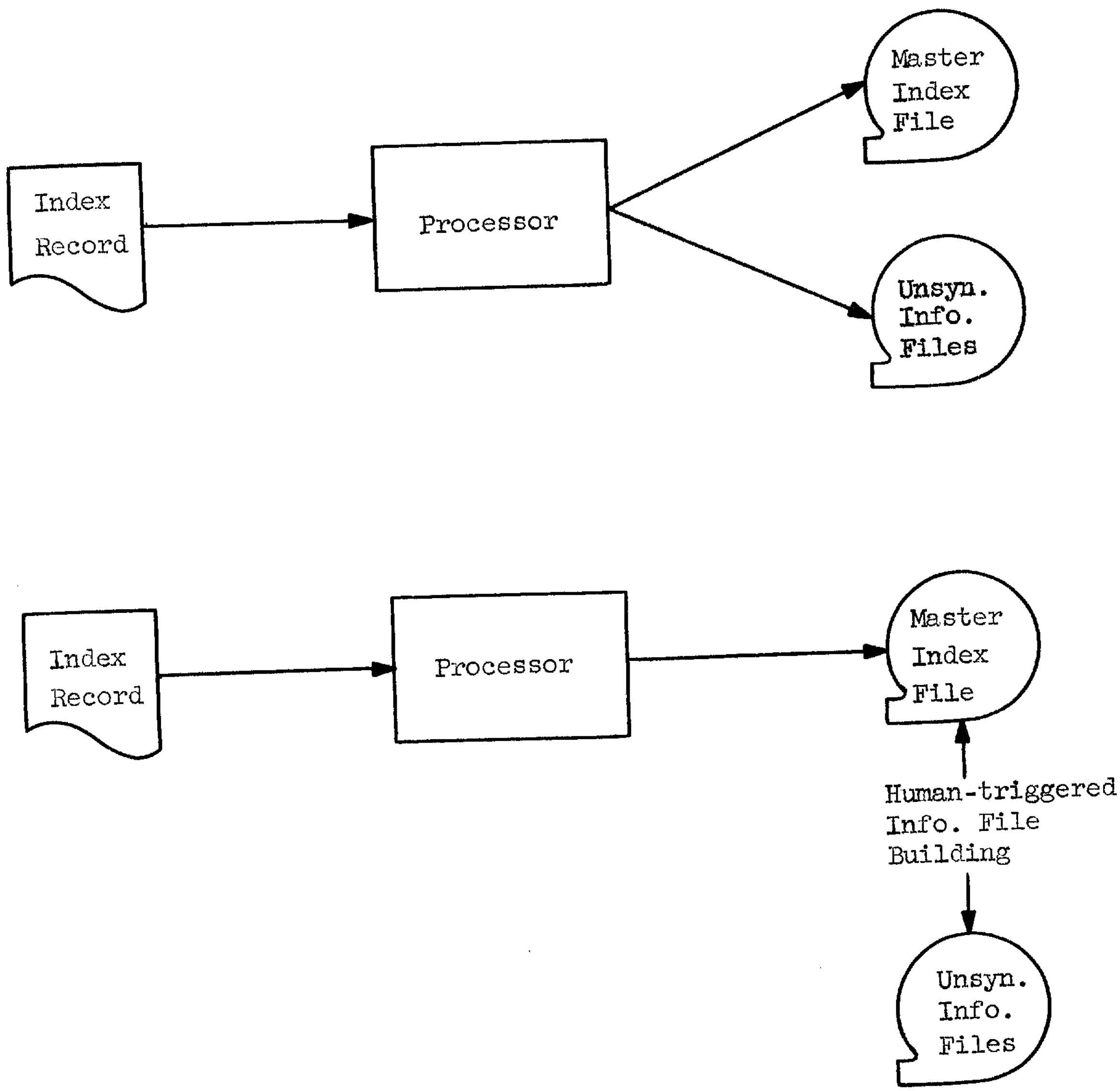
- It will create less of a burden on the machine processor which would otherwise have to examine every incoming record to determine if it contained data relevant to a particular information file.
- There is no real requirement to update the information files at the instant that the data is entered into the machine.
- By requiring some external action to be taken before data is transferred to an information file, management control is enhanced.

To facilitate file building, standing queries will be written, punched, and entered into the system. Thus, when the information analyst wishes to add new data to an information file, he can merely call for the pertinent query by name and the computer will make the necessary search and load the data into the relevant file.

The content of records in a UIF may consist only of a specified set of fixed elements of information which

Figure 5-2

UIF FILE BUILDING ALTERNATIVES



appear once in each data record or a combination of fixed and repetitive elements of information including some fields of variable length. For example, in a file

25X1B

[redacted]

elements in a record contained in this file might have a fixed number of characters or, alternatively, some may be fixed while others (e.g., "function attended") may be fields of variable length. Similarly, in the case of a

25X1B

travel file [redacted]

25X1B

Document references, while not vital to information files which ordinarily do not require consultation of the documents from which the data was originally extracted, can nevertheless be included in UIF records where required by citing the pertinent document control number. Similarly, the security classification of a record in such files is easily denoted since the record should typically carry the same classification as the parent record in the Master Index. In both cases, the transferral of the document control number and security classification from the header portion of a Master Index Record to a record in a UIF can be accomplished automatically at the same time that the content data is duplicated for storage in a UIF record.

Provision will be made to automatically identify those records in the Master Document Index whose content has been extracted for a given information file so that the same records need not be searched later. The system will also allow the analyst to specify both an "active" and a "history" file of information for any one functional area if it seems desirable to save the digital records representing a file for some indefinite period of time.

Outputs from the UIF will largely consist of periodic listings of the complete contents of a UIF arranged in various sequences. Such listings will be used to service customers as well as information analysts within CHIVE who will analyze and modify the listed records which will then be converted back to machine language and input to

Summary Information Files. Some of these listings will no doubt be required on a regularly scheduled basis, e.g., a leader appearance listing published weekly. Others will be demand products issued irregularly as the need arises.

In addition to the provision of hard-copy machine listings of an entire file for browsing purposes, a digital UIF file itself can be queried for records meeting the request specifications, in which case only the records satisfying the request will be output. This mode of man/file interface will probably be used less than hard-copy browsing; but, unlike the CHIVE Vocabulary Control Files discussed in section 5.5.4., the means by which the UIF records will be made available to the human in the system will not be exclusively through a hard-copy representation of the file contents.

5.5.6. SUMMARY INFORMATION FILES (SIF)

Summary Information Files (SIF), like the Unsynthesized Information Files, can be classed as formatted in nature since their specifications can be pre-defined and the data elements making up their content can be handled as logical entities in the system for purposes of input, query, and output processing. Pertinent input data is organized by major subject and formatted for ready retrieval and tabulation by content.

Summary Information Files consist of semi-evaluated data relative to specific classes of events or named objects. In format and content they are indistinguishable from UIF files, differing only in the fact that redundant, and, usually, contradictory information has been removed from the SIF files through a process of human analysis and synthesis of the raw data originally received. While they cannot be accurately described as containing only "finished intelligence" (if, by definition, this term is meant to apply only to the refined outputs of an intelligence research facility), neither is their content truly "raw" and, for this reason, the expression "semi-evaluated" has been used advisedly.

Certain of the Vocabulary Control Files which must exceed the boundaries of a typical dictionary in order to adequately "identify" a controlled term can also be properly classified (as pointed out in the introduction to this chapter) as Summary Information Files. These include the personality and organization/facility identifier files discussed in sections 5.5.4.1. and 5.5.4.2., respectively. Not only are these formatted files whose content is fixed and specific, they contain a variety of summarized, semi-evaluated facts about named objects which, when displayed, can serve a variety of information needs other than that of supporting the document indexer.

An example of an SIF would be a [redacted] 25X6 Officials File. In reality, a type of organization summary file, since the stored data would consist of a set of facts about significant public and private organs of society within the country concerned and not summary data about personalities as such, the file entries might contain in tabular form information on the name of an organization, its subordination, the names of its officers and perhaps ordinary members of the organization, their individual position titles, and the dates of appointment and/or earliest and latest dates of identification for each. The data would be retrievable on the basis of any of the categories which make up the file so that answers to such

25X1B questions as: [redacted]

25X1B

SYSTEM FILES
Summary Information Files
5.5.6.

25X1B

25X1B [redacted] can be readily obtained.

Summary Information Files can, of course, include fully automatic, semi-automatic, and manual data files ranging from the highly structured to the unformatted, narrative type. FIB's Installation Summaries File and the various types of biographic reports on file in BR are examples of manual summary information files which are only partially formatted if at all. BR's Who's Who Card File is a formatted, semi-automated summary file on personalities. The CHIVE system will also produce and maintain manual files of summary information, but the focus of discussion in this section is on the digital, and not hard-copy, summary files planned for the system.

The criteria dictating the establishment and maintenance of an SIF file are identical to those for a UIF file, i.e., the set of data elements making up the file must be definable and a "substantial" number of requests for such data must be anticipated. The means by which input and maintenance transactions will be obtained, however, will differ from that of the UIF files.

Unlike UIF processing, additions to and modifications of SIF files cannot be automatically generated since human judgment is required, if not to identify potentially relevant inputs, to make the conclusive determination that a record should be added to a file or a change made to an existing record.

The process by which these functions will be performed will vary depending on the nature of a particular SIF file. Figure 5-3 depicts the various approaches which can be used.

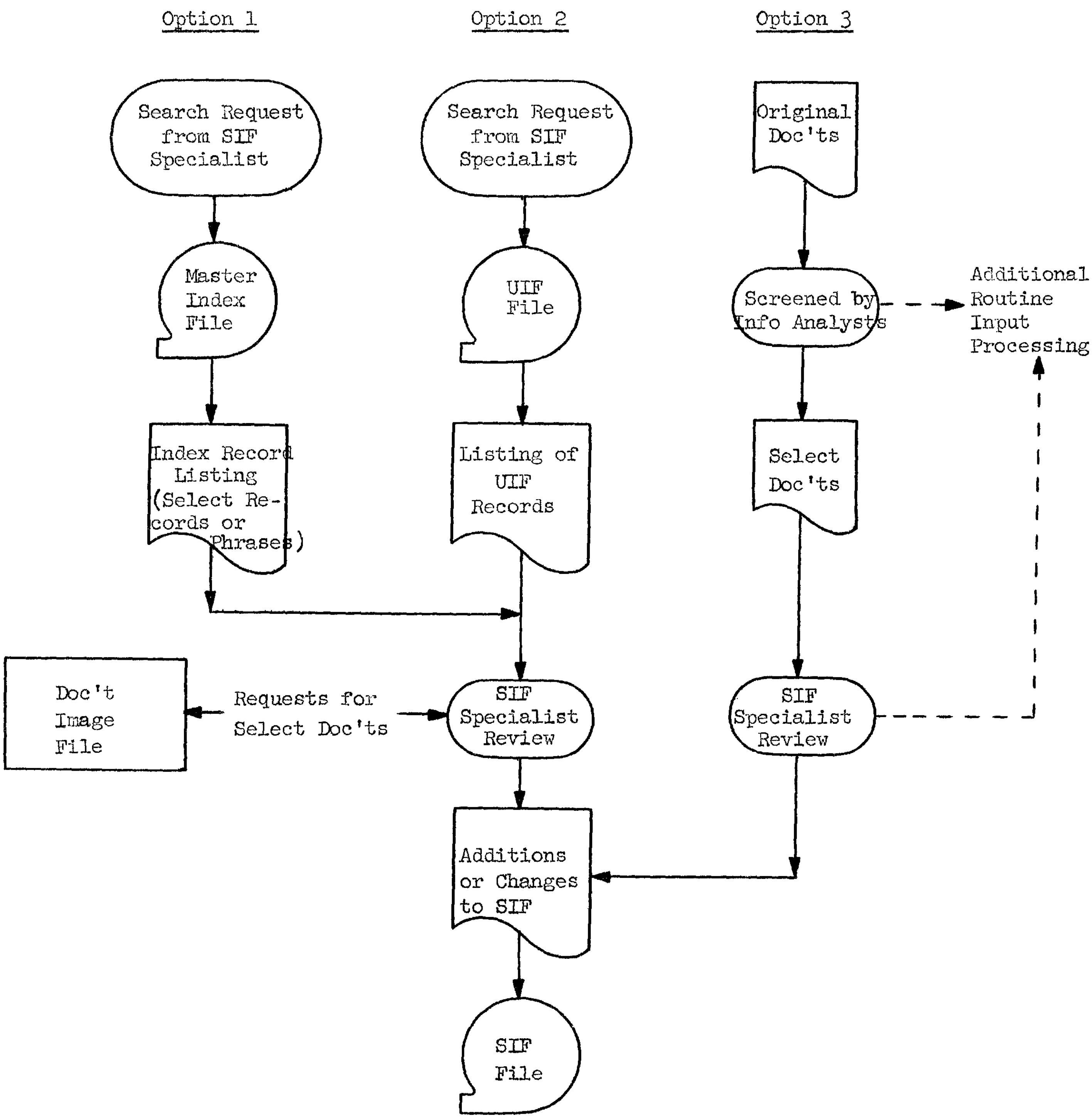
As indicated in the figure, the SIF file builder has essentially three options open to him as the means of inputting data to a Summary Information File: (a) he may query the basic index records to documents for data pertinent to an established SIF file (Option 1), (b) he may obtain a printout of a UIF file which serves as the raw data base for an associated SIF file (Option 2), or he may arrange with other information analysts to have all documents containing information pertinent to his needs routed to him (Option 3).*

---

*A fourth option, which has not been suggested because it would lead to completely duplicative document handling, would require the SIF specialist to examine all incoming documents for their possible relevance to a summary data file. Such an approach would never be required as long as the initial document indexing was sufficiently specific to enable the information file specialist to recover the pertinent index records and/or documents by a search of the Master Index.

Figure 5-3

SIF FILE BUILDING ALTERNATIVES



Options 1 and 2 both provide for the possibility that the SIF specialist may wish to examine certain documents for himself to clarify some fact reported in the records listed for him as a result of his inquiry. It is anticipated, however, that in most instances the listed records will speak for themselves and no reference to documents will be required in the SIF input process.

SIF files will be available in both digital and hard-copy (listing) form, and, like the UIF, can include active as well as history files. Source citations in the form of document control numbers may be listed at the end of each summary information record or referenced to each term (element of information) in the record.

5.5.7. SPECIAL PROJECT FILES

In the remarks made earlier in this chapter with regard to Special Project Files, it was noted that the limits of CHIVE responsibility for special file building, maintenance, and output processing have not as yet been satisfactorily determined. Furthermore, in a survey of existing OCR files which were either exclusively or in part [redacted] (see Appendix 5.D.), no files were identified which in the CHIVE context would be considered

25X6

SYSTEM FILES
Special Project Files
5.5.7.

as "special project" in nature. For both of these reasons, therefore, it is difficult to specify the characteristics of individual Special Project Files which might be included either in the initial or final CHIVE system. One must anticipate, however, that the need for such special files will be expressed by customers from time to time, and some remarks may be in order as to the features which distinguish these files from other system files and how requirements for such files might be accommodated.

Processing demands of a one-time nature which necessitate the input, manipulation, and retrieval of a peculiar set of data will not fall within the Special Project File category since these will be handled like any other request. However, if the system were asked to continue the activity on an indefinite basis, this method of handling would no longer suffice and a special project need would have been established.

Special projects will include any files obtained from organizations external to CHIVE which cannot be fully integrated with equivalent CHIVE files and which require machine handling. In this sense the term "special projects" could apply to certain EAM files inherited from OCR, as well as to files acquired from other agencies.

SYSTEM FILES
Special Project Files
5.5.7.

Special Project Files will also describe any customer input requirements which cannot be satisfactorily handled by the established system for representing the information content of documents. The data involved might be largely numeric in character or, if non-numeric, would require the extraction of items of information not planned for inclusion in the system and which, if accepted, would significantly add to total processing time.

It is well known that individual members and groups within the CIA customer population have a number of relatively unique and distinct information handling problems which cannot be met by a generalized information system attempting to serve only the common interests of the many. Examples of such requirements were uncovered during the earlier fact-finding survey of the DD/I. The following is but a partial list of some of the needs expressed:

25X1B [redacted]

SYSTEM FILES
Special Project Files
5.5.7.

25X1B

[redacted]

In the past when a research analyst, or group of analysts, developed an information control problem which was not being satisfactorily handled either in the appropriate production office or by the central reference system, one of the following courses of action was generally adopted:

- The problem was contracted out to some external organization which obtained the necessary source documents and did the input processing and retrieval required to respond to the identified need. Examples of this approach include OSI's Project [redacted] and LSD/SI's [redacted]
25X1A
25X1A

- A special group was set up within one of the research offices to perform either or both the input and data manipulation functions, depending on what was required. The [redacted] of ORR is a good illustration of this type of problem solution. 25X1A

- OCR was asked to expand its operations either by increasing the depth of its indexing or by broadening its document coverage, or both. OCR projects generated by a demand for increasing indexing depth beyond that normally provided by the basic indexing

SYSTEM FILES
Special Project Files
5.5.7.

systems employed include the SR Test Range Activity File and the BR Travel File. Illustrations of projects which required expansion of the OCR document base can be found in the [redacted], the Library's Science Information Service (SIS), the SR PI Reports File, the [redacted] etc. 25X1A

25X1A

- Mechanical, if not data extraction, assistance was requested either of OCR's Machine Division or of some external machine facility (governmental or private) where manual techniques for data manipulation and display were unsatisfactory. Examples [redacted] File for OBI.

25X1B

Today, the resources available to an analyst faced with an unresolved information processing requirement are much the same--but with one significant difference. If he is willing to prepare the input data of interest to him to the point at least where it can be transcribed into machine-recognizable form, he now has a powerful machine capability available to him to perform a host of operations on the data.

The CHIVE EDP System can, of course, increase this capability. The question, however, which remains is whether CHIVE should get involved in "special-project" applications at all, where only a limited set of customer interests are served, and, if so, whether its involvement should be restricted to the provision of EDP support or whether it should also assist in input preparation.

SYSTEM FILES
Special Project Files
5.5.7.

Currently, the OCS/Applications Division is performing a major role in supporting special project requirements of CIA research analysts. In all such projects, however, the data extraction responsibility has been assumed by the customer concerned.

It might be argued that to avoid confusion of responsibility, the role CHIVE should play in this area (if any) should be restricted to the assumption of the responsibility for those special projects where, for one reason or another, it seems most efficient to have the central reference organization, rather than research analysts, prepare the input data. This, however, heightens the risk of gradually proliferating the information processing responsibilities of the CHIVE system to the point where it might become simply a collection of special projects.

In the design concept presented here, the responsibility for CHIVE's undertaking certain special projects has been accepted, but the duty of preparing the input to Special Project Files is assumed to be the customer's and the report reflects this philosophy. In the final analysis, however, the matter can only be resolved by management decision. If the choice is to include special projects within CHIVE, including the function of data preparation, the approach suggested above may provide a modus vivendi for relating the respective roles to be played by CHIVE and OCS/Applications in the handling of these projects.

5.5.8. REFERRAL SERVICE FILES

Current manpower ceilings seriously limit the coverage of present central system operations. Even if available manpower can somehow be more effectively utilized, the volume of material of potential value is so great that complete coverage would still not be possible. Thus, the only alternative appears to be to develop support from other systems, including centralized as well as personalized (analyst-driven) file activities. To do this, CHIVE must identify information resources available in such systems and determine how best to tap these resources for the Agency consumer.

In earlier CHIVE documentation, reference was made to a "support mode" which envisaged not only the referral of customers to persons or files of possible interest external to CHIVE, but the actual acquisition of certain machine-language files and supporting documentation which would be searched in-house in behalf of CHIVE customers. The "support mode" concept remains valid in present CHIVE thinking, but in the further refinement of the design a distinction has been made between: (a) files outside of CHIVE's control which are actually available within the system in either manual or mechanized form, and (b) information resources not directly accessible to CHIVE to which customers may be referred. The former have been classified in this report as "Supplemental Files" (see section 5.5.1.), i.e., files neither built by CHIVE nor inherited from OCR, of which Special Project Files may be one class. The latter are now termed "Referral Service Files" and are the subject of this section.

The community's information resources are so vast and scattered that even the simple identification of all potential sources constitutes a major problem. For this reason, it is planned to concentrate initially on the identification of the many different human file resources scattered amongst the various service components and production shops within this Agency. Only after this is done will any attempt be made to obtain descriptions of information resources and repositories in other USIB components.

Perhaps the simplest means of beginning to build a referral service capability will be to derive a set of analyst profiles from requests levied against the central system. Using this technique, search terms chosen for query purposes will gradually form the set of subject identifiers descriptive of each analyst's interests. This approach will be supplemented by the circulation of questionnaires to analysts throughout the research (and select service) components of the Agency, which would solicit narrative statements of their areas of substantive knowledgeability, including their personal files or files maintained by their respective sections, branches, or other organizational component. Not all analysts, of course, can be expected to respond. However, experience with similar surveys in other organizations suggests that a response figure of about 80% is not beyond reason.

The returned questionnaires will be indexed in the vocabulary of the CHIVE system, including both the ISC classified schedule as well as key words. These descriptors

SYSTEM FILES
Referral Service Files
5.5.8.

will not necessarily be limited to conceptual-type subjects (although the emphasis will probably be on these kinds of topics) but will, in all probability, also contain on occasion named-object identifiers such as the names of persons, organizations, military installations, etc. It is not anticipated that the responses will include every specific subject heading in an analyst file. However, it is hoped that the principal categories of information contained in such files will be described and this alone would greatly assist analysts in seeking to exploit the Agency's human and documentary resources.

In addition to storing information descriptive of the subject matter in which a person or file specializes, the data which will be contained in these referral service records will include:

- Name of Individual
- Organization Identification (component to which analyst or file is attached)
- Address of Individual or File (room and phone number)
- Descriptive Title of File
- Overall Security Classification of File

SYSTEM FILES
Referral Service Files
5.5.8.

- Releasability
- Countries or Geographical Areas Covered by the Person or File
- Primary Intelligence Activity Supported by the Person or File (e.g., Missile OB, Ground Photography, [redacted] Intelligence, etc.) 25X1B
- File Storage Medium (documents, 5" x 8" cards, EAM cards, magnetic tape, etc.)

Assuming the cooperation of a reasonable number of analysts, it is likely that the collected records will be sufficiently voluminous and file order requirements so varied that a machine data base will be needed. It is not contemplated, however, that the Referral Service Files will be automatically searched at the time queries are levied against the substantive data files of the CHIVE system. Rather the content of such files will be made available in the form of a published Directory of Information Resources. This Directory would be issued to CHIVE system operators and perhaps, in a variety of classifications, selectively disseminated to Agency consumers.

When specifically requested to do so by a customer, CHIVE personnel, in addition to searching the basic files of the CHIVE system, will consult the Directory for the

SYSTEM FILES
Referral Service Files
5.5.8.

purpose of determining what other files or intelligence analysts might possess information pertinent to a given query. The type of service that will be provided, if and when a potentially relevant resource is uncovered, will vary depending on such factors as the location of the file or the urgency of the request. In some instances, CHIVE information analysts will act as the intermediary between the customer and the other information resource. In other cases, they will simply refer him to the appropriate system.

With regard to the provision of a referral service capability for files outside of CIA, advantage might be taken of efforts currently being sponsored both by DIA and by CODIB to collect descriptions of intelligence data files maintained in an automated form by Department of Defense elements and USIB member agencies, respectively. It is planned that a catalog of such files will be published periodically and may be interrogated on an _ad hoc_ basis. If these external collection programs prove successful, the data resulting therefrom might be merged with the product of the internal file survey to form a relatively comprehensive record of information resources throughout the Community.

SYSTEM FILES
Referral Service Files
5.5.8.

5.5.9. MANAGEMENT DATA FILE

The CHIVE Management Data File will contain two types of data:

- Data, obtained by computer methods, about the processes performed by the EDP portion of the CHIVE system.
- Data, obtained by manual methods, about the non-computer processes of the CHIVE system.

The following paragraphs discuss the sources and collection methodologies for these two types of data, the reasons for the dichotomy, and the use of this data to operational management.

5.5.9.1. Collection Techniques

As indicated above, the method employed in collecting the data (EDP data or manual data) determines the origin of the data and to a large extent the use of the data by CHIVE managers.

EDP data collection refers to an activity within the computer itself. The monitor program system (with its attendant bookkeeping functions) will supervise all computer operations. This is implied under the philosophy of a multiprogramming system. This method of operating provides a natural means of recording:

(a) Process times. The computer has timing mechanisms which the monitor can use to record computation, input, and output times as individual entities, as well as the total time the computer uses to process a transaction.

(b) Error rates and types. A variety of errors and malfunctions may abort an operation or degrade the output. Certain of these, e.g., misuse of the language, transcription errors, equipment disorders, and illegal file manipulations, may be more readily detected and recorded by the computer programs than by manual means.

(c) File activity. It is of significant importance to determine which files or parts of files experience a high rate of use. File system design, program system design, and language structure are just a few of the areas which affect the use of the files and are, in turn, influenced by usage statistics.

This dynamic data may be supplemented by such relatively static data as:

- Equipment availability
- Day, month, and year
- Priority of the transaction

As this data is recorded (either by the bookkeeping routines within the monitor or by specially produced CHIVE programs as adjuncts to the monitor) it should be entered into a file. This file is essentially a log of CHIVE EDP transactions and their associated management data. The normal method of labelling entries in such a file is by "job" or transaction number. Under the multiprogramming mode of computer operations, it is mandatory that each "job" which enters the computer be uniquely identified. This "job" or transaction number provides a natural storage and retrieval device.

Manual data collection refers to that activity outside the domain of the computer which collects management data about the processing of transactions. As presently envisioned, the process will begin when a transaction is initiated and will end when the transaction is completed. For example, a query against a file is a transaction which begins with the request and ends when the requester obtains the data and materials which satisfy his request. Between these two events many functions are performed in many organizational elements. The majority of the time-consuming and error-prone functions are performed by people. Data regarding these functions may be conveniently collected by manual techniques. It is suggested that data regarding each transaction accompany the transaction during the entire process. If feasible, a standard form should be used. Examples of manual data are as follows:

SYSTEM FILES
Management Data File
5.5.9.1.

- Name of Requester
- Requester's Organization
- Name of Analyst
- Analyst's Organization
- Type of Transaction
- Transaction Number
- Dissemination Code
- Time Received and Time Released (by each organizational unit which handles or is responsible for the transaction)
- Organizational Identifier (for each component which handles or is responsible for the transaction)

Not all of these are applicable to each transaction. However, the last two items--times and organizations--must be supplied for each component and each transaction for two reasons:

(a) To account for each transaction and its location in the system.

(b) To provide a complete file of data for process evaluation.

5.5.9.2. <u>Storage</u>, <u>Retrieval</u>, <u>and</u> <u>Processing</u>

The EDP data, due to the collection method, is naturally stored as a file of data within the CHIVE EDP system. As such, it can be processed and retrieved through

SYSTEM FILES
Management Data File
5.5.9.2.

the use of the CHIVE query language. It is suggested that initially no language capabilities be added for this specific purpose since the number and nature of reports on machine processes which CHIVE management will require is not completely predictable.

The manual data collected in the system should initially be stored, retrieved, and processed by manual methods. As the system is used and evaluated, the file of manual data and the number of management reports will increase. At some point, this data must be processed by the EDP part of the CHIVE system. For this reason, it is important to design the manual data forms so that, as volume increases and operational procedures become firm, the data may readily be input to the computer and integrated into the EDP management data file. When this point in system evolution is reached, all manually collected data regarding CHIVE operations will be retrieved and stored by initiating a transaction. Thus, data about the processing of the Management Data File is recorded in the Management Data File and constitutes a resource which management may use to study its own evaluative and analytic activities.

SYSTEM FILES
Management Data File
5.5.9.2.

5.5.9.3. Reports and Their Use

In discussing reports and their content, a distinction should be made as to when, during the evolution of the system, the reports are needed. This is particularly true in the case of those reports drawn from the EDP management data file prior to the incorporation of the manual data.

The purpose of reports based on data collected by EDP methods is to assist the CHIVE analysts and designers in improving, correcting, and modifying the EDP portion of the system. During the initial stages of operational testing it will be necessary to examine EDP operations carefully in order to eliminate bottlenecks and optimize equipment usage. Certain reports will be highly specialized, e.g., an analysis of disk storage use over some period of time, and will not be necessary as a regular product. Of continuing interest will be reports which provide management with an insight into the amount of time used on the computer and its various components. This has long-range implications regarding computer hardware acquisition.

Reports derived from the manually collected data will vary in frequency and detail as the system gains operational acceptance. In any new system, there will be imbalances which must be adjusted if the best results are to be obtained from the available personnel and equipment. The parameters which can be measured within the system are primarily concerned with rates and volumes. It is suggested that forms be designed and procedures instituted which will provide managers with raw data on how long a transaction stays in each component. This is the first step toward the elimination of delay points in the system. Shifting of manpower and new procedures will undoubtedly be necessary. This in turn will prompt another round of reports and analysis. And so on. Of interest to management in terms of long-range changes to the system will be reports on sources and types of transactions. Such reports are generated by the present system and will be produced by CHIVE. Data on the number of cards produced, number of file accessions, number of references generated, and number of pages delivered will also provide managers with the necessary background for making adjustments in the processing of transactions.

After all management data has been combined in the file maintained by the EDP system, reports can be

SYSTEM FILES
Management Data File
5.5.9.3.

generated on a regular or demand basis with much less expenditure of manpower. The nature of the reports will probably vary little after the shakedown period is completed. However, the volume of data which must be manipulated dictates an EDP mode of report generation.

SYSTEM FILES
Management Data File
5.5.9.3.

Chapter 5.6.

SYSTEM FLOWS AND TRANSACTIONS

This chapter provides a more detailed view of system flows and transactions, i.e., the more dynamic aspects of the data processing activity, including some descriptions of illustrative tasks. The document image storage and delivery portion of the system is covered in outline only, leaving the more definitive treatment of this subject to Volume VI. Similarly, only passing mention is made of the EDP design since it is fully discussed in Volume VII.

5.6.1. DOCUMENT INPUT

Referring to Figure 5-4, the input to the system will be described.

The principal categories of incoming documents will consist of (a) textual-type documents received in all source classifications ranging from Unclassified to T/KH, (b) select documents (principally SI Teletype)

Approved For Release 2000/05/30 : CIA-RDP78-03952A000100050001-7

received in machine language (as well as hard copy), (c) graphic images in the form of ground photography and films, (d) maps, and (e) machine language (ML) index records prepared by external organizations according to CHIVE rules and formats. Graphics and maps will continue to flow to GR and the Map Library Division (ML) through their existing acquisition channels. The only significant change in their operations will be that they will employ the CHIVE vocabulary in their indexing or cataloguing operations, and will transmit a copy of their index transcript sheets to CHIVE for conversion into machine readable form and entry into the Master Index File. CHIVE in turn will return to them a printed version of their index records for entry into their manual files where this seems desirable.

Documents selected by the information analyst which are available in machine language and have a formatted header and title (e.g., SI Teletype) will bypass indexing and transcription steps and go, in their machine language versions, directly to the EDP

System where the necessary conversion to CHIVE format will be performed. The hard copy versions of the documents will be sent simultaneously to microfilming for processing into the microimage store (Master Image File) Other machine language receipts, consisting of abstracts of foreign scientific and technical literature, bibliographic records, and formatted information extracts pertaining to named-object data appearing in open sources, may likewise be input directly to the EDP System. Printed versions of these receipts, however, may be passed to information analysts within the system who will thereby be afforded the opportunity to review their content, and, if desired, delete the corresponding machine record from the EDP file. Since the source documents will not accompany these ML inputs, no photoprocessing will be required.

The remainder of this section will deal with the principal input flow process depicted in Figure 5-4, i.e., that relating to all-source textual documents.

Upon their receipt in the mail room, these documents will be counted, batched by type, and assigned document control numbers where required. The batches will then be forwarded to a dessemination unit where the documents will be disseminated to other offices as well as to CHIVE. Documents to be distributed to CHIVE will be divided into two categories; (a) reports for which CHIVE has a repository responsibility, and, therefore, must be kept regardless of substantive content (hereafter referred to as "R" documents); and (b) non-repository ("NR") documents whose retention value can only be determined after examination by an experienced intelligence information analyst.

"R" documents (constituting the vast majority of incoming receipts) will be addressed to the appropriate CHIVE subcomponent, but will flow initially to a centralized Header Indexing Group which will index the bibliographic data on the documents. Once this operation is completed, the documents will be transmitted directly to the Document Delivery System for image processing into the Document Image File, while the header index would be sent to the EDP System for conversion to machine language. The "R" documents which, in all probability, would be the ones most often requested by Agency customers in the period immediately following their receipt, will (by this process) find their way quickly into the document store where they will be available for retrieval. Following image processing, they will be forwarded to the CHIVE analytical desks marked on the documents for content review and indexing where warranted.

"NR" documents will bypass the centralized Header Indexing Group, being forwarded by the Dissemination Unit directly to the analytical components within the CHIVE geographic divisions. Here a further redistribution of some of the "R" as well as "NR" documents might take place if the initial dissemination was not sufficiently precise. In any event, the ultimate recipient of both types of documents will be an information analyst specializing in an area or topic within area. His responsibility, relative to the "R" documents, will be to determine whether content indexing is warranted in addition to the header indexing already performed. If not, he will destroy the documents and send a notice to the EDP System that no content index will be forthcoming. If the documents, however, do warrant content indexing, he will mark the parts of the documents which he wants reflected in the index, and will pass the marked documents to a Content Indexing Group serving his Division. (The activity of these individuals is described below.)

"NR" documents will likewise be examined by information analysts and will either be destroyed or marked for some form of indexing. If indexing is required, the documents will be sent first to header indexing clerks functioning at the division or desk level. They will prepare header transcript sheets, like their counterparts in the centralized Header Indexing Group. Where content indexing is not required but storage is

SYSTEM FLOWS
Document Input
5.6.1.

desired, the "NR" documents will be sent to the Document Delivery System for microfilming, while their corresponding header transcript sheets will be passed to the EDP System. The remaining "NR" documents (and transcript sheets) which were to be content indexed will be forwarded to the Division's Content Indexing Group where they will rejoin the select "R" documents discussed above.

In the Content Indexing Group, semi-professionals known as content indexers will prepare content data transcript sheets by extracting and formatting the data identified for them by the information analysts. A selected portion of this work will be inspected and revised if necessary. Corrections and changes will be written on the data sheets.

Once the content data transcript sheets have been prepared, the marked-up copies of the "R" documents can be destroyed since an image of these will already be available in the Document Delivery System. The indexed "NR" documents, however, will be forwarded to the Document Delivery System for processing into the Master Image File.

Content data transcript sheets for both "R" and "NR" documents will be sent to a Data Transcription Group where they will be copied by typists.* The typed index entries, after sight verification, will then be fed to the EDP System for machine processing.

Within the EDP Subsystem, a Page Reader will convert the clear-text header and content indexes into machine language. Following this operation, punched Work Cards will be generated by the computer from a portion of the header data record which will be used in the Document Delivery System (see below) in the preparation of the microimage store. The complete digitalized records of the header and content indexes will be processed by computer programs which will check the records for format and certain types of content errors and add them to the pertinent system files.

---

*Header data sheets can presumably be typed by the header indexers who prepared them.

In the Document Delivery System, documents to be kept in hard copy for reasons of length, image quality, or other will be shelf-filed in an area contiguous to the microimage file according to their meaningful document control numbers. The remaining documents will be routed to a microfilm section. There they will be photographed, and, assuming the storage medium selected is the 35 mm. aperture card, the resultant product will be an aperture card with the document batch and serial numbers eye-visible in the aperture. After these numbers are punched into the aperture cards, the aperture cards will be mechanically collated on these numbers with the deck of Work Cards prepared by the computer from the header data records to the same documents. Following collation, other data punched in the Work Cards will be reproduced and interpreted for the Vital Materials Repository (VMR) and NSA as appropriate. Lastly, a master set of the cards will be filed in document control number sequence in the Master Image File.

5.6.2. DOCUMENT RETRIEVAL

Referring now to Figure 5-5, the recovery of information from the files will be discussed.

