

REFERENCE AID

DOCUMENT CLASSIFICATION

PAPERS PRESENTED  
AT THE  
CONFERENCE ON  
PHILOSOPHY OF DOCUMENT CLASSIFICATION IN OCR



CIA/CR-31  
January 1960

CENTRAL INTELLIGENCE AGENCY  
OFFICE OF CENTRAL REFERENCE

Approved For Release 1999/09/24 : CIA-RDP84-00951R000400070003-0

F-O-R O-F-F-I-C-I-A-L U-S-E O-N-L-Y

DOCUMENT CLASSIFICATION

PAPERS PRESENTED

AT THE

CONFERENCE ON

PHILOSOPHY OF DOCUMENT CLASSIFICATION IN OCR

21 November 1959

CIA/CR-31

OFFICE OF CENTRAL REFERENCE

CENTRAL INTELLIGENCE AGENCY

JANUARY 1960

Approved For Release 1999/09/24 : CIA-RDP84-00951R000400070003-0

F-O-R O-F-F-I-C-I-A-L U-S-E O-N-L-Y

CONTENTS

	<u>Page</u>
Foreword . . . . .	ii
Introduction to the Nature of Classification . . . . .	1
Panel I. The Intelligence Subject Code . . . . .	8
Discussion . . . . .	18
Panel II. Classification Tools . . . . .	19
Discussion . . . . .	29
Panel III. Supplements to the Main Classified File . . . . .	30
Discussion . . . . .	39
Panel IV. Contribution of Machines to the Classification Process . . . . .	44
Discussion . . . . .	56
Summary of Final General Discussion . . . . .	58
Appendix: Conference Schedule.	

Foreword

Until recent years the subject control of written information has been largely limited to the control of printed books. The present flood tide of documents and reports has resulted in a proliferation of specialized subject codes and classifications, each suitable for the control of a portion of this flow.

Intelligence documentation must cover virtually all fields of human knowledge at sufficient speed to bring the pertinent portions of millions of documents to bear on a problem demanding immediate solution. Advances towards this goal have been achieved by the liberal extension and modification of traditional information processing techniques. In many areas machines have superseded the manual searcher, bringing with them new capabilities and new limitations.

The papers that follow reflect some of the developments that have taken place within the Central Intelligence Agency to make documentary information usable. They were presented at an off-duty gathering of document analysts and reference personnel, sponsored by the Office of Central Reference. The views expressed are those of the individual writers and do not necessarily constitute OCR policy.

INTRODUCTION TO THE NATURE OF CLASSIFICATION

25X1A9a



I have been assigned the task this morning of providing some introductory remarks on the nature and use of classification.

The first problem, of course, will be to make sure that we agree on what we mean by classification. That calls for some definitions.

Secondly, it will help in our orientation to take a brief look at the history of this art and to try to identify some of the principal systems of classification which have evolved.

Finally, it will be appropriate and I hope useful to speculate on the use and value of classification in intelligence.

I should note at the outset that we are concerned here with the classification of knowledge, not with security classification which is a highly specialized application in the field.

Classification is not some inert device such as you might look at in a display case in a museum. It is a highly adaptable tool for solving problems. Every organization of individuals engaged in a common activity will inevitably require and develop a locally adapted classification for sorting and retrieving its information. How effectively the system operates cannot be judged in any abstract manner. It must be evaluated in the particular, local situation in which it has been developed and employed.

Fortunately I am talking to an audience this morning that possesses an advanced degree of experience in this field, even so it is not a little ambitious to discuss the general aspects of classification in twenty minutes when one remembers that library schools offer year-long courses on the subject. Now for some definitions.

"Classification is the grouping of various things on the basis of likeness." Classification is also described as a grouping or segregation into classes which have systematic relations usually founded on common properties or characteristics. I should insert a comment at this point concerning the sources I have been drawn on for this paper. I have relied at a number of points on the work of an Englishman, Mr. John Edwin Holmstrom. His book "Facts, Files and Action" published in 1953 is the most complete and satisfactory single discussion I have found on the general subject of documentation. Also useful was the book "Classification in Theory and Practice" by Thelma Eaton published in 1957.

Mr. Holmstrom makes the following observation at one point: A true classification is a map showing the interrelationship of ideas whereby a user can orient himself and make cross-country journeys from one idea to another in a more or less distant part of the field.

Should a person seek to devise a scheme of classification there are two conditions he must satisfy in order to make it workable: First, a distinctively labelled home must be provided (or providable) for every possible kind of item that is liable to arise; secondly, these homes must be so labelled as to make them mutually exclusive. A classification must have conciseness, orienting power, and specificity. If its terms cause a user to knock on the wrong door or to go past doors which in fact enclose what he is looking for, his scheme is inadequate in one or more of these respects.

Knowledge is gathered into classes so that causes and effects may be systematically examined. Where a cause invariably produces a certain effect we discover in this process a natural law. It has been pointed out that classification clarifies thought, advances investigation, reveals gaps in the sequence of knowledge and thus promotes discovery.

The type of knowledge classification with which we are concerned today was first applied to books. The development of mass produced books and of public education brought such tremendous increases in the rate of growth of libraries that no one individual could any longer hope to command personal knowledge of all of the books in a large library. Thus it became important both for librarians and for researchers that books be arranged on shelves so that they could be got at without endless searching.

Systematic preplanned classification of books according to a scheme workable in many libraries is a surprisingly recent development. An American, Melvil Dewey, in 1876 developed the first of the present-day widely used book classifications, usually referred to as the Dewey Decimal Classification. The Dewey concept proceeded by branching the whole of knowledge into ten main divisions. Each of these in turn was sub-divided into ten and so on to whatever number of decimal places might be necessary to specify the subject matter under analysis.

Dewey also integrated in a logical and concise manner the other components that make up the system of organization of the holdings of a modern library. We need to check our glossary at this point. The outline of a field of knowledge is called a schedule. An index in alphabetic order of the terms contained in the schedule is required to provide ease of entry for a user with a particular subject interest. The books are located on the shelves according to a notation, a scheme of numbers or letters or a combination of the two, patterned to reflect the hierarchy of knowledge in the classification schedule. Finally since books contain many subjects as a rule, yet only one of these can control the point at which the given book shall be shelved, a system of subject headings is required as a means by which all pertinent subjects in the book may be individually recorded in a subject catalog.

The Dewey system has enjoyed a remarkable success probably for reasons well stated in a comment by William Gladstone, the former British Prime Minister: "It is an immense advantage to bring the eye in aid of the mind; to seek within a limited compass all the works that are accessible in a given library on a given subject; and have the power of dealing with them at a given spot instead of hunting them through an entire collection."

The next major system to appear after Dewey was the Library of Congress classification first published in 1901. Perhaps its principal innovation was the use of many more classes and the development of an alpha-numeric notation to accommodate them. Furthermore many libraries were discovering even by that date that the Dewey structuring of classes in ten divisions was arbitrary and unsatisfactory in many subject fields.

The third modern classification scheme, the Universal Decimal Classification, appeared in 1905. Basically it was an internationally standardized extension of Dewey. It followed the same plan as Dewey and its main classes bore the same numbers. However, because it was designed to deal with the analytical indexing of miscellaneous detailed items of information, especially in scientific and technical journal literature, its categories were extended much further into detail and to date over 100,000 categories have been agreed upon against 11,000 for Dewey. There are some very familiar criticisms of the UDC. Listen to these from Holmstrom: "Despite its standardization it is not in fact the case that independent classifiers will always give the same item exactly the same class number and searchers will invariably know under which number to look. At many points the choice of numbers still leaves room for a considerable personal equation. Also, since expansions are centrally controlled the extension of class numbers to cover new developments always lags substantially behind the needs of libraries."

The systems we have talked about thus far are what we call pre-planned classifications because they seek to provide in advance for all knowledge. A rather remarkable occurrence of the past thirty years has been the appearance of what may be called self-developing classifications. These represent an attempt to avoid the difficulties experienced under pre-planned classifications in dealing with change and with the growth of knowledge. They seek a flexibility that will permit the addition of new terms and new intersections of knowledge without upsetting existing records. They avoid making a cataloger establish a correlation of knowledge and instead, make it possible for each researcher to proceed according to his own personal concept of the classification of a subject field.

In this country the best known schemes of this variety carry the label "coordinate indexing" as fought over by Calvin Mooers and Mortimer Taube. The system is applied roughly as follows:

1. A documentation staff develops a list of terms which are significant to its researchers and under which it wishes to establish a record of pertinent documents, books or other recorded information.

2. A separate record card is established for each of these terms.
3. Incoming documents are analyzed for subject content according to these terms.
4. The control number of a document is posted on the term cards which the indexer has determined are appropriate.
5. Now if one compares the document numbers posted on the record cards for terms A and B, a match of numbers will indicate that both terms are dealt with in the given document.

Knowledge comes in patterns. Here is an attempt to atomize such patterns on the theory that previous and new patterns can be reconstructed by the user of the system. Whether this is truly practicable or not is a highly controversial issue at the present time. Much thought has been given to the possibilities of symbolic representation of the terms and their manipulation according to the rules of mathematical logic.

In 1934 S. F. Ranganathan of India published the first edition of his now famous proposal for what he called colon classification. Unfortunately there just isn't time to dig into this system in detail. A major characteristic is the lack of a comprehensive hierarchical structure of knowledge. Rather, Ranganathan has sought to develop a method for analyzing knowledge. He conceives of five basic facets from which logical branchings of knowledge and indications of the intersections of knowledge should proceed. These are Time, Space, Energy, Material and Personality. The symbols for expression of facets are letters, numbers and punctuation. Relationships between facets must be expressed in a prescribed sequence, separated, or linked if you wish, by colons and various other symbols which specify role. In making a subject entry in the Ranganathan card catalog one places a card not only at the terminal point of a complete facet analysis, for example, at the end of a linkage base idea. I might add that various applications and tests of the system are underway; also, that the comment has been made that the whole concept is rather alien to American thought patterns.

The future of self-developing classifications is closely tied to machines. The original applications of coordinate indexing were embodied in simple manual card systems. Applications involving progressively sophisticated equipment have fairly mushroomed in the post-war period. We can mention edge-notched cards, the Batten or peek-a-boo perforated card devices, applications of IBM punched cards and the use of computers, for example at the General Electric gas turbine plant in Cincinnati.