The retrieval process will ordinarily begin with a customer external to CHIVE originating a request for data either on a form designed for this purpose, by telephone contact, or by personal visit to the system. He will be put in touch with an information analyst working on the geographic/topical area of concern. The information analyst will be familiar with the current reporting, having screened incoming documents to determine what should be indexed, and will also have had extensive training in the indexing vocabulary, the logical files available within the system, and the query language required to conduct the computer search.

After ascertaining the clearance level of the customer, the degree of sensitivity desired in the search, and the heterogeneity of the document base to be explored (e.g., "search document and photo

25X1B

Approved For Release 2000/05/30 : CIA-RDP78-03952A000100050001-7

indexes, but not maps or films"), the information analyst (assuming a machine search is required) will translate the request into a set of commands using the formal language developed by CHIVE (see section 7.A.). To prepare the necessary search criteria he will consult the various Vocabulary Control Files--e.g., MOFIF, ISC, etc.--in order to derive the proper terms on which the search should be conducted. This research might also reveal whether certain inherited files would be worth interrogating (see section 5.5.4.2.1.).

Having determined what descriptors to employ in the search, he will obtain a request number from a central control point and proceed to fill out an inter-leaved set of request forms on which he will identify himself (as well as his customer) by name and address, cite the file(s) to be interrogated, detail the logic and priority of the search, and define the output format required. One copy of his request statement will then be sent to the request control point to be added to the file of open requests. Assuming, however, that some inherited files must also be searched since the date span of the request encompassed the period prior to the initiation of the CHIVE system, the information analyst may be required to take one or more of the following additional steps:

a. Examine hard copy files of cards or documents co-located with his organization component.

b. Request the retrieval of hard copy records (e.g., MIRA, one-name cards, etc.) from the system's centrally-located, master document collection.

c. Consult with other information analysts familiar with the contents, vocabularies, and record formats of machine files inherited by CHIVE and obtain their assistance (where required) in preparing the special request forms to interrogate said files.

The formulated machine requests will be typed and sight verified, and then transmitted to the Page Reader via the pneumatic tube system. For those requests to be passed against the EDP files, the computer will check for such things as the completeness of the request statement and validation of the terms composing the query. All requests will then be queued for processing against the pertinent inherited and CHIVE-built files.*

Searches of unconverted EAM files will be conducted as at present, with the output taking the form of existing machine listings which cite documents, personality dossiers, installation numbers, or photo accession numbers relevant to the request. For files converted to EDP and the CHIVE-built Master Index File, the product of the search will also be a listing, albeit in a different form. On the first page(s) of the listing will appear the identity of the information analyst levying the request, the request itself, and the list of document control numbers which satisfied the search criteria. On succeeding pages will appear, depending on the output format requested, either the complete "hit" index records or select elements thereof. (Output of a statistical count of the number of documents which matched the search

---

*The periodicity of searches may differ between these files, i.e., inherited files may customarily be searched only once a day while the CHIVE-built files will be searched on a demand basis.

prescription, without the records themselves, is also possible if the information analyst so desires.) Codes appearing in the records would be translated into clear text for ease of understanding by the information analyst and customer (if the latter also reviews the listing directly).

The information analyst will study the various machine listings received to determine the relevance of the retrieved records to the search prescription, and, particularly in the case of inherited file outputs, will consult with other information analysts familiar with the contents and vocabularies of such files as required. In a certain percentage of cases the output records may, themselves, answer the request. If so, the retrieval activity will end with the information analyst transmitting the desired information by mail or phone to the customer. On the other hand, the response might have been such that he will wish to re-enter the request with improved criteria.

When the index record output is satisfactory but, in itself, does not supply the answer sought, the information analyst may order the pertinent documents from the Document Delivery System before transmitting the results of the search to his customer for review. If so, he will encircle the appropriate document numbers appearing on the first page of his listing and send this page to the Document Delivery System. Where inherited files, however, are involved he may be ordering personality or installation dossiers, as well as documents, and will, therefore, follow a slightly different procedure. Graphics and map index records uncovered during the initial search will be transmitted to the customer who will order these items for himself.

Dossiers, following their retrieval from the file, will be forwarded directly to the information analyst requesting same. A replica, rather than the file copy of all other documents, however, including those recovered from the existing Intellofax and SR collections as well as from the microimage and hard-copy files of CHIVE, will be prepared before being transmitted to the analyst.

The information analyst will review the output from the various document files, and, after removing those documents which do not appear to be pertinent, will transmit the response to the customer. Alternatively, the information analyst may be asked to respond to the inquiry by phone, memorandum, completion of a customer's response form, or by the preparation of a narrative report (e.g., a biographic summary). In the latter case, he would obviously have to supply information rather than documents, which might necessitate a more sophisticated analysis and synthesis of the materials at hand.

Lastly, the information analyst may update certain of his identifier records, as well as dossier files, to reflect the results of his analysis (see section 5.5.4.1.1.), or send a marked copy of his report (if it deserves retention) back through the input process for indexing and storage in the Master Image File. He will also return any master cards or dossiers to their appropriate files, and report the closing out of the request by completing his copy of the request form. The latter will be sent for processing into the Management Data Files.

5.6.3. INFORMATION FILE BUILDING, MAINTENANCE, AND RETRIEVAL

As has been pointed out, the CHIVE system, like the existing central reference operation, will require a variety of dictionaries and other support tools (given the general title of Vocabulary Control Files in this report). In addition, it will maintain substantive files of information either in unsynthesized or summary form. Since the procedures for building such files as well as retrieving data therefrom will differ substantially from the document indexing and recovery process, they are reviewed here separately. Moreover, these files, unlike the Master Index records, will require continual maintenance, i.e., the deletion of obsolete or useless data as well as the correction of, or addition of information to, existing records in the file. The Master Index File, on the other hand, will require little maintenance at the sub-record level as such--only the addition of new records to the file and the periodic retirement of segments of the file to a less accessible storage medium.

5.6.3.1. Vocabulary Control File Maintenance

Vocabulary Control Files (e.g., MOFIF, MLD, etc.) will be consulted by content indexers as well as header data indexers in order to select the approved term or code for representing a subject or named-object mentioned in a document.* These files, initially, will be represented in listing form although some alternative reference medium will be intestigated. If the indexer finds no suitable entry for the topic mentioned in the document, or if the entry is erroneous or incomplete, he will prepare a File Maintenance Transcript Sheet on which he will specify the changes to be made to the file in question,

*The maintenance of the personality identifier file (Master Dossier Index) is excepted from this discussion since, as the reader will recall from section 5.5.4.1.1., names will not be "identified" during the input process.

using a portion of the same command language employed in the retrieval of records from the Master Index File.

The File Maintenance Transcript Sheet will be passed to a dictionary editor who will be responsible for reviewing all changes made to this specific vocabulary control file. He will insure that the proposed transaction is legitimate and proper, and, after entering the proposed changes by hand in his master listing, will forward the transcript sheet to the Data Transcription Group for typing.

After the transcript sheet has been copied and any necessary corrections made, it will be processed in essentially the same manner as the Document Index Transcript Sheets, that is, the forms will be converted to machine language by the Page Reader and the resultant output fed to the EDP System for updating the pertinent machine files. A record of the changes made will then be printed out in the various arrangements required, and returned to the dictionary editor as well as all indexers using the particular vocabulary control file affected. The frequency of preparation of these printed supplements to master listings, as well as the frequency with which the master listings themselves will be rerun, will vary depending on the number of changes occurring over a given period of time. The initial period of CHIVE operation will permit time for some experimentation to arrive at the most satisfactory procedure.

5.6.3.2. <u>UIF and SIF Processing</u>

As indicated previously, formatted information files consisting of logical data units either in unsynthesized or summary form may be initiated either by: (a) analysts external to the CHIVE system having a pressing and continuing need for the retrieval of select facts (as distinct from documents) pertaining to a given subject or function; or (b) by CHIVE information analysts reacting to the accumulative effect of specific request patterns. Requirements of this nature, since they will increase both the human and machine burden, will be reviewed by managers at the branch or higher level to determine the anticipated load on the system and its capacity to respond to same.

Accepted requests for the establishment of UIF or SIF files will be assigned to one or more information analysts conversant in the subject matter involved, for initiation of the input as well as maintenance and retrieval processing. Assuming the data is to be stored in digital files, the information analyst responsible for the file will consult first with a specialist assigned to the EDP System known as an EDP File Analyst. The latter will be throughly familiar with the internal operations of the EDP System and, in particular, the method used to establish new digital files. His duties would be analagous to those of an individual in the Planning Staff of the Machine Division/OCR, i.e., he will design the format and record structure of the machine file required by the information analyst and see to it that the file is actually established.

In general, the approach of the area information analyst will be to use the document retrieval system

25X1B

to help build the required information files. If the file, however, is to have the characteristics of an Unsynthesized Information File (see section 5.5.5. above), the actual involvement of the information analyst in the input process may not be great since, presumably, the data requested is already reflected in the content of document index records (i.e., the UIF would be built directly from rearranged elements of index records).* Where this is indeed the case, the information analyst will periodically direct the computer to take such action by calling for the appropriate standing query and record generation job to be run.

SIF files, on the other hand, will require more activity on the part of the information analyst since they will consist of evaluated, summary records about named-objects or events. These can only be

---

*If the data is not already being captured, then the request must be classified as a "special project" which would require a procedure all its own.

generated (as suggested in section 5.5.6.) by the analysis of the output from a UIF file, from the Master Index File, or by the processing of the incoming documents themselves. Assuming the SIF is to be built from data in a UIF, the information analyst will review the listed product from a UIF, comparing it with a listing of any records already stored in the SIF. If he decides to make a change to the SIF either by adding new data, deleting what was there, or by replacing old information with new, he will prepare a File Maintenance Transcript Sheet (similar, if not identical, to that used to update vocabulary control files) on which he will describe the transactions to be performed. This form will follow the usual path to typing, thence to the Page Reader, and finally to the EDP System for computer processing.

The retrieval of data from either the SIF or UIF files might be initiated for a variety of reasons, the principal ones being as follows:

a. To provide a listing of changes to the master file in order to update the information analyst's printed version of the file.

b. To provide a listing of the complete master file either for reference use by the information analyst* or for periodic publication and distribution to interested customers.

c. To search, in response to a customer's request, for a specific fact or correlation of facts which could not be readily derived by human browsing of the printed records.

Whatever the reason for initiating a retrieval transaction the process will be virtually the same as that followed in the retrieval of document index records (using the same retrieval language), with the exception that no inherited files should be involved in the search and no documents will ordinarily need to be retrieved from the document image store. Schedules can, of course, be set up for the levying of standing queries which would cause the listing of all or a portion of a file on a periodic basis without any action being required on the part of the responsible information analyst.

---

*The listing will be the primary mechanism for analyst-SIF communication.

5.6.4. TASK TABLES FOR SYSTEM TRANSACTIONS

Examples of the step-by-step procedure by which some of the system transactions outlined above might be carried out using the equipment, file organization, program organization, and operator procedures described elsewhere in this report are provided below. Obviously, there are a variety of procedures that might be used to perform any of these tasks. What is suggested here must, therefore, be regarded as tentative and subject to modification as procedures are worked out in detail during Phase III.

With regard to the method of presentation, it should be pointed out that written descriptions of even the most routine human activities make difficult reading at best. And this is no less true of a data processing operation, especially when couched in the language of the systems analyst. Secondly, it is a fact that if flow charts were prepared of many current central-reference operations, the resultant products would also appear relatively complex. Yet, somehow, humans manage to carry out the operations involved.

Lastly, it should be recognized that some atypical problems are covered in the task tables which would not ordinarily be encountered in the average transaction. These, necessarily, further complicate the narrative discussion.

The tables which follow have four columns. The first column (STEP) contains the number of the operation. The number is used in the body of the table to reference deviations from the normal sequence of operations. The phrase, "go to step 10," will tell the reader that the next operation in the sequence is step 10. The second column (AGENT) identifies the person or equipment which is chiefly responsible for carrying out the operation. The third column (LOCATION) shows where most of the operation is carried out. The fourth column (OPERATION) has one or more sentences for each operation which describes what takes place in the operation. These are either processing operations, in which some action is taken on the data covered by the task table, or they are decision operations in which a question is asked and the consequences are given for the two or more possible answers. These consequences are usually in the form of "go to" statements. The statement, "STOP," is the last statement in the OPERATION column for a particular task and indicates that the task is completed.

Next 3 Page(s) In Document Exempt

Table 5-3

OVER-COUNTER DOCUMENT SEARCH

Step	Agent	Location	Operation
1.	Requester	Will vary	Communicate available bibliographic identifying data on document(s) wanted by phone, mail, or in person to Document Delivery System, and indicate response priority.
2.	Information Control Clerk	Document Delivery System	Prepare request form if not already made out.
3.	Information Control Clerk	Document Delivery System	If control number is available for the requested document, send one copy of the request form to the search unit responsible for the particular collection or sub-file in which the document would be stored; if the control number is not available, go to step 10.
4.	Document File Clerk	Document Delivery System	If the document would ordinarily be in the Microimage File, search the motorized card file for the document control number cited and proceed to step 5; if the document would ordinarily be in the Hard Copy File, go to step 18.

Step	Agent	Location	Operation
5.	Document File Clerk	Document Delivery System	If the document is found, remove document, replacing it with an "out" card, and send document with request form attached to reproduction; if document is not found, and it is in a category for which the system has a repository responsibility, forward request to Hard Copy File searchers and go to step 18.
6.	Reproduction Equipment Operator	Document Delivery System	Prepare paper copy of document on appropriate image-processing equipment.
7.	Reproduction Equipment Operator	Document Delivery System	Transmit paper reproduction of document plus request form to request receipt point, and return master image to appropriate files section for refiling.
8.	Information Control Clerk	Document Delivery System	Deliver copy of document (if found) to requester. Otherwise notify requester that document is either still in transit or not available in CHIVE (and why). If requester wishes, hold the request for a second search after a suitable time interval.

Step	Agent	Location	Operation
9.	Information Control Clerk	Document Delivery System	Record temporary or final completion of action on request form and transmit form to Data Transcription Group for typing and subsequent insertion (via the Page Reader) into the Management Data Files. End of Over-Counter Document Search. STOP.
10.	Information Control Clerk	Document Delivery System	Telephone, or send copy of request form to, EDP System.
11.	Information Control Clerk	Computer Center	If a priority request, deliver to console operator; if not priority, send to key punching and go to step 16.
12.	Computer Operator	Computer Center	Key the request into the computer using the inquiry console.*
13.	Computer Operator	Computer Center	Using the document identifying handles provided by the requester (e.g., post, airgram number, JPRS number, date, or other), search the header data portion of the Master Document Index File and print out the corresponding document control numbers.

*Cross reference listings, arranged in various sequences, will also be available for consultation and may be used in preference to machine queries to recover document control numbers where this approach would be equally effective.

Step	Agent	Location	Operation
14.	Computer Operator	Computer Center	Transmit results of printout to Information Control Clerk.
15.	Information Control Clerk	Computer Center	Telephone or transmit request form with list of document control numbers to Document Delivery System. Go to step 4.
16.	Key Punch Operator	Computer Center	Key punch search specifications and transmit cards to operations section to await batch processing.
17.	Computer Operator	Computer Center	Insert the request into the computer and go to step 13.
18.	Document File Clerk	Document Delivery System	Search the appropriate segment of the Hard Copy File. If document is found, remove document, replacing with an "out" card, send document with request form attached to reproduction, and go to step 6. If document is not found, so indicate on request form, send request form back to receipt point, and go to step 8.

Table 5-4

GENERATION AND INPUT PROCESSING
OF FORMATTED INFORMATION/INDEX
RECORDS PREPARED UNDER CONTRACT*

Step	Agent	Location	Operation
1.	Information Control Clerk	Contractor	Receive and log in the periodical, monograph, or other publication to be exploited.
2.	Information Control Clerk	Contractor	Obtain code designation (if a serial) from an official list and enter same on a routing sheet clipped to the publication.
3.	Information Control Clerk	Contractor	Sort and distribute publications to appropriate translators depending upon language or content of publication.
4.	Translator	Contractor	Scan content of publication for data of interest to CHIVE and determine elements of information to be extracted.

*This table illustrates the procedure which might be followed where the following conditions prevail: (a) CHIVE can influence the automation of data at the source; (b) the elements of information to be extracted lend themselves to a highly formatted record structure. Information of this type which enters the central reference system now, but only in hard copy, includes the Political and Scientific Biographic Cards from JPRS, Bibliographic Cards from the MIRA contract at the Library of Congress, abstracts of scientific articles from FDD, etc.

Step	Agent	Location	Operation
5.	Translator	Contractor	Type formatted transcript sheet for each article, monograph, or other, containing the pertinent information required. Enter data in English in the appropriate columns or spaces provided, and in the coding convention required where this does not require dictionary consultation. For the latter (e.g., organization names), enter descriptor in clear text. Type "remarks" - type information, the abstract body (if a scientific article), and similar unformatted text at the end of the index record.
6.	Translator	Contractor	Clip transcript sheet to publication and transmit both to coding group co-located with the Contractor or internal to CHIVE.
7.	Content Indexer	Contractor or CHIVE	Add codes, where required, on to transcript sheet in addition to clear text after consulting pertinent CHIVE dictionaries.
8.	Content Indexer	Contractor or CHIVE	Return publications to file and send transcript sheets to typists.
9.	Typist	Contractor or CHIVE	If typed product is to be read by CHIVE's Page Reader, type entries in form of hard copy; otherwise, generate paper tape as well as hard copy on Flexowriter-like device and go to step 11.

Step	Agent	Location	Operation
10.	Page Reader	CHIVE	Read typed copy and feed machine-language product to computer.
11.	Computer	CHIVE	Process records into Master Document Index File.
12.	Computer	CHIVE	If CHIVE area desk most concerned with input records generated by contractor does not desire to review additions made to the files, input process is completed. End of Input of Formatted Index Records Prepared under Contract. STOP. If opposite is true, print out (on a periodic basis) a hard copy listing of new records entering system, transmit listing to appropriate CHIVE area desk, and go to step 13.
13.	Information Analyst	CHIVE	Scan output listing for unwanted items.
14.	Information Analyst	CHIVE	Prepare a File Maintenance Transcript Sheet containing the usual job specifications (e.g., transaction originator, classification, file to be addressed, date, etc.), the numbers of the unique records to be addressed, and the operation (presumably a "delete") to be performed.

Step	Agent	Location	Operation
15.	Information Analyst	CHIVE	Send transcript sheet via typing and Page Reader to EDP System for processing.
16.	Computer	CHIVE	Delete unwanted records from the pertinent file.*

*An alternative approach to that taken in steps 13-16 would have the information analyst responsible for the file make use of a remote display device to screen additions to the file and make deletions thereto. Indeed, such a device could be introduced much earlier in the input cycle as the means by which codes would be added to the records and any unwanted entries deleted before file updating is actually undertaken by the computer.

Table 5-5

INFORMATION ANALYST ACTIVITY RELATIVE TO AN ALL-SOURCE, ALL-FILE SEARCH FOR A NAMED PERSONALITY

Step	Agent	Location	Operation
1.	Information Analyst	Will vary	Obtain available identifying data (e.g., name, citizenship, occupation, affiliation) on personality wanted by phone, mail or in person from requester.
2.	Information Analyst	C.G.D.	If request has been levied on right party, accept same; if request has been levied on right area desk but wrong Information Analyst (because, on this desk, there is more than one analyst and each specializes in a different topic), transfer request to correct individual.
3.	Information Analyst	C.G.D.	Obtain request number from central control point and enter in first section of interleaved request form the elementary data needed for logging purposes, i.e., name of requester, date, name of analyst handling request, etc.
4.	Information Analyst	C.G.D.	Send one copy of request form to control point for filing with other "open" requests.

Step	Agent	Location	Operation
5.	Information Analyst	C.G.D.	Search Master Dossier Index listing for references to inherited as well as CHIVE-built dossiers. If an entry for the personality is found, extract dossier number and date dossier identifier record was last updated.
6.	Information Analyst	C.G.D.	Enter in the query statement section of one copy of the interleaved request form the specific search parameters to be used in querying the CHIVE-built Master Index File. For example, if the Name Group Table is to be used, enter single spellings of both surname and personal names; if the name group feature is to be bypassed, enter the specific variant spellings to be included in the search; if FNU's are not wanted, so specify; if a dossier is available on the personality, exclude unwanted references already on file in the dossier by specifying that the date of preparation of any document index record containing the desired name should not be of a lesser value than the date the dossier identifier record was last updated. Also list any other factors which will serve to limit the scope of the search, e.g., citizenship, general or specific occupational category, date of birth range, etc.

Step	Agent	Location	Operation
7.	Information Analyst	C.G.D.	Assuming the Special Register (SR) name index portion of the Detail File to Comint Reports has not been integrated with the CHIVE Master Document Index, complete the copy of the interleaved request form used for searches of the SR Detail File consulting (as necessary) with an Information Analyst familiar with the vocabulary and file structure of this inherited file system. Refer to the printed version of the Name Group Table to help select the variant name spellings to be searched in this file, and also include any variant spellings required if the transliteration system employed in this file is unique.
8.	Information Analyst	C.G.D.	If a dossier was discovered on the personality in step 5, enter its number on the dossier retrieval copy of the request form.
9.	Information Analyst	C.G.D.	Forward the completed request form resulting from step 6 through typing and Page Reader to the Computer Center for retrieval of the pertinent index records from the CHIVE-built Master Index File; forward the request form resulting from step 7 directly to the Computer Center for manual retrieval and subsequent listing of the relevant name records from the punch card

Step	Agent	Location	Operation
			file inherited from SR; forward the dossier request form to the hard copy section of the Document Delivery System for recovery of the dossier desired.
10.	Information Analyst	C.G.D.	Telephone or communicate in some other fashion the details of the request to the Graphics Register (GR) for manual retrieval of photographs on the individual wanted from the inherited GR Personality Photo File. (Photos on the person processed <u>subsequent</u> to the initiation of the CHIVE system will be uncovered, initially in the form of index records, in the computer search of the Master Index File referred to above.)
11.	Information Analyst	C.G.D.	While awaiting receipt of the listed index records from the Master Index File and SR Name File, as well as the arrival of the hard copy dossier and photos, investigate any self-indexed card or document files on personalities inherited from BR which may be located either with the area desk or in the central hard copy files of the Document Delivery System. Also examine any Supplementary Files (e.g., Who's Who publications, commercial indexes, etc.) available at the area desk.

Step	Agent	Location	Operation
12.	Information Analyst	C.G.D.	Upon delivery of the index listings from the Master Document Index and SR Name File searches, examine the references printed out to determine whether they indeed refer to the person sought.* Consult again, if necessary, with an Information Analyst familiar with the SR system to interpret the output from the SR file.
13.	Information Analyst	C.G.D.	Assuming the request will not be rerun with improved criteria, identify the documents desired by encircling the appropriate document numbers appearing on the first pages of the listings. (Alternatively, the listing may be on a two-part form which will allow the Information Analyst to keep a carbon copy of the index record listing after using the original as an order for documents.)
14.	Information Analyst	C.G.D.	Transmit the document orders to the Document Delivery System, and any photo control numbers to GR, for retrieval and reproduction of the items desired.

*It is assumed, for the purposes of this table, that all material available on the personality being searched must be examined before a response can be made to the requester. For this reason, the search cannot end with the retrieval of an index record or card from a manual file.

Step	Agent	Location	Operation
15.	Information Analyst	C.G.D.	Assemble all material collected from the various document repositories (i.e., hard copy dossier . . . reproductions of documents from the CHIVE Master Image File, inherited Comint Document File, and GR Personality Photo File . . . original items pulled from self-indexed card or document files . . . and reference works from the Supplementary Files). Remove those items which, after analysis of the documents themselves, prove to be unrelated to the person in question, and prepare the response in the manner requested by the customer. End of All-Source Search for a Named Personality. STOP.

Chapter 5.7.

FILE CONVERSION

5.7.1. INTRODUCTION

Of the many types of extant central reference files which might be candidates for full or partial conversion to the CHIVE system, two are of primary concern. These are the document index and document image type files. In the former category are such files as the following:

- SR Detail Index File (Comint)
- SR Detail Index File (PI)
- FIB Active Installation Index File
- BR Dossier Index File
- IRS Document Index File
- GR Ground Photo Index

Inherited document image files include:

- IRS Document File (includes aperture cards and hard copy)
- SR Comint Document File
- BR One-Name File
- FIB Active Installation File (includes cards and folders)
- BR Dossier Folder File

There are, of course, many other types of central reference files in addition to those listed above, including some already in machine language.

Most of these, however, are either information files of such short-term interest that there would be little reason for converting the existing records, or are vocabulary control type files which, while they might be used to build analogous CHIVE indexing and retrieval tools, would not be converted *per se*. The discussion in this section, therefore, will cover only index and image files, in that order.

5.7.2. DOCUMENT INDEX FILES

5.7.2.1. Reasons for Conversion

One of the most important reasons for converting the inherited files to the CHIVE system would be to create a truly centralized source of reference data and information for the Agency. Conversion of the existing document index files to magnetic tape under the CHIVE system would provide a means of establishing effective data systems management.

The conversion of the inherited files would result in a reduction in the total number of document index files that would have to be maintained. In addition, conversion of these files would tend to simplify the operating procedures of the document indexing and retrieval system. By converting, only one set of procedures would be needed as opposed to a set of procedures for the inherited files and a different set of procedures for the CHIVE-built files if conversion were not undertaken. Furthermore, a reduction in the total number of personnel in the document indexing and retrieval system and a reduction in space should be obtained by converting the inherited files.

5.7.2.2. Degrees of Conversion

There are at least three different degrees or types of conversion that are possible. The first is a direct conversion and is probably the simplest and least expensive. Direct conversion means simply that the card image would be converted directly to tape. This type of conversion would not reduce any of the duplicative information existing in the card files. Moreover, it is the least desirable because it would provide the least amount of flexibility.

The second type of conversion is to convert the card files to the CHIVE format. This would eliminate any redundancy existing in the card files by pulling all data that was indexed on any particular document into one logical CHIVE record. This type of conversion is more desirable since it would provide good flexibility and would eliminate the built-in redundancy of the existing card files.

The third type of conversion would be a complete conversion, both syntactic and semantic. The syntactic aspects of the change would be similar to that described in the preceding paragraph. The semantic or vocabulary conversion, however, would require a considerable amount of intellectual participation by analysts from the respective areas where the inherited files originate. This type of conversion would be the most desirable and most flexible, but it would also be the most complex and difficult to accomplish.

5.7.2.3. SR Detail Index File Study

Of the various document index files described in Appendix 5.D., only one has been looked at in any detail to ascertain the conversion possibilities. That file was the SR Detail Index File--the study being performed to determine the advisability and feasibility of converting the file from cards to magnetic tape or to a direct access device. Some of the findings of this preliminary study are presented below. Whether these are representative of similar conclusions that might be reached vis-a-vis other inherited document index files after investigation of their individual conversion potential, one cannot say. Further study of the entire problem will be required during Phase III before any final recommendations can be made.

The following are the data that were collected from the SR study. The number of cards, in millions, that would have to be read to convert all of the Detail

FILE CONVERSION
Index Files
5.7.2.3.

Index File is as follows:

No. 1 File-Subject/Commodity	7.2
No. 4 File-Area	4.0
No.'s 2,3,6,7,8,9 Files-Organization and Personality	4.1
	15.6

This means that 15.6 million cards would have to be read to acquire all of the data in the current Detail Index File. This data applies to conversion to tape or conversion to a direct access device. Both approaches are discussed in the following sections.

5.7.2.3.1. Conversion to a Magnetic Tape File

For the first part of this study, it was assumed that the Detail Index File would be converted to one long file ordered on series-document number, with all data pertaining to any one document constituting a logical record. The file size converted to tape would be approximately 930 million characters. This would result in approximately 40 tapes for the master file, with that many as first backup also. This indicates that at least 80 tapes would be required at any one time to represent the file on tape.

Assuming a thousand cards per minute input rate with 20% allowed for manual handling, this results in 307 hours of 360/Mod 30 machine time to read the file in. This is equivalent to approximately 1.8 months of Mod 30 time (eight hours per day). Assuming the read-in is performed on extra shift, the minimum cost would be $3,000. In addition 30 to 35 hours of 7090 or 360/Mod 60 time would be needed for sorting, merging, and file building. This cost would amount to approximately $14,000. Programming and analysts costs are estimated at $10,000. Therefore, an initial cost of conversion would, at a minimum, cost about $27,000.

It would take a minimum of three hours to read a tape file of this size. An additional half-hour per day would be required for input request processing, sorting of input and output, output processing, output and maintenance. It was assumed that the Mod 60 would be used to do the search processing. This would amount to approximately $200 per hour. Assuming a once-a-day search, the approximate monthly machine rental to perform the maintenance and retrieval of the SR Detail

Index File would be $15,400. This is approximately two-and-a-half times the present EAM rental of SR's Machine Branch.

Turn-around time on requests would suffer by converting a large file of this nature to tape. The SR personnel contacted indicated that a 24-hour turn-around on all requests would be unacceptable. They further indicated that approximately 20% of the requests handled by SR require a two-hour-or-less response time. These priority requests are spread throughout the file, not just in a selected portion of the file.

The amount of space presently occupied by the SR Machine Branch (card files and EAM gear) is approximately 4,300 square feet. A reasonable value to place on this would be about $4 per square foot, per year. Assuming that 3,000 square feet of this area could be saved by conversion, this would result in an effective savings of $12,000 per year.

FILE CONVERSION
Index Files
5.7.2.3.1.

5.7.2.3.2. Conversion to a Direct Access File

Slightly different ground rules were chosen for this technique than were used on the "long tape file." Instead of trying to form one logical record from all the cards existing in the Detail File which originated from any one document, the existing file structure was assumed to be transferred to the Data Cell. Also, it was assumed that a directory or access file of a very simple nature would be maintained to enhance retrieval on this file. It was further assumed that the IBM 360/Mod 60 would be used to build the file and perform the operational activities required of the file.

The file would reside on a 2321 Data Cell which has a capacity of 400 million characters on line storage. However, the cells on a Data Cell Drive may be changed much in the same manner that tapes are changed on tape drives or disk packs on disk drives. Only one Data Cell Drive, which can have a maximum of ten cells on line, is required. The converted Detail Index File would occupy approximately thirty cells assuming about 75% packing.

The read-in of the file, assuming a thousand cards per minute reading rate and 20% handling, would take 307 hours on the 1402 attached to the Mod 60. The data could be read on to tape as an interim measure to save some rental on the Data Cell. However, some of these savings may be absorbed by additional programming costs. Assuming the Mod 60 would be operating in a multi-programmed mode, the cost of initial conversion would be as follows:

Reader (1402)	$ 1,600
Channels	200
Tapes	3,000
Data Cell	200
CPU	100
Analysis and Programming	10,000
	$15,100

As was mentioned earlier, the structure of the file would be the same as exists presently in cards. Therefore, no sorting for the input conversion is needed.

Retrieval on the file would take advantage of the directory to reduce the number of records that must be read to satisfy a request. A rough estimate of the average number of cards accessed from the existing file is in the range of 60 to 70 thousand per request. Therefore, 100 thousand cards per request was assumed as a very safe estimate for the direct access file. Assuming 10 requests per day (based on current usage), this results in approximately one million cards being processed per day. The average time of 137 microseconds per card was estimated for card processing. This results in approximately 0.83 hours per month CPU time. CPU time for retrieval and maintenance is 1.83 hours plus about 10% for handling which equals approximately two hours per month. This results in approximately $300-400 per month rental for the Mod 60 (for everything except the Data Cell). A range of costs are provided instead of more stable figures because of the difficulty in estimating for a multi-programming environment.

Estimated use of the Data Cell is approximately 20 hours per month for retrieval and 54 hours per month for maintenance if the entire file is passed each maintenance run. These two functions result in approximately $1200 a month rental.

5.7.2.3.3. Summary

The comments in this summary generally apply to both parts of the study except where specifically stated otherwise.

The following table of data was provided, with some modifications by SR personnel, from the [redacted] Report: 25X1A

Request Rates

File	Requests/Mo.	Requests/Day	Searches/Mo.	Searches/Day
No. 1	32	1.5	167	7.6
No. 4	62	3.0	443	20.0
No. 8	24	1.+	41	2.0
No. 7	14	0.6	88	4.0
No. 6	11	0.5	33	1.5
No.'s 2,3,9	73	3.3	1076	50.0
	215	10.0	1848	85.1

The table shows, as the headings indicate, the average requests per month and day. It should be noted that 90% of the requests against the No. 4 (Area) and No.'s 2,3,9 (Personality) files are selected by manually browsing the files. This means that 57% of the SR requests are handled manually. Further, from these facts, it is seen that the conversion to tape or direct access file would effectively replace an EAM system that is handling an average of only 93 requests per month. This usage rate is very low.

Even if the total request rate were used, it would still be a low usage rate for a computer driven file. The last statement is made for two reasons. First, if the actual number of requests (from a computer file point-of-view) were 215 per month, it would be highly questionable whether this would be large enough to warrant conversion. Second, the original 215 "requests" do not actually represent that many requests from a tape or direct access file standpoint. To explain--a sheet of paper entering the SR machine area containing instructions for searching a file may ask for references relating to pipes, paper, and cars. These parts are treated as three requests, not one, even though they all would go against the same file. However, this would represent only one request against the file from a tape or direct access point-of-view. Therefore, the total request rate of 215 per month would have to be divided by some factor to reflect how many requests this would represent in a tape or direct access system. Data on what this factor should be is not available at this time.

On the basis of these findings it is recommended that the total Detail File not be converted to magnetic tape. On the other hand, conversion of the Detail File to Data Cell storage appears to be economically feasible. The costs of performing the conversion and doing the required retrieval on a Data Cell attached to an IBM 360/Mod 60 are reasonable. Also, the turn-around time on a request is satisfactory since it would only take a little over five minutes to read and process the required 100,000 records to answer a request. This should leave adequate time for coding and outputting the request.

The decision to convert this file, however, cannot be based on these technical considerations alone. The usage rate must also be carefully appraised. Finally, it is important to remember that this conversion problem is but one of many CHIVE implementation tasks which must be addressed during the next 18 to 24 month period.

5.7.3. DOCUMENT IMAGE FILES

A comprehensive list of existing document image files is contained in Appendix 5.D. along with a capsule description of the function and activity characteristics of each. Also included for most files is an appraisal of each file's susceptibility to being segmented according to geographical area as a means of transition to the creation of an all-source document file.

This section will discuss the conversion alternatives and recommend a posture for concurrent operation of inherited and CHIVE-built document image files. It is felt that the approach presented will constitute a basis for orderly implementation of a new central document reference facility.

It is appropriate, first, to look at the reasons why conversion to a single document system should be considered. The overriding argument for such a step is to eliminate the multiple reference points that an analyst must currently consult and present to him a central reference point where a comprehensive response to his request can be provided. A further incentive for conversion to a central document system would be intra-Agency standardization of:

- File media and techniques
- Microfilm processing and reproduction equipment
- Hard copy quality and format

Conversion to a central repository and reproduction facility also presents a potential for reducing operating costs by combining similar clerical efforts, and by facilitating the use of more advanced processing devices.

Assuming then that there are advantages to be derived from converting to a centralized document reference facility, let us consider to what degree this could reasonably be accomplished.

Of about 25 document image files which are candidates for conversion (files enumerated in Appendix 5.D.), many can be excluded from consideration as candidates for conversion. A policy decision has been made to the effect that only textual documents are to fall within CHIVE repository responsibility. This immediately excludes all graphic files (i.e., photo, film, slide, and map files) which are to remain the respective responsibilities of GR and the Map Library.

Another major group of files are the dossiers which are subject-oriented folders relating to personalities, organizations, and installations. These files are maintained and referenced by information specialists who generally act as intermediaries between the consumer and the files. It has not been demonstrated that this type of information reference service can be improved by conversion of the existing files to another storage medium. Consequently, for the present it will be assumed that these files, which are primarily under the cognizance of BR and FIB, will be retained in their present form.

The foregoing exclusions restrict the discussion, then, to document image files currently maintained by the Library (Intellofax) and SR. These files are characterized by direct reference activity by the consumer, and, in most cases, respond by furnishing the consumer with a document. Primarily, they fulfill a document retrieval function rather than an information retrieval function, and, as such, are prime candidates for initial implementation as part of a centralized document reference service. Other files may prove suitable for incorporation into such a facility, but they should be evaluated on an ad hoc basis after a nucleus system has been established. Our recommendation, therefore, is that an all-source document reference facility consisting of document image files within Intellofax and SR be a design goal for the initial system.