It is time now, however, to break off from this line of discussion and to consider the use of classification in CIA. There will be time for only the most general of propositions.

I think you have got to approach the problems of information handling initially from the viewpoint of the researcher. This is interesting business. The typical analyst brings with him a university background of training in



scholarly research and a rather elaborate structure of formal knowledge of his chosen subject field. In CIA he finds a kind of newspaper world with its field collectors or reporters of information and its copywriters, editorial writers and editors at headquarters who estimate the news up to each edition deadline, follow developing stories through edition after edition and employ "successive approximation" as it has been called by Max Millikan, as the method of getting at the "truth." Today's conclusions may confirm, contradict or modify those of yesterday. Much depends on the brain power of the team, but good sources, good methods and just plain luck will also bear on the quality of the performance.

Now what does the analyst do on the job? He sifts incoming information from a variety of sources. What he can't keep in his head he records pretty much as he pleases in a first-stage external memory, his working file. This is a classification system. It may be formal or highly subjective in its structuring of knowledge. It represents a sort of capillary system for handling new knowledge and new language. It also constitutes the basic platform on which the analyst proceeds with his problem solving in his field of specialization.

In large systems, a second-stage externalized memory such as the Office of Central Reference is also required. This facility must serve the needs of all analysts in its audience. Unavoidably it must compromise the desires of individuals. It must perform what the analysts have neither the time nor the acquired skill to do. Each category of knowledge requires a discipline - rules for consistent processing and manipulation. Thus specialization of information storage occurs. And let me say that I think OCR does these jobs very well indeed, as well as they are done anywhere. We have developed "know-how" in defining, controlling and retrieving by category or within category, our data on names, area, photography, industrial plants, trade fairs or information on Communist Party activities. We need apologize to no one when we also acknowledge that we hope to improve in the future our methods for making these systems better serve our customers.

The intelligence process in which both the analyst and we play a part is certain to be deeply affected by automation and in the relatively near future. There is much evidence already at hand.

Print-reading devices will transfer information from document to machine.

A method is already proved for bringing field reports to headquarters on tapes ready for machine input.

A first-cut identification of information in the document will be made by auto-abstracting and language manipulation techniques. As you may know, ACSI hopes to begin a program of this sort in 1960.

Dissemination to the analyst's office by Western Union type ticker tape will be inaugurated based on computer matching of document contents with his requirements.

Analysts and field collection staff will communicate by voice and television and adjust requirements to "feedback" immediately as agreed upon. This communication will include instantaneous transmission of text, photographs, maps, et cetera.

The analyst will store at least some of his information under categories of his own choosing in the central facility. He will record his evaluation of each document with us so that we may correlate it with others for the benefit of future users.

The analyst will contribute to the construction of OCR's indexing dictionary, thesaurus and hierarchic classification scheme.

We will apply automatic indexing techniques in particular to our directory type programs.

Our indexing staffs will be primarily concerned with the subject control of complex subjects, with abstracting and consolidation of index data, with purging, and with the coverage of many categories of information we cannot now afford to process.

Our customers will learn to utilize our facility on a much more spontaneous basis, as they now use their personal files, because our system will respond promptly. It will return their own contributions and those of other experts including their evaluations of any report in the system.

I offer you these opinions in closing:

Our system does not mesh as well as it ought to with the information handling patterns of the intelligence researcher. In the future the researcher must be able to query our facilities as simply, quickly and directly as he does his own files, or his telephone directory, dictionary or encyclopedia.

We must be keenly aware, day in and day out of the researcher's interests, his language, the file system he uses and his opinions as to the effectiveness of our operations.

We cannot provide the value judgments on what we process. We may assist the analyst to discover significant relationships between facts but ours will never be the final authoritative judgment.

I am much puzzled by the problem of handling what I will call saturation reporting. I have seen 20 or more reports on the Paris information conference of last summer. These reports have been highly repetitive in content. It may be a reductio ad absurdum to try to subject-code each facet of their contents for fine-mesh retrievability.

Washington Platt has estimated that strategic intelligence information depreciates at the rate of 20 per cent per year and that facts

such as those concerning a port or manufacturing plant depreciate at the rate of 10 per cent per year. As our body of knowledge grows, we will experience increasing difficulty in selecting out, that is purging, this dead information.

We will have to be thick skinned towards criticism. A retrieval of unsatisfactory information or an answer of "no information" may find the researcher condemning us even though the system itself operated efficiently. Failure in a search may mean that no information has been received in our system. On the other hand we must not evade criticism when it can be shown that we have in effect hidden information from its potential user.

In the automated reference center of the future, the analyst will be able to a degree not now possible to orient himself on the map of interrelationships of ideas and to make cross country journeys from one idea to another anywhere in the field of knowledge.

Panel I

THE INTELLIGENCE SUBJECT CODE

25X1A9a

Panel Members:

5X1A9a  
(spokesman)

In carrying out their joint obligation for making intelligence documents available to consumers, the Document Division and the CIA Library have developed a system of classification which is the basis for the storage and retrieval of information through the Intellofax System. The experience of these two divisions reflects a continuous encounter with problems of input and retrieval. Actions taken in response to immediate needs are integrated into the system. Collectively, these may seem to reveal more hopeful pragmatism than systematic and logical progress. Yet each decision has been taken with the same object in view -- that of providing the user with the material he needs and excluding that which is irrelevant. While this paper is intended to examine the system from the viewpoint of its aims rather than methods, it is still necessary to rely upon a measure of description to explain how the classification scheme has taken its present form. The various factors which combine to form the problem of classification refinement will be discussed in general and in particular. These are: the classification system, the documents, and the requests for information.

While it is true that plain words are best, all technical operations develop a vocabulary which permits the substitution of word or phrase for a lengthy description. Certain terms used frequently in describing the processing of documents and the system itself should be defined here before they appear in context:

Intellofax System. A mechanically supported system in effect in OCR since 1948 consisting of an index file of IBM cards coded by subject and area; taped lists of bibliographic citations to documents; and microfilm aperture cards which are the Library's file copies of documents.

Index. The verbs "to index", "to classify", and "to code" are used interchangeably.

Nodex. The term used to indicate material which is not indexed.

Notation. The professional term for what we commonly call the "code." It may be alphabetic, numeric, or a combination. In our particular classification system, numbers are used to represent subjects.

Processing. Here refers to the entire activity concerned with receiving, disseminating, indexing, and microfilming documents; and punching IBM cards.

ISC. Abbreviation for the Intelligence Subject Code, the classification system used in the Document Division and the Library since 1948.

#### I. The Classification System

The need for a classification system was evident from the outset of operations if the incoming documents were to be handled in keeping with principles of systematic process and uniform control. At its inception in 1948 the ISC consisted of 8 chapters each representing one major subject category. The entire code contained 980 notations. Within each series, decimal breakdowns permitted refinement to the extent of six digits. This project originated in an attempt to consolidate various systems of classification into one comprehensive code. There was no particular requirement that the scheme devised at the time should form a pattern for other agencies forming the intelligence community. As a matter of fact, there was no meeting of minds at that time for the need for a common system. The expanded code of today with its 15,000 notations and the revised new issue to be published in 1960 represent the response by the Document Division to the need imposed by the increased flow of documents, the wider range of Intellofax patrons, and its more recently proposed USIB adoption of the ISC as a common classification scheme.

One example and comparison may illustrate how a single subject has been treated in response to intelligence needs. In the original ISC, Communism and the Communist party were covered by eight notations. There are now 190. For this same subject, the Dewey Decimal Classification uses only 12; the Library of Congress lists 13 headings. This one example shows clearly how detailed a special purpose index can become as compared with a universal system. Likewise, the particular needs of the Agency and the diversity of material received from other Government agencies have determined the direction of our efforts to provide the service required by consumers.

The chronology of the many adaptations within the present ISC shows that there have been two parallel developments. First, there has been a continual process of additions to the code. Some were in the form of major revisions of entire chapters, but most are single new notations, generally inserted as a result of internal decisions to take care of day-to-day needs not already provided for in the ISC. The major changes in the past 10 years have been new issues of 2 chapters, one for Air Force, the other on Scientific Research and Development. Both were reconstructed according to the wishes of the

office most concerned. The Air Force chapter was designed to fit Air Intelligence needs when the Air Force adopted the ISC for its own Minicard use in 1956. In consequence, the classification refinements of that section are more suited to Air Force needs than to those of this Agency.

Other Agencies and Offices have expressed active interest in revisions of the ISC with results similar to the example of the Air Force. Army submitted a revised code in 1956 which was so refined in subjects of purely military interest such as order of battle, logistics, and army organization that it was unsuited for general use.

In the case of the chapter on Scientific Research and Development, the demands for expansion have been many and varied. One of the most articulate was in 1949 from a plant biologist who succeeded in introducing 88 codes for diseases of plants. Although not one of these codes has been used in retrieval during the past two years, the section remains as a harmless curiosity of over-classification of no particular use to intelligence needs.

At the same time, while the classification scheme was growing more detailed, the number of documents received for processing increased rapidly. Therefore, the second major development and a logical consequence was the decision to exclude altogether certain types of documents from the coding system. This process of not indexing has been termed NODEX, a necessary limitation intended to allow more time for coding those documents selected for thorough processing because of their intelligence value.

In the present ISC, within a moderate size of 15,000 notations, a place is provided for very wide ranges of human knowledge and activity. This development, however, has not been altogether systematic and a glance at its varying degrees of fineness suggests that a number of influences have been brought to bear on its contents. The pressure from outside has generally been for detail expansion, but usually in a haphazard way. The result has been equally unbalanced, as certain sections have been accorded very fine breakdowns and others have remained relatively unchanged. Demands for over-specialization create a problem which invariably faces those who are trying to construct a classification scheme which fits the needs of input and retrieval. The fineness of the classification structure must reflect a compromise between the document analyst's need to apply the scheme to all types of intelligence documents and the librarian's and researcher's need to retrieve for very specific needs. Because the documents vary greatly in their degree of generality or specificity and because the Intellofax System serves a variety of customers whose needs often contradict each other, indexing standards must aim at being at once both uniform and flexible.