It should be pointed out that the document system is largely independent of the CHIVE computer/indexing effort and consequently could be implemented prior to placing the EDP system on an operational basis. The centralization goal could be attained either in one step or on a modular basis. Either all incoming documents from the two systems could be incorporated into the new system, or some portion of each (such as Chicom materials) could be assimilated into the CHIVE-built system. The latter approach offers the advantage of limiting the volume during an initial shakedown phase.

The question remains as to how such an all-source document reference capability could be instituted. Essentially, it involves the problem of somehow combining two diverse inherited systems and integrating these with a third, new CHIVE-built system. As a fundamental tenet, total conversion of the existing document image files to the newly adopted file medium is not warranted or practical. The inherited files are very large in volume, having been accumulated over a number of years. Conversion to virtually any new system would require a copy of the document to be completely re-photographed and reprocessed into the new file medium. Some partial conversion to the new system might prove advisable for any segment of the file where high reference activity, over a long term, can be anticipated.

FILE CONVERSION
Image Files
5.7.3.

However, because of the low activity rate of the total file, the cost of converting records which will never be active should be avoided. The recommended posture, therefore, is that inherited files will not be converted from their current form but will merely be co-located within a single area along with the CHIVE-built files. The appropriate processing equipment will be installed within this same area and a single reference point will be presented to the consumer. Requests will be serviced through the appropriate systems, and responses furnished through a single distribution point where the proper enforcement of security restraints will be administered. The inherited files will be retained for reference purposes only and will not be augmented. All new items introduced into the file will be assimilated into the CHIVE-built system.

It is recognized that the recommended approach perpetuates existing files and techniques while introducing one additional document system to operate concurrently. Nonetheless, this approach seems to be

FILE CONVERSION
Image Files
5.7.3.

the only feasible way to cut over to a single, standardized document system and also eliminate the extreme cost and effort associated with a large-scale retrospective conversion. Experience has shown that there is a bias of reference activity toward more recent materials which would effect a gradual phasing out of the inherited systems with the growth of the CHIVE-built document file.

Chapter 5.8.

COMPUTER INTERFACE

5.8.1. GENERAL

The EDP portion of CHIVE will perform the following functions:

- Build and maintain files
- Create sub-files from existing files
- Search files and retrieve data from them
- Display data

The techniques chosen to implement these functions provide a built-in flexibility that will also allow revisions in the definition of the content and structure of CHIVE-built files.

In a computer based system, special effort must be devoted to inputting data, searching for it, reorganizing it, and subsequently displaying it. An integral part of the EDP system is a command language that allows these types of manipulation. It is recognized that "unlimited" flexibility is allowed if the user can be persuaded to use machine language. More practically, a set of commands is provided that permits personnel other than programmers to use the EDP system.

The CHIVE command language is fully described in Appendix 7.A. The language allows the user to direct the performance of the four functions mentioned above. Full use of the commands requires good knowledge of the indexing procedures, logic, and the content and structure of the records and files to be manipulated. It is planned that only information analysts, dictionary editors, and, to some extent, content indexers, will be trained to use the language.

The responsibilities concerned with defining new files and modifying existing file definitions will be assigned to the EDP file analyst. (See section 5.2.3. for further description.) The EDP file analyst must be trained to a level similar to that of a programmer, since he must be able to specify files to the system, initiate jobs for the machine operations personnel and participate in subsequent check-out.

5.8.2. COMMAND LANGUAGE

The command language permits the information analysts to direct the EDP system to provide desired results and products. The first consideration of the user is to build and maintain files. The usual file maintenance operations are provided. They are:

- Adding new data to a file
- Changing existing data
- Deleting existing data

The user can control the file maintenance operations in either of two ways. The first way is the usual one of specifying a unique record identification and then having the desired maintenance performed on that record. The second way is to specify logical conditions that could qualify a single record or many records within a file for the specified maintenance operation. For example, it may be desired to change the names of all factories named the Stalin Works to Big Brother Industries. In such a case it is only necessary to set up the test condition with a replace command. The desired changes are made without requiring the user to know in advance the unique identifications of all of the records involved in the transaction.

The second concern of the user is to search the files. The CHIVE command language provides basic search operators and logical linkage. The available operators are: and, or, not, greater than, less than, and equal. In addition, a "scan" operator allows searches for a contiguous string of characters in a value field. Notation is provided for specifying that the character string can be in any position within the value field and in some relative position. For example, it may be desired to find all occurrences of the character string ACZN22 no matter where it occurs in the value field or only when it is the first six characters of a value.

Another capability provided by the command language is to allow indirect searches. Here we mean that the user can specify the results of one search to be used as arguments in a subsequent search. An example would be: "What universities or colleges were attended by engineers working at radar plants in Country A?" A first search is necessary to determine the names of engineers associated with radar plants in country A. A second search can then be made to associate these engineers with schools. The command language allows the researcher to specify that the names of the engineers be automatically used as input arguments to the second search. Thus the problem involved with routing an intermediate machine output to an information analyst, setting up a second search, and then submitting it to the system are eliminated.

New files can be created by preserving the results of extensive searches of large document files. In addition, the capability of restructuring records is provided by the HIT processing commands of the CHIVE language. These commands allow a user to manipulate records after they have been found to satisfy search criteria and before they are transmitted to an output file. The control available permits saving for output all or specified portions of the original records. In addition, computations can be specified and the resulting values can be appended to the new output records. The resulting files can in turn be searched and updated in the same manner as any other system data file.

The command language also governs printing and displaying data. Section 7.11. describes output processing in detail and Appendix 7.C. shows samples of the types of reports provided by the EDP System. To specify a report it is only necessary to use the print command and then to state the name of the file, the sort sequence, and the output format desired. The format type includes such parameters as number of lines per page, width of printed portion of page, top and bottom literals, pagination, etc. The current report capability is felt to be adequate at this stage of the CHIVE development. Additional features will be provided only after actual need is established in an operational environment.

5.8.3. FILE DEFINITIONS AND THE EDP FILE ANALYST

The CHIVE command language allows manipulation of data in existing files and also permits a way of creating sub-files which can in turn be processed by the EDP system. These features directly concern the information analyst.

The tasks and procedures associated with changing file definitions and adding new files to the system are the responsibility of the EDP file analyst. The CHIVE EDP programs are controlled by external descriptions of the data files to be processed. The data descriptions taken collectively are called File Format Tables. Each table describes a file and its constituent elements. If it is desired to process files other than those currently defined it is necessary to add new table descriptions to those already in existence.

The File Format Tables contain all the information about an item that is required to process it. Included are the terms allowed in a record, term groupings, which terms are used as identifiers, addressing parameters, occurrence data, how stored, and content legality parameters. Extensive revisions can be made to the tables. In addition to adding new files, terms can be added to or deleted from an existing file. Legalities can also be changed. It is important to note that revisions of this type do not require any maintenance to the EDP programs.

The external file definition concept requires a special maintenance system. There are two main functions involved: the first concerns generating file format tables, and the second involves restructuring existing file data records. File format tables are generated from descriptions supplied by file analysts. Some types of table revision will result in producing a table that is inconsistent with the existing file. In this case, the existing file is processed so that its item structure reflects the new table revisions. After this step it is possible for the EDP system to operate correctly on the revised file with the new file format table.

5.8.4. SUMMARY

The CHIVE EDP System can be viewed by the information analyst as a tool for manipulating data. In order to get at this information, he must learn the rules and procedures attendant with the CHIVE command language. Forms will be designed to aid and guide in transcribing the commands. The EDP system is designed to allow random transactions which will obviate to some extent the scheduling of input to the machine. Output will be sufficiently identified so it can be routed back to the information analyst.

It is recognized that the interaction of the man and machine is never smooth. For this reason two remote consoles will be included in the initial system. These consoles will permit experimenting, in an operational environment, with the problems of direct communication between the information analyst and the EDP System. They should be helpful in expediting search processing, reducing paper output volumes and in simplifying the problem of routing requests to and from the computer.

Appendix 5.A.

THE ORGANIZATIONAL PROBLEM

This appendix describes the reasoning which led CHIVE to recommend the geographic organization of input and retrieval personnel with additional topical specialization for certain priority countries. In it, various alternative organizational configurations are described and their advantages and disadvantages discussed. A formal report on the CHIVE Indexing Experiment which led to some revision of the organizational concept recommended here--namely, the removal of the coding responsibility as such from the information analyst's area of concern--will be published in the near future as an additional appendix to this Phase II Report.

5.A.1. ORGANIZATIONAL OBJECTIVES

In considering the overall problem of how best to organize the functions to be performed and personnel to carry out these functions in a future storage and retrieval system, it appears logical to address oneself first to the primary objectives of the contemplated system and to derive from these a subset of organizational or management requirements which, if met, could assist in the attainment of the ultimate system goals. A particular organizational and management framework, of course, cannot by itself insure the achievement of a system superior to that now in existence. On the other hand, it is equally clear that despite all the advantages of EDP hardware (including stored program logic, speeds, etc.) and new developments in the information retrieval state-of-the-art, these tools alone are as yet insufficient to provide any major breakthroughs, and indeed have inherent disadvantages as well as advantages which, in the final analysis, must be taken into account. For this reason the efficient organization and employment of personnel takes on added significance. In fact, it may well determine whether a major step forward is possible.

The principal CHIVE system design objectives which have been discussed in some detail in earlier

ORGANIZATIONAL PROBLEM
Objectives
5.A.1.

documentation may be summarized for the purposes of this discussion as follows:

Objectives derived from user needs

1. Broader document coverage
2. Increased indexing specificity
3. More exhaustive indexing
4. Capability to answer more complex questions
5. Reduction of retrieval time
6. Single-service point
7. Common system vocabularies
8. All-source output capability

Objectives derived from operator needs

9. Micro-storage medium
10. Increased transcription speeds
11. Increased file utilization
12. More efficient use of available manpower w/o unacceptable degradation of system performance
13. Reduction of index and support file query time
14. Reduction of manual labor involved in preparing system outputs (research aids, acquisition lists, etc.)

15. Improved communication with customer

16. Increased index record lengths so as to reduce file proliferation

17. Improved evaluative tools for management

Some of the above are themselves organizational objectives for CHIVE, e.g., items 6, 8, and 15. Other listed objectives, if they are to be achieved, have implications at least for the organizational side of the total system design effort as well as for other design tasks. Combining the former with some deductive reasoning about the latter which is oriented towards the personnel and management implications thereof, it is possible to form a list of what might be called CHIVE organizational requirements. This list follows, and it is important to this discussion since it sets the goals in terms of which various alternative organizational configurations are compared.

Objectives Influencing CHIVE

Organizational Structure

1. Specialization with minimum processing

duplication

Encourage specialization on the part of information analysts to the extent possible so as to improve the quality of inputs and relevance of outputs to customer needs. At the same time minimize duplicative processing activities--i.e., multiple readings of the same documents, expenditure of intellectual time in term selection, transcription, etc.

2. Minimum customer contact points

Facilitate direct interface between the user seeking information and the information analyst most knowledgeable on the problem. Provide a coordination capability where required, but organize analysts so as to reduce need for same.

3. All-source service from any point

Organize system so that requester, if he so desires, can receive all pertinent information from whatever source that bears on his search problem.

4. Close communication between input and query handlers

Enable person querying system store to be thoroughly acquainted with processed inputs. Similarly, keep indexers informed of requests being handled by the system. Ideally, input and query processors should be one and the same.

5. Close communication between system operators and users

Operators should be fully cognizant of intelligence needs and priorities of research analysts. This is especially important in the CIA application where the breadth of customer subject interests and responsibilities and the volume of the data base are so large as to prevent equal attention being given to all subjects or sources.

6. Document control--first priority

The primary responsibility of the central reference system, i.e., to establish a basic retrospective search capability for all positive intelligence documents of immediate or potential interest to the Agency, must not be diluted by the addition of special tasks which, if permitted to grow unrestrained, would prevent the achievement of fundamental goals. Elemental priorities must be established and adhered to, and personnel organized in a fashion to bar the drift toward serving specialized user interests.

7. Job satisfaction

Morale of the central reference personnel must be maintained to reduce turnover and attract high-quality persons to the staff. Information analysts positions should afford opportunities for career growth and offer sufficient intellectual challenge to interest professional employees.

8. Flexibility in personnel allocations

New processing requirements and shifts in intelligence interests and priorities should not unduly upset the central reference operations and organizational structure. Requirements for retraining should be minimal if standard vocabularies, input, and retrieval systems prevail throughout CHIVE. Ideally the shift of one or more persons to more pressing tasks would not completely destroy an existing activity assuming the assignment of more than one person to a given subject or geographic area to begin with.

5.A.2. ALTERNATIVE FIRST-LEVEL ORGANIZATIONAL CONCEPTS

Keeping in mind the above-listed objectives for organizing the central reference personnel and activities, what kind of organizational configuration would appear to offer the best hope of meeting most if not all of these aims? In this section we will review some of the possible alternatives without necessarily considering all variant approaches which might theoretically be envisaged. The focus here will be on the initial, or first-level, organizational breakdown. In a subsequent section we will address the problem of how to manage activities within the rough organizational framework selected.

5.A.2.1. <u>Alternative A - Retention of Present Configuration</u>

Under this concept the existing structure of OCR would be accepted as is. Input and querying would be organized by subject (Biographic Register,

25X1B [redacted], and Intellofax), by

subject within source (Special Register), and by information carrier (Graphics Register and Map Library). Specialized EDP systems could be developed which would be tailored to the needs and desires of each Register or Division which might well employ different vocabularies, input and output processes, document storage media, etc. Alternatively, all systems might be required to adopt common file formats, dictionaries, programs, document storage and delivery systems, and so forth in order to simplify management understanding and control of processing activities and reduce design costs.

The principal advantages of this approach are operator and management familiarity with administering such a system, the availability of trained personnel and established operational procedures, the avoidance of any drastic reshuffling of personnel and slots with all the attendant problems associated therewith, and the assurance of continuing a level of system performance at least as high as that which it now obtains. In summary, the retention of the existing configuration is attractive because it would be the easiest to implement, and because we know it works even if the efficiency and quality of its performance is perhaps less than might be desired.

The major reason for not following this route is that, while the risks are less, the system will always be constrained by the organizational structure within which it must operate. Thus the potential for real improvement will be limited. Specifically, it would be impossible to make any real progress toward achieving objectives 1-3 above and limits severely what can be accomplished on objective 8. Redundant reading and analysis of collateral documents could scarcely be avoided and the trend toward all-source information files might foster duplicative processing (already initiated by FIB's exploitation of Comint materials) in the SI area as well. Semi-duplicative document repositories, such as now exist in FIB, BR, the

Intellofax System, and to a minor extent GR, would probably persist because of the difficulty of identifying in advance which repository will choose to keep a given document. Customers seeking to exploit all the subsystems would still be faced with the necessity of interrogating each system separately unless an inter-system reference group were provided or the system contacted assumed the responsibility of querying all others. Either of the latter potential solutions, however, would interpose request "interpreters" between the customer and the ultimate respondent with consequent ill effects to the communication process.

In brief, while Alternative A is appealing because of its familiarity, its inherent disadvantages are sufficient in number to influence a search for something better if such can be found.

5.A.2.2. Alternative B - Single, All-Source Document Retrieval System; Separate Biographic Information Facility

Between the extremes of a completely centralized, all-source, all-topic storage and retrieval system and the existing decentralized configuration of OCR many variations and alternative combinations can be conceived. That which has attracted the most attention perhaps is the concept of merging Intellofax, the Special Register, and the Foreign Installations Branch but leaving the Biographic Register as a separate activity. Proponents of this approach (some of whom would also except FIB from the merger) generally point to the "unique character" of the BR operation, its "analytical" responsibilities, its production of finished intelligence, the fact that it is not a document retrieval system at all but rather an information file, and so forth.

Most of those favoring this compromise approach are somewhat vague on the organizational details. Some, apparently, would establish an all-source BR, removing the responsibility for personality control of Comint materials from the conjoined Intellofax-Special Register operation.

Others would not make this transfer of responsibility arguing, inter alia, that most BR customers are not cleared for Comint anyway. Some would retain the all-source FIB system as a separate file as well, presumably with installation indexing remaining a part of the Intellofax-SR document input activity. The redundant analysis of documents common to each of these systems has either not been considered by those who have recommended this approach or has been accepted as a necessary evil.

Of those favoring Alternative B or some variation thereof, most do so in the belief that there are indeed advantages to be gained from the all-source approach, integrated indexing, system standardization across OCR, common vocabularies and other reference tools, and other CHIVE goals. Most would, therefore, adopt CHIVE's system recommendations if biographic data handling at least were excluded.

What appears, however, to disturb people the

ORGANIZATIONAL PROBLEM
First-Level Concepts
5.A.2.2.

most about the prospect of including biographic intelligence in a centralized system is the index transcription problem. It is pointed out first of all that, while the necessity for filling out transcript sheets has long been accepted by Intellofax and SR analysts, it would not be readily accepted by BR personnel who, in recent years, have employed a file system (sometimes referred to as a "Collectanea" by documentalists)* which requires no transcription at all. Second, there is the fact that any transcription requirement, no matter how limited, would diminish the number of personality references which could be processed by BR since it would necessarily add to processing time. Third, there is the argument, frequently expressed, that BR's need for multiple access points to personality data has fallen off steadily over the

---

*This term refers to any file system that used the general approach of lifting sections from a single source document, reproducing these excerpts, and physically filing them under each of the categories or key words of interest.

ORGANIZATIONAL PROBLEM
First-Level Concepts
5.A.2.2.

past several years following the assumption of community responsibility for political personalities. Why have more than name control over files, the reasoning goes, if the majority of requests are for specific named individuals?

The transcription argument might, indeed, justify leaving BR outside the central system concept were it not for the fact that following such a course helps none at all to resolve BR's storage and retrieval problems. Examined realistically, it appears clear that there are only two fundamental ways of processing biographic or any other kind of information: (a) by creating an index to documents containing the pertinent information (which index is then screened prior to the recovery of the documents themselves); or (b) by filing (and, if necessary, reproducing) the documents under the terms which constitute the desired search parameters (i.e., by establishing a "self-indexed" document collection). If the choice is to take the index path then certain elementary requirements must be met if retrieval from the system is to be successful. In the case of large personality record collections it means the index must carry sufficient identifying information about the personality to enable the searcher to distinguish between personalities bearing similar names. The more identifying information extracted from the document the better, but at the price of increased transcription time. Alternatively, the more abbreviated the index the less the transcription burden, but at the cost of more irrelevant documents retrieved.

The "collectanea" (or self-indexed document file) approach offers the user a reverse set of advantages and disadvantages. On the one hand, it virtually eliminates the function of having to transcribe words from documents. On the other hand, it vastly increases the physical storage requirements of the system by virtue of the fact that each document must be multiplied by as many file headings as one chooses to store the document under. Since no system has unlimited space, this usually means that the means of access to the document collection are severely limited in comparison with document index systems. In addition, the filing problem is exaggerated by the explosion of the original document population

25X1A
25X1A

(witness BR's assignment of [redacted] file clerks full-time to its central biographic card file and [redacted] clerks to its dossier system).

The point of this brief detour into the world of personality data handling is to make clear that nothing is really gained by leaving BR outside the central system framework unless it has already been concluded that biographic data will not be controlled by an index *per se*. Even this would not necessarily dictate the exclusion of biographic processing since it would be perfectly possible for the input analyst, after indexing the remainder of the document's content, to have the document or selected pages therefrom reproduced and filed (in hard copy or microimage form) under the personality names of interest. If, on the other hand, the decision is to index biographic information then there are certain very real benefits in integrating this index activity with the representation of other subjects discussed in documents.

As for the remaining arguments deployed in the cause of keeping BR outside the integrated processing activity, they have little bearing on the manner in which biographic data should be stored and retrieved. Rather, they relate to the analytical functions to be performed, i.e., interpretation, correlation, synthesis, etc., <u>after</u> the raw material has been recovered from the files. Admittedly this intellectual process could be carried out by a separate group altogether, as indeed often occurs when a customer (e.g., a scientific intelligence analyst) chooses to review and interpret the basic documentation himself. But it can also be performed, perhaps equally well, by persons who also index and retrieve biographic information. Whichever path is chosen it need not affect where and how documents are processed.

5.A.2.3. Alternative C - Co-located Organizational Configuration

A radically different organization concept from those discussed thus far, one which deserves at least brief consideration, is the notion of decentralizing document processing in the Agency by dispersing the activity amongst the research and production components. Among the arguments for upgrading the so-called "analyst files" versus attempting to improve the central reference system are the following:

- Analyst files will continue to be maintained whatever is done centrally. Since they are a major information retrieval resource why not make them even more effective and efficient?

- Providing analysts with manpower support in the form of information assistants physically co-located with research personnel in the production offices would relieve the analyst of most of his file maintenance problems and enable him to devote more time to research.

- Analysts could more readily control what goes into the files thus reducing input chaff and providing semi-evaluated retrieval.

- Full-time information specialists could index more material than analysts can process into their files today thus improving the breadth and depth of coverage.

In the decentralized as in the centralized system approach, it is possible to think of many ways in which the processing activity might be organized. The following, however, are perhaps the most logical alternatives:

a. Decentralized input and files/central directory of files

Under this approach OCR would virtually disappear with the exception of the Library, FDD, and possible the Graphics Register. Analysts would continue to process materials into their own files but might be provided some machine assistance in the areas of file manipulation, storage, and reproduction. In addition, a master profile or directory of analyst files would be created and maintained at some central location. Analysts with a search problem would consult the directory, determine which file(s) to peruse, and then either exploit the file directly or work through the analyst who maintains the file. Personnel formerly attached to OCR could either be assigned to the research analysts as information assistants where they would perform the bulk of the input and retrieval activity, or the research analyst population might be increased by converting the slots to intelligence production positions.

b. <u>Decentralized Input/Centralized Files</u>

This scheme would be much the same as the above in that input processing would still be performed on a decentralized basis. The difference would be that, in addition to maintaining decentralized analyst files, research analysts and/or their information assistants would be required to transcribe their indexing in such a fashion that a record thereof could be passed to a central storage and retrieval facility. Similarly, reproductions of the documents they wished to store or the pertinent citations thereto would be sent to central storage. Adoption of this approach would greatly increase search specificity over the directory technique and greatly simplify the problem of gaining access to the data files themselves.

c. Decentralized input and files for select subjects/centralized input and files where interests overlap

This system is perhaps best represented in the real world by NSA where files of restricted interest are co-located with the most appropriate customer offices, while files of interest to many are maintained centrally.

d. Centralized input and files/information specialists co-located with research components

This system would continue the central refer-

ORGANIZATIONAL PROBLEM
First-Level Concepts
5.A.2.3.

ence activity without prejudice to decentralized analyst files, but representatives of the central system would serve on permanent or rotational assignments in the customer offices. Their function would not be to index material for analysts, nor to actually search and retrieve material from the central system, but to improve communications between the analysts and the central storage and retrieval operation. They would provide advice to analysts on the reference services available to them, transmit their queries to the proper components, identify unnecessary and/or duplicative data files, inform the central service of current intelligence priorities and anticipated retrieval needs, and in general insure that both sides of the house achieved a full understanding of each other's problems, capabilities, and requirements.

There is much that is attractive about all

ORGANIZATIONAL PROBLEM
First-Level Concepts
5.A.2.3.

the above alternatives primarily because all provide better user definition and control of what the Agency should be retaining in its record collections, and because all provide a means for the analyst to exploit potentially useful files maintained by others. With the exception of system 3.d., however, which appears to offer some significant advantages which might well be tested on a limited basis, all suffer from one or more of the following disadvantages which are sufficiently serious to recommend the rejection of the decentralized organizational concept as a practical solution:

- Elimination of part or all of the existing central processing activitivies would inevitably give rise to increased record keeping by Agency analysts. Indexing by these analysts would be highly duplicative and inefficient because of overlapping interests amongst Agency components. Even today the duplication of analyst file activity is sufficiently widespread to cause some to seek ways in which the situation might be ameliorated. In a recent study* one research analyst reported that "the files of

25X1A

* [redacted] The Analyst's Inbox in the DD/I Area: Help or Hindrance?, 30 June 1964, OTR/IPC, Confidential.

several offices within OCI and ORR practically mirror each other, if not in totality, then at least in certain subjects." Among the reasons for this situation, the same analyst observed, is the failure of management to properly define the exact responsibility of the analyst beyond his geographic area, the necessity for the analyst to be aware of the "big picture," fear of requests from Agency officialdom whether they fall within the analyst's assigned mission or nor, physical distance from other potentially useful files, etc. Whatever the truth of these remarks (and all were noted during the CHIVE Fact-Finding Survey of the DD/I), any enlargement of the analyst's filing responsibilities would result in a corresponding increase in duplicate files.

- It would be virtually impossible to establish and maintain inter-analyst consistency in indexing, and to enforce adherence to standard rules and practives. The many components involved, each responsible to a different line of command, would make coordination and management most difficult.

- Analysts regard file maintenance as a necessary evil. Any suggestion that they expand their input activities, especially if it requires them to prepare index records in a fashion which can be "captured" for storage at a central location, would meet with great resistance.

- Analysts select only a small percentage of incoming documents for filing. This fraction of collected intelligence information ordinarily reflects a current problem bias or that material pertinent to an analyst's production assignments for the coming year. Moreover, some information which would be filed by an analyst with less experience on the job would be ignored by the more senior type who has already stored such information in his head. Unfortunately, the analyst's cranium, although a well-recognized part of the Agency's institutional memory, is not easily accessed by information seekers and is lost when the analyst leaves the Agency.

- Analysts almost universally state that they want and need a central system for retrospective search and file back-up. They do not feel that their own files, nor even the sum of all files of all research components even if they could be made readily available to them, would fully satisfy their requirements.

- The possibility of co-locating select central reference files with the primary users, as suggested in 3.c. above, is practical only for intelligence organizations having clear demarcations of subject and area responsibility. Regrettably, no such pattern prevails in this Agency, as pointed out in the study referred to above.

- Agency reference responsibilities to other USIB components, whether imposed by DCID directive (e.g., biographic) or the result of tradition and historical precedent, could be met only with great difficulty if the centralized file concept were abandoned. Interface problems of indescribable complexity would inevitably arise.

In summary, there appears to be no acceptable alternative to a central reference system for a consumer population as large and complex as that represented by the DD/I and other CIA and non-CIA components.

5.A.2.4. Alternative D - Centralized, Geographically Organized Configuration

Assuming the organizational objectives listed on pages 224-227 are indeed the controlling parameters in selecting a management framework for a future information storage and retrieval system for the Agency, it is difficult to conceive of any better way of organizing the personnel involved than by grouping them initially by geographic area. While this would not overcome all operational problems that can be envisaged, of all the systems considered it comes nearest to meeting the requirements outlined above.

In a geographic organizational arrangement there would be, perhaps, five major geographic divisions reporting directly to a single manager, presumably at the Assistant Director level. Most of the existing central reference repositories (i.e., BR, FIB, SR, and DD) would be abolished and their personnel transferred to the new geographic components. Previous area assignments would be taken into account in relocating personnel.

Documents would be disseminated to the geographic divisions by an external dissemination group which would also handle dissemination to the research offices. These documents would include all materials of whatever classification, format, or mode of presentation. International documents (those dealing with subjects or events occurring in more than one country) would be routed to each of the geographic desks concerned when the application of area expertese in the indexing process seemed justified by the character of the subject matter dealt with in the document. The majority of documents, however, would be processed by one desk only. A single master file would be maintained of all documents indexed by the central reference system.

Most requests would be levied directly on the geographic unit having responsibility for the area of concern. Occasional requests would have to be coordinated between the divisions when more than one country was involved, but this would be the exception rather than the rule. The respondent, under the new configuration, would be familiar with reporting from all sources on the matter of interest to the customer, and could thus insure that the data retrieved reflected the full response potential of the system.

The proposed configuration would lose the advantage of source specialization in processing and would pose occasional problems of geographic overlap in document indexing and query coordination. However, these disadvantages are not felt to be

ORGANIZATIONAL PROBLEM
First-Level Concepts
5.A.2.4.

serious. The system would come very close to achieving all of the organizational goals set forth earlier as the following review of said objectives demonstrates:

a. Processing duplication

There would be a minimum amount of redundant reading and expenditure of intellectual effort in input processing since the majority of documents would be completely processed by the person to whom they were sent. While the international document problem will arise, there are fewer international documents than there are documents dealing with multiple subjects (i.e., persons, organizations/installations, commodities, etc.). Nor must Information Analyst specialization necessarily be surrendered. Instead of concentrating on biographic, installation, or other data, they could specialize in certain topic areas of interest to intelligence--e.g., military, economic, political affairs, etc.--within the country to which they are assigned. In addition, the extant duplication of document files would be eliminated with concomitant benefits in terms of storage space, reproduction loads, and filing requirements.

B. Customer Contact Points

Analyst inquiries normally relate to a particular geographic area of the world, although the information sought is frequently diverse in character and not restricted to any particular collection resource. Under the configuration proposed, there would ordinarily be no need for the requester to interrogate more than one component of the system since the organization of service personnel would mirror the manner in which user organizations are themselves organized, i.e., by topic within country.

ORGANIZATIONAL PROBLEM
First-Level Concepts
5.A.2.4.

c. <u>All-source service</u>

One of the principal advantages of geographic organization is that, in addition to the establishment of all-source files, there is an extra benefit to be derived from the bringing together of information analysts who have specialized source background. This pooling of knowledge will make for more informed reference personnel and will help remove gaps and ambiguities in the data files and authority lists developed in separate source environments.

d. <u>Input-output communication</u>

The geographic organization of central reference personnel does not, in itself, assure or encumber communication between input and query handlers. Rather, this is affected by the communication processes built into the system, and by the extent to which personnel specialize in the various functional areas of input and output processing. These

ORGANIZATIONAL PROBLEM
First-Level Concepts
5.A.2.4.

matters will be discussed in the next section.

e. Operator-user communication

While it may seem that geographic organization *per se* offers no inherent benefits over the present central reference configuration in terms of insuring better communication between information and research analysts, in fact the information analyst in the proposed system, by virtue of the fact that he has access to a wider variety of sources and shares a subject/area assignment similar to that of his research counterpart, should be more cognizant of the latter's resources and problems and, therefore, be able to offer him better service. This, to be sure, is not enough, given the separate physical and operational environments in which each operates, and for this reason experiments such as locating certain Information Analysts in the research components should be tried as well.

f. Processing priorities

Geographic organization at the upper management levels cannot prevent information personnel from being assigned to respond to narrow user interests. Within the geographic divisions an organizational structure reflecting processing concerns (e.g., document control vs. special file projects) might help clarify who is doing what, but since personnel can always be shifted around it is management control which, in the final analysis, will determine the direction and continuity of the effort.

g. Job satisfaction

It would appear that the system proposed offers a richer and more meaningful environment for the information specialist than that now available to him in the majority of OCR registers. He would not be assigned one function only as, for example, the input analyst in the Intellofax System; he would have access to a greater variety of documentary materials; he would be able to specialize in a substantive area of intelligence concern; he would have contact with users of the information store and thus gain some appreciation of the problem to which his effort was addressed; and, not least important from the Agency's point of view, he would be better able to assume a research position if the opportunity arises for him to make such a move--as it often does.

h. Flexibility

Common system standards and procedures across the geographic divisions as well as the increased availability of personnel on any geographic area should lessen the problems entailed in re-allocating personnel to accommodate changes in user needs. In a sense, the bringing together of all persons working on the same country--persons now scattered amongst the various OCR registers--is analagous to the establishment of a medical clinic composed of specialists in various subject areas versus the continuation of individual medical practice. The assemblage of these various skills increases overall flexibility and assures the highest quality service.

Before concluding this section of the discussion, some additional facts may be worth noting.

In a report to the Critical Collection Problems Committee of USIB, the Director/SCIPS observed that "information processing activities, as contrasted with collection or research, generally are not oriented to area or country organization." However, he went on to point out, "most of th 340 information handling activities surveyed [b SCI S] are concerned with peripheral descripti data rather than the substantive content of the information items and are, therefore, organized on a

ORGANIZATIONAL PROBLEM
First-Level Concepts
5.A.2.4.

functional basis rather than a geographic coverage basis." A different situation exists, he said, "where the process is dependent upon substantive content such as . . . deep indexing." In the latter case "then the lowest organization level is more apt to be structured on a geographic area basis, like collection and research activities are prone to be."*

In fact there can be little doubt that the processing of multi-source documents by geographically-organized personnel will work. Within our own agency we have the Analysis Branch of the Document Division organized on this basis to process inputs into the Intellofax System. As is well known, the system deals with a wide variety of intelligence report series and other documentary media. BR and FIB are similarly arranged and, though they confine themselves to restricted subject areas, are faced with an even

---

25X1A [redacted]

ORGANIZATIONAL PROBLEM
First-Level Concepts
5.A.2.4.

greater diversification of documentary inputs, including books, periodicals, newspapers and even photos. Outside CIA there is the DIA document storage and retrieval system which recieves inputs from all USIB agencies and indexes persons, organizations, locations, as well as other subjects. Thus, the issue is not whether input processing organized on geographic lines will work, or whether a multiplicity of document types can be handled by a single organization, but what the tradeoffs are versus some other approach to the problem.

5.A.3. ORGANIZATIONAL ALTERNATIVES WITHIN A GEOGRAPHIC DIVISION

The preceding section was addressed to the issue of the first-level organization of the central processing activity. The problem, however, does not end here since, even if the geographic division concept is excepted, each geographic division would be so large that some division of personnel into more manageable administrative units would be required.

Referring back to the organizational objectives listed earlier it appears that if the geographic arrangement makes sense as the first cut, it would likewise be the preferred approach at every succeeding management level within the organization until the country level itself is reached. For example, if it did not seem desirable to group persons by document source or by the subject matter in documents they were assigned to store and retrieve because of the effects this would have on processing overlap, interface with the customer, capability for providing all-source service, and so forth, then it would make equally little sense to permit them to creep back into the system, although at a lower level, if the effect on the system's performance was still the same.

The geographic concept begins to break down, however, when the volume of activity (input as well as requests) on a single country is characteristically so great that a relatively large number of information analysts must be assigned to the same country. It would be possible, of course, to have both the documents as well as the requests distributed indiscriminately amongst these analysts, but specialization is always advantageous if it can be achieved at minimum or no cost to other system goals.

Since not enough is known at this point about the input/output traffic that can be expected on every country in the world, nor what the manpower requirements and constraints will be on the CHIVE system, it is impossible to state with any degree of certainty where a division of personnel within a given geographic area will be required. For some areas it seems logical to predict that an analyst will have complete responsibility for a country, e.g., one of the emerging states in Africa which is of little consequence in international affairs and, therefore, engenders little in the way of intelligence reporting or analyst interest. On the other hand, many information analysts will be required for the larger countries such as the USSR and China and thus the organization of these analysts becomes a matter of serious concern.

The most reasonable alternative ways of grouping personnel assigned to one country would seem to be the following:

5.A.3.1. Organization by Document Source

Adoption of this approach would mean that separate groups of analysts would be established for each major document category. These categories might be the open literature, collateral intelligence reports, Comint and T/KH, etc. The principal advantage to be gained from this method of organization would be the availability of personnel trained on a document source basis. It would have stronger selling power if the indexing systems used were to differ by source. However, the latter will not be the case. Its disadvantages are that almost every request would have to be coordinated among the different source-oriented units since customers would customarily want more than one source searched; Information Analysts would operate in different worlds and none would have a complete picture of reporting in his particular area of concern; the tendency would be to maintain separate rather than integrated all-source files; and the multiple service-point problem would remain. On balance, it does not seem to be a desirable approach.

5.A.3.2. Organization by Function

This system would allocate to certain information analysts assigned to a country the responsibility for indexing all documents received on their area, to others the responsibility for answering all requests on said country, and possibly to a third group the task of maintaining "special project" files and establishing and periodically updating information files consisting of summarized data about a particular person or group of persons, installation, or activity.

The notion of distinguishing input from retrieval personnel is not a new one. Libraries have traditionally followed this approach in separating the cataloguing from the reference librarian function. Many EDP-supported information retrieval systems have also chosen this route, the original 438-L system of SAC being a prime example in which so-called "query specialists" (as distinguished from "coding specialists" and "file modification specialists") were to handle all searches directed against the system.

The advantages of separating personnel by the functions named are the following:

- It heightens the job satisfaction of those assigned to the output end of the activity, thus reducing personnel turnover and enabling the system to recruit and retain higher-quality personnel.

- Persons unqualified to deal effectively with requesters can be separated therefrom with less embarrassment to management. Similarly, persons who have neither the interest, background, nor temperament to become effective indexers can be given assignments more in keeping with their qualifications.

- New personnel can be trained more quickly if the job responsibilities are more narrowly defined. This will reduce the total amount of unproductive time expended by the system, a matter of no small significance if the turnover rate is reasonably high.

- By encouraging specialization the quality of the system's performance is enhanced.

- It permits processing to go on undisturbed by request interruptions with some consequent increase in operational efficiency.

- By formally separating the document storage and retrieval responsibility from special and general-purpose information file maintenance, system functions would be better defined and management would have a clearer picture of their investment in either area. This would bar the often unnoticed drift of centralized retrieval systems toward increased special f...e-building activities to the detriment of establishing a basic retrieval capability over the documents entering the system.

The principal disadvantages of functional separation are:

- Query specialists would be unfamiliar with the inputs to the system except those they retrieved as the result of searches levied against the files. As a result they would tend to lose touch with current intelligence reporting unless some mechanism was provided for them to read select incoming documents, review the product of the indexer activity, or other. In addition, all persons who index documents as well as answer requests retain a great deal of information in their heads which is never reflected in the index representation of documents. Subtle though this advantage may be, it makes for more effective service to customers in ways too numerous to mention. And it is most difficult to acquire this knowledge through any other mechanism than participating in the input process itself.

- Input specialists would have little appreciation of customer needs. Being barred from dealing with requesters, they would not know what subjects to stress in their input processing, nor how to distinguish the significant from the insignificant.

- The inevitable tendency would be to consider the query specialist a cut above the indexer to the detriment of the input person's morale. As experience has shown in OCR, the request handler would be regarded as having the more interesting job primarily because, having contact with users, he could understand better what contribution the entire activity was making to the intelligence mission. Those indexers who were unable to make the change from input to request handling because no vacancies developed would ultimately take positions elsewhere. Those who remained would tend to represent the less capable and imaginative until, ultimately, the entire input staff would take on these characteristics.

- This approach would conflict with the mode of operation in most OCR components. With the exception of the Intellofax System, most O C R systems have chosen to have the same individuals handle queries who handle input to the files. Both the Special Register as well as sections of the Biographic Register have actually operated for varying periods of time on a functional basis but reverted back to the integrated configuration. Certainly, the majority of experienced OCR staff members would prefer to have information analysts operate in both modes and would resist the other approach.

- Peak request or input loads would require the temporary assignment of personnel to the duty which was not their prime responsibility. Indexers who performed the retrieval function would thereafter be able to claim, and rightly so, that they were able to do the job otherwise they would not have been called on in the first instance. This would tend to weaken management's argument for continuing the distinction.

As can be seen, while a good case can be made for either configuration, we tend to favor not making a formal division of central reference personnel along functional lines. While there will, inevitably, be some persons in the system whose functions will be more or less unique, and others who because of personality or other limitations will be confined to a restricted area of operations, these will be the exceptions rather than the rule and, in the latter case at least, would not be reflected in the formal organizational structure.

5.A.3.3. Organization by Named Object

This configuration would organize the information analysts by the major classes of data stored and retrieved by the system. For example, within the USSR Division there might be a Personalities Branch, an Organization/Installation Branch, and a Subject/Commodity Branch. The Special Register is divided on this basis today and, in a sense, the collateral repositories of OCR, i.e., BR, FIB, and the Intellofax System, are reflections of the same concept except on a larger scale.

We know that this approach will work since it has been proven over many years of operating experience. Furthermore, by introducing this kind of division at a much lower operational level (namely the country desk) than is the case today, many of the ills of the existing system such as conflicting vocabularies, overlapping document files, diverse input/output procedures, and so on might well be eliminated. It also offers the advantage of immediately identifiable manpower trained in these particular areas and, in addition, permits a high degree of analyst specialization.

What makes this solution unattractive? The principal objection is, of course, the fact that it would be a rare document that would not have to be read and indexed by all three groups. While Comint materials would be less troublesome in this regard, collateral documents and open literature are not typically oriented to any single type of named object. Attempts to coordinate the input effort so as to reduce duplication would be extremely difficult to implement, and document dissemination would in all likelihood take the form of dissemination of the same documents to all three points. Finally, there would remain the problem of coordinating the response to queries. A significant proportion of the requests would relate to all three subject areas and require a coordinated response.

In summary, while this configuration is preferable in many ways to the existing central reference organization, it would be less efficient and economical than what might be desired. That there may be a better alternative was suggested earlier, and it will be the subject of the next section.

5.A.3.4. Organization by Topic

The major selling points for a topic approach to the organization of the central reference activity beneath the country level are: (a) that it corresponds more closely than any other configuration to the kinds of requests we can anticipate will be levied on the system; and (b) that, while it does not eliminate entirely the problem of the multi-subject document, it would seem to confine the problem to reasonable bounds. If the former statement is accepted, then organization along topic lines would lessen the need to coordinate the search activity in order to provide the customer with a complete response to his query. Similarly, if documents tend to relate to a single, though broad, subject area of intelligence concern (for example political affairs, scientific and technical intelligence, military activities, or economic matters), then the need for multiple routing of documents should be diminished and processing duplication minimized.

A preliminary examination of documents entering the current system, as well as a review of queries levied on the system by analysts, indicates that both do tend to concentrate on one or another of these basic subject areas. This is not too surprising since these are the classic divisions of strategic intelligence, and collection as well as production organizations within the intelligence community reflect this fact. It also appears that there is a reasonable balance of documents as well as queries in each of these topic areas such that there would not be a preponderance of personnel assigned to any one field.

As to whether it might be desirable to further refine the topical breakdown within political affairs, economics, etc., this would depend on the number of information analysts assigned to any one topic. Additional subdivisions are clearly possible and could be advantageous in that they would permit increased analyst specialization and lessen the span of control problem for supervisors. On the other hand, these benefits might ultimately be offset by the inability of the system to separate documents cleanly on the basis of these increasingly narrow subject categories.

Documents dealing with two or more major topics would, of course, be received by the system. However, this need not cause any undue concern. Multi-processing of a single multi-subject document by different topical specialists is less important than multi-processing of an international document by geographic specialists. Such documents would be directed to the unit which seemed principally concerned for complete indexing even when the choice seemed rather arbitrary. If it appeared that the information reported seemed of more than average significance, this would not preclude an information copy of the same document being routed to another unit.

Research analysts should prefer to deal with topic-oriented information specialists since they

ORGANIZATIONAL PROBLEM
Geographic Division
5.A.3.4.

would find them better able to understand their search problems. Indeed, such information specialists might in time become more factually knowledgeable than their customers since they would have fewer extraneous responsibilities and could concentrate their exclusive attention on the subject at hand.

ORGANIZATIONAL PROBLEM
Geographic Division
5.A.3.4.

Appendix 5.B.

PRELIMINARY EVALUATION OF THE CHIVE INDEXING EXPERIMENT

5.B.1. SUMMARY DESCRIPTION OF EXPERIMENT

A joint OCR/CHIVE indexing experiment was conducted from about 16 November 1964 to 15 January 1965. Approximately two months training preceded the indexing phase of the experiment, while the query and evaluation phase is expected to extend through May. The personnel involved included 16 indexers, 4 senior indexers, 3 clerk typists, and 3 project monitors. The data base consisted of some 5,000 all-source documents on [redacted] collected during the period 1 July - 30 September 1964.

25X6

25X6

The experiment was held to test certain organizational and indexing techniques proposed by CHIVE. Specifically, it was desired to test the following major concepts:

- That with adequate supporting tools, a person can satisfactorily index all of the information contained in documents, i.e., people, organizations/installations, areas, subjects, etc.

- That all-source materials (including Collateral, SI, and T/KH) not only can be processed and retrieved in one integrated system but that certain advantages will accrue therefrom.

- That personnel organization by geographic area and, if necessary, by topic is feasible and desirable.

- That the CHIVE indexing approach will provide at least as many entry points to documents as that now obtainable from the sum of the individual indexes and other controls established in the various registers of OCR.

- That header data (bibliographic) indexing can be performed by clerical personnel with a minimum of guidance.

To test these concepts, an experimental [redacted] 25X1A Branch was established. The Branch was organized into four topical sections: Political, Economic, Military, and Scientific and Technical. Each section was headed by a senior indexer. More than half of the OCR personnel assigned to the project had some previous indexing experience, but less than half were currently full-time indexers. Each section was allotted personnel who had experience in working with SI materials, a
25X6 [redacted] background, or familiarity with the Intelligence Subject Code. Some of the individuals had more than

one of these attributes. Unfortunately, few of the indexers had previous topical specialization similar to that employed in the experiment.

The indexing tools used during the experiment included:

- The Intelligence Subject Code
- 25X6 A listing of [redacted] on whom the Biographic Register maintains dossiers
- 25X6 The Special Register [redacted] Manual
- 25X6 The NIS Gazetteer [redacted]
- The Special Register Code Book Supplement to the ISC
- The CHIVE Indexing Manual
- Miscellaneous dictionaries and other reference works.

The Intelligence Subject Code and the SR Code Book Supplement were used to index subjects and commodities. The BR dossier list, SR Organization Manual, and the NIS Gazetteer were used as authorities for entering people, organizations, and place names--that is, whenever a significant person or organization was encountered, the indexer had to refer to the dossier list or organization manual, find the correct entry, and enter the code assigned by BF or ,R. All place names were checked in the NIS Gazetteer for the correct entry form. The CHIVE Indexing Manual contained the explanation of the indexing techniques, the method of transcription, and some preliminary indexing rules and procedures.

The data base was all-source and consisted of Collateral intelligence reports, translations, the FBIS, newspaper articles, Comint, [redacted] 25X6
[redacted] T/KH materials, and miscellaneous 25X6
other series. Each document category was represented in proportion to the total documents currently received in that category during a year. All of the documents concerned [redacted] 25X6
tions with other countries.

For documents which contained multi-country/subject content, the rule was established to index that material which would normally be processed by
25X1A an operational [redacted] Consistency was dif-

ficult to obtain here because the indexers had slightly different interpretations as to what a

25X1A [redacted] would process.

It was decided not to apply any selection criteria, but to index all of the information

25X6 concerning [redacted] As a result, many low-level personalities and installations, as well as fragmentary subject matter, were indexed which would not be captured in an operational system. No selection criteria were applied because it was felt that realistic criteria could not be established prior to the experiment and that artificial criteria would affect the experimental results. It was further felt that great indexing depth would aid in establishing future criteria--that is, that the experiment would show that redundant indexing of many subjects is unrealistic. However, despite the lack of criteria, an indexing consistency test following the experiment showed that each indexer tended to apply his own criteria based on his views of what was important.

The documents were broken out into the four topical categories mentioned above. Each senior controlled the flow of material to his indexers thus assuring that each processed a variety of sources. Upon completion of the indexing, the seniors reviewed the transcript sheets for accuracy and logic. However, many errors were not caught because neither the indexers nor the seniors were as well versed in the system as would be desirable in an operational system. In fact, it would be fair to say that it was not until the end of the experiment that the indexers and seniors were beginning to gain confidence in what they were doing. In addition, several of the indexers were not suited to the task and would have to be given other assignments in an operational system.

Following review by the seniors, the documents and transcript sheets were transmitted to the three typists for header data transcription. One of these clericals acted as a senior for resolving problems. In addition, an OCS system analyst who had planned the header data transcription task oversaw this phase of the operation. The documents were then filed by a CHIVE accession number, and the transcript sheets were transmitted to key punching. Computer processing resulted in a print-out of index records which contained errors. These listings were reviewed by one of the project monitors and final corrections were made.

5.B.2. PRELIMINARY FINDINGS

The final results of the experiment await the conclusion of the query phase. However, preliminary findings relating to indexer reactions, selection problems, indexing times, etc., can be described, and these are perhaps the critical factors affecting the organization of the proposed CHIVE system.

5.B.2.1. Personnel Considerations

The personnel involved in the experiment were college-graduate professionals and less than half had worked in jobs that involved full-time indexing. Even those with an OCR indexing background, had worked on allied tasks such as querying or dictionary building, or had served as experts on some aspect of indexing. In this experiment they did nothing but index and found the tools and rules with which they worked inadequate and frustrating. Although there was good morale and rapport throughout the experiment, most found it a tedious experience.

The following are the major problems which seemed to contribute to this negative reaction:

- Few professionals desire to pursue a full-time indexing career.

- During the experiment, there were no special projects or queries for the indexers to engage in to relieve the tedium of indexing.

- There was insufficient training prior to initiating the experiment. Several months training would have made the indexers more confident.

- There were not enough rules prior to the initiation of the experiment, and the formulation of procedures often involved considerable discussion and change on a daily basis. This was a necessary procedure, but it encouraged inconsistency and lack of confidence.

- The lack of selection criteria was harmful psychologically. The indexers were told to index material they believed to be worthless based on their previous information processing experience. This affected their indexing rates.

- The indexing tools were inadequate. The ISC needs revision for an operational system. The BR and SR identifier lists do not contain enough information for proper identification, and the NIS Gazetteer is weak in its coverage of [redacted] names.

25X6

- The use of the identifier lists was time consuming. The lists used by each indexer took up approximately two feet of space and several of the indexers commented that they seemed to spend most of their time referring to one authority or another.

Some of the above problems are a necessary adjunct to an indexing operation, but most would have to be resolved or their effects lessened to have a properly motivated organization. The main problem is to find people who will accept indexing as a career or, at least, as an interim occupation, assuming they are supported with the proper tools and procedures.

5.B.2.2. <u>The All-Source Issue</u>

A preliminary evaluation suggests that there are no severe problems in processing all-source materials and a number of definite advantages. From the retrieval point-of-view, the principal benefit is that an all-source response can be obtained from the system. There are also advantages from the input processing point-of-view. Each of the major sources contains information which fills gaps in the reporting of other sources. Even during the short span of the experiment, there were occasions when Comint would positively identify the activities of an installation that were reported in an entirely different context in collateral.

The indexers did not express any particular problems in coping with multiple sources, and some seemed to feel that Comint was easier to handle because it is more factual and straightforward. The main problem encountered was in the use of indexing tools which were designed for Collateral or Comint, but not both. For example, there was difficulty in using the SR Organization Manual when processing Collateral due to the different terminology used to reference organizations in Comint material. This problem would have to be overcome through the development of all-source tools.

5.B.2.3. <u>Indexer Specialization</u>

One of CHIVE's earliest and most fundamental concepts has been that an indexer/analyst specializing geographically and topically can satisfactorily process all of the information in a document. The geographic/topical concept held up very well during the experiment, and most of the indexers felt they gained some knowledge [redacted]. This is not surprising since even the OCR subsystems which are currently organized topically (e.g., BR on the basis of biographics) at the first level, are organized geographically at the second. The main question is whether the experimental first-level geographic organization is superior to first-level topical organization. It will be difficult to get substantive proof of this out of the experiment. Intuitively there seems to be a significant advantage in having all of the people dealing with one area located together, and this feeling is strongly reinforced by the remarks of the indexers in their critiques.

25X6

5.B.2.4. Indexing Depth and Exhaustivity

Depth of indexing refers to the generic level at which subjects or concepts are described, e.g., whether to index personalities by general topic, by name only, or by name and other attributes, such as birth date. Exhaustivity is concerned with the extent to which every concept dealt with in the document is reflected in the index, e.g., should all people be indexed or only selected persons.

The experimental indexing was both deep and exhaustive. Many attributes of personalities, organizations, and locations were transcribed and key words were applied to subject indexing to provide greater specificity. The results of the query phase of the experiment will provide guidelines as to the requirements for precision in an operational system.

As stated earlier, no selection criteria were applied. Consequently there was no control over the exhaustivity of the indexing. This would not be practical in an operational system. Not only would it result in high costs, it would load the file with data which would rarely, if ever, be retrieved. During the experiment, its effect was to lower motivation and increase indexing time significantly. The average of 10 documents indexed per day per indexer achieved in the experiment could not be tolerated in an operational environment. Arriving at reasonable and responsive selection criteria is one of the most critical problems facing an operational system.

5.B.2.5. <u>Indexing Tools</u>

The indexer aids used in the experiment were in general developed for other applications and were not well suited to the system employed. Moreover, insufficient training time was allowed to become completely familiar with their use. Indexers received three or four weeks training on the ISC as used in the Intellofax system. It takes several months to become familiar with the ISC, and learning Intellofax techniques which had to be unlearned proved a hindrance. The SR supplement to the ISC proved largely duplicative of ISC subjects and tended to further confuse ISC use. The ISC was employed to index organization types and occupations, an application for which it was never intended. As a result, consistency was very low in this area. The SR organization manual was intended for use on SI documents and proved difficult to use on Collateral materials. The BR dossier list was not informative enough for non-biographic specialists, making it difficult to match people referred to in documents with those in the list. The NIS Gazetteer supplement did not have sufficient place-name depth or adequate cross-referencing.

It would not have been practical to develop better tools for the experiment in view of the time and effort required, but the experiment did show the weaknesses of these tools and gave an indication of the amount of effort needed to modify them for an all-source, all-subject system.

5.B.2.6. Header Data Indexing

Header data indexing was accomplished by three clericals after several weeks practice during which uniform entry standards were established. This was one of the most successful aspects of the experiment, and the CHIVE contention that this task could be performed by non-professionals with a minimum of guidance was upheld. Despite the fact that the header data indexing was more specific than that practiced by any current OCR system, the three clericals were able to average better than 15 documents an hour. This figure may not seem particularly high, but it could easily be doubled with better transcription sheet design and the development of continuous roll forms. Moreover, production was affected by the learning process and by a multitude of other tasks performed by the clericals, e.g., batching and filing. Although none of the clericals particularly enjoyed typing--as stated in their written critiques--they found the header-data typing more challenging and interesting than most typing assignments.

5.B.3. FEASIBLE ALTERNATIVES IN INDEX DESIGN

Based on the foregoing, the following CHIVE concepts were reinforced by the experiment:

- All-source system
- One-time indexing
- Geographical/topical organization

- Header-data indexing by clerical personnel

Problem areas, however, which must be overcome in an operational system include:

- The CHIVE contention that medium and high-grade professional personnel could satisfactorily index all of the information contained in documents was not disproved, but personnel of this caliber are not motivated by an indexing career.

- The CHIVE view that greater depth of indexing is required on more sources was not disproved, but in-depth indexing must be selectively applied to obtain reasonable indexing rates. Without good selection criteria, in-depth indexing is impractical given existing manpower constraints.

- Indexing tools and rules need revision for an operational system.

Alternative configurations which might overcome the most significant problems uncovered during the experiment, while retaining those features which promote a more efficient input and retrieval system, are briefly reviewed below. The need to revise indexing tools is not considered further in this discussion since this requirement would exist for any integrated OCR system. All of the alternatives incorporate a geographically-organized, all-source system supported by computers.

5.B.3.1. Alternative A - Reduce the Complexity of the Indexing

This approach would retain the frist-level geographic organization, but indexing time would be lessened by reducing the number of elements of information to be captured and/or by limiting the degree of linkage between terms in index records. Both of the latter steps would, naturally, severely restrict the system's recall capability. Reduction of the number of term types would mean that less control would be provided over people, organizations, and subjects which would ultimately lead to some loss of information on retrieval. Reduction of linkage would result in mismatches or false combinations of terms on retrieval which would increase the output volume as well as analytical time required to review the products of the search.

Advantages of Alternative A

- Less average indexing time per document.

- An integrated geographic system using common input tools and providing good communication between system components and between system and customers; an all-source, all-subject retrieval capability from a minimum of service points, usually one; and a minimum of overlapping processing and files.

- Greater uniformity in the application of selection criteria because of the geographic organization of the input activity.

- Broad index control over information.

Disadvantages of Alternative A

- Index control in no more depth than is available today.

- A reduction in job satisfaction compared to the analytic functions performed by BR and FIB analysts.

- Does not overcome the selection problem.

5.B.3.2. Alternative B - Reduce the Number of Documents Input to the System

This system would be similar to the previous configuration in that it would retain the integrated organizational advantages and lessen indexing time by reducing the indexing workload. In this case, however, the amount of information indexed rather than the indexing system itself would be restricted. The plan assumes that reasonable selection criteria can be developed which would obviate the need to consider many series of documents for indexing, and that the selection of information to be indexed in other series would be restricted. The notion that many series of documents would not have to be reviewed may not be entirely realistic based on past experience, but there is little doubt that what should be indexed has to be less than the total information reported. It is a matter of degree. We all recognize that OCR must have good selection criteria--this proposal assumes very stringent criteria.

<u>Advantages of Alternative B</u>

- Total indexing time reduced.
- An integrated geographic system using common input tools and providing good communication between system components and between system and customers; an all-source, all-subject retrieval capability from a minimum of service points, usually one; and a minimum of overlapping processing and files.
- Uniform selection criteria.
- In-depth control over that information which is indexed.

EVALUATION OF EXPERIMENT
Feasible Index Design
5.B.3.2.

Disadvantages of Alternative B

- A severe reduction in the amount of information brought under control--perhaps less than is controlled by OCR today.

- A reduction in job satisfaction compared to the analytic functions provided by BR and FIB analysts.

5.B.3.3. Alternative C - Divide the Responsibilities for Input Selection and Analysis from Index Preparation

Under this approach, analysts would review the incoming mail, mark the documents or portions of documents to be indexed, maintain information files, prepare written reports, and be the primary point for dealing with customers and answering queries. The amount of analyst specialization would largely depend on the size, volume of reporting, and priority of the various geographic areas. For large Bloc countries there might be a high degree of specialization. For smaller countries it is likely that one analyst would process all of the information on one or several countries. The most appealing aspects of this concept are that it has associated with it a high degree of professionalism, with resulting job satisfaction, and that selection would be managed on a geographic basis by an analyst in close rapport with his customers and knowledgeable about the reporting on his area.

Indexers would index those materials chosen by the analyst. In theory the analyst would be highly selective because, with his familiarity with the document traffic and customer needs, duplicative and low-level reporting could be eliminated. The selection problem would, therefore, be alleviated, and only those materials would be indexed which were most responsive to customer requirements.

Indexers could be organized in one of two ways--either as a central indexing group organized geographically with a T/O separate from the analytic division, or integrated with the analytic divisions. The advantage of the separate group would be the increased likelihood of achieving greater consistency in the application of the indexing rules and better control over the adoption and use of indexing tools. The advantage of the integrated group would be closer rapport between the indexers and information analysts.

This organization differs from earlier CHIVE proposals which assigned to one person the responsibilities for input selection, querying, and indexing. In this configuration, the information analysts would operate in a mode similar to the ORR or OSI analysts and would have commensurate grades and responsibilities. Indexers would be semi-professionals in the GS-5 to GS-7 range, i.e., bright people but not necessarily college graduates. A fairly high turnover rate could be tolerated among the indexers, as is true in most OCR indexing operations today. Indexer career development would involve moving to senior indexing positions, or, for those who possessed the right background and capabilities, moving into analytic positions.

Advantages of Alternative C

- An integrated geographic system using common input tools and providing good communication between system components and between system and customers; an all-source, all-subject retrieval capability from a minimum of service points, usually one, and a minimum of overlapping processing and files.

- High job satisfaction among the analytic group and potentially high job satisfaction among the indexing group assuming the right kind of personnel are placed in or recruited for these positions.

- The most practical solution to the selection problem.

- In-depth control over that information which is indexed.

Disadvantages of Alternative C

- Potentially more expensive than the other alternatives since it involves not only in-depth indexing but also duplicative reading by analysts and indexers.

- Since selection is more personalized and perhaps less uniform than in the other alternatives, there is a potential loss of information, particularly information of marginal value.

5.B.3.4. Conclusion

The first and second alternatives discussed both offer some advantages over the experimental organization by reducing indexing times. However, the OCR professional would still spend a significant proportion of his time as an indexer and low job satisfaction would result. The final alternative seems to offer considerable promise in that it appears to be the most reasonable solution to the selection problem, and it retains the analytic and indexing functions of OCR in an integrated organization. It is recommended that this alternative be investigated further and be subjected to follow-on experimentation.

Appendix 5.C.

CHIVE INDEXING GUIDE

5.C.1. INTRODUCTION

This indexing guide should be viewed as an initial draft of an ultimate CHIVE Indexing Manual. Many of its provisions will, of necessity, be changed as other elements of the system design become more fully developed. Transcript sheet formats, for example, will be significantly influenced by design decisions in the data conversion area regardless of whether a page reader or some other data conversion device is used. It is obvious, therefore, that the transcript forms presented here will undergo revision. Similarly, other forms and procedures will be revised to reflect the final conclusions resulting from the CHIVE Indexing Experiment.

In its present form, the guide represents some modification of the manual used in the CHIVE Indexing Experiment. Deleted from the guide are those instructions which related to the organization of the experiment, the philosophical justification of the CHIVE system concept, and sections which dealt with the issue of indexing depth. The guide is generalized and does not relate to any specific geographic area. In an operational system, each geographic organization would be expected to develop certain of its own procedures (for example, how to enter personal names or how to process document series peculiar to an area) which would serve as supplements to the basic guide.

5.C.2. CONTENT INDEXING SYSTEM

5.C.2.1. <u>Elements of Information and Index Entries</u>

The CHIVE Indexing System defines 59 kinds or elements of information which are to be indexed (see Tab B). These elements include NAME OF PERSON, NAME OF ORGANIZATION, NAME OF CONFERENCE, etc. Each element of information has been assigned a three-letter mnemonic tag for identification purposes. The tag PNO, for example, represents the element, NAME OF PERSON; the tag ONO the element, NAME OF ORGANIZATION; the tag CNC the element, NAME OF CONFERENCE; and the tag LAC the element, COUNTRY OR PROVINCE LOCATION.

The tags are always enclosed within parentheses whenever they are used. The first letter of a tag suggests the element being identified, that is, a P suggests a Person, an O suggests an Organization, a C suggests a Conference and an L suggests a Location. The use of the N as the second letter in three of the above tags suggests the concept, Name. Wherever possible a mnemonic structure has been used.

An index entry consists of one of the 59 tags followed by the appropriate "value." For example, consider the following three index entries:

(PNO) [redacted]

(ONO)

(LAC)

25X6

The use of PNO indicates that the value which follows is the name of a person; the use of ONO indicates that the value which follows is the name of an organization; and the use of LAC indicates that the value which follows is a geographic place. The terms which may be used as a value and the format for entering these terms are governed by specific tag rules (see Tab B). For example, the values for the POH tag, HEAD OF ORGANIZATION, and the PHT tag, HONORIFIC TITLE, consist of terms taken directly from text:

(POH) PRESIDENT

(PHT) GENERAL

The values for the PDB tag, BIRTHDATE, and the LSN tag, STREET NAME, consist of terms taken directly from the text. However, they must be entered according to a fixed format:

(PDB) 240512 (meaning 12 May 1924)

(LSN) MARDOV STREET, 163

The values of other tags such as the ONO must be established by referring to authority files or identifier lists.

5.C.2.2. Identifier Lists

5.C.2.2.1. NIS Gazetteer

The NIS Gazetteer is the authority for entering place names. If the place name cannot be located in the gazetteer, enter it as it appears in the document, and, in addition, place a question mark (?) between tag and entry so that it can be captured for supplemental listing.

5.C.2.2.2. Organization List

This list should always be consulted to obtain the approved entry number (code) for organizations or installations. If an organization cannot be found in this list, enter it in clear text as spelled in the document, and, in addition, place a question mark (?) between tag and entry so that it can be captured for supplemental listing. The organization list is provided in three sort orders: (a) organizations within place name, (b) alphabetically by organization name or number, and (c) hierarchically by subordination within the country administrative structure. Some types of organizations will not be entered by name either because of their relative unimportance or because name control is not feasible. In such cases, an ISC subject code will be used in lieu of name, preceded by an OTF tag. (See section 5.C.2.9.3 for a list of organization types which fall in this category.)

5.C.2.2.3. Personality List

A personality identifier list, comparable to that maintained for organizations, will be available within the system. However, it is not intended that this list will be used to control names on input. Rather, names will be entered in clear text with the PNO tag substantially as reported in documents. Following the servicing of queries, however, the identifier list will be used to record identifying data about personalities on whom physical or logical dossiers have been established, and to provide a table of contents to documents known to be pertinent to a given personality.

5.C.2.3. <u>The Intelligence Subject Code</u>

5.C.2.3.1. Subject Indexing Authority-the RSC Tag

The ISC is used as the authority for indexing subject elements of information. These elements are distinct from the "named object" elements of information discussed above. The subject elements correspond with those subjects which are contained within the seven chapters of the ISC. The RSC tag is used to identify these elements. For example:

(RSC)621000 (METALS)

(RSC)319625 (ATOMIC STRUCTURE)

5.C.2.3.2. Occupation Type and Organization Function Codes-the PJP and OTF Tags

Separate code schedules will be used to index organizations by function (e.g., steel plants) and personalities by occupation (e.g., chemists, physicists, etc.). These code schedules are separate from the ISC, although they may eventually be incorporated into the ISC as appendices. These code schedules will be used with the PJP and OTF tags.

5.C.2.3.3. Subject Modifier List-the RMO Tag

The ISC is also used as the authority for the selection of the subject modifiers which are combined with ISC subjects. These subject modifiers are used to indicate the aspect or the role of a specific tag value. The modifiers can be freely applied so that any modifier which appears in the ISC may be applied to any subject within the ISC. In addition to the modifiers listed in the ISC, there are so called "indicator attributes" which refer primarily to personalities (e.g., 21(= FAMILY DATA) or to organizations (e.g., 026 = UNDERGROUND INSTALLATION). The RMO tag is also used to identify these modifiers (see Tab B).

5.C.2.4. Phrase Construction

As previously stated, the CHIVE index entry consists of a tag, such as a PNO, ONO, or RSC which identifies the particular element and the value which represents the actual information found in the text. The terms which are used as values may in some cases be identical to the terms used in the document. In other cases, the terms used are subject to the format rules and the authority files discussed above.

The index entry is used to represent a basic element of information. It follows then that several index entries must be used to represent several elements of information. The problem, however, arises as to how the index entries should be linked together in order to retain the syntactical properties of the original elements of information. Consider the following information:

25X6

[redacted]

Each underscored term, or set of terms, represents an element of information. Four indexing terms must be used to represent these elements. These are:

26 (PNO) [redacted]

26 (POH)

26 (ONO)

26 (LAC)

25X6

The syntactical properties of the original elements of information are retained by including each of the above index terms in the same logical phrase.

The number on the left in the above example is the phrase number and indicates that each index term belongs in phrase 26. By definition, the fact that these terms are in the same phrase allows the index terms to retain the same syntactical relationships as the elements of information in the original English sentence. Ideally,

INDEXING GUIDE
Content Indexing System
5.C.2.4.

a phrase with its index terms may be likened to a simple English sentence. However, care must be exercised in translating the elements of information in the original sentence to the index terms in the logical phrase. Consider the English sentence:

[redacted]

25X6

The six information elements have been underscored. Note that the two persons each have different positions with the institute. There is no way of distinguishing this fact in one logical phrase. Hence the following two phrases must be used:

[redacted]

25X6

Effectively, this sentence with a compound subject has been reduced to two sentences each having a simple subject, and then each sentence has been translated into

INDEXING GUIDE
Content Indexing System
5.C.2.4.

a logical phrase.

5.C.2.5. <u>Use of Line Items</u>

Phrase numbers 26 and 27 above carry redundant [redacted]

25X6

These terms which may be factored out, as it were, from the two phrases need be written only once. This is accomplished by entering the two terms as <u>line items</u> instead of entering them as phrase entries. The line items are then added to the phrases. Consider the following:

PHRASE NO.	LINE	ENTRY	LINKAGE
	A	(ONO) [redacted]	
	A	(LAC) [redacted]	
26		(PNO) [redacted]	
27		(PNO) [redacted]	

25X6

The linkage indicator on the far right in the above example indicates that the line item A must be reintroduced into phrases 26 and 27 in order to complete the logical phrases.

INDEXING GUIDE
Content Indexing System
5.C.2.5.

The reintroduction of the line items to complete the logical phrases is accomplished in machine processing and hence the indexer need not transcribe the terms into each phrase.

5.C.2.6. Example of Transcript Sheet

A sample of a completed CHIVE transcript sheet is enclosed (see Tab C). On this example, index terms have been entered to represent the following information:

25X6

[redacted]

INDEXING GUIDE
Content Indexing System
5.C.2.6.

line item (line B) for the same reason. The index terms entered in line C will be used in more than one phrase in order to show the organizational affiliation of the two individuals. The question mark used with the (ONO) is a special character which indicates that the organization name was not taken from the Organization Dictionary (i.e., it could not be found) so it was entered as it appeared in the document. (See section 5.C.2.9.7. for use of question mark.)

[redacted] 25X6

done by entering the index term (POH)PRESIDENT and then entering the line designators A and C in the spaces on the right side of the sample sheet. By convention, this indicates that those index terms on line A and on line C must be introduced into the phrase in order to complete the phrase. Hence phrase 26 consists of the index term (POH), the index term on line A, and both index terms on line C. In summary, phrase 26 has the index terms: (POH)PRESIDENT, (PNO)[redacted], (ONO)?RSCH 25X6

INDEXING GUIDE
Content Indexing System
5.C.2.6.

25X6

at the institute.

In phrase 29, the tag/value (LAC)1252 signifies travel to Indonesia and (PDT) the date span of travel. As in phrase 28, the names and nationality of the persons travelling are also shown. Note that the (PLC)1013 index term was entered into both of these phrases. One could have made this index term a line item.

5.C.2.7. Transcription Procedure (see transcript sheet, Tab C)

a. Columns 1-2-Enter phrase number (26-99). If columns 1-2 are filled in, column 3 is blank.

b. Column 3-Enter line designator (A-Z). If column 3 is filled in, columns 1-2 are blank.

c. Columns 4-61-Enter index terms.

- Multiple entries can be made on one entry line as long as they are included in the same phrase number or alphabetic linkage line.
- Tags must always be enclosed in parentheses.
- Follow entry rules included in tag listing (Tab B).
- When transcribing data, do not leave spaces unless spaces are required, e.g., between words.

- To prevent keypunch error, write the following letters and numerals as illustrated:

One	- 1	Seven	- 7
Eye	- I	Letter F	- F
Zero	- 0	Letter U	- U
Letter o	- ø	Vee	- V
Five	- 5	Letter S	- $

d. Columns 62-72-Enter alphabetic symbols for linkage line entries which are being added to phrases. Order of entry is not significant. Left justify.

e. Columns 73-74-For documents over five pages, enter page number where indexed information is located in document. Append page numbers to phrases only. If phrase relates to several pages, cite earliest page number. Right justify with leading zeros.

f. Columns 75-78-Enter CHIVE document accession number. Fill out accession number with leading zeros, e.g., document 99 = 0099.*

g. Columns 79-80-Enter item accession number for documents which contain multiple articles (items), e.g., the FBIS. The item number (01-99) is assigned by the indexer for each item indexed. Fill out the number with a leading zero, e.g., item 2 = 02. Fill in zeros if not used.

h. Indexer should fill in his name in the space

---

*This instruction will no doubt be changed since current thinking is to use the USIB common control number (when available) for documents rather than some arbitrary accession number.

provided at the top of the index form, indicate number of transcript sheets used per document (e.g., PAGE 1 OF 1, PAGE 2 OF 3, etc.), and check classification of filled-in transcript sheet (same as document) in boxes provided.

5.C.2.8. Title Expansion

In some cases the documents or the individual items within the document will have very terse indicative titles, for example:

- USSR Steel Shipment to China
- Somali Labor Leader
- Western Military Developments Reported

These titles should be made more complete and useful by expanding or modifying them so that they more closely reflect the index terms applied to the document. This expansion may be made by adding certain informative subjects to the title, for example:

[redacted] 25X6

This expansion may also be made by adding certain indicative subjects to the title, for example:

- [redacted] 25X6

25X6 [redacted]

For short, single-subject documents the informative title expansion is most practical and useful. For long documents, containing a variety of information, the indicative expansion will normally be used. If adequate space is available, the expansion should be written on the document. Otherwise, write the expansion on a separate sheet of paper and attach it to the document.