The subject specialist's approach to code structure and indexing is sometimes termed "over-sophistication". It commonly reflects a close familiarity with one subject and only certain aspects of this subject, and quite naturally assumes a particular importance to the individual concerned. A better term might well be "over-simplification" for it can lead to a degree of specificity which requires indexing every substance and function by name. If carried to its logical conclusion, this fineness of classification would require a code of dictionary size while at the same time narrowing the base of application. Coding by words as opposed to ideas may serve fairly well in dealing with commodities and installations. But when the element of judgment is removed altogether from coding, the end product becomes a mass of unmanageable size and much unrelated data.

This conflict over specificity is carried over into an unresolved disagreement about who may best be assigned to the task of coding documents. There are three possible candidates: 1. The subject specialist with professional status in indexing. This type is hard to find, particularly so at moderate salary levels. 2. The subject specialist with no indexing experience and no great desire to index. 3. The trained indexer with a general educational background. The only experiment of any size which was intended to form some conclusions on this problem has been conducted in Great Britain. Its value is exceedingly limited by the selection of one type of technical document as the entire body of test material. Here in OCR in 1957, a team of library consultants surveyed operations and made many recommendations. Task Team #1 assigned to answering the consultants' observations on the Intellofax System stated that "it did not recommend the hiring and maintaining of true subject specialists (such as an organic chemist, an inorganic chemist or biochemist) but rather the division of the coding universe into large subject groups, and specialization only within these groups. Even though the coder would be a generalist compared with the subject specialist in the consumer office, rough specialization would result in many factors capable of improving the coding."

## II. Nature of Documents

Other considerations which bear upon classification stem from the nature of the documents and the requests which are expected to be placed upon retrieval.

The volume of documents received in the Document Division ranges between 700 and 1,000 a day and the variety is extremely diverse. There are many short factual attaché or foreign service information reports which are 1 or 2 pages in length and can be covered with 1 or 2 codes with little depth in analysis required. There are the longer raw information reports, such as many OO-B reports, which cover technical research, and therefore require more specialized subject knowledge. In these cases, the language of the document often does

not match the language of the classification scheme, and the document analyst, who is a generalist and not a subject specialist, may have difficulties with too fine a classification. Here it is better to use a broader approach which then places the burden upon the technical researcher to read many related documents in order to pull out the specificity he needs.

Many raw information reports cover political and economic subjects - some factual, others abstract and theoretical. It is the latter type of report which is the most challenging and which requires effective analysis. This is the information which is also the most difficult to retrieve for it allows the minimum reliance on mechanical aids. The consumer may not be clear in his request and the classification cannot always provide the proper clues.

Finished intelligence reports prepared by evaluating components of the intelligence community are normally longer than raw information reports. The emphasis is different and because much of the factual data has been culled from raw information reports, the depth of coding is not the same. Although it may be necessary to assign many codes, the fineness of classification is not so vital since broader aspects are used.

### III. Nature of Requests

In the construction of the chapter on economic subjects and commodities provision is made for discrete coding of materials and fabricated objects. It was assumed that analysts would ask for all material on one or more items, for example - copper. It was soon evident that even this one subject divided by further decimals into stages of processing from ore to finished product would be too general for the customer who asked only for production statistics. With the addition of supplemental modifying prefix codes, it is now possible to select any or all 35 functions of the copper industry (or any selected industry). This is the finest classification we have attained in our present Intellofax System with the use of prefix codes. However, it is also the least analytical. It requires no, particular skill and certainly no interpretation. Likewise, the newer mechanical or electronic devices which will retrieve this type of document are better equipment only in the sense that they can produce references more rapidly.

Experience has shown that searches in response to requests for information of this sort are frequently the least successful in terms of customer satisfaction. The difficulty stems from the fact that the analyst is asking for statistical data and he receives a list of documents which contain some coded reference to that subject. Available equipment does not permit the accumulation of information to be issued in the form of direct answers to queries, nor would this necessarily be a desirable substitute for the source documents.



Too often in intelligence reporting the circumstances of origin are essential to accurate and logical evaluation of information. With these supporting facts removed, the reliability of pure information cannot be determined, while it is perfectly possible that different documents would give different figures in answer to a given question.

A second category of specific subjects falls within the general description of target information, that is data about installations and cities. Information of this sort is more often requested from users outside the Agency who are not well acquainted with the facilities or services of OCR. What may appear to be a very simple request such as "major industries" in a selected area, while it could be retrieved would be highly impractical and uneconomical to translate into an ISC search. Here the detailed coding which the ISC provides becomes an obstacle to efficient retrieval, and the Industrial Register becomes the proper source for such information. Target information is further hindered by the limitations of area coding on the present Intellofax card. Locations are not coded beyond the level of countries with the exception of Russian oblasts and Chinese provinces. There are occasions in searches for installations when an IBM punch for city coding or clear text of city names would be valuable. However, the Intellofax index does not stress finer classification of this type of information because the Registers have been established to service such requests.

The fineness of a classification scheme may also be dependent upon the use of clear text in conjunction with codes. The Intellofax System uses no clear text punch columns today because of limitations of the IBM card as it is designed for our present use. Some specialized files have used clear text for many years and with great success, and any new system devised for Intellofax will without doubt consider the possibility of adopting clear text. Panel II will discuss in greater detail clear text as a necessary classification tool. It should be mentioned here, however, that many of our present problems with the classification structure could have been solved and the requester more adequately served if we had been able to use clear text.

So far we have examined the coding of concrete subjects, which have proved to be the least demanding in terms of coding experience or substantive knowledge. Here fine classification has proven to be useful for the most part only in reducing volume in retrieval.

When we turn to the use of subjective codes, the handling of reports which discuss political and economic conditions and affairs, the criteria for both codes and coders must be altered to accomplish efficient retrieval. Here the classification tool and the ability of the indexer must join to produce a useful product capable of supplying the information needed for consumers. The chapter of the ISC which covers world politics has also been expanded greatly from its original 152 notations, but its growth has been more systematic

and controlled. Here the problem has been the difficulty in maintaining a balance between the language of documents and the ideas which they reflect. Moreover, as the need for interpretation increases, uniformity in coding becomes more and more difficult to maintain. Supervision and review are the available means of control, but much individual analysis of documents remains unchecked as the large volume of documents flow through processing each day. Yet the problems of ensuring uniform coding of ideas are fewer than those which arise when each and every word is indexed, and the concepts or ideas are removed. Without a lengthy excursion into the subject of semantics, it is still possible to observe some of the problems which are implicit in reconciling words and meanings. In theory, the meanings which may become attached to words are not subject to fixed limits, yet the words when printed become a measurable piece of information. If the material is to be of any intellectual value, the general context in which a sentence is placed must be taken into account. The alternative is a very limited form of classification, which reduces effectiveness of the system to the minimum. With a document before him a coder can achieve some measure of understanding of the total meaning created by the words in print, consisting of their literal values plus their significance as presently used. This in turn becomes the basis for the interpretive coding which the ISC classification for political subjects allows.

The ISC's principle of mutually exclusive subdivisions within larger categories serves equally well in subjective classifications but it does limit the fineness of classification which can be employed. From the viewpoint of retrieval, this is not the handicap that it may appear to be. Requests for political information are commonly made in search of material bearing on a research topic for selected countries such as "Present situation and outlook for Austria," or "Relative importance of military, labor, Church, and intellectual forces" in Spain. In efforts to serve such problems, the need is for a collection of documents reflecting these various subjects not only with a set of isolated facts but complete with the observations of the report writers whose work is based on local experience. At best it is difficult for a research analyst to reconstruct the variables of any recorded event or situation. He must be especially careful not to assign meaning to what he reads which may stem from his own bias or imagination. To assist him in making estimative judgments, in weighing one political faction against another, meanings rather than isolated words provide the clues. So in classification, broad categories serve the need more effectively than a close word-indexing system which may provide good statistics but no ideas.

Returning to our original problem, it is clear that classification is the means employed for the systematic storage and retrieval of information contained in intelligence documents. The operational system by which this information is brought to the requester, and

the form of the end-product, while they may enhance the speed and scientific appearance, do not contribute to the intellectual quality of the research reports to be written. The emphasis must remain therefore upon the documents themselves and the consumers who will use them. The diverse contents of the first and the varied circumstances of the second create the essential problem of classification. If the two are to be brought together with any degree of success, some variety of treatment is needed. For material objects, the finer classification which records statistics such as: how much? and by whom? is most likely to provide the answers. There is little room for marginal material, none for the irrelevant in subjects of this nature, and we should be able to provide only such documents as will contain the desired information. Whatever is more or less must be regarded as a poor product. For abstract and interpretive subjects a classification scheme must be specific enough to permit fairly rapid indexing within a uniform pattern which will allow for discrete retrieval searches. Yet it must also be capable of reflecting concepts and ideas which may not be found in the direct language of the documents.

#### IV. Theoretical Problems of Fine Classification

Specificity of classification is sought in order to pin-point a species inside a class, e.g., to distinguish men from all other animals inside the class vertebrates. By means of such specificity the specialist can go to his field of interest immediately and without regard for all other coordinate fields. However, the end-result of such specificity is, oddly enough, the creation of a new broad category, particularly if the specific subject is elaborated from its original status as a differentiated species into a pseudo-class. This can be seen by a consideration of the guided missile.

The missile most probably made its earliest appearance as a piece of rock which fitted easily into a man's hand and was just the right weight to be thrown a short distance and to bash a man's skull. Modern technology has changed the piece of rock into an intercontinental ballistic missile with an atomic warhead, and the skull into a city or a nation, but the principle is still the same: the guided missile, like the rock, is a weapon. However, inside the large including class of military weapons it seemed logical to give a special notation for the species, guided missiles, to distinguish them from boomerangs, for example.

Soon there was the anti-missile missile, and then the anti-anti-missile. Finally there will be a missile which is itself an anti-missile missile or which will carry its own anti-missile missiles. Since it no longer seems possible to consider a missile without regard for the possibility of an anti-missile missile specifically designed to frustrate it, the differentiated species of guided missiles should necessarily include the anti-missile; the anti-missile should necessarily include the anti-anti-missile. However, if the

specificity has been carried to the point that, from the beginning, the anti-missile and the anti-anti-missile are equal to the missile in the classification as co-ordinate species, even as the ape and the chimpanzee are equal to man in the large class of vertebrates, there is another problem in classification. Double or triple coding must be used to compensate for the excessive specificity by creating a new large class or broad category which might be called guided missile warfare and weapons. It is not beyond probability that, since so many other military devices, each given as a differentiated species, e.g., radar, are involved in guided missile warfare, eventually the guided missile warfare and weapons class would be as large as or equivalent to the original broad category of military weapons, particularly in a document collection specializing in current military technology.