5.C.2.9. <u>Special Indexing Instructions</u>

5.C.2.9.1. Areas

(1) The area entry is, in general, the key or controlling element in a phrase, particularly in phrases containing named objects. All organizations and personalities in a phrase must be logically related to the area (LAC) in that phrase. Therefore, in phrases involving the following organizational relations:

(ONO) Name of Organization,

(PEO) Higher Educational Institute Attended, (POT) On Temporary Duty at an Organization,

(POL) Organization Visited,

the location of the organization is the proper area entry. This also means that two organizations (or people) cannot be entered in the same phrase unless they are located in the same area. For example:

25X1B

[redacted]

An exception to the above is the CSS entry which refers to the sponsoring organization of a conference. This need not apply to the same area as the location of the conference. However, if the document contains significant information about the sponsoring organization, it should be re-entered with an ONO tag and the LAC area for the organization location in a separate phrase.

(2) The area tags LAF (COUNTRY FROM), LAN (NATIONALITY), LAR (COUNTRY COMMENTING OR REACTING), and LAT (COUNTRY TO) can be used only with ISC subject indexing. The area rules printed in the ISC apply to the use of these tags, e.g., Primary Area = LAF, Secondary Area = LAT, Nationality = LAN, etc.

(3) Every phrase must contain an area entry.

(4) The LSN tag (STREET NAME AND NUMBER) may also be used to represent non specific locations such as: "next to the library," "east of the Shan mountains," "at the confluence of the Han and Ho rivers," etc. These terms are entered as clear text, for example:

(LSN) NEXT TO THE LIBRARY

The use of the LSN tag does not preclude the use of the LAC tag. The latter must be used in the same phrase with the LSN tag.

(5) The LAF and LAT tags are used primarily to name the countries which are mutually engaged in certain activities, e.g., shipment of goods, political agreements, foreign aid, etc. However, for certain specific activities, such as international air routes or movement of foreign aid materials, place names may be appended to the country code. For example, consider an international air route:

(RSC)563200 (LAF)[redacted] 25X6

(LAT)5103,TIRANA

The use of this technique is limited to those ISC codes so marked.

(6) Foreign travel is indexed by use of the LAC tag to show the country of travel, and the PLC tag to show the nationality of those who are travelling. The PLC tag may be used to show travel without having the names of persons involved. For example, if it is known

25X6 that [redacted] had representatives at a certain meeting [redacted]s fact would be indexed as follows:

(CNC) Conference Name

INDEXING GUIDE
Content Indexing System
5.C.2.9.1.

(LAC) 1205

(PLC) 1013

5.C.2.9.2. Organization Function (OTF) and Occupation (PJP)

(1) Whenever an organization type or occupation is stated or can be derived for an ONO (ORGANIZATION NAME) or PNO (PERSON'S NAME) entry, it should be entered with the OTF or PJP tags followed by the appropriate code. A PJP should be carried with a PNO whether the person is in his own country, at a foreign conference, traveling, or other.

(2) Communist Party membership is not considered an occupation unless the individual is an officer of the Party or, for non-officers of the Party, it is their full-time occupation. If a CP officer holds another job (e.g., plant manager, chief of a military region, etc.), use an additional PJP entry.

(3) Clear text may be used with the OTF and the PJP tags. For example:

(PJP)123000, MAYOR

Clear text is used to provide greater specificity to organization function and occupation titles.

INDEXING GUIDE
Content Indexing System
5.C.2.9.2.

(4) Some organizations will not be entered by name because they are too low-level to be of significance for name control (see section 5.C.2.9.8.), while the names of other organizations may not be reported. In these cases, minimum control over these organizations will be obtained by indexing the organization function, OTF tag, and location.

5.C.2.9.3. Biographic Information (PNO)

(1) For those documents which contain particularly detailed biographic information on an individual, only the following should be indexed:

- All current organizational affiliations.
- The last previous organizational affiliation (only if this previous affiliation occurred within the last 5 years).
- All other biographic data relative to education, current addresses, current travel, current awards, current occupation and titles, current leader appearances, etc.
- Names of persons with whom the individual has been associated only if there is significant information about these persons (e.g., positions held, universities attended, etc.).

5.C.2.9.4. Use of Subject Modifier (RMO)

(1) The ISC modifiers are freed for use with any

ISC code. They are also freed for use with ONO and OTF whether or not an RSC is used. However, care must be exercised since, whenever an RSC, ONO, and OTF appear together in a phrase with an RMO, the RMO must apply logically to all three index terms.

(2) The subject modifier (RMO)023 is used for equipment descriptions, i.e., the description of specific equipment at a facility. The modifier (RMO)025 is used for plant descriptions in general. The modifier (RMO)040 is used for the description of a product (of a plant).

5.C.2.9.5. Phrasing

(1) The same organization, conference, or person may appear in several different phrases for the same document with the same tag for each entry. This is required in order to prevent logical inconsistencies in phrasing. However, the attributes of these named objects, other than country and place name location, need to be entered only once. In other words, do not repeat such attributes as organization type, honorific title, parent organization, etc., in multi-phrases for the same named ob-

jects since this would result in a number of redundant entries. For example, the reporting of the death of an individual or the appointment of an individual to a position need be indexed only once for a specific document. It should not be repetitively indexed every time his name is entered.

5.C.2.9.6. Abbreviations

(1) When entering an organization name, place name, position title, clear text with an ISC code, or any other English text, use abbreviated forms for the text only if these forms are on the approved list. This list (see Tab E) includes such forms as: PL = Plant, CP = Communist Party, etc. If a term in the document is in abbreviated form, enter this form only when the unabbreviated form is not known.

5.C.2.9.7. Use of the "?" Character

(1) The special character "?" is used to mark certain index entries for special maching listings. These listings may be used as an aid in thesaurus building or as an aid in retrieval. For example, the "?"

will be used with the organization name tag whenever the organization's name cannot be found in the Organization Dictionary. In such a case, the name is entered as it appears in the document. For example:

(ONO)?RESEARCH INSTITUTE

A special listing of all organizations not found in the Organization Dictionary can be prepared from these records.

(2) The "?" is also used with the LAC tag whenever a place name found in a document cannot be found in the NIS Gazetteer. In such cases the place name is entered as it appears in the document. For example:

25X6 (LAC)[redacted]

The tags that are to be used in this manner have been so designated in the tag list (Tab B).

5.C.2.9.8. Identifier Lists

(1) The organization/facility types listed below are not entered in the Organization Identifier List and will not be entered in clear text with an ONO tag entry. However, they should be indexed by type of organization (OTF).

Approved For Release 2000/05/30 : CIA-RDP78-03952A000100050001-7

(2) Attributes of organizations and place names which appear in identifier lists need not be indexed if the same information is reported again in a document. If the reported information differs from that in the identifier lists, it should be entered, e.g., a document might cite a different parent organization for a facility from that which appears in the Organization Identifier List.

5.C.3. HEADER DATA TRANSCRIPTION GUIDE

5.C.3.1. Introduction

Header data entries will be typed by clerical personnel directly on the Header Data Transcript Sheet (see Tab D). In section 5.C.3.2. are general instructions and definitions for header data indexing. In section 5.C.3.3. are specific instructions relating to entry rules for each field on the transcript sheet. Various code schedules used for specific header data fields are included in Tab A.

The authority or source cards referred to in section 5.C.3.2. are a group of 3"X5" cards which contain format rules for entering source references. Because of their physical makeup, they have notbeen included in this guide. For security reasons, the [redacted] Technical Manual has also be excluded.

5.C.3.2. General Instructions and Definitions (see transcript sheet, Tab C)

Abbreviations: Do not enter abbreviations for title words unless they are abbreviated in the document, either by the originator or by the indexers. Titles will be permuted by a machine program and abbreviations will not be as meaningful as the full words.

Authority: An authority or source card must be created for each unique document type. This authority card must show the complete organizational position of the originator and necessary cross-reference cards are also required. For example, an [redacted] would be sourced and cross-referenced as follows:

25X1C

Cross Reference: [redacted] see CIA, OFFICE OF OPERATIONS, CONTACT DIVISION

25X1C

Source Authority: CIA, OFFICE OF OPERATIONS, CONTACT DIVISION

Line and Field Number	Entry
02-01	CIA, [redacted]

25X1C

Line and Field Number	Entry
02-02	[redacted]
03	3210000064
04	BLANK
05	AS GIVEN

25X1C

It is essential that entries in the specific fields conform exactly to those shown on the authority card down to and including abbreviation, punctuation, and spacing. If a card is already in the source card file, follow the transcription rules outlined. If no card is in the file, a new card must be prepared. Follow the format used for other source cards and pass the new card to the senior for review. Cross reference cards will be prepared as appropriate.

Blank Fields: When no data is provided for a particular field leave it blank unless the instruction for that specific field requires the entry of zeros or some other supplied entry (see section 5.C.3.3.).

Classification: The transcription sheet is unclassified until filled in. It then takes the highest classification of the source document being transcribed. If Codeword, check both Codeword and classification.

Clear Text: Record exactly as it appears in the source document being transcribed except where more specific rules are given elsewhere in this manual, or in the source authority card file.

Content Transcript Sheet (Definition): The transcript sheet used by professional indexers to record subject, area, personality, and organizational information from the document.

Coordinates: Type without spaces, e.g., 2305N11424E. Separate from rest of title by commas.

Document Numbers:

Common Control Number: Drop any slashes appearing in the 10-digit common control number. Certain common control numbers include both alpha and numeric characters. Type these as one number with no spaces separating the alpha from the numeric portion.

SI Numbers: Type SI numbers exactly as they appear on the document. Retain all punctuation and special symbols.

All Other Numbers: Drop all punctuation except slashes from these numbers. Occasionally a dash will be used to separate the number from the year of the report. In these cases substitute a slash. Separate the alpha portion of these numbers from the numeric portion with a slash. Substitute a slash for all spaces appearing in document numbers. If there is any question on how to transcribe a given number, refer to the appropriate source authority card.

Field (Definition): Each box or group of boxes on the transcript sheet identified by a heading regardless of length. The one character box for CLASSIFICATION is a single field as is the 5-line, 360-character entry for ITEM TITLE-ENGLISH AND EXPANDED.

Multiple Item Report (Definition): A document or publication which contains more than one individual article. A separate Header Data Transcript Sheet will be created for each item indexed within a multiple-item report or publication.

Multiple Document Numbers: Multiple or secondary document numbers are handled as cross references. All such items will be identified by a separate cross reference card.

Multiple Source (Definition): A document which can be attributed to more than one source, e.g., an Army docu-
25X6
ment with a [redacted] as an enclosure. In this case, the base source would be Army
25X1A
and the secondary source [redacted] Usually, a multiple source document involves an enclosure attributable to a source different from that of the covering or base document. The base document does not receive an item number. Each multiple source will be assigned a separate item number. Separate header sheets will be prepared for each multiple source item.

Punctuation:

Commas: Follow all commas with a space regardless of the card or field in which they are used.

Dashes: Dashes used in the title will be copied but will be preceded and followed by a space. The only exception to this rule is for words which are normally hyphenated. These will be typed exactly as they appear in the document. Spaces preceding and following dashes are necessary to avoid artificial hyphenation of separate words by the machine programs which will process header data.

Hyphenation: When insufficient space is available to complete a word within a box of a multi-box field, continue typing the word to the end of the first box, do not hyphenate, and begin the rest of the word in the first space of the next box in this field. The computer program which will process this data will automatically assemble such split words.

Parentheses: Do not type the parentheses which are used to set off the expanded title from the regular document title. Continue the expanded title as if it were simply a continuation of the given title.

Single Item Reports: Header Data Transcript Sheets will be created for all single item reports, e.g., IR's, SI reports, monographs, etc.

Truncation (Definition): As used in this manual, the term means that an entry in a specific field has been cut off so it will fit in the allowed space. When it is necessary to truncate an entry in any field enter a question mark (?) in the last position of that field.

This will serve as an indicator that the field is too short and may need to be enlarged.

5.C.3.3. <u>Specific Field Instructions</u> (see transcript sheet, Tab C)

The instructions for header data transcription in this section are presented in the same order as the header fields on the transcript sheet. The first number preceding each instruction refers to the line number on the left hand side of the transcript sheet. The second number is an arbitrary number assigned to consecutive fields within each line. This number does not appear on the transcript sheet and is used solely for ease of lookup within this instruction manual.

<u>Definitions</u>:

<u>Line Number</u>: Identifies the line number(s) on the left hand side of the Header Data Transcription Sheet.

<u>Field Number</u>: The number used to identify a particular field on the transcript sheet. In this manual entries on the transcript sheet are referred to by line and field number. For example, the entry for INFORMATION DATE is referred to as 01-07.

Field Name: The clear text name of a given field as it appears on the transcript sheet.

Number of Entries: The legally allowable number of entries which can be entered in any one field, e.g., the field COUNTRIES REPORTED ON may have a maximum of three area codes entered.

Code/Clear: An entry on this line indicates whether the data for this particular box is to be entered in clear text or is to be transcribed in code from one of the code schedules given in Tab A of this manual.

Maximum Record Length (MRL): The maximum number of characters which can legally be entered in any one field.

Required Entry: Identifies those fields in which an entry must be made if the transcript sheet is not to be in error.

Justify: This term means to move the entry to the right or to the left within a given field. For example, in the field ITEM NUMBER at the top of the transcript sheet, numbers from 1-9 are right-justified with leading zeros as 01-09.

Line: All Field: 15 Field Name: ACCESS NR
Instructions
- No. Entries: 1 to
- Code/Clear : Numeric characters recorded from document
- MRL : 4 characters
- Req. Entry : All documents
- Type Docts.: All documents
- Justify : Right justify with leading zeros
- Definition : An arbitrary number assigned to each unique item within the document corpus
- Example : For 101st document enter as 0101
- Comments : The document storage system adopted for CHIVE will determine the need for and length of this entry

Line: All Field: 16 Field Name: ITEM NR
Instructions
- No. Entries: 1 to
- Code/Clear : Numeric characters recorded from document
- MRL : 2 characters
- Req. Entry : All documents containing items
- Type Docts.: All multiple item documents
- Justify : Right with leading zeros
- Definition : Arbitrary numbers assigned consecutively to individual articles within a larger publication
- Example : Third article of the 101st document enter as 03

Line: 01 Field: 01 Field Name: CAT
Instructions
- No. Entries: 1 to
- Code/Clear : Two-character numeric code (see Schedule #1, Tab A)
- MRL : 2 characters
- Req. Entry : All documents
- Type Docts.: All documents
- Justify : Right with leading zeros

Definition : This entry is used to identify the major category of documents to which the particular item being processed belongs

25X1C Example : An [redacted] is considered an IR and would be coded 01

Line: 01 Field: 02 Field Name: CLASS

Instructions

No. Entries: 1 to

Code/Clear : 1 character alpha code, (see Schedule # 2, Tab A)

MRL : 1 character

Req. Entry : Required. Enter Unclassified if no classification given

Type Docts.: All documents

Justify :

Definition : The security classification of the report in hand. This will be either Top Secret, Secret, Confidential or Unclassified

Example : Top Secret, Enter T

Comments : Always enter the classification of the overall report. Do not use this block for the classification of articles within a report.

Line: 01 Field: 03 Field Name: CODEWORD

Instructions

No. Entries: 1 to

Code/Clear : 1 character alpha code

MRL : 3 characters

Req. Entry : Required on all SI material (blank if Collateral)

Type Docts.: SI

Justify : Specific subfield, see comments below

Definition : The word(s) used on SI to indicate the level of protection which must be accorded this material

INDEXING GUIDE
Header Data
5.C.3.3.

Example : KH, Enter K
Comments : The extreme sensitivity of these sources requires the exercise of caution in entering data in this field. If different levels of material are included for the same category within the same document enter the code for the highest level. If more than one category of material is involved, enter the codes for each. Specific subfields: A-E in first column, T in second column, K in third if used. No leading or trailing zeros required.

Line: 01 Field: 04 Field Name: DISSEMINATION
Instructions
No. Entries: 1 to 5
Code/Clear : 2 character numeric code (see Schedule #3, Tab A)
MRL : 10 characters
Req. Entry : Required when applicable
Type Docts.: IR's, SI material, virtually all except open pubs.
Justify : Left justify
Definition : Control markings placed on documents to limit their distribution and use
Example : NO FOREIGN DISSEM, Enter 09
Comments : No trailing zeros required in blank subfields

Line: 01 Field: 05 Field Name: EVAL
Instructions
No. Entries: 1 to
Code/Clear : Code, 1 alpha and 1 numeric character (see Schedule #4, Tab A)
MRL : 2 characters
Req. Entry : Required when applicable
Type Docts.: IR's and SI material
Justify : See comments below
Definition : The originator's evaluation as given in the document, Alpha character refers to source, and numeric to content validity

Example : C-3 indicates "fairly reliable" source and content which is possible true.

Comments : SI material requires only a one-character entry. Refer to SI Technical Manual for definition. If no entry or double source entry, leave blank. Enter either alpha or numeric if only one is given. Alpha always in the first position, numeric always in second position. If statement is worded exacely as in Schedule 4 of this manual, substitute code for statement.

Line: 01 Field: 06 Field Name: ITEM DATE (yr, mo, day)

Instructions

No. Entries: 1 to

Code/Clear : Numeric characters recorded from document in formatted form

MRL : 6 characters

Req. Entry : Required when applicable

Type Docts.: All multiple item reports

Justify : Left justify with trailing zeros

Definition : Varies with type material being processed. For example, for open publications, the date of article within publication or date of the foreign publication from which the article is taken; for intelligence-type summary reports, the date of the article if given.

Example : Jan 3, 1964 = 640103; Jan 64 = 640100; 1964 = 640000

Comments : If any part of the date is entered, fill in balance of field with zeros. If no date, leave blank. Item date is usually the issue date of SI material. Year only, or year and month only, are legal entries. Year and day, or month and day, are not legal entries.

INDEXING GUIDE
Header Data
5.C.3.3.

Line: 01 Field: 07 Field Name: INFORMATION DATES

Instructions

No. Entries	:	2 to
Code/Clear	:	Numeric characters recorded from document in formatted form
MRL	:	12 characters
Req. Entry	:	Required when applicable
Type Docts.	:	All documents
Justify	:	Left, in first and last subfields
Definition	:	The inclusive dates or date span during which the action described in the report occurred.
Example	:	Follow pattern and rules given for item date
Comments	:	If no information date(s) are given, leave blank. First or beginning date must be entered in first six columns, last date in last six. If partial dates are used, they must be entered in columns reserved for year. Possible entries include first year only, last year only, or first and last years. A single date will be entered as beginning date. Typist will not look through the body of the document for this date.

Line: 01 Field: 08 Field Name: TOTAL PAGES

Instructions

No. Entries	:	1
Code/Clear	:	Numeric characters recorded from document
MRL	:	3 characters
Req. Entry	:	All documents
Type Docts.	:	All documents
Justify	:	Right, using leading zeros
Definition	:	The total number of pages in the report or publication in hand, including enclosures and attachments
Example	:	A 1-page document with 2 enclosures, one containing 3 pages, the other 8 pages. Enter 012.

Comments : Some articles within reports and publications may not be indexed. The total number of pages for the publication will, nevertheless, still be entered. To insure standardization, count any page with printing on it in the total page count.

Line: 01 Field: 09 Field Name: ITEM PAGES

Instructions

- No. Entries: 1 to 2
- Code/Clear : Numeric characters recorded from document
- MRL : 6 characters
- Req. Entry : Required when applicable
- Type Docts.: Multiple-item documents
- Justify : Right, with leading zeros in each subfield
- Definition : The inclusive pagination of an article within a larger report or publication.
- Example : For an article beginning on page 13 and ending on page 45, enter 013045.
- Comments : When an article is completed on one page enter that page number in first subfield. Leave last subfield blank. Item pages applies only to the field titled ITEM TITLE-ENGLISH AND EXPANDED. When entering pagination for FBIS articles enter the identifying letter in first column (Column 42) of first subfield, and follow with pagination as given.

Line: 01 Field: 10 Field Name: PUB DATE

Instructions

- No. Entries: 1 to
- Code/Clear : Numeric characters taken from document
- MRL : 6 characters
- Req. Entry : All documents
- Type Docts.: All documents
- Justify : Left with trailing zeros

Definition : The date document was published
Example : Follow the format given for ITEM DATE and INFORMATION DATES
Comments : Publication date is the date the base document was printed for distribution. A base document is the document in hand, e.g., an IR, a translation of a foreign publication, the foreign publication itself, etc. Fill out partial dates with zeros.

Line: 01 Field: 11 Field Name: VOLUME
Instructions
No. Entries: 1 to
Code/Clear : Numeric or alphabetic characters taken from document
MRL : 2 characters
Req. Entry : Required when applicable
Type Docts.: Open publications
Justify : Right, with leading zeros
Definition : The volume number of a publication
Example : Vol. VIII, enter 08

Line: 01 Field: 12 Field Name: EDITION
Instructions
No. Entries: 1 to
Code/Clear : Numeric characters taken from document
MRL : 2 characters
Req. Entry : Required when applicable
Type Docts.: Open publications
Justify : Right, with leading zeros
Definition : Usually referred to as 1st, 2nd, etc., edition
Example : First printing is the 1st edition, enter 01.

Line: 01 **Field:** 13 **Field Name:** ISSUE

Instructions

- **No. Entries:** 1
- **Code/Clear:** Numeric characters entered from document
- **MRL:** 3 characters
- **Req. Entry:** Required when applicable
- **Type Docts.:** Open publications
- **Justify:** Right, with leading zeros
- **Definition:** The publication number given to periodical-type publications. Numbers are usually consecutive within year, but may be consecutive from the beginning of the series without regard to year.
- **Example:** "Datamation," Volume 10, Number 8, August 1964. Enter 008.
- **Comments:** An error was made in this field when the transcript sheet was designed. Many issue entries require 3 characters and this particular field is limited to 2. The correct number of characters will be allowed for in a reissued transcript sheet.

Line: 01 **Field:** 14 **Field Name:** COUNTRIES REPORTED ON

Instructions

- **No. Entries:** 1 to 3
- **Code/Clear:** Code (see code schedule #5, Tab A) ISC Area Code
- **MRL:** 12 characters
- **Req. Entry:** All documents
- **Type Docts.:** All documents
- **Justify:** Left
- **Definition:** The country or countries which are discussed in the document

25X6

- **Example:** [redacted] Enter 5416.
- **Comments:** Three countries can be entered in this block. Use entry given in country line. When more than three countries are entered in this line, use appropriate geographic bloc code. Second and third subfields will be left blank when only one country is entered.

Line: 02 Field: 01 Field Name: REPORT PRODUCING COMPONENT

Instructions

No. Entries	:	1
Code/Clear	:	Clear text (see Source Authority File)
MRL	:	40 characters
Req. Entry	:	Required when applicable
Type Docts.	:	Most
Justify	:	
Definition	:	The name of the component originating the report
Example	:	State, INR
Comments	:	Generally the name of the organization appearing on the "originator line" on information reports. Originator appears in code on SI reports. Refer to Technical Manual for name of originating component. See Source Authority Card for exact entry.

Line: 02 Field: 02 Field Name: DOCUMENT/SERIES NAME

Instructions

No. Entries	:	1
Code/Clear	:	Clear text (see Source Authority File)
MRL	:	32 Characters
Req. Entry	:	Required when applicable
Type Docts.	:	Most
Justify	:	
Definition	:	The name of the document or report series as it appears on the first page or the cover of the report
Example	:	FBIS, FE DAILY/250
Comments	:	This entry refers to series names and number when given. Do not enter common control number in this field. Do not use for volume and series numbers of open publications.

INDEXING GUIDE
Header Data
5.C.3.3.

Line: 03 Field: 01 Field Name: DOC/SERIES IDENTIFICATION NR.

Instructions

- No. Entries: 1
- Code/Clear : Numeric characters (see Source Authority File)
- MRL : 29 characters
- Req. Entry : Required when applicable
- Type Docts.: Most
- Justify :
- Definition : The 10-digit common control number for IR's. The variable length number (up to 29) for SI. The 7-digit common control number assigned by Document Division to unnumbered [redacted] and Defense Reports.
- Example : Refer to sample header data sheet for examples on document and publication numbering
- Comments : This field will not be used for volume and series numbers of open publications nor for specific series numbers of intelligence publications. See Document Identification Numbering Systems for interpretation of numbers. See also SI Technical Manual for explanation of SI numbering systems. See also Source Authority Card for exact form of entry.

25X6

Line: 04 Field: 01 Field Name: CROSS REFERENCE

Instructions

- No. Entries: Variable
- Code/Clear : Clear text (see Source Authority File)
- MRL : 50 characters
- Req. Entry : Required when applicable
- Type Docts.: Intelligence type reports
- Justify : Left, and separate multiple entries by a semicolon
- Definition : A second and sometimes a third number representing the base document

Example : Series name: Mil Info on EE No. 91, Cross reference entry: [redacted] TT/64/41516. Keep order of entry the same as order on Source Authority Card. STATSPEC

Line: 05 Field: 01 Field Name: CITED REFERENCE
Instructions

No. Entries	:	Variable
Code/Clear	:	Clear text (see Source Authority File)
MRL	:	50 characters
Req. Entry	:	Required when applicable
Type Docts.	:	All
Justify	:	Left, and separate multiple entries by semicolons
Definition	:	A reference to another document. Appears in the reference line of the document in hand.
Example	:	Document in hand: [redacted]; cited reference: State Airgram, Rome 57, 64. 25X1C
Comments	:	References appearing within the body of the report will not be picked up. See Source Authority Card for correct entry.

Line: 06 Field: 01 Field Name: LOCATION OF SENDER
Instructions

No. Entries	:	1
Code/Clear	:	Code first subfield, clear text second subfield (see Schedule #5, Tab A)
MRL	:	30 characters: 4 country, 26 city
Req. Entry	:	Required on SI documents when applicable
Type Docts.	:	SI material only
Justify	:	Left within subfields
Comments	:	Country must always be used. City may be left blank.

INDEXING GUIDE
Header Data
5.C.3.3.

Line: 06 Field: 02 Field Name: LOCATION OF RECEIVER
Instructions
- No. Entries: 1
- Code/Clear : Code first subfield, clear text second subfield (see Schedule #5, Tab A)
- MRL : 30 characters: 4 country, 26 city
- Req. Entry : Required on SI documents when applicable
- Type Docts.: SI material only
- Justify : Left within subfields
- Comments : Country must always be used, city may be left blank.

Line: 07 Field: 01 Field Name: LANGUAGE
Instructions
- No. Entries: 1
- Code/Clear : Clear text (see Schedule #6, Tab A)
- MRL : 15 characters
- Req. Entry : Required when applicable
- Type Docts.: Foreign language material and translations
- Justify :
- Definition : The foreign language used in the document in hand or the language from which translated

Line: 07 Field: 02 Field Name: TRANSLATING AGENCY
Instructions
- No. Entries: 1
- Code/Clear : Clear text (see Source Authority File)
- MRL : 40 characters
- Req. Entry : Required when applicable
- Type Docts.: Translations
- Justify :
- Definition : The name of the organization which translated the material in hand

Example : FDD, JPRS, etc.
Comments : If there is an entry in LANGUAGE and no entry in this block it can be assumed the material in hand has not been translated. Refer to Source Authority Card for entry form. Do not enter for Bio Cards.

Line: 08/09 Field: 01 Field Name: PUBLICATION TITLE-ENGLISH

Instructions
No. Entries: 1
Code/Clear : Clear text as reported
MRL : 144 characters
Req. Entry : Required when applicable
Type Docts.: Open publications
Justify :
Definition : The title of an English-language open publication or the translated title of a foreign language publication.
Comments : Do not enter transliterated title in this field. Transliterated title, if given, will be entered in PUBLICATION TITLE-FOREIGN, cards 18 and 19.

Line: 10-14 Field: 01 Field Name: ITEM TITLE-ENGLISH AND EXPANDED

Instructions
No. Entries: 1
Code/Clear : Clear text as reported
MRL : 360 characters
Req. Entry : All documents
Type Docts.: All
Justify :
Definition : The English language title of an IR or SI document, an individual article title taken from a multi-article intelligence report, or the title of

an individual article taken from a publication. The latter includes translated titles.

Comments : Item titles and expanded titles are treated as a single title. Do not use transliterated title in this field. Transliterated title, if given, will be entered in ITEM TITLE-FOREIGN, cards 20 and 21. See Source Authority Card for specific processing rules. See also general rules on punctuation, section 5.C.3.2.

Line: 15-17 Field: 01 Field Name: ENCLOSURES

Instructions

No. Entries: Variable

Code/Clear : Clear text as reported

MRL : 216 characters

Req. Entry : Required when applicable

Type Docts.: IR's, SI

Justify :

Definition : Attachments, annexes, appendices, etc.

Example : Title, PP55, R or NR

Comments : Enter title of enclosure and total number of pages, and show whether received or not received. Separate enclosures with a semicolon. Expand enclosure titles where necessary to convey the sense or meaning of the given title. Enclosure title block must also identify repository where enclosure can be located.

INDEXING GUIDE
Header Data
5.C.3.3.

Line: 18-19 Field: 01 Field Name: PUBLICATION TITLE-FOREIGN

Instructions

No. Entries	:	1
Code/Clear	:	Clear text as reported
MRL	:	144 Characters
Req. Entry	:	Required when applicable
Type Docts.	:	Open publications, translations, etc.
Justify	:	
Definition	:	The foreign language title of a publication or transliterated version thereof

Line: 20-21 Field: 01 Field Name: ITEM TITLE-FOREIGN

Instructions

No. Entries	:	1
Code/Clear	:	Clear text as reported
MRL	:	144 characters
Req. Entry	:	Required when applicable
Type Docts.	:	Open publications, translations
Justify	:	
Definition	:	The foreign language title of an article within a publication, or the transliterated version thereof

Line: 22 Field: 01 Field Name: AUTHORS

Instructions

No. Entries	:	Variable
Code/Clear	:	Clear text as reported
MRL	:	72 characters
Req. Entry	:	Required when applicable
Type Docts.	:	Open publications, translations
Justify	:	
Definition	:	The author or co-authors of publications or of individual articles within publications
Example	:	Smith, John, J; Jones, Henry, L

INDEXING GUIDE
Header Data
5.C.3.3.

Comments : Enter last name, first name and middle initial if given. Separate by commas. If more than one author, separate individual authors with a semicolon. If an illustrator has contributed to the publication, enter as an author or co-author. In case of speeches, enter speaker as author.

Line: 23 Field: 01 Field Name: PUBLISHER
Instructions
- No. Entries: 1
- Code/Clear : Clear text as reported
- MRL : 40 characters
- Req. Entry : Required when applicable
- Type Docts.: Books, monographs
- Justify :
- Definition : The name of the organization responsible for production of the publication

Line: 23 Field: 02 Field Name: PLACE OF PUBLICATION
Instructions
- No. Entries: 1
- Code/Clear : Code country subfield, clear-text city subfield
- MRL : 30 characters: 4 country, 26 city
- Req. Entry : Required when applicable
- Type Docts.: Open publications
- Justify :
- Definition : Self explanatory

Line: 24 Field: 01 Field Name: EDITOR
Instructions
- No. Entries: 1
- Code/Clear : Clear-text as reported
- MRL :
- Req.Entry : Required when applicable
- Type Docts.: Open publications
- Justify :

Definition : The name of the individual when given in the publication

Comments : Enter only when some individual is specifically designated as the editor. Follow entry procedure under authors, line 22, field 01.

TAB A

CODE SCHEDULES

Schedule 1. Item Category

Notation	Document Categories
--	I. Government (U. S.)
--	A. USIB Agencies
01	1. Raw Intelligence Reports (Collateral IR's)
03	2. SI Messages and Teletype
05	3. Finished Intelligence Monographs Reports, Compilations
09	4. Cables, Teletype
[redacted]	
17	6. Translations
21	7. Maps (Including Charts, Graphs, Blueprints, etc.)
--	8. Photos
25	a. Stills-Ground
29	b. Stills-Personalities
33	c. Aerial
34	9. Motion Pictures (Film)
--	B. Non-USIB Agencies
38	1. Information Reports
42	2. Finished Reports, Monographs, Compilations
46	3. Maps
--	4. Photos
50	a. Ground
54	b. Personalities
58	5. Motion Pictures
59	6. Translations

25X1A

INDEXING GUIDE
Code Schedules
5.C., TAB A

Notation		Document Categories
--	II.	Non-Government (U. S.)
62		A. Contractor Produced
66		B. Commercial Publications including Periodicals, Monographs, Newspapers, Brochures, etc.
70		C. Maps
--		D. Photographs
76		1. Ground
80		2. Personality
84		E. Motion Pictures (Film)
85		F. Miscellaneous
--	III.	Foreign Sources
86		A. Governmental (Intelligence Reports, etc.)
87		B. Finished Intelligence
88		C. Commercial Publications (Newspapers, Brochures, etc.)
90		D. Maps
--		E. Photographs
92		1. Ground
94		2. Personalities
96		F. Motion Pictures (Film)

INDEXING GUIDE
Code Schedules
5.C., TAB A

Schedule 2. Security Classification

Notation	Classification
U	Unclassified
C	Confidential
S	Secret
T	Top Secret

INDEXING GUIDE
Code Schedules
5.C., TAB A

Schedule 3. Dissemination Control*

Notation	Control Name
01	Background Use Only (not a distribution warning)
02	CIA Internal Use Only
03	CIA Internal Use and Background Use Only
04**	Discreet
05	Intelligence Components Only
06	Limited Distribution or Limit Distribution
07	Limited Official Use
08	No Dissemination Abroad
09	No Foreign Release No Foreign Dissem
10	No Foreign Release except name of country No Foreign Dissem except name of country
11**	Not for Attribution
12	NSC Participating Agencies Only

*See CHIVE W-14, 8 June 1964, for details on definitions and regulations.

25X6 [redacted]

INDEXING GUIDE
Code Schedules
5.C., TAB A

Notation	Control Name
13	Restricted
14	[redacted]/Eyes Only
15	USIB Reproduction Prohibited
16	Warning Notice--Sensitive Sources and Methods Involved
17	Controlled Dissem
18	Official Use Only (use for Government Use Only-GUO); Not for Publication
19	Continued Control
25	U.S. Eyes Only

25X6

Schedule 4. Originator Evaluation

Source	Information
A - Completely reliable	1 - Confirmed by other sources (confirmed by other sources considered reliable)
B - Usually reliable	2 - Probably true
C - Fairly reliable	3 - Possibly true
D - Not usually reliable	4 - Doubtful
E - Unreliable (not reliable)	5 - Improbable (probably false)
F - Reliability cannot be judged	6 - Probability cannot be judged

INDEXING GUIDE
Code Schedules
S.C., TAB A

25X1B

Next 35 Page(s) In Document Exempt

Approved For Release 2000/05/30 : CIA-RDP78-03952A000100050001-7

Appendix 5.D.

INHERITED FILES

5.D.1. INTRODUCTION

This appendix consists of an enumeration and discussion of existing [redacted] which are candidates for use in CHIVE.* The "inherited file problem" as such is discussed at some length in section 5.5.9. under the heading "File Conversion."

25X6

The information contained in this appendix has been drawn together for two basic reasons. First, the files described herein constitute a corpus of considerable size and intelligence importance, one which is likely to be a significant repository of information for some time to come. From this point of view, the files listed here are an accounting of the files with which CHIVE information

* The information in this appendix has been assembled from CHIVE data-gathering activities and through interviews with OCR divisional personnel. Owing to deadline pressures in the production stage of this report, only limited coordination or review of the file descriptions has been attempted. It is recognized that exceptions to some of the generalized statements exist, and that discrepancies in statistics used in interpretation may be developed. Since the purpose here is to provide an initial view of CHIVE's approach to these files, such discrepancies can be corrected in later, more detailed analyses.

analysts must deal in one form or another. Second, it was felt desirable to give some attention, however abbreviated, to the conversion problems that might be encountered in each file in order to develop a better appreciation of the conversion problem as a whole.

The candidate files have been divided into two broad classes. These are Document Index Files and Document Image Files. The former are searched to locate relevant documents, dossiers, or other data sources. Document Image Files are files of source documents. Thus, index files are machined, or are printed on listings which are produced by machine; document files are hard copies or microfilmed hard copies, the targets of machine searches.

Most of the registers of OCR are represented--SR, FIB, BR, GR, and the CIA Library each have files of potential use.

Approximately 17 index files and 20 document image files are examined. This is a gross count, in which individual components or sub-files of larger files are not enumerated. In fine count, 26 index files and 25 document image files are enumerated.

In most instances, each file is discussed first in a general description and then from the point of view of use in CHIVE. Where possible, a specific conversion recommendation is made.

"Use" in this context is defined to mean two things. First, it signifies the use to which the file can be put in CHIVE, that is, the mode of operation by which it will be utilized. The discussion in this sense is preliminary because the operational design of CHIVE is preliminary and because use and mode of operation depend upon each other.

Second, use means conversion potential, that is, the degree to which the file can be transmuted into a form suitable for utilization in an automatic or semi-automatic environment. Obviously, some files can be used in CHIVE without any conversion, especially document image files. Some of these need only be sorted into a geographic partition to be rendered usable.

In selecting candidate files for study, the scope has not been limited to [redacted] lthough all the

25X6

files discussed con[redacted] Broadening of the scope has been necessary because (a[redacted] 25X6 are often interrelated physically with other materials,

INHERITED FILES
Introduction
5.D.1.

and, (b) sometimes use can be enhanced by accepting a file *in toto* even if this imposes a responsibility for total service. Fortunately, accepting an entire file would not in all cases impose such a total responsibility.

Some, but not all authority files are discussed. An authority file was rejected for study if it would only be used as an input to the construction of a CHIVE-built authority.

Certain underlying usage problems cut across all file boundaries. These problems include vocabulary conversion, vocabulary commonality, fielding unfielded data, recombining related unit records now scattered in a file, and reorganizing file sequence for CHIVE searching. When the implications of these issues are fully understood for each questionable file, it will be possible to develop a unified and coherent conversion technique equally applicable to every file. Exceptions to this technique can then be made intelligently, and the impact of alternative solutions upon CHIVE system operations can be defined.

Figure 5.D-1 lists the subject files under study in this appendix, their function (that is, index or document image), and the information system in which they reside. Figure 5.D-2 lists inherited files other than index or

document image files which are under consideration for use by CHIVE. Some of the files listed in Figure 5.D-2 are also in Figure 5.D-1; these belong to more than one file class. The remainder of the files entered in Figure 5.D-2 are candidates for use by CHIVE, but have not been studied in this appendix.

5.D.2. INDEX FILES

5.D.2.1. SR Files

5.D.2.1.1. Subject/Commodity File

5.D.2.1.1.1. General Description

The Subject/Commodity File is the machine index to all subjects and commodities referenced in depth-indexed SR documents. The file is prepared by SR/OCR and is maintained on punched cards in the SR Machine Branch. These cards contain both coded and clear text information. They are sequenced by the SR 9-digit subject/commodity code.

Of a total of close to 8 million cards in this file,
25X6 approximately 1.5 million refer to [redacted]. The
25X6 exact size of this [redacted] is unknown. It can be determined only by re-sorting the entire file according to

INHERITED FILES
Index Files
5.D.2.1.1.1.

25X6 area and counting the [redacted]. The total file is growing by approximately 230,000 cards per year. Estimated yearly growth of the [redacted] 25X6 per year.

The Subject/Commodity File is consulted daily. According to the June 1964 SR monthly report, there were 32 requests made against the file during that month. A request is defined as a question posed by an information consumer. These requests generated 167 searches (that is search terms) and produces 4,374 document references. (This number does not include the output of some special requests which result in lengthy subject/commodity listings).

All 80 columns of the punched card are available for data. The information found on each Subject/Commodity card can be seen in Figure 5.D-3.

5.D.2.1.1.2. Use in CHIVE

The SR series number, document identification number, publication date, message date, and page number can be used without difficulty in the header of a CHIVE record. However, conversion to a CHIVE code scheme of the subject/commodity codes found in the SR Subject/Commodity record represents an imposing task, for the following reasons.

INHERITED FILES
Index Files
5.D.2.1.1.2.

First of all, SR uses a modified version of the old Intelligence Subject Code (ISC), while subjects are indexed by CHIVE using the new ISC. There are about 4,000 subject codes used by SR as compared to about 5,000 codes in the new ISC. There is no significant correlation between these code systems, so that any conversion would involve a considerable intellectual, manual effort to create tables which equate these codes. Such a conversion would result in some loss of information because there is no one-to-one relationship between a large number of the codes. Moreover, the ISC is a 6-digit code, while the SR subject code is a 9-digit code, further adding to the incompatibility.

Secondly, SR uses a 9-digit commodity code based on a revised Standard Commodity Classification Code issued by the Bureau of the Budget. There are approximately 4,000 of these commodity codes, so conversion to a CHIVE commodity code which is compatible with the new ISC would encounter problems similar to those encountered in subject conversion.

Conversion of the Use/Nationality code and the Action code can be accomplished without significant effort, but with some loss of information due to the lack of one-to-one correspondence. (Of course the ISC could be expanded at some cost to attain such a correspondence.) There are about 60 SR Use codes to be associated with 6 CHIVE Use parameters, and about 200 nationality codes to be associated with appropriate ISC country codes. About 60 SR action codes must be translated into about 95 similar CHIVE codes.

The area code used in the Subject/Commodity file consists of a joint country code and a political subdivision code, based on an SR version of the Industrial Register area code scheme. Because of the relatively small number of countries to be converted, intellectual aspects of conversion of the country code portion of the joint code to the ISC code used by CHIVE would not be difficult. Loss of information would result for countries

25X6 other [redacted], however, since the new ISC represents the current country situation, while the SR area code scheme does not. SR political subdivision codes, although not as numerous as SR country codes, would be difficult to convert to CHIVE equivalents, since political subdivisions are not always applied in the same way in both code systems. When they are, country boundaries do not necessarily correspond.

SR's clear-text place names are controlled by the SR area dictionary and the NIS Gazetteer. Since CHIVE does not use this dictionary, these names would have to be manually converted to CHIVE equivalents.

In SR, several areas are indexed with one subject/commodity, the 6-digit area code (col.41-46) and the 12-digit city name fields (col. 48-59) are combined into six 3-digit country code fields. An "R" code in the TFA (to-from-at) field (col. 47) shows the presence of several areas on the card.

It should be pointed out that all area references punched on cards in the Subject/Commodity file have also been punched on cards entered in the Area file (see section 5.D.2.1.2.). The Area file also uses the IR area code scheme. However, the area codes in the Area file have been kept up to date, while corresponding area codes in the Subject/Commodity file have not. Hence, it might be advisable to use in CHIVE only clear-text place names in Subject/Commodity cards (when they are present) and not to convert area codes as they appear in the Subject/Commodity file. The more complete and more recent Area file then would serve as the source of CHIVE area code information. Then only the up-to-date area code scheme used in the Area file need be converted and combined into CHIVE index records.

5.D.2.1.2. Area Detail File

5.D.2.1.2.1. General Description

The Area Detail File is an index to all locations encountered in SR depth indexing. Indexers simultaneously extract areas, subjects, and commodities from a given document and copy these on a single transcript sheet. However, subjects, commodities, and areas are entered onto punched cards in a different format. Cards in the Area file are maintained in Machine Branch of SR in order by area (sequentially by country code and the first digit of subdivision code, then alphabetically by place name).

25X6 The [redacted] ch is an index to all document [redacted] is an integral part of the total Area File, but, because of the area sort, the

25X6 [redacted] are contiguous and can easily be located and extracted.

The total Area File numbers some 3,960,000 cards and is growing by about 400,000 cards per year. The

25X6 [redacted] cards number 159,000. During one period, an average of 62 requests per month against the entire file resulted in 433 searches and 51,103 references. "References" includes both individual document numbers and listings of numbers. Growth and activity statistics

25X6 are not available for the [redacted] of the area file.

There is room on each card for 3 areas. During keypunching, a card which contains more than one area is

INHERITED FILES
Index Files
5.D.2.1.2.1.

reproduced until there is one card for each area. Each card has a central area field. Each area is entered in this central field once, with the remaining areas surrounding it. Thus, a card containing more than one area is filed once for each area, the area punched in the central area field.

If there are more than 3 areas referenced in the one document, an "x" overpunch in column 79 is made to show the existence of additional area cards. This "x" is used to re-link the areas punched on different cards when such linking is required by the search parameters.

5.D.2.1.2.2. Use in CHIVE

Area file series number, document number, publication date, message date, and information date can be readily incorporated "as is" in a CHIVE header record.

Area codes used by SR are based on the IR code scheme, and would have to be converted to the ISC for inclusion in the CHIVE system. (See Section 5.D.2.2.4.). The area codes used in the Area File, unlike area codes in the Subject/Commodity File, have been kept updated to correct for country changes and reorganizations.

Conversion of political subdivision codes poses a challenge because: (a) either the ISC and/or the SR area code scheme may not employ a subdivision code for a given country, and (b) in some cases, boundaries of subdivisions represented by the two different code schemes do not correspond exactly (that is, subdivision boundaries differ).

City names are not always provided on the card, but when they do appear they are represented by controlled text, mediated by the area dictionary and the NIS Gazetteer. Since CHIVE does not use this area dictionary, any text and abbreviations controlled by it would have to be translated to the CHIVE system.

For use by CHIVE, an effort should be made to retain the linkage between areas appearing on a card. The 9 possible TFA codes could be transformed into about three of the CHIVE location tags. The tags act as role indicators which can express the TFA relationships in a CHIVE phrase.

An effort also should be made to retain the linkage between areas indexed to more than one punched card yet derived from a single document. These areas could be linked only by re-sorting the entire area file by document identification number.

25X6 5.D.2.1.3. [redacted]

5.D.2.1.3.1. General Description

The SR Organization file is an index to all organizations and organizational data referenced in SI documents processed by SR's Organization Section. The file is maintained in three separate portions:

25X6

- [redacted]
- Soviet Organization File
- All Other Organization File (ALLO)

Each portion has similar, although not identical, card formats. Each is ordered by organization code. The discussion which follows will refer only to the [redacted] 25X6

25X6 [redacted] organizations regardless of location.

This file is maintained on punched cards in SR's Machine Branch in sequential order according to the SR 9-digit organization code.

25X6 The [redacted] consists of over 50,000 cards and is growing by 20,000 cards a year. It is consulted weekly. Figures from the June 1964 SR monthly report show that 14 requests were made against the file that month. There were 88 searches which resulted in 209 document references. The number of references does not include listings. (The field arrangement of the

25X6 [redacted] card can be found in Figure 5.D-5).

INHERITED FILES
Index Files
5.D.2.1.3.1.

Any organization cards with an entry in the personality field (columns 48-61) are duplicated for inclusion in the SR Personality File. If any of these cards have an entry in the nationality field, another card copy is made to be filed in the Personality Foreigner File.

5.D.2.1.3.2. Use in CHIVE

Series number, document number, publication date, message date, information date, and page number can be readily accommodated in the CHIVE header record. SR country and political subdivision codes are based on the IR code scheme. Their conversion to CHIVE is discussed in connection with the SR Subject/Commodity File (Section 5.D.2.1.1.).

There are only 8 columns set aside for city name. When necessary, the name is truncated on the SR card. Since this clear-text name is controlled by the NIS Gazetteer and the SR area dictionary, the codes must be converted to CHIVE standards for entering controlled clear-text city names. The card entry must be untruncated, and the accepted clear-text version selected.

If an organization card references a [redacted] personality and the standard telegraphic code for this [redacted] is available, the telegraphic code is entered in columns 36-46. These codes are used by the 25X6 25X6 25X6

CHIVE system where they are tagged by the mnemonic PNT. Hence, the SR codes can easily be used. The personality name on the card can also be associated directly with the PNO CHIVE tag.

SR uses a unique occupation code which assumes about 300 different code values. These would have to be equated with associated ISC codes by an intellectual analysis which establishes a correspondence dictionary. The dictionary would then be automated and used to convert from SR value to CHIVE code value. Each code as entered in an [redacted] card would be converted to the CHIVE equivalent in the CHIVE [redacted] organization record.

25X6

The 200 possible SR nationality codes could be converted to ISC codes and entered in CHIVE in the same way.

The TFA code in column 71 indicates the nature of the association between the personality and the area on that card. In CHIVE, this linkage could be used as above by establishing a TFA code-to-tag transformation dictionary.

The organization code used by SR for the [redacted] 25X6 Organization File is a 14-digit (9-digit optional) hierarchical code with 4 levels of subordination. Not all 4 levels are required for a given code entry. The code is entered onto the card from left to right (columns 72-80), with the most generic code on the right. Because of the size of this field, either the third or fourth level organization code can be entered on a card but not both. For every third level code entered, a similar organization card is produced which contains the corresponding fourth level code but lacks the third level code. If CHIVE used the SR

25X6 [redacted] it could enter the full 4 code levels in one record. In order to use the SR card data in CHIVE it is necessary to recombine both

25X6 SR [redacted] cards if more than one exist, and to construct the complete organization code from the split entries.

5.D.2.1.4. Personality File

5.D.2.1.4.1. General Description

25X6 The SR Personality File is derived in part from the [redacted] Soviet and All Other Organization files by

25X6 reproducing those organization cards which also have entries in the personality field. The resultant [redacted] and Soviet personality cards have the same format as the

INHERITED FILES
Index Files
5.D.2.1.4.1.

organization cards from which they stem (see Figure 5.D-5 and Figure 5.D-6). The resultant All-Other personality cards are in a different format than their source organization cards (see Figure 5.D-7 and Figure 5.D-8).

The Personality File also includes entries that do not have associated organizations. These cards are punched at the same time, in the same format, and from the same transcript sheets as the organization cards. (See Section 5.D.2.1.3.).

The total Personality File is sorted in alphabetical order by personality name (columns 48-60 of all three formats). It consists of more than 2.5 million cards. In addition, about 500,000 cards (those references to persons mentioned in 1953-1959 select series documents) have now been sent to Records Center. The size of the
25X6 [redacted] portion of the personality file cannot be determined unless the entire file is re-sorted by location.

The total Personality File is increasing by about 300,000 cards per year. Some 9,700 of these originate
25X6 in the [redacted]. An average of 73 requests per month against the total file result in an

INHERITED FILES
Index Files
5.D.2.1.4.1.

average of 106 searches and 27,209 document references and listings.

A byproduct of this file is the Soviet Personality File which is an index to references on all Soviet personalities residing in the USSR. This file is derived by reproducing only the Soviet personality cards with no entry in the nationality field, and then ordering these cards by area.

5.D.2.1.4.2. Use in CHIVE

The total Personality File consists of cards in three different formats, intermixed and ordered by personality, a common field in the three formats. The entire file can be used only if the three different formats are considered.
25X6
Partitioning of the cards originating in the [redacted] Organization File can be accomplished by sorting the entire file on column 1. Code conversion for these cards has
25X6
been discussed in conjunction with the [redacted] Organization File because the card formats are identical (see Figure 5.D-5).

25X6
25X6
If the [redacted] Organization File were used by CHIVE, most of the [redacted] portion of the Personality File would automatically be available, since cards with personality and organization entries would appear redundantly in both files. Additional conversion would have to be done on

25X6 [redacted] personality cards without organizations, however.

25X6 [redacted] persons are entered on personality cards derived from Soviet and All-Other organization cards

25X6 when these contain references to [redacted] abroad. A way to partition these cards is to extract all cards

25X6 which contain [redacted] nationality codes and

25X6 all cards which name [redacted] areas but have no nationality codes. This can be done by re-sorting the entire card file using nationality as the primary sort field and area as the secondary sort field.

5.D.2.1.5. Personality Foreigner File

5.D.2.1.5.1. General Description

The Personality Foreigner File is a derivative file which consists of all Personality File cards which have an entry in the nationality field. The three card types have three different formats (see Figure 5.D-5, Figure 5.D-6, and Figure 5.D-7). The file is ordered first by nationality and secondarily by location. Nationally and location fields appear in the same position in each card type.

The Foreigner File is used to recover references to persons traveling outside their own country. This use is supported by the file structure, in which the primary sort is nationality. It is not well-suited to recover foreigners located in a cited country because the secondary sort, namely location, is not available as an entry to the file.

25X6

The cards which stem from the [redacted] Organization File are interspersed throughout the Foreigner File. They cannot be segregated without re-sorting the entire file.

25X6

However, all cards which reference [redacted] traveling abroad are contiguous because of the primary nationality sort and can easily be extracted.

25X6

Cards which reference [redacted] traveling abroad number 10,600 out of a total volume of 800,000 Personality Foreigner cards. Figures are not available for the growth

25X6

of the [redacted] portion, but the entire file is increasing by over 160,000 cards a year.

Statistics available for August 1964 show an average of 0.5 requests, 100 searches, and 150 references per month. The Foreigner File was created in July 1961, so is relatively new. For this reason also the figures stated may not necessarily be representative of the future activity of the file.

INHERITED FILES
Index Files
5.D.2.1.5.1.

5.D.2.1.5.2. Use in CHIVE

The Foreigner File is a subset of the Personality File which has been created first by extracting all personality cards which have an entry in the nationality field (columns 68-70), then by sorting on nationality and location. Conversion for use by CHIVE could be accomplished by the same technique.

25X6

[redacted]

INHERITED FILES
Index Files
5.D.2.1.5.2.

25X6

If the [redacted] portion is identified by the first definition, conversion is analogous to the conversion of the [redacted] Organization cards. If identified by the second or third definition, the [redacted] portion would include cards with all 3 formats, so further work would have to be done to partition the file into its distinctive formats and then to convert each format separately.

This evaluation of use in CHIVE is aimed primarily at [redacted] Conversion feasibility of the codes found on the Soviet and All-Other Organization cards remains to be evaluated.

25X6
25X6
25X6

5.D.2.1.6. PI Subject/Commodity File

5.D.2.1.6.1. General Description

The PI (Type A) Subject/Commodity File is the index to all subjects and commodities referenced in PI documents. This file is analogous to the No. 1 Subject/Commodity File for SI material but contains PI data. The PI File is ordered by the 9-digit Subject/Commodity code and is maintained on punched cards in the Machine Branch.

The total volume of the file is about 19,000 cards. The [redacted] is estimated to be 1,700 cards. Each month 2,600 cards are added to the total file. The PI

25X6

Subject/Commodity File is consulted weekly. The total weekly number of requests for both this file and the PI Area File average 4 per month. An average of 15 document and item references are produced from these 4 requests. The low activity of these files may be due to the fact that they are new.

Because of the ordering of the PI Subject/Commodity File (on subject/commodity code), the [redacted] 25X6 cannot readily be isolated. Thus, statistics on the growth and activity of the [redacted] are not available. 25X6 (See Figure 5.D-7 for the format of the file card).

5.D.2.1.6.2. Use in CHIVE

The CHIVE header record can accommodate PI File data such as security code, agency originating report, document type and number, publication and information dates.

The use of subject/commodity codes, use codes, nationality codes, and action codes in the PI Subject/Commodity File is analogous to their use in the SI Subject/Commodity File. (Section 5.D.2.1.1. contains a discussion of the use in CHIVE of these codes.)

The country codes and political subdivision codes used by the PI and SI Subject/Commodity files are identical. If the PI subject relates to coordinate information, these coordinates are entered on the punched card in the city name field (columns 48-59). In this case the Army Map Service map number, consisting of 2 alphabetic characters and 4 numeric digits, is placed in columns 41-46 in lieu of the country and subdivision codes. This coordinate information can be incorporated into CHIVE with no difficulty.

5.D.2.1.7. PI Area Detail File

5.D.2.1.7.1. General Description

The PI Area File (Type B) is the index to all area references in PI documents. It consists of duplicates of PI Subject/Commodity cards ordered by location. In all respects except format, this file is analogous to the SI Area Detail File (see Figure 5.D-9).

Location is indicated by a country code, political subdivision code, and a clear-text place name if available. When the subject refers to coordinate information, these coordinates are entered in the city name field and the Army Map Service number replaces the country and subdivision codes.

INHERITED FILES
Index Files
5.D.2.1.7.1.

Of a total of 14,000 cards maintained in the PI Area Detail File, there are an estimated 1,300 [redacted] entries. The total file growth is 30,000 cards a year. 25X6

Monthly requests and references from the combined Type A and Type B file average 4 and 15 respectively. The Type B file is consulted weekly.

5.D.2.1.7.2. Use in CHIVE

The PI Area File has the same format as the PI Subject/Commodity File. Use in CHIVE would be identical. This area file contains not only duplicate subject/commodity cards but area cards without subject commodity codes. Hence, a CHIVE-usable PI Subject/Commodity File could be obtained directly as a byproduct of the conversion of the PI Area File.

5.D.2.1.8. Job 3 (Permuted Title Index) File

5.D.2.1.8.1. General Description

The Job 3 (KWIC) File is an index to SI report titles. It "is a machine controlled and printed clear text index (in alphabetical order) of all 'keywords' or semantically important word groupings appearing in the titles of monographic reports or in the tables of contents of periodicals received in the Special Register."* The clear-text file is referred to as "Job 3." It is a permuted title listing whose title words are controlled by an accepted vocabulary and by well-defined procedures for accepting, rejecting, or modifying title words. "KWIC" means "keyword in context."

PI reports also receive permuted title processing,** as do selected sensitive SI periodical articles, e.g., the DIA Summary and OCI Weekly.***

The Job 3 publication serves many purposes. It provides control over reports which are not detail-indexed. It is used to locate an individual report by title or controlled key word when the report number (documentation) is unknown. It is used as a convenient tool for subject searches, the key words serving as subject indicators.

---

*SR, SC#20505/61, 10 January 1961, Appendix A, "Index to SI Report Titles," pp. 88-98.

**The keyword technique applied to all PI reports is essentially the same as for SI. The discussion here applies to both PI and SI unless otherwise stated.

***There are additional sensitive SI indexes described in TCS 4724-64. Essentially the same processing procedures apply to all types.

INHERITED FILES
Index Files
5.D.2.1.8.1.

The punched cards are used primarily to prepare standard listings which are periodically accumulated. The punched card can be used for retrospective searching, but this is seldom done since a machine search can create only what is already available in the printed listings. Such searches are made only to print out a selected portion of the Job 3 listing.

The punched card file is organized into 3 main decks. One is in sequence by SI documentation (series and document number); one is in alphabetical order by keyword; and one is in sequence by cross reference.

Job 3 files contain 925,485 punched cards. Monthly file growth is approximately 13,200. A breakdown of this data is given in Figure 5.D-12.

Output products are machine listings generated from the punched card index. Two primary keyword lists are created. The first is an SI Index to Report Titles. It is issued and cumulated each month up to six months. The second is a similar listing for PI reports. It is issued and cumulated quarterly up to one year. Listings of document acquisitions in series order are also generated from these records.

SR analysts develop the title record by entering selected words in the subject segments of the transcript sheet (see Figure 5.D-13).* The IBM cards (see Figure 5.D-14) are punched from this transcript sheet. Next they are exploded by machine to create a "printing deck" in which the documentation and each keyword appear in the sort order print field position. One such title is printed for each keyword in a title. The permuted listings are printed from this exploded deck.

The format of the transcript sheet is identical to the punched card format. All 80 columns are reserved for documentation, keyword data, and other elements of information (see Figure 5.D-15). The elements are printed on the transcript sheet by SR analysts. They can be equated to similar data elements in the CHIVE header record.

5.D.2.1.8.2. Use in CHIVE

The Job 3 card file is used in SR primarily to create and cumulate a printed index, not for retrospective searching. Printing of the cards already on

---

*Slightly modified transcript sheets are used for other Job 3 projects.

file has been accomplished when the Job 3 is published. Reprinting by computer would be unnecessary. New titles which are input to CHIVE will at least be header indexed. These titles will be published in a CHIVE-built permuted title index comparable to the Job 3 (Permuted Title Index). Hence, no substantial gain could be made through incorporating existing, already published Job 3 files in CHIVE.

A possible reason for converting would be to save space. The present file at Headquarters consists of some 150,000 punched cards, enough to fill 2.5 card cabinets. The anticipated space savings, however, would hardly justify the expense of conversion.

In summary, even without conversion, the Job 3 (Permuted Title Index) listings are available to service CHIVE requests. The Job 3 card file cannot be used profitably by CHIVE, and should not be converted.

5.D.2.1.9. Summary of Use in CHIVE of SR Files

For CHIVE purposes, it is helpful to think of SR files as being composed of three groups:

(a) Subject/commodity and area files (see Figure 5.D-10)

(b) Organization and derivative files (see Figure 5.D-11)

(c) Job 3 file

If one of the files from a group is converted, with only a little extra effort, possibly just re-sorting, the other files probably can also be converted (with the exception of Job 3) and used by CHIVE.

Although conversion from the IR-based SR area code to the ISC-based CHIVE area code is a major analytical and data processing task, the area code is present in every SR file discussed. Doing the job for any one file is equivalent to doing the job for all.

File group (a) consists of:

(1) SI Subject/Commodity File

(2) SI Area Detail File

(3) PI Subject/Commodity File (Type A)

(4) PI Area File (Type B)

The SI Subject/Commodity File card and SI Area File card are in different card formats but are produced from the identical transcript sheet. Although areas represented in the Subject/Commodity File are also represented in the Area File, it is not possible to derive the Subject/Commodity File from the Area File because the association between location and subject/commodity code is not retained in the Area File.

The (3) and (4) PI files are analogous to the (1) and (2) SI files, but differ in that they reference PI documents. Files (1) and (2) also differ from (3) and (4) in this significant respect: the PI Subject/Commodity and Area Files both use the same card format, while the SI counterparts use different formats. This is because area code and subject/commodity code association is retained in the PI Area File. The PI Subject/Commodity File is generally a subset of the Area File [A cards with a 3 in column 47 are unique], so conversion of the type (b) Area File can directly produce a converted type (a) Subject/Commodity File.

File group (b) consists of:

25X6 (1) [redacted]

(2) SI Personality File. This file consists of:

a. #2 cards, duplicated from #7 file

b. #3 cards, same data as in #6 file (All Other Organizations) but in slightly different format. There is a modified #3 card called the C Card.

c. #9 cards, duplicated from the #8 file (Soviet Organization File)

(3) SI Personality Foreigner File. This consists of:

a. #2x cards, duplicated from #2 cards (x indicates overpunching)

b. #3x cards, duplicated from #3 cards

c. #9x cards, duplicated from #9 cards

25X6 The [redacted] is one of three organization files, the other two being:

(4) Soviet Organization File (#8)

(5) All Other Organization File (#6)

All three organization files have different formats and are maintained as separate entities, but only conversion of the [redacted] has been directly studied here. 25X6

Relative to the group (b) conversion, note that the SI Personality File is made up in part of all organization cards with an entry in the personality field, regardless of country. The SI Personality Foreigner File, which is a subset of this file, consists of all personality cards with an entry in the nationality field, regardless of nationality. Thus the portions of these files that originated from only the [redacted] Organization transcript sheets can be segregated only by re-sorting the b. and/or c. file. For this reason, conversion of the portions of the b. and/or c. file which reference 25X6
25X6 [redacted]
suggests conversion of the 3 organization card formats.

INHERITED FILES
Index Files
5.D.2.1.9.

If such transformation is accomplished for b. and c., it can be accomplished as a byproduct [redacted] 25X6 Organization File as well. In this manner, the conversion of file group (b) can be achieved.

The fixed field SR card format information such as series and document number, publication date, message date, and information date would be used in a CHIVE header record. Some reformatting and expansion of data codes that utilize overpunches is required to achieve this conversion.

Dictionaries which would convert SR code schemes to CHIVE-used code schemes would have to be devised. Then these dictionaries would be automated. In automated form, they would be used to transform SR card data to CHIVE record data. Vocabulary-converted data and any data that can be used "as is" without vocabulary conversion must then be associated with the appropriate CHIVE tags or field names. For example, PNO must be associated with "personalities," LAC with "location," ONO with "organization." In some cases a code found on the SR card will determine what tag or field name will be associated with another code on that card. For example, the TFA code will determine whether to associate the LAF, "country from," or the LAT, "country to," tag with the country code chosen.

When more than one punched card in a given SR file refers to a single document, an option exists as to whether or not to recombine these cards into a collection of CHIVE phrases. If combined, the phrases and header together would constitute a CHIVE record. In most cases, one file card will generate one CHIVE record.

File group (c) consists of the Job 3 file. This file produces a permuted title listing which can be used without conversion to support CHIVE reference activities. The card file itself is not needed to support retrospective searching. Hence the card file need not be used by CHIVE.

5.D.2.2. FIB Files

5.D.2.2.1. Active Installation Index File

The FIB Active Installation Index File is a file of punched card records selected from the IR Inactive Installation Index File, otherwise known as the Industrial Card File (ICF). It is a basic index to the Industrial

25X1A

Installation 5" x 8" file maintained by the [redacted]

25X1A [redacted] It is physically composed of three types of cards or sub-files: (a) the Town/City Information Card, (b) the Installation Information Card, and (c) the Location Cross Reference Card (see Figures 5.D-16, 5.D-17, and 5.D-18).

The file is normally maintained in sequence by the following elements of information:

- Location (country and location codes)
- Installation Name
- Industrial Category Code
- FIB Identification Number
- Coordinates

The file serves as a basic reference search tool to priority installations at foreign locations. Locations are specified through inclusion of chart numbers, geographic coordinates, and variant or alternate name entries. References to installations are amplified by use of specific names (if available), FIB identification numbers, industrial category codes, installation use/association indicators, and coordinates. All <u>locations</u> from the IR Inactive Installation Index File are covered

INHERITED FILES
Index Files
5.D.2.2.1.

25X6

File is estimated to contain about 20,000 Town/City Information and Location Cross Reference Cards, and about 35,000 Installation Information Cards.

There are four primary products of the active index file. These are in the form of listings used to identify foreign locations and related priority installations. The first listing contains locations and installations arranged alphabetically by location name, the second contains installations arranged alpha-numerically by installation name, the third contains installations arranged numerically by Industrial Category Code, and the fourth contains installations arranged numerically by FIB Identification Number (Dossier Number).

5.D.2.2.2. FIB Inactive Installation Index File

Functionally, the FIB Inactive Installation Index File, that is, the Industrial Card File, is a punched card index to the former IR Installation file. No new cards are being entered. The Inactive file is a composite punched-card file composed of Industrial Card File (ICF) coordinate cards, ICF city cross reference cards, and ICF name cards. Information punched into these cards was obtained from a single transcript sheet prepared by

FIB (IR) indexing personnel (see Figures 5.D-19, 5.D-20 and 5.D-21). This file was used to prepare four reference listings, called "A," "B," "C," and "D" listings.

The "A" listing is a printed record of the entire card file ordered on country, place name within country, and installation name within location name. The "A" listing includes information on geographic coordinates, WAC number, firm number (plant folder), and Industrial Category Code.

The ICF name cards were used to produce the "B," "C," and "D" listings. The "B" listing is sorted on country, firm name, and city. It can be used to identify duplicate firm names. The "C" listing is ordered on country and Industrial Category Code. The "D" listing is a listing of the ICF name cards in firm number (plant folder) order.

This FIB index file is used today to:

- Access a manual store of hard-copy documents in plant and town folders, and industrial category files.

- Generate working aids in support of file input and output.

- Confirm file coverage by location in terms of country, political subdivision, place name, coordinates, and WAC number.
- Show relationships between place names and variant spellings, and between plant names and industrial categories.
- Control input processing in terms of format, area codes, spellings, and coordinates.
- Provide summary information in the form of listings.

The file has been in existence since 1947 and has reached a total size of over 500,000 cards (all three card types). Those portions of the file dealing with the Soviet Union and Soviet-bloc nations have been selectively converted to the Active Installation Index File and are no longer being updated in this file. Those portions of the Inactive Installation Index File dealing with the remaining countries are currently being maintained, pending a decision on their conversion into the new file. Approximately 20,000 of the cards in this file pertain to

25X6 [redacted]

5.D.2.2.3. FIB Model-Type/Brochure Index File

The FIB Model-Type/Brochure Index File is a punched card file which serves as a dictionary of notation on FIB holdings of references to foreign materiel and as an index to supporting documentation stored in dossiers. Each model-type/brochure produces one entry. Entries in this file cover the entire spectrum of the Industrial Category Code. The entries are provided in a single punched card format (see Figure 5.D-22). Emphasis is placed upon materiel of USSR origin, but not to the exclusion of other countries. All items noted by an asterisk in the ICC manual are priority items and must be included in this index file when encountered in documents. Other items are non-priority and are included at the discretion of the FIB analyst.

Four basic listings are produced from this punched card file. These listings are arranged:

- Numerically by Dossier Number (Firm #)
- Numerically by Industrial Category Code
- Alphanumerically by Model-Type Designation
- Alphabetically by Descriptive Name

These listings are normally ordered by country, expressed as country codes. The USSR holdings may also be further broken down by area code (USSR economic region).

For each entry, cards in this file indicate the approved Model-Type/Series designation, the descriptive name for this designation, the appropriate Industrial Category Code, and the Dossier Number (Firm #) that identifies the producer and shows the file location of any documentation pertaining to the item. A Technical Material Notation (*) is also included to show possession of more detailed information on the item. This information may range from basic technical notes to complete descriptive brochures on the item.

The file has been in existence since 1951 and presently contains 56,000 cards, primarily oriented toward the USSR. The file includes approximately 4,000 cards pertaining to

25X6 [redacted]

5.D.2.2.4. Use in CHIVE

Since coding conventions and record structures are common in the three discussed FIB punched card files, their use in CHIVE can be treated as one subject. Generally speaking, most of the coding systems used in these three files are peculiar to FIB or to FIB and the Special Register.

- The FIB country and political subdivision code is derived from codes used by the Foreign Funds Division, Treasury Department, and is used in a revised form by the Special Register. The country

25X6 is a three-digit numeric code (e.g., [redacted] 491 for USSR). The political subdivision code is a three-digit numeric code registered in cards with an "x" punch over the first digit for numbers exceeding 999. The thousands portion (columns 1-12) of this code is also used in the USSR area code (columns 79-80) of the Model-Type/Brochure Index Card to indicate the economic region of the USSR. These elements have analogies with the numeric and alphabetic ISC Area Codes, and a conversion system could be established (see Section 5.D.2.1.2.2.). 25X6

25X1B

- [redacted]

- The FIB Dossier Number code is also known as the Dossier Number, Firm Number, or the FIB Identification Number. It is an element peculiar to the Foreign Installations Branch and is derived from the system used to number their collection of installation dossiers. This is a locally-developed and used system which could be readily incorporated into CHIVE records "as is."

- The Installation Identification code is another name for the Industrial Category code which was established by FIB when it was known as the Industrial Register (see Section 5.D.2.2.1.). This code denotes activities of foreign industrial, scientific, and technical establishments and their interrelationships. For use in CHIVE, the code would be converted to a CHIVE industrial category code.

- The Location Identification code is a local FIB coding scheme for machine sequencing and identification of covered locations and installations. It performs the same function in the Active Installation Index File that the Sequence Number performs in the Inactive Installation Index File. This code provides a function which is not needed by CHIVE.

The other elements of information contained in the three FIB files are essentially entries pertaining to: actual name designations of locations, installations, and items of equipment; designation of AMS and 200 Chart Series associations; coordinates; and various file controls. Two elements of possible future use by CHIVE are the Town/ Installation "C" Code and the Category Designation Code. These last two elements are contained in the Town/City Information and the Installation Information cards of the Active Installation Index File.

5.D.2.3. BR Files

5.D.2.3.1. Dossier Index File

5.D.2.3.1.1. General Description

The Dossier Index File is a punched card index to dossiers of selected foreign personalities. The file is maintained by Machine Division/OCR. It provides EAM listings of personality data with dossier number citations.

The file is divided into five subfiles to facilitate searches. These files are as follows:

- Organization Name File (sequenced by organization code within area)

INHERITED FILES
Index Files
5.D.2.3.1.1.

- Organization Member Main File (sequenced by organization code within area)
- Basic 3-Card File (sequenced by dossier number within area)
- Location File Other than USSR (sequenced by dossier number within area)
- Qualification File (sequenced by occupation code within area)

There are additional file sequences for other sub-files. The Who's Who File, not listed above, is used as a summary information file (with its printed biographic brief), not as a dossier index file. For this reason, it is not included in the conversion study.