Furthermore, consideration of the anti-missile missile without regard for the missile could lead a researcher to think of it as something which had sprung full-grown into being. An apt analogy can be found in space medicine. It did not spring into being as a fully defined and specific field of knowledge. It has its gestation in aviation medicine, which grew itself from army medicine, which was an outgrowth of general medicine. It was subsumed in the class for aviation medicine until suddenly it had such a sturdy self-existence that it seemed to require a class for itself. However, what about the earlier material classified as aviation medicine?

The easiest way to handle the problem is to double-code space medicine, that is to assign to it both its own code and the code for aviation medicine, treating it as if space medicine were both a new class and an old class. In this way it is possible in the ISC to relate the present state of knowledge to its antecedents, to show the hierarchical relationship, not by means of classification but by means of an additional subject heading. Even though classification has been ignored, there is built into this method the seed of a new specific class, i.e., that portion of A, e.g., aviation medicine, which is also part of B, e.g., space medicine or space aviation medicine. The alternative to this third new specific is to reclassify as much of the class for aviation as is exclusively space medicine. In this way A is A, B is B, and the two can be made coordinate subdivisions of AB, flight medicine. This is an expensive business, but it is good classification.

As the field of knowledge increases in size, it is necessary to readjust the classification, i.e., the arbitrary, formal, conventional scheme of organization. The possible ways of adjustment are four. First: Each new field can be made into a discrete class. Second: As the specificity of the discrete classes explodes like the population of the earth, larger and broader categories can be constructed to pull the specific small classes into another large class. Third: The whole classification scheme can be altered on a continuous basis so that the specific small classes will be simultaneously specific and related in a hierarchical order. Fourth: Double,

triple, and quadruple coding can be given to the same subject in a vain attempt to classify into two or more classes simultaneously. One can feel more and more sympathy with the oriental belief that the "All is one," and "One is all," or even with the "Bellum omnium contra omnes."

The foregoing observations reflect the exceptional complexity of general intelligence document collections as opposed to the relatively definitive problems of more specialized and limited collections. This distinction accounts in great measure for the lack of literature applicable to intelligence document problems. Opinions regarding new approaches to classification vary from those which may be termed unreasonably doubtful to the other extreme of the unwisely confident. Both are subject to some adjustment when confronted by the realities of daily practice which must take its form from the materials available and the demands which are made upon them.

Mr. Jesse Shera, Dean of the Library School at Western Reserve University, has offered one brief analysis which can be applied to our situation: "The pattern of classification appropriate to a given library situation is conditioned by (a) the volume; (b) the characteristics; (c) the pattern of thought of the field; (d) the pattern of thought of the individual user." We have discussed all the points Mr. Shera has made. We have indicated the large volume of approximately 1,000 intelligence reports a day received for processing. The characteristics of these reports are found to be as varied as the types of people who are hired to do the processing. The pattern of thought of the writer of the report may not be the same as the pattern of thought of the indexer or document analyst, nor as a matter of fact, as the classification system itself. And last, but certainly not least, there is the thought pattern of the individual consumer whom we are trying to serve. Certainly all these factors, subject to such varied degrees of control, have played and will continue to play an important role in determining the nature of classification.

There will undoubtedly be some powerful and complex machinery devised for future use in document and information retrieval. Its success will still rest in great measure upon the skill acquired by those who perform the indexing phase of the system. These people when thoroughly trained are specialists in their own right, responsible for constant and discreet judgment upon the documents. Neither broad coding nor fine coding has any intrinsic value unless accompanied by sound reason and systematic control for both input and retrieval. We have found that there are inherent deficiencies in any system or classification, and there are ambiguities in requests which often resist a ready solution. It is the obligation of those assigned to retrieval to make up for these deficiencies.

Panel I

DISCUSSION

QUESTION: What have been the positive responses to requests for the changes to the Intelligence Subject Code (ISC)?

25X1A9a [REDACTED] The revised ISC incorporates many suggested changes. The proposed Army ANSI code has been accepted in part; however the detailed specificity of the Order of Battle and logistics codes would have been too many to include in toto. Likewise Air Force suggestions are included.

QUESTION: Who is using the present ISC?

25X1A [REDACTED] The present ISC is used by Air Force Intelligence in its Minicard coding project and by two AF commands; Strategic Air Command (SAC) and Shepherd Air Force Base. The code is used in part by the Army Signal Corps Intelligence Agency. In addition to the above the ISC is used within the military organization of SHAPE and is known as the SHAPE Intelligence Code (SISC). Five NATO countries are using it today as a national intelligence subject code.

QUESTION: In addition to the obsolescence of the system what about obsolescence of the information itself.

25X1A [REDACTED] The system reflects the present stage of knowledge but the older material also remains in the collection. The search for retired material is too dependent upon the memories of people.

[REDACTED] It has been ascertained that about 22 per cent of our retrieval is for retired material, that is, material more than five years old. The Minicard Coding Group, which is working with a discrete corpus of documents, was asked to assess the half life of the documents handled. The preliminary conclusion is that we cannot estimate the half life of a document because of its historical value and the nature of research.

QUESTION: How is the document analyst kept current with the needs of researchers?

25X1A [REDACTED] Participation on a monthly rotation basis of senior analysts on the Composite Group (Library/DD Intellofax Processing Team) alerts them to needs of the researcher which they relay to junior analysts. In addition, rotation provides a means of keeping the permanent Library member apprised of current coding practices.

[REDACTED] It may seem extravagant to have two persons doing the work instead of one as in the past, but we have found the cost of two on the Composite Group in interrogating the requesters has been more than offset by the savings in processing time and the end product.

25X1A9a

CLASSIFICATION TOOLS

Panel Members

(spokesman)



A. Introduction

1. The views and problems presented in this paper represent the experience and experimentation of the Intellofax System, the Special Register, and the Minicard Coding Group.
2. Intelligence document storage and retrieval present complex problems. Intelligence documents vary from highly polished, well organized, single topic dissertations to disorganized, multi-topic fragments. The questions asked of an intelligence index vary from the generic NIS type of request to specific problems, the nature of which no indexing system so far envisioned could possibly anticipate with index categories. In addition, a general indexing service such as Intellofax or Special Register must be able to cope with all fields of knowledge using document analysts with limited subject specialization. As recently stated by a Library of Congress consultant, we are facing problems never before experienced or anticipated in the field of documentation. Other documentation services face these problems in part, but the totality of factors mentioned above is unique to the intelligence community. For this reason it is difficult to get competent advice from experts in the field of indexing. Most of the experts' experience is limited to book cataloging or specialized, technical document collections. It sometimes seems there is entirely too much furor over which indexing system should be applied to the average, small, specialized document collection held by various U.S. industrial concerns. It would seem that subject heading indexing, the simplest form of all, would suffice except for highly complex subjects such as chemistry.
3. There are two problems critical to any storage and retrieval system which are particularly applicable to an intelligence system. They are the need for uniformity of input and specificity of retrieval. The indexing tools or techniques discussed in this paper are attempts to resolve these problems.

B. Intelligence Subject Code

1. There are in present use three main systems of indexing. They are:
  - a. Subject Headings - Subject headings are a simple alphabetical

arrangement of recognized words or compound terms which are familiar to the users of the indexing system for which they are designed. Complicated subject heading schemes tend to take on many of the characteristics of a classification system. The most notable example of an index using subject headings is the Reader's Guide to Periodical literature.

- b. Coordinate Indexes - Coordinate indexing carries the subject heading system a step further by allowing, in the retrieval process, the coordination of ideas or index terms which refer to the same document. Unlike subject headings, coordinate indexing is more effective if used with some sort of mechanical equipment. Uniterms, key words, and descriptors are all used in coordinate indexing systems.
  - c. Classified Indexes - Classification systems attempt to classify knowledge into broad groupings and sub-groupings. The botanical classification of plant life, Dewey Decimal Classification, and Library of Congress Classification are examples of classification systems as well as the systems used in OCR, namely the Intelligence Subject Code and the schemes used in the specialized registers.
2. Subject headings are in general not applicable to intelligence documents. A very specific subject heading list tends to get complicated and difficult to use, and generic searching is extremely laborious. In addition, the use of subject headings does not provide for the coordination of ideas which is extremely necessary to specific retrieval in an intelligence organization.
  3. Coordinate indexing, which seems to be gaining the most popularity, also has serious drawbacks when applied to intelligence documentation. Coordinate indexing has been applied almost exclusively to limited fields of knowledge, particularly scientific knowledge. The language of these limited fields is usually fairly stable and concrete. When new terms do arise they generally have an entirely new meaning and do not conflict with previously accepted terms. However, when coordinate indexing is applied to broad fields of knowledge it encounters many semantic difficulties. A word does not have the same meaning in one field of knowledge that it has in another, e.g., stability has a different meaning for the chemist, the physicist, the aeronautical engineer, and even the political scientist. The problem of synonyms is obvious and very difficult to overcome. In addition, as in the case of subject headings, generic searching is laborious or impossible without complicated techniques. Coordinate indexing seems to work very well in limited subject fields, particularly well disciplined scientific fields, but it presents some seemingly insurmountable problems when applied to a large collection covering all fields of knowledge, especially fields such as politics and sociology, which include many abstract concepts.
  4. It has become very popular in the documentation field to criticize



classified indexes. They are said to be structurally complicated, difficult to use, too rigid for easy incorporation of new subjects, and not specific enough or too specific. It is also maintained that they do not change fast enough and quickly become outdated. Many of these criticisms are true when one of the standard classification systems is applied to a document collection. The basic principles of classification systems originally designed for books have been used in the development of systems applicable to document indexing. Any classified scheme has the one great advantage, when properly indexed and cross referenced, of gathering all the subjects in a particular field together in a minimum number of places. It greatly facilitates generic searching and it alerts the indexer to subjects of index interest. When correctly constructed and indexed it need not be overly complicated and difficult to use. When designed to handle a particular documentation problem it need not suffer the criticism applicable to the general classification schemes designed for books. It would appear that the classified index and its auxiliary tools are more applicable to the intelligence document problem than the other index choices.