Inputs to the dossiers consist of documents and extracted biographic information from raw and finished intelligence reports and from open sources. These documents, or extracts therefrom, are physically added to the hard copy dossiers. The dossiers, not the documents, are indexed. Retrieval of documents takes place by searching for dossier number.

As of 30 June 1964, the Dossier Index File contained slightly in excess of 2.2 million punched cards.* These

---

*Excluding 297,000 Who's Who cards.

provided access to over 375,000 dossiers. In FY 64, a total of 72,800 punched cards were added to the file. The Chicom portion of the Dossier Index File is estimated to be 186,025 punched cards, which provide access to 16,500 Chicom dossiers.

File activity averages 25 requests per month, including Who's Who requests. Monthly requests utilize an average of 1,640 search terms and produce some 41,000 references. Who's Who cards are not prepared for political personalities, nor are all the elements of information listed below included in the index to the dossier.

Despite considerable redundancy of information, a total of 24 unique data elements are identified with the Dossier Index. These include the following:

- Card Type
- Date Punched
- BR Organization Location Code
- BR Organization Code
- Name of Organization
- Citizenship Code
- Dossier Number
- Date of Last Known Organizational Affiliation

- BR Occupation Code
- BR Occupation Abbreviation
- Name of Person
- Date of Birth
- Variant Name
- BR Political Affiliation Code
- BR Religious Code
- BR Location Code of Person (temporary)
- Date of Person's Location
- Soundex Code
- X, D, or P entry
- Political "P" Indicator
- Temporary Location Indicator
- Street Address Indicator
- City Code (USSR and Berlin)
- T.O.I. (Type of Individual) Code

5.D.2.3.1.2. Use in CHIVE

Many of the elements of information listed above are similar to those entered in the CHIVE record. They are associated with the depth index, that is, the content portion. With the exception of dossier number, there are virtually no header record elements. The elements most like CHIVE index terms are: (a) personality and organization names, variants, and event dates; (b) location, organization, citizenship, and occupation codes based on the BR-established code schedules; and (c) dossier number, assigned by BR.

The conversion potential of the Dossier Index File depends on the established commonality between CHIVE and BR vocabularies. CHIVE plans to use clear text personality names. However, organization names and their variants will be encoded. The BR code, then, must be converted to the corresponding CHIVE code.

The manual effort to prepare conversion tables is largely a function of the number of entries in a given code schedule. Thus, the most difficult to convert of the BR codes is the Organization schedule, which contains about 15,000 entries. This code is a 10-digit, 3-level hierarchical code (parent plus two subordinate levels) preceded by a functional prefix to indicate the type of relationship of the individual to the organization. A final assessment of the BR organization conversion task must wait until CHIVE designs or adopts an organization schedule. A byproduct of conversion could be a standardized organization abbreviation scheme. The BR abbreviation scheme could then be converted to the corresponding CHIVE schedule.

The BR location code consists of 8 digits (3/2/3) representing country, political subdivision, and city, in that order. Only the first three digits are used to show a person's country or citizenship. Location and citizenship codes are equal except when a person is traveling outside his country of citizenship. Location and citizenship are controlled by CHIVE using the ISC area code. The 8-digit BR code can be used to represent the location of CHIVE organizations or persons down to the city level.

The BR occupation code is a 7-digit hierarchical code structured so as to identify profession and area of specialization within profession. A letter prefix may be added to indicate the nature of the subject's relation to the profession. If other than the BR code is adopted or modified, it may become necessary to prepare a conversion table.

Several dates in the dossier index can be used by CHIVE. These include (a) date of birth, (b) latest date for organization affiliation, and (c) most recent location date, temporary or permanent. These are 3-digit, fielded (e.g., mo/yr) entries.

There are several BR-indexed elements which have no CHIVE counterpart. These include the religion and Soundex codes. Each of these would be suppressed during conversion for the reason that their intelligence value is considered limited. Some political affiliation is controlled in CHIVE phrasing. The [redacted] name grouping technique is considered a satisfactory substitute for the Soundex code.

25X1A

To summarize, the BR Dossier Index File offers much information for use in CHIVE. The file's organization by area simplifies the removal of [redacted] or other geographical entities. The principal dictionary representing organizations is in machine language, which would aid the preparation of conversion tables.

25X6

5.D.2.4. Intellofax Reference System (IRS) Files

5.D.2.4.1. IRS Document Index File (New)

5.D.2.4.1.1. General Description

This file is the principal machine index to collateral

INHERITED FILES
Index Files
5.D.2.4.1.1.

documents. It dates from mid-1960 to the present. The file is stored on punched cards controlled by the Machine Division/OCR. It provides selected references to intelligence information reports by means of subject searches. The file is organized for direct access by area and within area by subject. Further refinement is possible on the basis of other punched data such as organization name. In addition to punched data, each card contains printed extract data.* After the machine search, the extract is manually screened by the Intellofax Reference Group (IRG). Selected references are prepared in the form of bibliographic listings and reproduced extracts which are given to the customer for further screening.

The file contains in excess of 2 million punched cards. It is growing at the rate of approximately 500,000 cards per year. The [redacted] of the file now contains about 93,600 cards. It has an estimated annual growth rate of 24,000 cards. There is an average of 4 punched cards per indexed document.

25X6

---

*This discussion predates the conversion to the DARE technique which is now used to record indicative information on the cards.

INHERITED FILES
Index Files
5.D.2.4.1.1.

File activity averages 158 _ad hoc_ requests per month. These produce an average of 183 references per request. The references are normally screened by the IRG, and occasionally by customers, to reduce the final system output to roughly 107 references per request. In addition to _ad hoc_ requests, standing requests number between 40 and 50 per month. These generate 400 to 500 references.

There is also significant file building activity. This activity centers in the building of the CIA Library's Source Card File and the dissemination of Source Cards to other agencies. Source Cards are copies of the index punched card which are used for acquisitions and holding library functions.

The elements of information controlled by the file are identified below and in Figure 5.D-27. Because of the punched card surface area taken up by the printed information, punched information is limited to a total of 42 characters in each punched card record. Only a single entry is possible per field on a single punched card, but for some fields more than one entry can be made for a given document by punching more than one card per document.

Possible entries include:

- ISC Subject Code
- ISC Subject Modifier Code
- Tags (Intellofax)
- ISC Area Classification Code
- Subject
- Organization Abbreviation
- Place Name
- Source Code (Intellofax)
- Document Number
- Publication Date
- Security Classification Code (Intellofax)

5.D.2.4.1.2. Use in CHIVE

Each of the above elements of information relate to header and content elements of the CHIVE index record. The principal controlled elements--subject code, subject modifier code, and area code--are elements of information found in both Intellofax and CHIVE. Both systems employ common vocabularies for these. Three of the remaining fielded inputs--subject, organization, and place name--are also controlled entries, but are represented in clear text. These elements are controlled by the CHIVE system as well, but in a different record format and vocabulary. In the CHIVE record, a tag is used to identify the type of element under control. Conversion of the clear text vocabulary will prove to be a problem. The Intellofax code vocabularies are already available on punched cards.

Since the subject modifier code and area code fields may contain a variable number of entries, more than one IRS punched card may have to be combined to form one CHIVE record.

In the IRS record, the information controlled and the code schedule from which the entry was taken are identified by a preceding Intellofax tag. All but one of the tags have CHIVE tag counterparts. Thus a one-to-one relationship can be established between IRS and CHIVE tags, thereby minimizing the loss of information.

Each of the Intellofax header elements has a counterpart in the CHIVE header record. Two of these--document number and publication date--are document-derived entries, as in CHIVE. Conversion tables are required for the remaining two elements--source code and security classification--but both have a small range of values.

INHERITED FILES
Index Files
5.D.2.4.1.2.

In summary, the IRS File has a high degree of potential use in CHIVE. This is because:

- It is already in machine language
- It contains compatible elements of information
- There is a commonality of vocabularies
- Loss of punched information will be small
- 25X6 The file is primarily organized by area, which facilitates the removal of the [redacted] or other geographical entries
- Any removal on an area basis would not affect the integrity of the parent file

An important loss of printed extract information could result from conversion. However, screening of output for customers and the inclusion of a document summary, extract, or DARE image of the first page of a document could be achieved in several ways. The document number might be used to locate the source card, which contains the sought bibliographic information. Or, the entire Intellofax Document Index File could be sorted on document number. Used for this purpose, the file would no longer require more than one card per document, and could be reduced to one-fourth its present size.

INHERITED FILES
Index Files
5.D.2.4.1.2.

5.D.2.4.2. IRS Document Index File (Old)

5.D.2.4.2.1. General Description

This file is the largest machine index to collateral documents in the Agency. It covers the date span from 1948 to mid-1960. It consists of a punched-card file maintained by the Machine Division/OCR which is used to provide selected references to raw and finished intelligence reports by means of subject searches. The file is organized for direct access by subject and within subject by area. Further search refinement is possible on the basis of other punched data. In addition to punched data, the card contains printed data used by IRG to screen the punched-card output. Printed data also is used in the preparation of bibliographic listings for customers.

The file contains about 9 million punched cards. There is no growth. The [redacted] 25X6 contains 453,400 cards. There is an average of four punched cards per unique document reference.

File activity averages 58 requests per month, which produce an average of 15,130 references per month or 263 references per request. These references are screened by IRG, and occasionally by customers, to reduce the final output to an average of 87 references per request. (These figures do not include the effects of direct use of extracts by customers.)

Each punched-card record contains 32 columns for recording content and header elements of information. More than one element may appear in the same field as over- or under-punches. The elements of information contained in the record are as follows:

- ISC Subject Code (old)
- ISC Prefix Action Code (old)
- AMS Area Classification Code
- Intellofax Security Classification Code
- Intellofax Source Code (agency)
- Intellofax Locator Number Code (post)
- Intellofax Related Area Code
- Publication Date
- Document Control Number

5.D.2.4.2.2. Use in CHIVE

The elements of information identified above can be related to similar data elements within CHIVE. The basic content elements--subject, prefix action modifier, and area--are controlled-vocabulary entries. They are based on three separate code schedules: (a) the old ISC Subject Code Schedule, (b) the ISC Prefix Action Code, and (c) the AMS Area Classification Code. Similar subject and subject modifier elements are controlled by CHIVE, but they are controlled using the new ISC Subject Code Schedule.

Area information is also controlled by CHIVE. Area is identified by the appropriate CHIVE location tag linked to the ISC Area Code. The 36 prefix action codes can readily be converted to the CHIVE vocabulary.

It is highly improbable that the old ISC Subject Code can be used in CHIVE. There are 15,000 code values controlled by the old Subject Code Schedule in contrast to 5,000 code values contained in the new ISC Code Schedule. It is estimated that, at best, as much as 70% usage could be achieved at considerable sacrifice of conversion manpower and computer time. Significant potential loss of information would occur.

Header elements represented in the record are readily usable. Three of these--security classification, source, and location number--are controlled-vocabulary inputs created by the Intellofax system. The number of possible values per code schedule is extremely small, so that little manual effort is needed to equate these

INHERITED FILES
Index Files
5.D.2.4.2.2.

elements with comparable CHIVE values. The remaining header elements--publication date and document control number--are document-derived inputs which can be used "as is" in CHIVE.

Although this file represents one of the larger machine indexes to intelligence documents, it offers very little, if any, conversion potential for the following reasons:

- Lack of commonality of vocabularies which precludes a one-to-one conversion of the subject code.

- Loss of information resulting from this conversion effort. The loss would be three-fold. It would include, first, loss of some subject control. This has already been discussed. Second, it would include the effects of the removal [redacted] 25X6 index files from the parent file. The integrity of the remaining file is violated, since only about 50% of the cards containing related areas are cross-referenced for filing under each area included in the punched card record. In cases where file expansion has taken place, in order to differentiate the primary and secondary areas for proper CHIVE tagging, considerable machine searching and matching might be needed. Third, loss of information would also result from omitting the printed information on the punched card from the CHIVE record, although some solutions could be effected at an undetermined cost (see above).

- The great cost of conversion. Since the file is organized by subject, cards pertaining [redacted] 25X6 or other geographical areas could not be sorted out without searching the entire 9 million cards in the file. A heavy investment would be required in terms of manpower and machine time to convert the subject codes found in the 9 million punched cards.

INHERITED FILES
Index Files
5.D.2.4.2.2.

- Many of the documents represented by this system are outdated. There is a relatively low intelligence yield for many IR series. There is less usage of this file than of the new IRS File. The system relevancy factor is low. These facts suggest a relatively low use potential in CHIVE.

5.D.2.5. GR Files

5.D.2.5.1. Film Index File

5.D.2.5.1.1. General Description

This file is a machine index to 8mm, 16mm, and 35mm motion picture films, film strips, video and audio tapes, and disk recordings. The file is stored on punched cards maintained by Machine Division/OCR. It is used to provide selected references to motion picture films and related sound and video recordings on the basis of subject and area searches. Further search refinement is possible on the basis of header data punched in the record. In addition to punched data, each card contains a printed extract or summary of the film, used to prepare bibliographic announcements. The file is arranged by area and within area by subject. Inputs to the film/tape/disk file are from Collateral and open sources.

As of 30 June 1964, the file contained 235,600 punched cards. During FY 64, a total of 19,000 cards were

25X6

[redacted] are indexed per year. An average of 9 subjects are indexed per film, which adds 9 punched cards per film to the index file.

File activity is relatively light. During FY 64, requests averaged 14 per month. These produced an average of 313 references per request. The somewhat low activity rate is attributed in part to (a) a manual IBM card file (Intellofax type) maintained by subject and by area to answer ad hoc requests, and (b) an inventory control file maintained on punched cards to produce EAM listings on a demand basis.

The punched card record contains 27 columns used to control content and header elements of information. The data fields are as follows:

- ISC Subject Code (old)
- AMS Area Classification Code
- Language Code (GR)
- Film Type Code (GR)
- Holding Agency Code (GR)
- Publication Date (year)

INHERITED FILES
Index Files
5.D.2.5.1.1.

- Classification/Control Code (GR)
- Control Number
- Availability Code (GR)

5.D.2.5.1.2. Use in CHIVE

Most of the above-named elements of information are found in the CHIVE system. Only two of the elements--film type and availability--are simultaneously unique and essential to the control of films.

Content control over films is limited to subject and area indexing, both of which are controlled inputs. Subject entries are based on the old ISC. (See Section 5.D.2.4.2.2. for a discussion of old ISC conversion problems.) Others have been deleted, and many new ISC codes have been added. A one-to-one relationship between the AMS and ISC area classification codes can be manually developed and is practicable. Both area codes denote the same level of political hierarchy for the USSR [redacted] 25X6

Old ISC header elements are both controlled and uncontrolled. Uncontrolled, or document-derived elements, are publication date and control number (GR assigned). These require no conversion effort. The film type and availability codes currently excluded from the CHIVE header could be

INHERITED FILES
Index Files
5.D.2.5.1.2.

added "as is" to the CHIVE system. Each code schedule contains a limited number of one-character values. Security classification/control and language require some manual conversion effort, but neither imposes a difficult or time-consuming task. No loss of information is anticipated in achieving a one-to-one code mapping. The holding agency code is no longer used by GR and will be omitted.

The Film Index File offers little conversion potential because of:

- Lack of commonality of subject vocabulary needed to create a one-to-one relationship of values.
- Potential loss of information resulting from any conversion attempt, including subject information and printed bibliographic code film summary or extract.

However, if the subject codes could be equated to comparable CHIVE values, the file could be used in CHIVE. Perhaps equation could be made practical by converting only the subject codes actually appearing in the index records. This hope is based on the observation that a smaller range of subject codes are used to control films than documents. If this proves true, a further reduction in range of codes could result from [redacted] 25X6 area from the parent file. This removal is possible

INHERITED FILES
Index Files
5.D.2.5.1.2.

without affecting the integrity of either file. The number of punched cards, estimated at [redacted] 25X6 242,000 in all, is small compared to most OCR machine files.

5.D.2.5.2. Personality Index File

5.D.2.5.2.1. General Description

This file is a machine index to personality photos indexed and maintained by GR in the Personality Photograph File (see Section 5.D.3.5.1.). The file media are punched cards controlled by Machine Division/OCR. The cards provide EAM listings of personality photographs retrieved by searches for nationality and profession or name. The file is arranged by area (nationality) to simplify searches. Inputs to the photo file are from Collateral and open sources.

The file contained 246,100 punched cards on 30 June 1964. It is growing at the rate of about 16,000 cards
25X6 per year. [redacted] this file is estimated to contain 6,800 cards. No growth rate statistic is available.

Although file activity averaged only 14 requests per month in FY 64, the number of references per request averaged 3,010. Each of these references represented a unique photo number.

All 80 columns of the punched card record are devoted to the content and header control of personality photographs. The elements of information included in cards are:

- Nationality Code
- Profession Code
- Name and Title
- Holding Agency Code
- Photograph Number
- Date of Information (year)
- Quality Code
- Security Classification and Control Code

5.D.2.5.2.2. Use in CHIVE

Most of the elements of information controlled by this file are used by CHIVE. Content elements include nationality code, profession code, and clear-text name and title. These elements equate to CHIVE citizenship, general occupation, name, and position title. Nationality and profession are GR controlled-vocabulary inputs. Nationality is represented by an area code based on the

AMS Area Classification Schedule. Profession codes are derived from a modified version of GR's subject code, which is used to content-index ground photos. The personality name is an uncontrolled entry derived directly from source documents. Thus, machine listings of personality photos may contain various name spellings for the same individual. In addition to surname, given name and middle name or initials, rank, title, or position are entered, space permitting.

Aliases may appear on the filed photo, but aliases are not key punched. The nationality codes for persons who have more than one nationality are not key punched. Instead, by an early GR policy, photos are cross-referenced in the hard-copy print file, where they can be retrieved by a manual search.

In order to be fully exploited by CHIVE, name, rank, title, and position as extracted from the GR index record for conversion must be brought under control by a manual inspection of the index record. In the GR record, these elements are unfielded, and are confused with the variable length names to which they are linked.

The basic header elements--photo number, information date, and security classification--are used in CHIVE header index data. Of these, classification is the only controlled input requiring conversion. This is expressed in a simplified GR code schedule. Photo number assigned by GR is synonymous with CHIVE document number. In CHIVE, the photo number can be preceded or followed by a "P", as in GR, to indicate a personality photo. The holding agency code is no longer used by GR and therefore can be omitted.

Although the controlled inputs are based on GR vocabularies, no severe conversion difficulty or loss of information will be encountered. The AMS area classification code schedule is the most complex, time-consuming vocabulary to convert. This code schedule also controls inputs to several other OCR files under discussion, and a one-time conversion could have many-fold applications.

Although the Personality Index File provides minimum control over biographic intelligence, it is believed that this file should be used by CHIVE. Although file activity is low, this file represents the only machine index to personality photos in the Agency.

INHERITED FILES
Index Files
5.D.2.5.2.2.

5.D.2.5.3. Ground Photograph Index File

5.D.2.5.3.1. General Description

This file represents a machine index to ground photography which is indexed and maintained by GR. The file storage medium is punched cards. The file is controlled by Machine Division/OCR and is used to provide EAM listings of ground photography coverage. Searches are made on the basis of coordinate, subject, or area. Further search refinement is possible using other punched data in the record. The file is arranged by area and within area by coordinates. This arrangement expedites searches. Inputs to the file are from Collateral and open sources.

As of 30 June 1964, the file contained 1,224,100 punched cards. Roughly 193,000 cards were added during 25X6 FY 64. [redacted] cards. No growth data is available for this portion. An average of three subject codes are used to describe the contents of a single photo. The result is three punched card records for an average photo, or one card per subject.

INHERITED FILES
Index Files
5.D.2.5.3.1.

File activity is low because most searches are performed on the photograph file itself (see section 5.D.3.5.3.). Total requests during FY 64 numbered only 7. Most of these were in support of internal GR activities rather than customer service. During the same period an average of 5,326 photo references were retrieved per request. However, this average figure is misleading because one request produced 25,299 references.

The elements of information contained in the punched card record are listed below. These elements occupy all 80 card columns.

- Coordinates (latitude and longitude)
- Area Code
- Subject Code
- Place Name and Subject (clear text)
- Classification and Control Code
- Photograph Number
- Date of Information (year)
- Type Code
- Quality Code
- Card Type

INHERITED FILES
Index Files
5.D.2.5.3.1.

5.D.2.5.3.2. Conversion Potential

Except for type code, quality code, and card type, each of the above elements of information are used by CHIVE. Coordinates, area code, subject code, and clear-text place name and subject are comparable to CHIVE content indexing.

Code entries are made in controlled GR vocabularies. The area code closely resembles the DIA digraph code, slightly modified to accommodate recent internal file changes. The subject code is a GR product that has expanded as the need dictates. Coordinates are usually obtained from established gazetteers but occasionally are document- or map-derived.

Place names are the only uncontrolled or clear-text entries. An attempt is made to standardize or legalize name spellings, but in practice these are often extracted from the source document without change.

Clear-text place names are used in CHIVE in conjunction with a location tag. Thus, in lieu of converting GR's coordinates, location information as found in the punched card can be combined with the GR place name. Together, these can be accepted "as is," to be used by CHIVE for future gazetteer development. If GR coordinates

are not retained by CHIVE, area searches against the
25X6 USSR [redacted] will be limited to country. In contrast, GR now performs area searches using coordinates, oblast, or province.

Neither the subject nor area codes are intellectually complex to convert, but the expenditure of manual effort can be significant. Both GR vocabularies contain a relative small number of code values. It will not be practicable in CHIVE to search on uncontrolled subject entries.

Among the header elements are security classification, type, and quality. These are controlled inputs based on GR code schedules. Classification/control is limited to about 6 values. Type and quality, two single-digit codes,are unique to the control of photographs. These can be used in CHIVE. Photo number is assigned by GR, but this number functions like the CHIVE document number and can be accepted "as is." The last column in the punched card indicates the location of duplicate punched cards, an entry serving no purpose in CHIVE.

CHIVE can use the Ground Photo Index File. Conversion costs are high, however, in light of the extremely low file activity.

INHERITED FILES
Index Files
5.D.2.5.3.2.

5.D.3. DOCUMENT IMAGE FILES

5.D.3.1. SR Files

25X6 5.D.3.1.1. [redacted] Serial Traffic File

5.D.3.1.1.1. General Description

This file contains special source documents, printed 5" x 8" cards, aperture cards and microfilm reels stored in the SR library for central reference purposes. Contents of the documents pertain to international political, economic, military and scientific topics.

The file is sequenced (from major to minor order) by source, area, subject, year, and document number.

The file size and activity are as follows:

File Size	1,786,781	items in total file (with no available estimate of [redacted] materials) 25X6
Annual Accretion	96,600	items (total file)
Annual Purge Rate	0	
Reference Frequency	daily	

This file is not purged, although partial reports for the years 1950 through 1958 have been relocated to the Records Center.

INHERITED FILES
Document Image Files
5.D.3.1.1.1.

5.D.3.1.1.2. Use in CHIVE

25X6 [redacted] materials are scattered throughout the files, making their segregation difficult.

25X6 [redacted]

5.D.3.1.2.1. General Description

This file is composed of special source documents and 35mm aperture cards containing [redacted], 25X6 scientific and political information. It is stored in the SR library and used for reference purposes. These documents are distinguished from those in the [redacted] 25X6 Serial Traffic File in that they have been assigned a

25X6 [redacted] indicator.

The file is sequenced (in major to minor order) by source, area, subject, document number and year.

File size and activity of the file are as follows:

File size (items)	35,735
Annual Accretion Rate (items)	2,400
Annual Purge Rate	0
Reference Frequency	daily

This file is not purged, although partial reports for the years 1950 through 1958 have been relocated to the Records Center.

INHERITED FILES
Document Image Files
5.D.3.1.2.1.

5.D.3.1.2.2. Use in CHIVE

This file is made up entirely of [redacted] 25X6
Therefore, no selection would be required to isolate this file for conversion.

5.D.3.1.3. PI Hard Copy File

5.D.3.1.3.1. General Description

This file consists of hard-copy documents used as the SR library reference file for sensitive PI and non-sensitive NPIC PI documents.

The file is sequenced by source and document number.

File size and activity of the file are as follows:

File Size	13,403 items in total file (with no available estimate of [redacted] materials) 25X6
Annual Accretion rate	2,403 items (total file)
Annual Purge Rate	0
Reference Frequency	daily

This file is not purged.

25X6 [redacted]

5.D.3.1.4.1. General Description

This file consists of hard copy and teletype sheets

INHERITED FILES
Document Image Files
5.D.3.1.4.1.

25X6

[redacted]. It is located in the SR reference library. The file is ordered by title.

File size and activity of the [redacted] are as follows: 25X6

File Size	88,763 items
Annual Accretion Rate	11,000 items
Annual Purge Rate	1,000 items
Reference Frequency	daily

Hard-copy documents are reviewed each month and may be purged after one year's retention. Purged documents are relocated at the Records Center. All electrically received documents are purged after one year.

5.D.3.1.4.2. Use in CHIVE

There is no problem in isolating [redacted] in this file. 25X6

5.D.3.1.5. Teletype File

5.D.3.1.5.1. General Description

The Teletype File is composed of two sub-files. These are:

25X6 [redacted]

INHERITED FILES
Document Image Files
5.D.3.1.5.1.

(b) Electrical and Non-Sensitive CRITICOM Teletype File

25X6 The Serialized [redacted] contains current documents bearing a [redacted] designator. 25X6 The documents are filed in serial order within this area designator. Teletype materials are retained for one year only, with six months' collection of material kept at Headquarters, and the remaining six months of material filed at the Records Center.

The file is in sequence by source, area, subject and month.

The file size and activity statistics are based on 1964 volumes.

File Size (items)	4,416
Annual Accretion (items/year)	4,416
Annual Purge Rate (items)	4,416
Reference Frequency	daily

The purging criterion is date of receipt. The file is purged monthly of all materials over twelve months old.

The Electrical and Non-Sensitive CRITICOM File makes up the balance of the SR Teletype File. It contains teletype materials other than those with a [redacted] 25X6 designator. It includes documents having [redacted].

INHERITED FILES
Document Image Files
5.D.3.1.5.1.

materials classified under other major area designations as well as documents which are not [redacted] 25X6

The file is sequenced by source, area, subject and month.

File size and activity statistics are based on 1964 volumes.

File Size (items)	279,079
Annual Accretion (items/year)	279,079
Annual Purge Rate (items)	279,079
Reference Frequency	daily

There is no estimate as to what percentage of this total

25X6 [redacted]

Purging is based strictly on date of receipt. The file is purged monthly of all materials over twelve months old.

5.D.3.1.5.2. Use in CHIVE

25X6 [redacted] is composed ent[redacted]efore, there is no partition problem.

The Electrical and Non-Sensitive CRITICOM File is not so composed. Extensive effort is required to browse

25X6 each item to determine its [redacted]. In addition, these materials rapidly become obsolete. Therefore, conversion of the existing file is not warranted. However, newly-received teletype could readily be screened and partitioned as received. This approach is practical, and would effect segregation of the entire file contents after twelve months of screening.

5.D.3.2. FIB Files

5.D.3.2.1. Active Installation File

This file is composed of three sub-files. These are the:

- Installation Folder File
- Installation 5" x 8" File
- Codeword [redacted] Installation Folder File

25X1A

5.D.3.2.1.1. Installation Folder File

5.D.3.2.1.1.1. General Description

This file supplements the Installation 5" x 8" Card File (see Section 5.D.3.2.1.2.). It contains all-source information relating to priority installations. It is made up of the overflow which consists of various-sized documents which are too large for inclusion in the 5" x 8" file. Access is by means of index listings which provide

INHERITED FILES
Document Image Files
5.D.3.2.1.1.1.

the FIB dossier number as the search output. The file is ordered by assigned FIB dossier number.

The file size and activity of the [redacted] 25X9 are as follows:

File Size (dossiers)	500-600
(items)	2,000
Annual Accretion Rate (items)	200
Annual Purge Rate (items	100
Reference Frequency	daily

A limited amount of purging is done on a continuing basis as a dossier is activated or superseded, or as redundant documents are removed from a dossier.

5.D.3.2.1.1.2. Use in CHIVE

Since the dossiers are in accession order, [redacted] 25X9 items are scattered throughout the file. It will be necessary to manually select these dossiers if they are to be separated from the body of the file for use by CHIVE.

5.D.3.2.1.2. Installation 5" x 8" File

5.D.3.2.1.2.1. General Description

This is an all-source file of documents, photos, and cards folded wherever necessary to 5" x 8" size, containing

INHERITED FILES
Document Image Files
5.D.3.2.1.2.1.

information on priority installations (including most

25X1A [redacted] The items are included in installation dossiers which are filed by assigned dossier number. This file is actively maintained. Access to the file is through tab listings by installation name, commodity, or area breakdown.

Dossiers are filed by assigned FIB dossier number. The file size and activity of the [redacted] are as 25X6
follows:

File Size (dossiers)	2,541	
(items)	58,000	(estimated)
Annual Accretion Rate (items)	5,000	(estimated)
Annual Purge Rate (items)	1,500	(estimated)
Reference Frequency	daily	

The file undergoes limited ad hoc purging whenever reviewed by the analyst. Only redundant or incorrect items are removed.

5.D.3.2.1.2.2. Use in CHIVE

25X6 [redacted] are interspersed throughout the file. Segregation can be accomplished via a manual selection using the geographic area listing.

25X1A

5.D.3.2.1.3. Codeword [redacted] Installation Folder File

5.D.3.2.1.3.1. General Description

This file is made up of special materials relating to [redacted] targets. It is used as a supplement to the active Installation 5" x 8" File and the Installation Folder File. These files have not been merged. Physically, the file is made up of legal-sized dossiers containing documents, cards, photos, maps and computer listings. The dossiers are arranged in sequence by assigned [redacted] number.

25X1A

25X1A

The file size and activity of the [redacted] as follows:

25X6

File Size (items)	3,000
Annual Accretion Rate (items)	2,000
Annual Purging Rate (items)	1,400
Reference Frequency	daily

Superseded documents are removed as new information is received.

5.D.3.2.1.3.2. Use in CHIVE

25X1A

25X1A

25X6

[redacted] dossiers within the Sino-Soviet Bloc are maintained in [redacted] number order. Some manual selection of [redacted] materials will be necessary to provide a breakout.

INHERITED FILES
Document Image Files
5.D.3.2.1.3.2.

5.D.3.2.2. Inactive Installation File

5.D.3.2.2.1. General Description

This is made up of two physically distinct files composed of inactive segments of the Installation Folder File (5.D.3.2.1.1.) and the Installation 5" x 8" File (5.D.3.2.1.2.). The inactive files are used as an adjunct to their active counterparts. They contain collateral information only on non-priority installations (located in large cities) and are no longer actively maintained. They are still referenced and are used as a basis for creating active dossiers whenever necessary. Both of the inactive files are ordered by the assigned FIB dossier number within city. They are neither augmented nor purged but are referenced on a weekly basis. The combined file consists of approximately

25X6 [redacted] containing an estimated 30,000 items (cards or documents).

5.D.3.2.2.2. Use in CHIVE

Because of the low reference rate and the low priority nature of the contents of these inactive files, they are not considered candidates for conversion.

INHERITED FILES
Document Image Files
5.D.3.2.2.2.

5.D.3.2.3. Town Folder File

5.D.3.2.3.1. General Description

This file consists of selected documents of intelligence value relating to installations of a non-priority nature. It is all-source and contains miscellaneous textual and graphic documents maintained in dossiers by town (or city). No effort is made to group items by installation within the town folders. The folders are not machine indexed. The file is ordered alphabetically by town name. File size and activity are as follows:

File Size (dossiers)	739	
(items)	22,000	(estimated)
Annual Accretion Rate (items)	2,000	(estimated)
Annual Purge Rate (items)	1,500	
Reference Frequency	daily	

Redundant and superseded items are purged from the file periodically.

5.D.3.2.3.2. Use in CHIVE

Because of the town-oriented organization of this file, [redacted]

25X6

INHERITED FILES
Document Image Files
5.D.3.2.3.2.

5.D.3.2.4. General Reference File

5.D.3.2.4.1. General Description

This file contains background information oriented to subject/commodity. It consists of various-sized textual documents organized by broad subject categories (e.g., guided missiles, textile mills) within location (province). [redacted] kept in open-shelf storage. The file currently is neither accreted nor purged. It undergoes very infrequent reference usage. It is anticipated that this file will eventually be reviewed and reduced in size. 25X6

5.D.3.2.4.2. Use in CHIVE

Because this is a low-activity, small-volume reference file used for browsing, it is not considered a candidate for a CHIVE conversion effort.

5.D.3.3. BR Files

5.D.3.3.1. One Name File

5.D.3.3.1.1. General Description

This file consists of original documents, extracts, abstracts, and reproductions of Collateral and open-source materials collected on political, scientific, technical, economic, and cultural personalities. It is used as a source of biographic information by authorized personnel within the intelligence community. The biographic data is recorded in the form of typed or printed 4" x 6" and 5" x 8" cards, as well as larger documents folded to an appropriate size. Documents are stored in seven-drawer files within a vaulted area.

The file is in alphabetic sequence by name within country. There is no index file associated with this document file.

File size and activity are as follows:

25X6

	Total File	[redacted]
File Size (items)	9,224,900*	428,400 (est.)
Annual Accretion (items/yr.)	882,000*	78,000
Annual Purge Rate (items/yr.)	--	1,200
Reference Frequency	daily	

Duplicate information is purged from the file on a continuing basis as the material is pulled to service a request. Analysts often summarize the data and remove duplicative items prior to refiling.

---

*From OCR Annual Report for Fiscal Year 1964 (volumes reported as of June 1964).

INHERITED FILES
Document Image Files
5.D.3.3.1.1.

5.D.3.3.1.2. Use in CHIVE

25X6 [redacted] n

Conversion to microform for use by CHIVE would offer advantages of (a) standardization--elimination of the separate filing systems for 4" x 6" and 5" x 8" materials, and (b) economy--reduction in storage space.

5.D.3.3.2. Dossier Folder File

5.D.3.3.2.1. General Description

This is a central biographic intelligence file, a consolidation into individual dossiers of materials pertaining to selected foreign scientific, technical, political, economic, and military personalities. The file contains information reports, newspaper clippings, bibliographic data, photostats, Form 214 Biographic Information Sheets (which summarize information contained in the dossier), and the latest copy of Form 844 Biographic Code Sheet. The documents are filed in open shelves in legal size folders. The file is used in conjunction with the BR punched-card Dossier Index File (see section 5.D.2.3.1.). It is ordered numerically by dossier number, which is a composite of year and accession number.

File size and activity for the [redacted] are as follows: 25X6

File Size (dossiers)	16,498
Annual Accretion Rate	
(dossiers/yr)	300
(documents/yr)	3,600
Annual Purge Rate	
(documents/yr)	900
Reference Frequency	daily

The file is purged on a continuing basis as dossiers are pulled for customers or used in the preparation of written reports. Prior to refiling, analysts remove a limited number of redundant documents as the dossier is analyzed.

5.D.3.3.2.2. Use in CHIVE

The file is arranged in numerical sequence regardless of nationality. However, a machine listing of the index file can be used as [redacted] 25X6
dossiers from the file.

5.D.3.3.3. Scientific Organization File

5.D.3.3.3.1. General Description

The file referred to here is the Scientific Organization File peculiar to [redacted] This file contains byproducts of the biographic function. It is used to maintain information on key scientific and educational institutes. It contains information on the location of S&T institutes, their history, and the nature of the work carried out. The file is used by analysts in their daily processing and in answering requests. The information is recorded primarily on 5" x 8" replicas of the cards that are filed by name in the One Name File. The file is sequenced by organization name.

25X1A

25X6

[redacted] and activity are estimated as follows:

File Size (items)	40,000
Annual Accretion Rate (items/yr)	1,000
Annual Purge Rate (items/yr)	500
Reference Frequency	weekly

The file is purged annually. A limited number of outdated or superseded items are purged.

INHERITED FILES
Document Image Files
5.D.3.3.3.1.

5.D.3.3.3.2. Use in CHIVE

25X6

[redacted] of the file since the nationality of the S&T institutes is known.

5.D.3.3.4. Scientific Background File

5.D.3.3.4.1. General Description

This is a relatively small file of scientific documents used in [redacted] to help answer requests and to write reports. It consists primarily of bulky reports containing important historical information which is not always available in other BR files. The documents are both letter and legal size. The file is in sequence by scientific field of interest.