5. The Intelligence Subject Code, which has been in use in the Intellofax System since 1948, is currently under revision for publication in early 1960. The ISC has been criticized for having the following weaknesses:
  - a. There is no guide on how to apply the ISC, and its structure is difficult to understand without knowledge of the interpretations placed on the various sections by CIA.
  - b. The repetition of the same commodities in several different sections is confusing and unnecessary in light of developments such as the subject modifiers.
  - c. The ISC is unbalanced in subject coverage. Important subjects such as space travel and artificial satellites have limited coverage, whereas an extensive section is allocated to plant diseases on which there is little reporting.
  - d. Its index is unreliable and outdated.
  - e. It does not have enough cross references and explanation of individual code meanings.
6. The revision attempts to overcome these weaknesses by:
  - a. Providing an introduction explaining the ISC's content and how it should be applied.
  - b. Placing commodities, including military weapons and equipment, in one chapter and assigning appropriate subject modifiers

(action codes) to distinguish the various actions affecting commodities. Also, the three former separate chapters for the armed forces are combined into one chapter with appropriate subject modifiers.

- c. Updating the subject content and deleting unnecessary subjects.
  - d. Providing a complete index prepared on IBM cards which can be kept current.
  - e. Providing complete cross references, liberal scope notes, and other annotations to aid the analyst and the reference librarian.
7. The revision is no panacea, but it will overcome many of the objections to the present code. In most respects the revision does not go into great subject depth, but greater depth can be added as needed. With the addition of clear text, classification depth is not as critical a problem as it is with the present Intellofax System. The revised ISC should ensure much greater uniformity of input and with the addition of other techniques discussed below there should be much greater specificity of retrieval.

C. Subject Modifiers (Action Codes, End Use Codes)

It was found in early retrieval experience in both the Intellofax System and Special Register that requesters were not interested in all aspects of some subjects, e.g., commodities, but that they wanted certain modifications or actions only, e.g., production, export, etc. It was also found that in the commodity field, for instance, these same actions were requested repeatedly. One solution to this problem would have been to add these modifications as subject subdivisions to all the subjects to which they applied. This was impracticable because there were many modifications and they applied to many subjects. Their addition as subject subdivisions would have increased the size of the code book tenfold. These modifiers finally evolved as two or three digit action codes which can be combined with various subjects as appropriate. The subject modifier or action code as applied in OCR is a new development but the idea itself is old. It is very similar to the Universal Decimal Classification system of auxiliary tables and bears some resemblance to faceted classification. Its use greatly facilitates specificity of input and retrieval.

D. Area Codes

1. Intelligence analysts usually have an area responsibility in addition to their subject responsibility. An overwhelming number of machine run requests are for information on a specific country only and in some cases on a subordinate area within a country. Area codes are

not difficult to construct aside from the problems of digital limitations and whether the code should consist of numbers or letters. The construction of the code should in general conform to the area interests of the users, e.g., Middle East, South East Asia, and should also be able to show limited geopolitical concepts, e.g., Communist versus Non-Communist. In some systems specificity is important enough to require an area code based on longitude and latitude.

2. A primary consideration in area coding is one of depth. It is obvious that the interest in Russia and China is so great that these areas should be broken down to at least the oblast and province levels. The need for fine area coding in other parts of the world is not so obvious. There are occasional requests for areas such as Lower Saxony in Germany, but it is questionable whether the additional coding time involved in fine area coding could be justified in view of the few requests received. Cities also present a problem. It does not seem feasible to construct an area code for cities, even one limited to Soviet and Chinese cities, because there is no particular criteria for choosing those cities considered important. A Soviet settlement consisting of 100 people becomes very important if a guided missile site is discovered nearby.
3. An area code consisting of all the countries and other major areas of the world, e.g., international waters, and the major political subdivisions of Russia and China would seem to be adequate. In addition, for the Intellofax System it seems necessary to be able to code in clear text at least Soviet and Chinese cities and other bloc or non-bloc cities deemed important.
4. One further consideration regarding area coding is file arrangement. A completely reversible subject-area file is the most useful, i.e., one file arranged by subject with area subdivision and a duplicate file arranged by area with subject subdivisions. The area file should include related (secondary) areas as well as main (primary) areas. The above arrangement is desirable because some searches are more easily accomplished through entry into the area file and conversely some searches are feasible only through the subject file. Very broad subject searches, e.g., everything on science, for a particular country are almost impossible without an area file approach.

E. Direction, Nationality and Reaction Codes

1. It is often necessary to show area relationships in order to ensure specific retrieval. The indexing of information such as export-import data is of little value unless both the origin and destination of the shipment are shown. Area relationships are expressed in the Intellofax System by the use of a two digit code called a

related area which can be selected by the IBM machines in conjunction with the main area. The related area code is fairly satisfactory, but it has taken on a variety of meanings. Usually it is used to show the direction of an area relationship, but it is also used to express nationality, e.g., French troops in Morocco, and comments and reactions, e.g., Soviet reactions to a U.S. nuclear bomb test. The multiple use of the related area has led to the need for strict rules for its application, a variety of memos to handle specific coding situations, and some retrieval confusion.

2. The Minicard Coding Group has adopted a very simple device for overcoming this problem similar to that used in SR. This was done by entering a 1, 2, N, or R, in the fourth position as an extension of the three digit area code. "1" equals the concepts of sending, from, whence, or source country. "2" equals the concepts of receiving, whither, target, or destination. "N" and "R" stand for concepts of nationality and reactions. This is a valuable coding technique which should be included in any future indexing system.

#### F. Clear Text Coding

Clear text coding is the entering of words, abbreviations, and numbers into a machine system to give more specific meaning to subject, area, and modifier codes. It is an auxiliary device which allows for any degree of coding depth desired. It has been used successfully by the MCG and SR and it would have been highly desirable in the Intellofax System had there been space available on the IBM card. It is presently being used in the Minicard experiment to specify subjects for which there is no exact code, and to cite the names of people, organizations, installations, and geographic place names. Clear text coding is the ultimate key to the specificity problem. Clear text and phrase coding (see below) used with a classified index allows for the organizational and generic values of classification plus the specificity advantages of coordinate indexing. It is an essential auxiliary to an index such as the ISC which for practical reasons cannot go into great depth on all subjects.

#### G. Phrase Coding

Phrase coding is inverse coordinate coding. In phrase coding, subjects, areas, modifiers, and clear text codes are all linked together by logic on input to express an idea or phrase rather than a single subject. The phrase can then be retrieved as a unified idea. The main advantage of the phrase is that it prevents the so called false drop. If index terms were entered into a Minicard type system (or IBM for that matter) without a phrase linkage and retrieval involved a request for two different subjects linked together in the same document, e.g., aluminum and aircraft,

false answers or false drops would occur since there would be a number of documents which discussed aluminum and aircraft unrelated to each other. Phrase coding does not limit searching since subjects can also be searched without regard to the phrase. Phrase coding is an integral part of any computer type system and there would seem to be some very real advantages to linking several subjects together in an IBM punch card system in order to eliminate false drops on coordinated searches. It might also simplify IBM searches.

#### H. Coding Dictionary

1. The coding dictionary should contain those aids to the classification scheme, necessary or valuable to the coding operation. It need not be bound in one volume.
2. It has often been stated that the key to uniformity of classified indexing input is the alphabetical index to the classification scheme. Classified indexes by their nature must place similar subjects in more than one place in the classification scheme in order to maintain the classification of knowledge pattern, e.g., locomotive production would normally not fall in the same subject series as locomotive engineering. The alphabetical index to the classification scheme points up these distinctions or at a minimum gives the various places in the classified index where locomotives are indexed.
3. No classified index can specifically include all of the subject matter which it must index, but it generally has subject categories broad enough to blanket almost any subject which may be reported, e.g., the index may contain pharmaceuticals, but no specific types. Specific types would be entered under the broad subject pharmaceuticals. When subjects such as specific types of pharmaceuticals are identified and their place in the classification scheme is located, an entry should be made in the index to the classification scheme so that when indexers encounter the same specific pharmaceutical in future reports, they can easily determine the previous decision and ensure uniformity of input.
4. There is also a third type of entry which should be included in the coding dictionary; namely, coding rules applicable to specific happenings, e.g., Berlin crisis. Generally a coding pattern has to be established to handle these situations. This coding pattern may consist of several subjects and areas. One way of informing the coding group of these coding patterns is to circulate a memo and establish a central authority card file. This has been the Intellofax practice. A superior method is to include these decisions in the coding dictionary with the other index entries, thereby lessening the number of places the indexer must search.
5. For those indexing operations which have a clear text entry capability,

the coding dictionary should also contain the form and authority for the clear text entry. Uniformity of clear text input is vital, and therefore has to be rigidly controlled.

6. The coding dictionary should contain then the index to specific entries in the classified index, the index to entries which do not specifically appear in the classified index, clear text entry authority, and other valuable coding rules or aids. The Special Register and the Minicard Code Group have both proved that the coding dictionary can be efficiently managed on IBM cards and issued in the form of an IBM printout. IBM cards can be added or deleted from the authority as necessary and printouts can be easily obtained.

#### I. Standard Operating Procedures

1. Growth of the classification scheme - There are three main sources of suggestions for subject addition to the classification scheme:
  - a. Suggestions for code additions arise from the document analysts who feel that certain subjects are not represented or that reporting on certain topics is so voluminous that further subject breakdown is needed.
  - b. When reference traffic indicates that certain subjects are difficult to search, consideration should be given to subject additions or changes to ease the search problem.
  - c. Research analysts may feel that their interests are not fully represented and that further subject breakdown or rearrangement is desirable.

Any logical and necessary subjects should be added to the classification scheme after due regard has been given to the possibility of using clear text in place of subject additions. Subject expansion, however, requires the strictest management to ensure that the subject does not already exist in the classification scheme in a different form, that the subject expansion requested is not more extensive than required, and also to ensure that when an addition is made, it is placed in the proper place in the classification scheme. Great caution should be exercised in considering the expansion needs of research analysts. It should be insured that there is actual reporting on the requested expansion. Also it is extremely important that the expansion not be too technical, otherwise it may be beyond the comprehension of the average document analyst.

2. General Coding Procedures - Aside from specific coding rules, there are a number of general procedures on which the document analyst should be instructed. These procedures include such things as

depth of coding for certain types of documents, a list of types of documents which should not be indexed, how to prepare abstracts and title expansions, how to fill in a code sheet, etc. The analyst should not be expected to retain all of these procedures in his head, rather he should be given a separate coding manual incorporating these procedures. Supplements should be issued as needed in form suitable for filing in the manual.

3. Informing the Document Analyst - In order to keep the analyst informed so that he can do a better job of subject analysis, there should be available to him a number of easily understood classified and unclassified reference works on difficult subject fields. If the reporting contains many abbreviations, an abbreviations file should be established (the Intellofax System has had such a file since 1950). Briefings by staff and non-staff members should be arranged to clarify the subject content of the code book. Anything which keeps the analyst better informed improves the accuracy of the input and makes the analyst's job more interesting.
4. Review - It is desirable to have total review of each analyst's work in order to assure quality and uniformity of input. The reviewers should of course have unquestioned coding competence. If 100 per cent review is unrealistic, there should be a definite program of review. The analyst should be made to feel that the purpose of the review is to better the input rather than to maintain a constant check on him.

J. Selection Problems

1. Inclusion of the tools and techniques discussed above will insure a high degree of input uniformity and retrieval specificity. There is one area of intelligence indexing, however, that bears heavily on these problems and on which there are no guidelines. What do you index and how much do you index? The managers of the Intellofax System determined that there were certain types of documents which had little intelligence value, or did not fit into the indexing system, e.g., fragmentary order of battle and State Department housekeeping reporting. These documents fall within a nodex or "no index" category. These nodexes which are not entered in the indexing system are, however, disseminated. This nodex category has been extended to the point where it is occasionally criticized by research analysts, yet we can get little guidance from using offices as to what we should index.
2. There are certain reports which are considered very important by research analysts which are not included in the Intellofax System, for example, FDD Summaries and FBIS Daily Reports. For a number of reasons these reports do not easily lend themselves to Intellofaxing. However, should not perhaps more attention be given on incorporating these reports into the system and less attention to reports of marginal value? This question, of course, raises again the basic

question: How can OCR determine whether a report is of marginal value?

3. Summary reports are another problem. Should finished intelligence be indexed in great depth or indexed only very broadly? The problem of indexing depth for intelligence summaries arises constantly and decisions are often made on the basis of the work involved rather than that of the value of the document.
4. On various coding uniformity tests there is usually agreement as to the central theme codes which should be assigned, but there is wide disagreement as to how much of the peripheral information should be indexed. Every indexer develops his own patterns which he can justify but which do not arise from any specific direction from the user.
5. Greater participation by the user in the form of briefings and actual selection of materials for indexing would improve feedback. Until this interaction between the indexer and user is achieved, the system cannot reach its full reliability and effectiveness.



Panel II

DISCUSSION

QUESTION: Is the Area Code being revised?

25X1A9a

[REDACTED] and I were members of the CODIB Working Group which developed a new Area Code. This code will be issued as a part of the Revised ISC and will carry a 4 digit numeric notation as well as a 6 character alphabetic notation.

QUESTION: Will dictionary building affect the organization of coding activity and/or the distribution of documents within the activity?

Panel members were in agreement that a "Dictionary Building" activity does not necessitate reorganization or effect appreciably the distribution or flow of documents. Recognition of the need for a dictionary entry and the recording of a term of concept according to a pre-planned format rests with the desk analyst. The dictionary card, together with the document on which it has been based, must then be routed to the Review Officer, who is responsible for standardization of all entries and for keeping updated listings on the desks of all Classification Analysts.

QUESTION: Explain the term "Minicard".

25X1A

[REDACTED] "Minicard" is a means of storage and retrieval of information in binary form on film. OCR has an experimental set of equipment in the 3rd Wing of M Building. The Minicard Coding Group (MCG) has been working in support of this experimental group of machinery for almost a year. A decision to use Minicard equipment involves the expenditure of 1.5 million dollars and therefore must be based on all evidence that can be accumulated as to the ease of input and retrieval, file organization, dependability of equipment, mechanical problems, etc. OCR has already learned enough to more than justify its experimental group and should the equipment itself be rejected the system of classification as used by the MCG in the coding of the corpus could be used with other equipment.

Panel III

SUPPLEMENTS TO THE MAIN CLASSIFIED FILE

Panel Members:

(spokesman)  
25X1A9a

25X1A9a

This paper deals with the classification philosophy underlying the existence of information systems which supplement the main classified file in OCR. Some might prefer to call these supplemental systems "special collections," "auxiliary files," or "special libraries." Whatever expression we use, we are all aware that they exist (the more obvious examples being the Industrial, Graphics, Biographic or Special Registers), although we may never have thought too much about the "why" of their existence. In the following paragraphs, I hope to outline some of the reasons why these files exist as entities separate and distinct from the main file, as well as some of the issues which relate to their maintenance or management.

It might be advisable to point out, first of all, that most files or information collections in the intelligence community are supplements to other files in one sense or another. This Agency's Office of Central Reference and all the Registers contained therein might be considered a special file created to serve the special needs of this Agency. Similarly, the RI file might be regarded as a supplemental file to OCR, specializing as it does in data of counter-intelligence significance.

At the other extreme are the files that an analyst, section, or branch might keep for whatever purposes they have in mind. There are countless such supplemental files. Almost every office has one, and new ones are born every day. The Document Division, for example, has an Abbreviation File which is maintained for the obvious purpose of identifying abbreviations found in documents or used in abstracting documents. The Library used to maintain a finished intelligence file which indexed by area and subject the intelligence reports of a finished and evaluated nature. It also has a bibliographic card file on the publications and speeches of Marx, Lenin and Stalin. ORR's Geography Division has a file on the Kurdish problem; the Industrial Register has a special file dealing with certain trip reports; while Graphics Register has a file which controls films containing information on the tradecraft of intelligence, as well as a file in which photographs of naval vessels are controlled by class of vessel rather than by the location thereof. My own Register has many special files which supplement our main file system. Logically, we cannot exclude any of these files, small though they may be, from being designated "supplements to the main file." For the size of the file has

nothing to do with it. A supplement is simply, as the dictionary defines it, something which supplies a want, fills the deficiencies of, or makes an addition to something already organized or set apart.

If there is agreement then that these supplemental files include a wide variety of files both great and small, the next question that might be asked is, "Why are these files separate from the main collection?" Why do we have a Biographic Register, an Industrial Register, a Historical Intelligence Collection, or an analyst file on Communist front organizations? Why aren't they part and parcel of the main collection where these and other kinds of data could be analyzed, stored, and retrieved in one centralized operation?

If we reflect for a moment on the world outside, we become aware of the obvious parallels between our auxiliary collections and the information libraries maintained by industrial concerns, the special collections within general libraries, libraries devoted to a single subject (such as the Folger Library on Shakespeare), and so on. In 1953 there were some 2,489 special libraries in the United States, covering about every subject field.

A number of these libraries developed because of caprice. Perhaps a wealthy benefactor wanted to bring together in one place all the books with a certain kind of binding, and supplied the funds to see that it was done. No doubt some special intelligence collections have been founded at least in part because of the caprice of a high-level official (or even a medium-level analyst), and either should never have been created at all, or at least have long since fulfilled what useful purpose they once had.

But caprice does not fully explain the phenomenal growth of specialized libraries in the outside world, nor is it an acceptable explanation for the creation of most auxiliary collections in the field of intelligence.

The reason that is most often given for this extraordinary development is the tremendous increase in recorded knowledge. Until a few centuries ago, information control problems, as we know them today, were entirely unknown. Few books were written, and these could be easily classified in one or more of the few then recognized sciences or fields of human endeavor. The industrial revolution, however, created a new body of knowledge entirely different in nature and much larger than all preceding knowledge. This knowledge was not only published in many forms other than books, but it also contained bits of information related to or of potential importance to many different subjects.

We in the intelligence reference business would certainly admit that the sheer size and diversity of application of the information in our collections has been a factor contributing to the development of supplements to our main file. But there is another factor also which has contributed to this development. The business of digging out information has become so involved and time consuming that a librarian can no longer remain a mere "custodian of knowledge," as Webster once defined him. He

can no longer merely collect and guard data. He is asked to assess it and often called upon to summarize it. In brief, he is asked to file information rather than material and this has necessitated the introduction of special filing and retrieval techniques.

We should also not forget the problem of the physical character of our documentary materials. No one as yet has found one acceptable solution for cataloging books, journals, maps, photographs, films, and so on -- or for filing them. Each medium raises problems different from the other, as evidenced by the growth of special manuscript collections, photo libraries, etc.

Yet even if we admit all these facts to be true, it still does not entirely explain why certain files have been separated from the main system and not others. And it is even more perplexing when one considers such auxiliary files as the Biographic or Industrial Register, where the physical nature of the material does not differ greatly from what is filed in the central document collection. Why can't these registers, who read many of the same documents received and classified in the Document Division, become a part of the main file?

I would suggest that the real reason is none of those that have been cited -- neither the growth in recorded knowledge, nor the increasing demand for information service, nor the physical nature of the material itself -- though all no doubt play their part. What actually causes a special or auxiliary file to be established -- I think we all would agree -- is consumer demand. But it is more complex than that, for consumers demand many things but they do not all occasion the creation of a special file or register. If one attempted to embrace all the complex factors that enter into it in a single sentence or formula, it might read as follows: Given the present state of information storage and retrieval theory, the magnitude of the requirement for establishing an auxiliary file system is a function of such factors as the size, nature, and organization of the collection, the scope of requests, and the comprehensiveness and form of the answers to be provided.

Note that I have qualified this statement with the words, "Given the present state of information storage and retrieval theory." It would, of course, be most desirable to have one central information system where all could go to get the data they wanted. It would eliminate the very real danger of duplication of effort, overcome the problem of attempting to define mutually exclusive subject areas, and achieve greater efficiency. But the science (or art) of indexing and machine filing and manipulation of data has not advanced to the point where such centralization is possible. Conceivably, the day might come when we will have one universal classification system applicable to all kinds of data of whatever depth, and some mechanical storage device into which all types of graphic materials could be filed. With such a development, we would theoretically not have to worry about the size or nature of our special materials, nor about the kind of information we would be expected to provide. But that day is not here, and even if it comes it may not dispense with the need for specialization on the part of

the classifier or information officer. Some human analysis and judgment will probably still be required, both in classification and in retrieval, even if auto-abstracting and similar tools become available. If so, then it would be immaterial whether the entire information system and all its personnel could be housed under one roof, all the data classified by one index system, and all stored in one machine -- for there would still be a need for industrial analysts, if not an Industrial Register, for graphic analysts, if not a Graphics Register, and so on.