25X1A

25X6

[redacted] and activity are as follows:

File Size (items)	500
Annual Accretion (items/yr)	25
Annual Purge Rate (items/yr)	0
Reference Frequency	weekly

The file is not purged because of its relatively small size and because of the historical nature of its contents.

INHERITED FILES
Document Image Files
5.D.3.3.4.1.

5.D.3.3.4.2. Use in CHIVE

This file, which occupies only a single four-drawer safe, is believed to be best-suited to its purpose in hard-copy form. Conversion to microfilm is not warranted.

5.D.3.3.5. Political Background File

5.D.3.3.5.1. General Description

This file contains biographic documents used in
25X1A [redacted] They are of historical value or are lengthy documents which cite more than one low-level personality of potential interest. Documents are either letter or legal size. The information in the file is categorized topically into political, economic, or military subjects. It is used in answering requests or writing reports. The file is sequenced by political subject category.

File size and activity of [redacted] are as 25X6
follows:

File Size (items)	2,000
Annual Accretion (items/yr)	100
Annual Purge Rate (items/yr)	0

INHERITED FILES
Document Image Files
5.D.3.3.5.1.

This file is not purged because of the historical nature of its contents.

5.D.3.3.5.2. Use in CHIVE

25X6 [redacted] documents are intermixed with others in the file and are difficult to disassociate. Because of the bulkiness of the documents and the relatively low activity, conversion to another medium or form is not warranted.

5.D.3.3.6. McBee Card File

5.D.3.3.6.1. General Description

This file consists of McBee edge-notched 5" x 8" cards prepared about scientific, technical and economic personalities on whom information is limited. Each card contains on the front side a thumbnail sketch of the individual identical to that contained on the Biographic Information Sheets filed in the Dossier Folder File, while the reverse side contains typed data extracted from documents. Currently, no new items are entered into this file, but the existing cards are actively referenced and maintained. This file is used as an adjunct to the Dossier Folder File and is accessed through the punched card index. It is used as a source of biographic information by authorized personnel in the intelligence community. The file is ordered numerically on case number.

File size and activity of the [redacted] as follows: 25X6

File Size (cards)	2,000
Annual Accretion Rate	0 (accretion discontinued)
Annual Purge Rate (cards/yr)	50
Reference Frequency	daily

As additional information becomes available on an individual, a dossier is created and added to the Dossier Folder File (see section 5.D.3.3.2.), allowing a corresponding reduction to the McBee Card File.

5.D.3.3.6.2. Use in CHIVE

Although the file is arranged in numerical sequence, a machine listing of the index file can be used to select

25X6 [redacted] items from the file.

5.D.3.4. <u>CIA Library Files</u>

5.D.3.4.1. IRS Intelligence Document File

5.D.3.4.1.1. General Description

This is a file of collateral intelligence reports, both raw and finished. The file is used to fill requests for documents derived primarily from Intellofax reference searches, but also from other sources. The file is composed of two media: (a) microfilm, and (b) hard copy (paper).

The microfilm portion is composed of aperture cards (Mil 'E' Spec) containing 16mm microfilm images of collateral intelligence reports. Some actifilm cards are interspersed throughout the file. Actifilm cards are used wherever oversized document enclosures occur. They contain up to eight 35mm images per card. The file is used to produce hard copy for consumers on demand, either in response to Intellofax request tapes or Form 1395 order. The file is indexed by the IRS Document Index File, a punched card file which is searchable by subject/area keys (Section 5.D.2.4.). The file is ordered numerically by document number.

INHERITED FILES
Document Image Files
5.D.3.4.1.1.

File size and activity of the microfilm portion are as follows:

	Total File	[redacted] 25X6
File size		
(aperture cards)	3,175,600*	323,851**
(documents)	2,886,910***	294,410***
Annual Accretion Rate		
(aperture cards/yr)	203,600*	24,200
(documents)	185,090***	22,000
Annual Purge Rate	0	0
Reference Frequency	daily****	daily

The file, which has been in existence since 1948, has never been purged.

The hard copy portion consists of both raw and finished intelligence reports located partially in the Library/Circulation Branch and partially at the Records Center. It is used primarily to fill document requests. Copies of documents are usually released. Occasionally, the file items themselves are loaned.

---

*OCR Annual Report, Fiscal 1964 (as of June 1964)

**Actual tabulation as of October 1964

***Approximation based on MD/OCR estimate of 1.1 cards per document

****OCR Annual Report (FY 1964) indicates the annual number of pages reproduced to be 454,600

INHERITED FILES
Document Image Files
5.D.3.4.1.1.

The file may be consulted directly by authorized users. It consists largely of legal and letter-size documents, although some oversized pages or even photos and microfilm reels may be included as enclosure material. The file is organized by document number within series within source. In general, material is stored in the hard-copy portion (rather than microfilm) whenever the material is not suitable for microfilming.

File size and activity for [redacted] 25X6 the hard-copy holdings are as follows:

File Size (items)	561,000*
Annual Accretion (items/yr)	15,000-20,000
Annual Purge Rate (items/yr)	2,000
Reference Frequency	daily

Purging is accomplished at irregular intervals in order to conform to space limitations. At headquarters, purging is done on the basis of use patterns and availability of material from other sources. At the Records Center, purging is accomplished on the basis of the Library Records Control Schedule. Despite occasional purging efforts, some items in the file date back to 1948.

---

*This estimate is based on the assumption that [redacted] 25X6 materials constitute 10% of total file which consists of 5,610,000 documents.

5.D.3.4.1.2. Use in CHIVE

Relative to use of the microfilm portion, aperture cards containing [redacted] images are intermixed with those of other areas throughout this file. An area search, however, of the IRS Document Index File could be used to select [redacted]. This could be followed by a manual selection of references from the aperture card file which would physically assemble a separate [redacted] image file.

25X6

Relative to the use of the hard-copy portion, physical separation of [redacted] from the hard-copy file could only be accomplished by direct examination of over 5 million file items, since much of the material is NODEXed.

25X6

5.D.3.4.2. Treaty File

The Treaty File is maintained by the Information Section of the Library. It is used as a reference tool. It is made up of information reports, largely State airgrams, which have been disseminated to the Library by Document Division on the basis of requirements. The total file occupies ten linear feet and includes some [redacted].

25X6

5.D.3.4.3. Speech File

The Speech File is also maintained by the Information Section of the Library. It is used as a reference tool. It is made up largely of pamphlets and clippings from open literature, but may include a few information reports. The total file, which consists of speeches made by public figures from all nations [redacted] occupies only six linear feet. 25X6

5.D.3.5. GR Files

5.D.3.5.1. Personality Photograph File

5.D.3.5.1.1. General Description

The Personality Photo File is a collection of individual and group personality photographs of significant international personalities. It may be categorized into three segments or sub-files according to the level of processing applied in filing the photographs.

- Processed Photo File--Each photo is given a CIA Photograph Number and is filed in a 4" x 6" folder under the individual personality's name within country. Group photos are reproduced as necessary and filed redundantly under the name of each person appearing in the photo. The file is indexed by the Personality Index File (see section 5.D.2.5.2.).

INHERITED FILES
Document Image Files
5.D.3.5.1.1.

- Semi-Processed Photo File--These photos, considered of limited value, are filed in the 4" x 6" folders under the name of the principal personality in the photo. They are not assigned a CIA Photograph Number nor are they filed redundantly. These photos are given full processing if they are used to satisfy a requirement. The Processed Photo File and the Semi-Processed Photo File are logically separate files but are physically stored in the same set of 4" x 6" folders.

- Unprocessed Photo File--Photos of marginal intelligence value are filed within a physically separate source or history file which is used as a backup, browsing file for filling photo requests. These photos are filed alphabetically by first letter of the surname (of the principal personality) within country. Some photos are also filed categorically according to events such as NATO or SEATO conferences. Items in this file that become active are given full processing and are filed in the 4" x 6" jacket file.

The file size and activity of the [redacted] 25X6 portion of the combined files are as follows:

File Size (items)	12,500
Annual Accretion Rate (items/yr)	1,750
Annual Purge Rate (items/yr)	120
Reference Frequency	daily

These files undergo a limited purging on a monthly basis. Purging selection is based on criteria such as photo quality, date, classification, and camera angle.

INHERITED FILES
Document Image Files
5.D.3.5.1.1.

5.D.3.5.1.2. Use in CHIVE

The processed and semi-processed photos are ordered by name within country, permitting ready segregation of [redacted]. The unprocessed photo file is predominantly organized along country lines suitable for geographic partitioning.

25X6

5.D.3.5.2. Personality Photograph Negative File

5.D.3.5.2.1. General Description

This file consists of 70mm film negatives filed in individual jackets which are ordered by an assigned CIA Photograph Number. It is used to produce finished photographs on request and is cross-referenced from the Personality Photograph File. It contains approximately 12,500 individual and group photos of [redacted] interspersed with photos of personages from other nations. The [redacted] file grows at an estimated rate of 1,000 photos per year. There is no effort to purge the file.

25X6

25X6

5.D.3.5.2.2. Use in CHIVE

Selecting [redacted] photographic negatives will require examination of the entire file.

25X6

INHERITED FILES
Document Image Files
5.D.3.5.2.2.

5.D.3.5.3. Ground Photograph File

5.D.3.5.3.1. General Description

This file is a collection of 4" x 6" ground photos, mounted on 5" x 8" cards with visual color tabs denoting the subject content of the photos. Subjects of the photos, in decreasing order of importance, are economic, terrain, military, scientific, political, and sociological. The file is ordered alphabetically by place name within area, country, WAC number, or province.

File size and activity for the [redacted] are 25X6 as follows:

File Size (items)	50,000
Annual Accretion Rate (items/yr)	5,700
Annual Purge Rate (items/yr)	300
Reference Frequency	daily

The file is purged on a continuing basis as determined by date, quality, classification, and source.

5.D.3.5.3.2. Use in CHIVE

The geographic organization of this file is suitable

25X6 [redacted]

INHERITED FILES
Document Image Files
5.D.3.5.3.2.

5.D.3.5.4. Ground Photograph Negative File

This file consists of 70mm negatives filed in individual jackets which are sequenced numerically by CIA Photograph Number. It is used as the source from which requested photo copies are produced, in conjunction with the Ground Photograph File. [redacted] 25X6 filed interchangeably with those of other countries making a breakout of [redacted] This file 25X6 is not purged.

5.D.3.5.5. Film File

This file contains positive prints of all motion pictures (mostly 16mm) acquired by the Graphics Register, Film Branch. They are filed on open shelves in order of accession, and are identified by an assigned Motion Picture Number, which consists of a letter and four digits.

There are approximately 400 reels of [redacted] 25X6 with an estimated growth rate of 65 reels annually. In addition, an indeterminate number of reels contain some 25X6 information on [redacted] but are attributed to areas other 25X6 [redacted]. Although the overall file is purged annually, 25X6 [redacted] because of their limited volume and high use potential. Reference to this file is on a day-to-day basis with activity in [redacted] segment occurring about once a week. The file is used in conjunction with the Film Index File (see section 5.D.2.5.1.) maintained in Machine Division. There is usually a 35mm duplicate negative for each film, which is retained at the Records Center. The file of duplicate negatives is called the Film Negative File. 25X6

5.D.3.5.6. 35mm Color Slide File

This is a relatively small file of 35mm color slides of ground photographs from all over the world. The file is ordered by subject within country, but physical separation of the [redacted] is inadvisable because of the small quantity of these items plus the fact that reference activity characteristically crosses country lines. There are approximately [redacted] in the file, which is consulted only about 3 to 4 times per week. 25X6 25X6

Next 1 Page(s) In Document Exempt

Figure 5.D-3

Format A

SR SUBJECT/COMMODITY FILE CARD

Col. 1	File Designator Number:
	1 = SR Subject/Commodity File
2-7	SR Series Number
8	Blank
9-13	Document Identification Number
14	Publication Year (overpunching used)
15-16	Message Date - month
17	Message Date - year
18-19	Information Date - month
20-21	Information Date - year
22-24	Page Number (1st page of information indexed)
25-33	Subject Code (Modified ISC) or SCC Commodity Index Code
34-36	Use or Nationality Code
37-38	Action Code
39	Moscow Column
40	Blank
41-46	IR Area Codes:
	Col. 41-43 Country Code
	44-46 Political Subdivision

Figure 5.D-3 (Cont'd)

47	TFA Code (TO, FROM, or ABOUT)
48-59	Clear Text City Name or Additional Areas
60	Descriptor Column (used in conjunction with Cols. 62-79)
61	Type Code
62-79	Alphabetic and/or Numeric Clear Text
80	Process Date

Figure 5.D-4

25X6

Format B

[redacted] AREA DETAIL FILE CARD

Col. 1	File Designator Number = 4
2-7	SR Series Number
	(prior to 1954, Col. 2-4 only)
8	Blank
9-13	Document Identification Number
14	Publication Year (overpunching used)
15-16	Message Date - month
17	Message Date - year
18-19	Information Date - month
20-21	Information Date - year
22-27	IR Area Codes:
	Col. 22-24 Country Code
	25-27 Political Subdivision
28	TFA Code (TO, FROM, ABOUT)
29-40	Clear Text City Name
41-46	Area Codes:
	Col. 41-43 Country Code
	44-46 Political Subdivision

Figure 5.D-4 (Cont'd.)

47	TFA Code	
48-59	Clear Text City Name	
60-65	Area Codes:	
	Col. 60-62	Country Code
	63-65	Political Subdivision
66	TFA Code	
67-78	Clear Text City Name	
79	Moscow Column (obsolete)	
	An x overpunch in this column indicates the existence of additional area cards for the same document	
80	Process Date	

Figure 5.D-5

Format C

[redacted] ORGANIZATION FILE CARD
PERSONALITY FILE CARD
FOREIGNER FILE CARD

25X6

Col. 1	File Designator Number:	
	7 = [redacted] Organization File	
	2 = [redacted] Personality File	25X6
	2 (x overpunch) = [redacted]	25X6
2-7	SR Series Number	
8	Blank	
9-13	Document Number	
14	Publication Date (overpunching)	
15-16	Message Date - Month	
17	Message Date - Year	
18-19	Information Date - Month	
20-21	Information Date - Year	
22-35	Location of Personality and/or Organization	
	Col. 22-24 IR Country Code	
	25-27 Political Subdivision	
	28-35 Clear Text City Name	
36-47	[redacted]	25X6

Figure 5.D-5 (Cont'd.)

48-61	Personality Name
62-64	Occupation Code for Personality
65-67	Page Number (1st page of information indexed)
68-70	[redacted]
71	TFA Code (indicates whether document is TO, FROM OR ABOUT person named)
72-80	Organization Code:

Col. 72	Organization Element Type
73-76	Organization Element (3rd or 4th level)
77-78	2nd Level Organization
79-80	1st Level Organization

Figure 5.D-6

Format D

SR SOVIET ORGANIZATION FILE CARD

SR SOVIET PERSONALITY FILE CARD

SR SOVIET FOREIGNER FILE CARD

Col. 1	File Designator Number:
	8 = Soviet Organization File
	9 = Soviet Personality File
	9 (x overpunch) = Soviet Foreigner File
2-7	SR Series Number
8	Blank
9-13	Document Identification Number
14	Publication Year (overpunching)
15-16	Message Date - month
17	Message Date - year
18-19	Information Date - month
20-21	Information Date - year
22-24	Country Code
25-27	Political Subdivision
28-32	Clear Text City Name
33-34	Organization Type
35-38	Organization Number (3rd or 4th level code)

Figure 5.D-6 (Cont'd)

39-43	Blank
44-45	Chief Directorate
46-47	Ministry Name
48-60	Personality Name (surname first)
61	Blank
62-63	Identification of Sovnarkhoz
64-67	Page Number
68-70	Nationality
71-78	Title
79	TFA Direction Indicator
80	Process Date

Figure 5.D-7

Format E

ALL OTHER ORGANIZATION FILE CARD

Col. 1	File Designator Number - 6
2-7	SR Series Number
8	Blank
9-13	Document (Message) Number
14	Publication Year
15-16	Message Date - month
17	Message Date - year
18-19	Information Date - month
20-21	Information Date - year
22-24	Area (Country) Code
25	Sub-Division
26	FTA Code
27-33	Clear Text City Name
34-40	Title
41-43	Nationality of Person
44-49	Organization
50-52	Nationality of Organization
53-54	SR Association Indicator

Figure 5.D-7 (Cont'd.)

55-57	Page Number
58-60	Occupation of Person
61-65	Blank (or continuation of Personality Name)
66-78	Personality Name
79	Blank
80	Process Date

Figure 5.D-8

Format F

ALL OTHER PERSONALITY FILE CARD

ALL OTHER FOREIGNER FILE CARD

Col. 1	File Designator Number: 3 = All Other Personality File 3 (x overpunch) = All Other Foreigner File
2-7	SR Series Number
8	Blank
9-13	Message (Document) Number
14	Publication Year
15-16	Message Date - month
17	Message Date - year
18-19	Information Date - month
20-21	Information Date - year
22-24	Country Code
25	Subdivision
26	FTA Code
27-33	Clear Text City Name
34-40	Title
41-43	Page Number

Figure 5.D-8 (Cont'd.)

44-47	Blank (or continuation of Personality Name)
48-60	Personality Name
61-66	Organization
67	Blank
68-70	Nationality of Person
71-73	Nationality of Organization
74-75	SR Indicator
76-78	Occupation of Person
79	Blank
80	Process Date

Figure 5.D-9

Format G

PI SUBJECT/COMMODITY FILE CARD

PI AREA FILE CARD

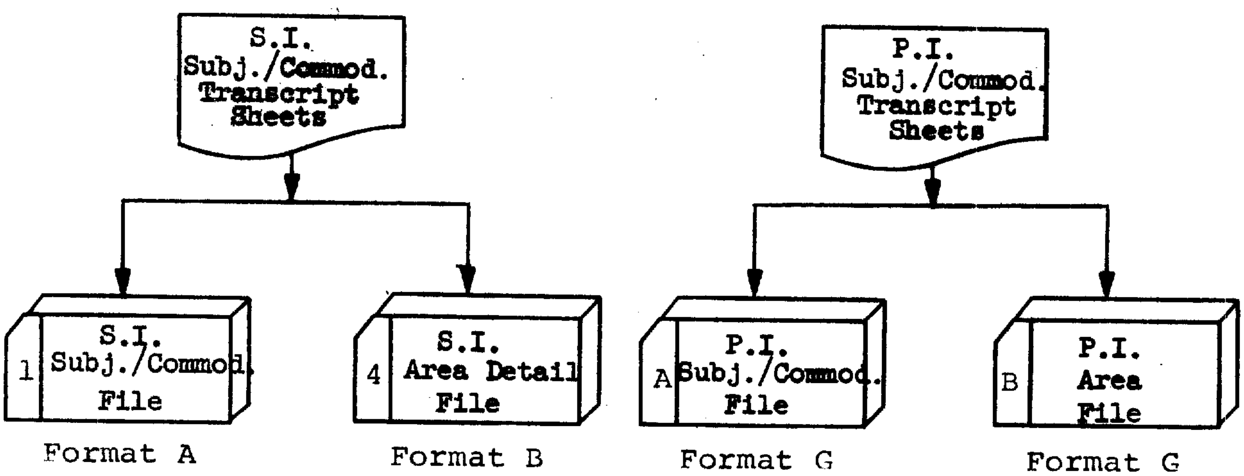
Col. 1	File Designator Number
	A = PI Subject/Commodity File
	B = PI Area File
2-3	Security Code
4-5	Agency Originating Report
6-12	Document Type
13-18	Document Number
19	Publication Date - month
20-21	Publication Date - year
22	Information Date - month
23-24	Information Date - year
25-33	Subject/Commodity Code
34-36	Use/Nationality Code
37-38	Action Code
39-40	Page Number
41-46	Location Code
	Col. 41-43 Country Code
	44-46 Political Subdivision
	(41-46 contain map number if no city name given)

Figure 5.D-9 (Cont'd.)

47	Area Tag (indicates whether to reproduce card for the Area File)
48-59	Clear Text City Name when 41-46 used for country and subdivision codes; otherwise used for coordinates
60-61	Tag (modifies columns 62-79)
62-79	Clear Text
80	Tag (indicates whether card is to be reproduced for filing under clear text in the Check File)

Figure 5.D-10

SUBJECT/COMMODITY AND AREA FILES



25X6

Figure 5.D-12

JOB 3 FILE STATISTICS*

File	File Size	Monthly Growth
Key Word Cards	724,668	
SI Cross Reference Cards	4,000	
Series Cards	134,000	
Sub Total	862,668	9,700
Key Word Cards	39,198	
PI Cross Reference Cards	4,053	
Series Cards	19,566	
Sub Total	62,817	3,500
Total	925,485	13,200
25X6 [redacted]	45,000 (estimated)	5,820 (annual)

File Location

Headquarters Building	6 Month Holdings 150,000 Cards**
Records Center	Balance

---

* Data as of 31 October 1964.

** This is an estimated figure and includes cards for cross references, authority file, SI, and TKH--the latter including both Codeword and Non-Codeword. It also includes cards in three other special decks.

REPORTS TITLE INDEX

Figure 5.D-13

FILE	DOCUMENT IDENTITY		DATE FROM		DATE TO			SUBJECT SEGMENT 1		SUBJECT SEGMENT 2		SUBJECT SEGMENT 3		SUBJECT SEGMENT 4		SEC
	SERIES & NUMBER	YR	MO	YR	MO	YR										
1	2-15	16	17 18	19	20 21	22	23	24-36	37	38-50	51	52-64	65	66-78	79	80

1 DECEMBER 1961

Figure 5.D-14

JOB 3 CARD FORMAT

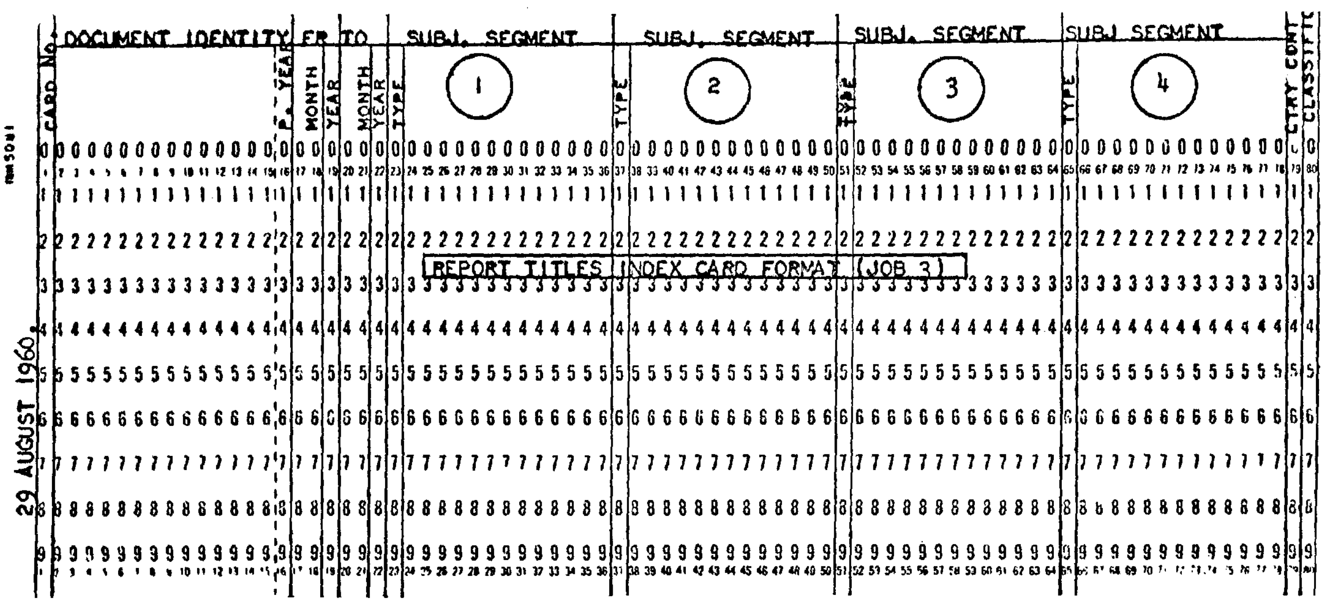


Figure 5.D-15

JOB 3 (KWIC) ELEMENTS OF INFORMATION

Col.	Field Name	Content Description	Corresponding CHIVE Field
1	File	The numbers (1-4) which identify the exploded punch card records. Numbers are suppressed in printout.	None
2-15	Document Identity, Series & No.	Clear-text transcription of the document series and the document number. Series and number are separated by a blank space.	Card 03 - Doc/ Series identification no. includes series, no. & year. The digraph subject portion of the series is also carried in field 02-02.
16	Year	Year of publication of the report taken from the documentation. The numbers 0 to 9 are combined with over punches to develop 2-digit year on printout. X overpunch for 4 No overpunch for 5 R overpunch for 6 X overpunch for 7*	
	Date of Information**		
17-18	From Month	0-12	01-07 first subfield
19	Year	as in 16 above	
20-21	To Month	0-12	01-07 second subfield
22	Year	as in 16 above	

* All year fields for Job 3 are one column and a 2-digit year is developed in the same manner for all.

** No information date on PI publication date appearing on the date line is entered in cols. 20-22.

Figure 5.D-15 (Cont'd.)

Col.	Field Name	Content Description	Corresponding CHIVE Field
	Subject segment		
23,37, 51,65	Keyword Code	3 or 0 are used to show whether or not the word following is a keyword and, therefore, to be printed in the alpha list of keywords which comprise the index. 3=Keyword. 0=Non-Keyword*	
24-36 38-50 52-64 66-78	Clear Text	The clear-text words taken from the document. Some but not all are dictionary controlled.	
79		a. SI Reports. The number 2 identifies Cuban reports. The number 3 identifies UAR reports.	01-14
		b. PI reports. One of 6 codes C,K,S,T, Z,N used to indicate the security channels in which the document is being handled. For multiple channels, highest indicator is used.	01-02 and 01-03 No one-to-one relation with CHIVE code.
80		Distribution control symbols 1-7.	01-04 No one-to-one equivalency. CHIVE code could be easily expanded to accommodate these entries.

* All Keywords are dictionary controlled. See sections 5 and 6 for sample pages from China area book and Job 3 dictionary.

Figure 5.D-16

FIB TOWN/CITY INFORMATION CARD FORMAT

Col. 1-3	1.	FIB Country Code
4-6	2.	FIB Political Subdivision Code
7-30	3.	Location Name
31-32	4.	200 Chart Series
33-35	5.	B. E. - WAC Number
36-40	6.	B. E. - Town Number
41-42	7.	Degrees (N/S)
43-44	8.	Minutes (N/S)
45-46	9.	Seconds (N/S) (South "X")
47-49	10.	Degrees (E/W)
50-51	11.	Minutes (E/W)
52-53	12.	Seconds (E/W) (West "X")
54-55	13.	Date of Latest Information (Yr)
56	14.	Source Code
57-64	15.	AMS Chart Number
65-69	16.	Location Identification Code
70	17.	Town Card Indicator ("X")
71	18.	Town Information Indicators
72-74	19.	Cat. Design. Code
75-80	20.	Town "C" Code

Figure 5.D-17

FIB INSTALLATION INFORMATION CARD FORMAT

Col. 1-3	1.	FIB Country Code
4-6	2.	FIB Political Subdivision Code
7-11	3.	Location Identification Code
12-35	4.	Installation Name
36-40	5.	B. E. Installation Number
41-42	6.	Degrees (N/S)
43-44	7.	Minutes (N/S)
45-46	8.	Seconds (N/S) (South "X")
47-49	9.	Degrees (E/W)
50-51	10.	Minutes (E/W)
52-53	11.	Seconds (E/W) (West "X")
54-55	12.	Date of Latest Information (Yr)
56	13.	Source Code
57-64	14.	FIB Identification Number (Firm #)
65-70	15.	Installation Identification Code (ICC)
71	16.	Installation Use/Assoc. Indicators
72-74	17.	Cat. Design. Code
75-80	18.	Installation "C" Code

Figure 5.D-18

FIB LOCATION CROSS REFERENCE CARD FORMAT

Col. 1-3	1.	FIB Country Code
4-6	2.	FIB Political Subdivision Code
7-11	3.	Location Identification Code
12-14	4.	"See"
15	5.	(Blank)
16-35	6.	Location Cross Reference Name
36-40	7.	(Blank)
41-42	8.	Degrees (N/S)
43-44	9.	Minutes (N/S)
45-46	10.	Seconds (N/S) (South "X")
47-49	11.	Degrees (E/W)
50-51	12.	Minutes (E/W)
52-53	13.	Seconds (E/W) (West "X")
54-69	14.	(Blank)
70	15.	Cross Reference Card Indicator("12")
71-80	16.	(Blank)

Figure 5.D-19

FIB ICF COORDINATE CARD FORMAT

Col. 1-7	1.	Sequence Number
8-28	2.	Location
29	3.	Country Code (Target "X")
30-31	4.	Country Code
32-35	5.	Political Subdivision Code
36-40	6.	(Blank)
41-42	7.	Degrees (N/S)
43	8.	(Blank)
44-45	9.	Minutes (N/S)
46	10.	"N" or "S"
47	11.	(Blank)
48-50	12.	Degrees (E/W)
51	13.	(Blank)
52-53	14.	Minutes (E/W)
54	15.	"E" or "W"
55-58	16.	"APPR" (If Approximation)
59	17.	(Blank)
60-63	18.	WAC Number
64-69	19.	(Blank)
70	20.	Control "X"
71-72	21.	(Blank)
73	22.	Control "X"
74-75	23.	(Blank)
76	24.	Town Folder Indicator
77-79	25.	(Blank)
80	26.	Card Type "1"

Figure 5.D-20

FIB ICF CITY CROSS REFERENCE CARD FORMAT

Col. 1-7	1.	Sequence Number
8-28	2.	Location
29	3.	Country Code (Target "X")
30-31	4.	Country Code
32-35	5.	Political Subdivision Code
36-38	6.	"See"
39	7.	(Blank)
40-63	8.	Location
64-79	9.	(Blank)
80	10.	Card Type "2"

Figure 5.D-21

FIB ICF NAME CARD FORMAT

Col.	1-7	1.	Sequence Number
	8-28	2.	Location
	29	3.	Country Code (Target "X")
	30-31	4.	Country Code
	32-35	5.	Political Subdivision Code
	36-63	6.	Firm Name
	64-67	7.	Plant Number
	68	8.	Status "X"
	69-75	9.	Firm Number
	76	10.	Plant Folder Indicator
	77-79	11.	Industrial Category Code
	80	12.	Alpha

Figure 5.D-22

FIB MODEL-TYPE/BROCHURE INDEX CARD FORMAT

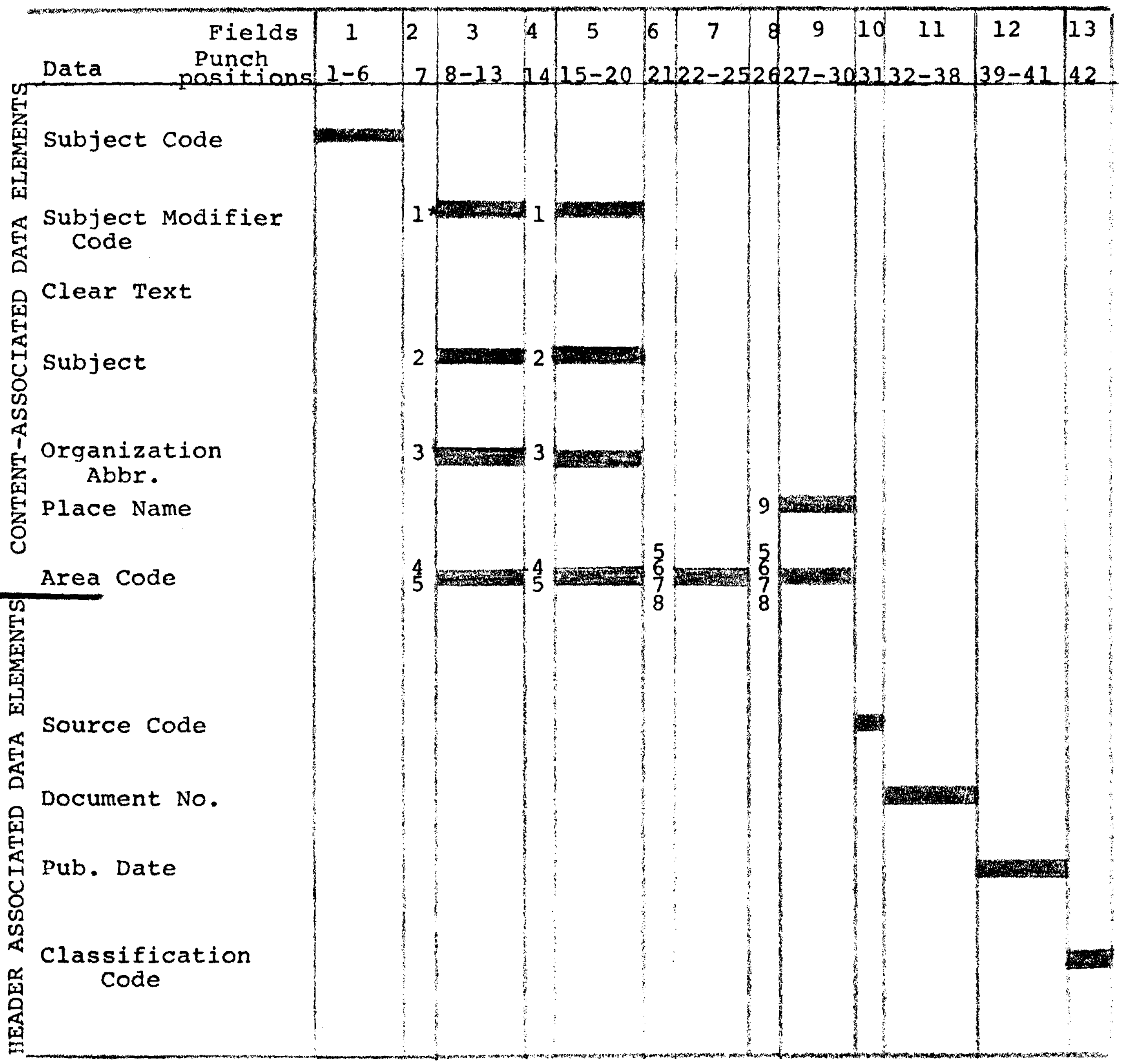
Col. 1-20	1.	Model Type/Series
21-59	2.	Descriptive Name
60	3.	Tech. Material Indicator
61	4.	Tech. Material Language*
62-63	5.	Industry) ICC
64	6.	Category Code) ICC
65-71	7.	Dossier Number (Firm #)
72-73	8.	Date (Month)
74-75	9.	Date (Year)
76-78	10.	Country Code
79-80	11.	USSR Area Code

*Admissable Entries are:

(1) English
(2) Native Language
(3) Other

Figure 5.D-23

PUNCH CARD CHARACTERISTICS of the IRS Document Index File (New)



* NOTE: Numbers indicate action codes. These are literal entries.

Figure 5.D-24

PUNCHED CARD CHARACTERISTICS OF THE IRS DOCUMENT INDEX FILE (OLD)

Data \ Fields / Punch pos.	1 1 - 6	2 7 -11	3 12	4 13-14	5 15-18	6 19-20	7 21-22	8 23-25	9 26-32
Subject Code	■								
Subject Modifier (Action Code)	■								
Area Code		■							
Classification Code			■						
Source Code				■					
Locator No.					■				
Related Area Code						■			
Related Area Code							■		
Pub. Date								■	
Control No.									■

Approved For Release 2000/05/30 : CIA-RDP78-03952A000100050001-7

Figure 5.D-25

PUNCHED CARD CHARACTERISTICS OF THE FILM INDEX FILE

Data \ Fields / Punch pos.	1 1 - 6	2 7-10	3 11-12	4 13	5 14-18	6 19-20	7 21	8 22-26	9 27
Subject Code	■								
Area Code		■							
Text (Language) Code			■						
Type Code				■					
Holding Agency Code					■				
Pub. Date (Yr)						■			
Classification Code							■		
Title No. (Control No.)								■	
Availability Code									■