I have said that consumer demand is the most important factor affecting the establishment of an auxiliary file system, because it seems self-evident that we do not ( or should not) store information for the sake of storage alone. Even a public library, which is not an information system in our sense, tends to reflect the interests of the community in which it is located. It is true that we may collect and store material in which there is no immediate interest, but we will certainly not index it to any great extent nor separate it from our main collection. Whether we will ever do so depends first of all on consumer demand, and secondly on a combination of certain other factors which I inserted in the formula offered above.

One of the items referred to was the scope of requests. And by scope I mean depth as well as breadth. Let us take, as an example, a document dealing with the Ukrainian Academy of Sciences which might be acquired by our central collection. Presumably the document would be abstracted for the Intellofax System and classified under the Intelligence Subject Code by such subjects as history, the Ukraine, science, and so on. If this kind of general classification satisfies the users of the main file, if they are only rarely interested in obtaining information about particular institutes within the various Soviet academies, then it would be foolish to index such a document in greater detail, much less to divert it from the main file to some auxiliary collection. Moreover, it would not be necessary for the classifier to have any special knowledge of the subject in order to catalog the document adequately.

Let us suppose, however, that the information system frequently receives requests on various matters related to Soviet science, including the general organization of scientific research in the Soviet Union. If the main file's classification system is too cumbersome or too general to enable one to locate documents dealing with this subject quickly and easily, a small desk file or background folder will be set up to make the information easier to locate. We now have the beginnings of an auxiliary collection which might easily expand into a large supplementary file operation -- an "Organization Register" -- employing dozens of specialists. It would all depend on how many customers required this kind of data, how much information was received, whether the depth of requesters' interest was such that they might require information on research conducted in a specific Ukrainian laboratory, whether the information specialists would have to learn Ukrainian or some other foreign language to perform their job properly, and what form of response would be required. I have used the example of organizational information because it is a subject which has, in fact, become of such interest

to the intelligence community, that within the USSR Section of BR we have created an organizational file which is auxiliary to our main biographic file, which in turn supplements the central intellofax system. We have persons who specialize in this kind of data, and they have even gone so far as to write organizational studies for the NIS and other intelligence programs.

I am told that somewhere there exists a file on the agents of a certain intelligence service. These agents habitually use aliases which may consist simply of a given name -- such as "Frank" -- which they often change. It is important that information obtained on these individuals be collected and filed for counter-intelligence purposes. But how can one classify and retrieve the pertinent data on one of these persons if he is continually reported under different or incomplete names?

The method employed in this case is to classify by physical characteristics -- by scars, birthmarks, moles, and other distinctive features of a person's physiognomy. Presumably, material on a certain "Henri" with a scar across his nose would be filed with information on a "Gustav" who is said to have a similar scar. Can anyone imagine such data being mixed in with our central document collection with any hope of retrieval? Think what it would do to the Intelligence Subject Code. Think too of the poor classifier who would have to leap from reflecting on how to code intra-bloc fiscal policies to those persons he has indexed as having scars on their faces.

Having examined some of the reasons for the existence of collections which supplement the main file, let us now consider some of the problems we encounter in their maintenance.

One of the inevitable consequences of a compartmentalized information system is duplication in some sense between the auxiliary files and the main file, and among the auxiliary files themselves. This invariably disturbs management. Duplication connotes waste and waste must be eliminated. An effort is often made to centralize the various information processing and retrieval operations.

Recently, I and two other OCR representatives were invited to study the information handling activities carried on by the various branches of a division in another CIA office, with a view to the possible centralization and standardization of these activities. Each of the branches of this particular division had certain specialized substantive interests, but all were concerned with the same general subject. Each branch was coding and classifying material that flowed into the division in a way that it felt best. None of the classification systems were the same, and there was no single place where a person could go to get all the data on a given subject. Naturally there was a certain amount of duplication among these specialized file systems, and pressure was being exerted for uniformity of processing, if not centralization.

In the course of this investigation I had occasion to visit another

division of this same office. Here the situation was exactly the reverse of the one we were studying. In order to avoid duplication and the other ills of decentralization, the responsible officials of this division had, some-time in the past, concentrated their information activities in one branch. I learned, however, from one of the information specialists in this branch, that this trend was actually being revised. Intelligence officers in the other branches were beginning to compile their own files again in their own ways, and it had reached the point where one could no longer rely on the central information system.

It may be that this kind of problem could be avoided if there was a better understanding of what we mean by duplication. Duplication in what sense? When is it permissible and when is it not?

All of us in the reference business, I am sure, can sense when duplication is good and when it is bad. I feel certain that my colleagues in IR, BR, and FDD would all agree that much of our work on organizations and institutes is duplication, and it is bad. Why? Not simply because we are processing the exact same data out of the exact same documents. But because we are all processing and storing in anticipation of certain needs of our customers -- in anticipation of present or future retrieval requirements -- which in this instance happen to be identical. The same cannot be said of IR's and GR's coding and storage of industrial photos. Although the photos are the same, each division feels that its retrieval requirements differ.

There is duplication too between the main classified file and some of the registers in that they are coding the same data. But it has been said, and I think rightly so, that neither activity can substitute for the other. The central system must supply the necessary generality in indexing so that it can handle intangible or abstract subjects, without delimiting file categories to a degree that might hamper future searching from a new approach. Since it is these abstract subjects which are most susceptible to change, reflecting as they do the thought patterns of the time and a particular researcher, they must not be coded in any great detail lest they be unable to generate the answers to new problems.

Let us imagine for a moment that the Library was told to stop answering requests having to do with Soviet scientists since this is a duplication of work done in BR. A requester might then approach BR for information on the number of Soviet physicists who received the doctorate degree in 1958. BR, like any specialized information system, attempts to exploit all sources of information which have any bearing on its reference mission, and to index such data to the greatest depth required.

It is conceivable that BR could answer this request by the laborious process of first selecting out all the Russian physicists in its files, and then reviewing these dossiers to see which physicists held the doctorate degree, and, of those, the number that received their degrees in 1958. But a reference facility could have answered this kind of general question much more easily since, unlike a specialized file, it is not imprisoned by the detail of its own classification system. The central system, like BR, may

have received information pertaining to the training of physicists in Russia, but instead of indexing names that might have appeared therein, it would have cataloged the material under "Science Education -- Russia," or some such subject.

In addition to this problem of duplication, there is also the issue of whether the classification problems of the subsystem differ in character from those of the main file or are any less difficult to resolve.

It has been said that it is the degree of diffuseness of information that is the heart of the classification problem. If this is true, then it seems logical to carry the reasoning further, as some do, and state that where an information collection is assembled for special purposes the problem becomes less severe, since the indexing need cover only a fraction of the full potential of the information.

Supporting this line of reasoning is the argument which some of our own people used a few years ago in replying to the criticisms of the Library Consultants. Their reply emphasized that it is much easier to classify specific named objects, such as people, plants, geographic place-names, and so on, than to classify abstract subjects. For the classification of named objects, they said, lends itself to specificity, detail, and relative stability when compared with abstract or intangible subjects.

At first thought, this view appears to make sense. BR's classification problems, for example, seem fairly simple -- its business is people, and people are specific enough. As the poet said, when asked to explain geography and biography:

"Geography is about maps,  
Biography is about chaps."

Nevertheless, we who have been concerned with the maintenance of the special collections have found that it is not quite that simple. That, in fact, as time goes on, every subfile of the main file system ultimately encounters most of the same classification problems which cause such difficulty for classifiers associated with the main file.

The reasons for this are not hard to determine. No respectable specialized collection which deals with named objects would be content to index by these named objects alone. If they did so they would soon be servicing only a fraction of their potential customers. In truth, the user of the information system would like to have every item of information, named object or otherwise, indexed by every conceivable category into which it may fall. This, of course, is impossible, since we are limited not only by cost considerations, but also by our incomplete mastery of the science of information storage and retrieval. But we do make some effort in this direction and it inevitably leads us into the indistinct world of abstract ideas and patterns of thought. For soon we are not simply indexing the name of the plant or the person, but the economic, scientific, or social scientific subjects with which these named objects are connected. We are not merely indexing the name "Dmitriy



Yemelyanov," we are also saying that we think he is a cyberneticist and that perhaps he should be connected with the subject of aid to under-developed countries. This is one of the reasons why we have "Snag Files," and why we develop complex hierarchical coding systems which look very similar to the main collection's Intelligence Subject Code -- although differing in content since they have been formulated to meet our own peculiar requirements. It also explains why one large specialized collection which at one time attempted to expand beyond mere name-index control of its material, found that its subject files had become catchall repositories, and has now concluded that the semantic problem is too great to overcome.

Another question that has disturbed some of us is whether there are any logical limits to the number of items of information that should be classified by the auxiliary file system. Must we index our primary subjects -- whether they be plants, people, photos, or other -- by every known fact which can be applied to them?

One of the reasons for indexing in detail is, of course, to enable one to find a specific piece of information quickly in a large mass of material. Of course, this does not mean that you have to index everything in your file system in order to find what you want. There is, however, a second advantage to the kind of detailed coding and indexing done by an auxiliary file system, and that is that it enables data to be synthesized at a later date in such a way that it may reveal items of intelligence information that might otherwise never have been discovered.

To put the matter in another way -- one of the best reasons for an auxiliary file to index in great detail is that it permits statistical analysis of a whole range of intelligence problems. And on occasion this kind of analysis will lead to significant intelligence breakthroughs. Undoubtedly we will see this kind of technique being used even more in the future, especially as we acquire faster and better machines to do the job. But while statistical analysis does require a vast body of detail to work on, this does not mean that we must index all incoming information by every classification category possible. Detailed classification is justified only when there is sufficient data to have statistical significance and when there is likelihood that there will be inquiries that can be answered by conclusions from this data.

This may seem to be an obvious point, but it is one which we tend to forget. Too often, in our zeal to satisfy the wishes of our consumers, we begin to classify what they want (or what we think they may want) even though we will never have enough data in these subject categories from which any significant conclusions can be drawn. Whether the subject is license plate information, the age of factory buildings, or the domestic travel of Soviet nationals, the classifier has the responsibility to decide, on the basis of his intimate knowledge of the materials with which he deals, whether indexing would be worthwhile. It is in this way, as in others, that he fulfills his true role as an information specialist.

In summary, this paper has argued: that when we talk about supplements

to the main classified file we must include in our thinking any file that fills the deficiencies of another file; that we have these files because specialized research requires specialized information; that they are born of consumer demand and shaped by its needs; that file duplication is to be condemned only when the retrieval objectives are identical; that while an auxiliary file may begin with a narrow field of operations, in the attempt to win complete mastery of its information content it meets the same classification obstacles as the general file; and that the application of index controls to data must always be governed by the quality and quantity of material coming in, and by the good judgment of the classifier.

Panel III

DISCUSSION

25X1A9a

[REDACTED] Do the members of the Panel desire to add to what said with respect to supplementary files?

25X1A9a

[REDACTED] Subject which I didn't cover with much detail is the snag file. Wayne, do you want to comment on that?

25X1A9a

[REDACTED] I propose that a Snag File may be understood to control selected information under criteria which are not adequately served by formal files already in the environment and which continue to conflict with the criteria of the formal files. To put this a little more simply, if you don't like the system, don't resign -- start a snag file.

Snag files are the competing rudiments of future specialized registers. Everybody, who is conscious of a problem which irritates him in his duties and which is not adequately controlled by the apparatus already available, has the responsibility (because of his awareness) to begin doing something about it. The physical equipment with which he records his evidence on his chosen subject is his snag file. I regard most of the information files, desk files and specialized accumulations of records in the possession of analysts throughout the intelligence community as snag files, supplementing (in a regretfully discoordinated way) the main file. We need a much better organized means of feeding back the quality that these files possess into the main file. For example, there are a great many files on the subject of organizations, in a great many different hands. An "Organization Register" could be supported with an immense amount of data, if it were assembled in one place.

Referring to the subject under which we assembled: the philosophy of classification, let us not forget that philosophy involves controversy. It is not usual for philosophers to agree. And we may all claim the right to maintain our own point of view, and document it and support it by building our own snag files (to the extent that we can get the energy and means to do so) because compromises which defeat our own point of view are not necessarily losses for all time. The occasion which justifies our special point of view may be coming. Of course, there is survival of the fittest in this game. Not everybody always wins.

The difference between the main, central file and all the many little snag files, at two polar opposites of the information control activity, may be summarized:

In the main file, everything received is thrust into some category or other. It amounts to an array of 5,000 or 15,000 boxes. A requester is presented with a selection of boxes in which he must rummage among the nuts and bolts to pick out what he wants. He is participating in a stage of the information control process because he has nobody else to do it for him.

In his snag file, his selections have already been made.

We must remember that the indexes we build, whether snag files or central files, are accumulations of our perceptions. We could repeat the effort two years hence, or we could have the indexing done by a dozen people (the White Stork method) and come up with a wide range of perceptions. But this process is never finished because perceiving is never finished. Variety of perceptions provides opportunity for snag files.

25X1 [redacted] Don't you think these snag files are actually files that are closer to the consumer, files that meet his needs in the most immediate way, because the snag filer knows what the consumer really wants? He is not guided by official missions or descriptions of missions. He talks to consumers all the time and he recognizes that they want a certain kind of control set up apart from the main file, or a supplement to the main file. He is really providing a more  
25X1A9a more intelligent service than that available from the main file.

[redacted] Yes, I agree. The snag files you are describing are the ones which accumulate around input desks. I've been emphasizing that they are not the only snag files. Around input desks we can have maverick files -- that is, files that fail to conform to team requirements. When thirty or forty indexers perform a major input program as a team they can't behave like thirty or forty mavericks. The freedom to be a maverick is available only in the snag file. But I propose not to overlook the snag files in the user community. A sufficiently active dissatisfaction with the detailed service that he can get from the main file is the proper license of every  
25X1A9a user. It is within his means to start a file which goes some distance in remedying his problem.

[redacted]: I might ask Mary this question: do you think there should be more special collections rather than less -- should more of the named objects be separated from the main file and made the subject of  
25X1A9a elemental register?

[redacted] Well, as you said, such files would have to be born of consumer demand and shaped by its needs. I think if we have one bond in common here today it is our problem with the consumer,

trying to ascertain what he wants and what we can do to give him better service. So we might well have a separate register for organizations. As brought out in the discussion, the coverage of organizations is scattered and there is different emphasis in retrieving information on organizations. So we'd have to ask the consumer three questions:

(1) We'd want to know how great is his need. Does the unequal depth of coverage cause any great problem to him?

(2) We should want to know what, exactly, he hopes to find in an "Organizations Register." Just what information does he hope to find? Because it would be possible to just give him back specific organizations but very often we find that the consumer is using an organization as an approach to another type of information. For instance, in the USSR he may be tracing the changes in subordination of organizations. He may find that a plant has changed in subordination from one ministry to another, that this indicates a change in production, and change in emphasis of a whole industry or an armament program. Or he may be interested in approaching a particular subject. There may be many cards in an IBM index of a certain subject and it might be much simpler to take an organization, or a few organizations, involved in this subject, if that's what he really wants to get. Or the consumer may be interested in finding out the names of particular persons associated with particular industries, plants or organizations. If we had an organization file we would want to emphasize the thing the consumer hopes to find by placing a request with us.

(3) Then we'd want to know in what form he hoped to get this information. Does he want documents, IBM listings, or a synthesis of data? Or a combination of all three? The answers to these questions would guide the operation of the register. A synthesis of data could be obtained as a by-product of the classification system. Particularly if you use a punched card system, you could determine which factors you were interested in, what you wanted to know about a particular organization. As this information became available in indexing on a daily basis, it could be recorded, kept up to date, corrected and changed, and the information could be arranged. Now, such a synthesis of data would speed up service to the consumer because you could arrange information, or, by checking your file, you could establish whether you had ever seen such information. (There would be no point in searching through 1,500 documents to find the street address for a particular factory if you had controlled that street address in your information file.) Arranging the information file in several different ways might suggest further channels for investigation. It would help achieve consistency in classification to have the information file on hand for your classification analysts as well as for the consumer. It would help to identify vague references, and pull them together.

So I think we could, or the consumer could use an organizations register. But, of course, it would be up to him in the final analysis.

25X1Aga  
[REDACTED] Aren't there two possible interpretations of the term "snag file"? Are they by their nature supplemental to a register, or are they files of information which should be handled by the register? I'm thinking of a file that the Industrial Register might have that identifies certain types of installations. Shouldn't the other snag files which represent failures of the system be incorporated into the system in some way?

25X1Aga  
[REDACTED] Mr. Campbell: Yes. I think we all share the employee's suggestion that analysts should please come forward with their half-done files and arrange for an improved degree of community accessibility by central listing of them. This kind of call has been made many times, we know. I'd estimate rather more than a thousand little files around the community that might be called snag files. I think a very small proportion of these will ever develop into, or be incorporated into a register. Some of them are not going to achieve this (perhaps desirable) central responsibility. If you mean there is vagueness in the term "snag file", that is granted. There may be a better term. But we are talking about voluntary files, the result of active dissatisfaction with the degree of control of information in a subject field where certain services are already available.

25X1Aga  
[REDACTED] Were the Registers created to process data and supply informational answers, as distinct from the Document Division's of providing for document retrievability?

25X1Aga  
[REDACTED] A requester asking for material relating to a specific individual might be provided with a dossier, which is nothing more than a collection of documents which refer to the individual. This is not much different from the provision of documents from the main file pertaining, for instance, to the subject of underdeveloped areas. Admittedly, in one case there is reference to a named object and in the other to a more abstract idea, but I don't see much difference in the way the questions were answered. True, the Registers produce research aids intended to save requesters time. These amount to research for them, in a sense. In the Register there is greater emphasis on providing the requester with information, but you can't carry that idea too far. The main file provides information, too, as well as providing

25X1Aga  
[REDACTED] Graphics Register is somewhat unique since the material on file is the photographs themselves. We have found, through experience, that the consumer prefers to come in and use the photographs as they are found in our files. Governed by this preference, we generalize in our coding. We have two classification systems. In the Film Branch, we use the ISC, and in the Photo

Branch, we have a photo-intelligence code, much more generalized than the ISC.

25X1A9a [REDACTED]s: Perhaps some of these "supplements to supplements" cannot be merged with the auxiliary file, desirable as that might be. For example, why couldn't you merge your little file of photographs of naval vessels, for control, with your photographic file system?

25X1A9a [REDACTED] Our photographic file system just wouldn't handle it. Our photo analysts are expected to select pictures which add to photo intelligence on the naval vessels. This file is handy for us, and it can be used, on occasion, by a requester who is searching for a photo of a particular vessel. The present general filing system for photographs is by country, province, city and then by a numerical control number. The mounting card bears a 32-subject set of broad categories which we call "selection by compartmentation". When a photograph of a naval vessel in Odessa comes in, we don't have the physical means of filing the master photo in two places. It is filed in Odessa, but this file for naval vessels gives a cross-reference, by number, type, name, etc., to the master copy. The photo is coded also in the ISC system, but not with the same degree of fineness as in the naval vessel file.

25X1A9a

[REDACTED] In the Industrial Register, in using a three-digit code, do they also use clear text as a means of entry to the indexed information?

X1A9a

[REDACTED] No, the primary emphasis is on the three-digit code. But we do have little reference files and subject files, as on mining within a country, transportation facilities, other economic factors, filed by country, but not a clear-text index.

25X1A9a

[REDACTED] Is the three-digit code card-punched?

25X1A9a [REDACTED] It is card-punched, and recoverable by machine methods by industrial categories.

[REDACTED]: Is there space on the punched card for other information, such as clear text, if you chose to use it?

25X1A9a

[REDACTED] Yes, I'm sure there are many, possibly eight spaces, available. Further punching would create several problems, however, perhaps longer listings and an unwieldy working tool. We deem the three digits sufficient.

PANEL IV

CONTRIBUTION OF MACHINES TO THE  
25X1A9a CLASSIFICATION PROCESS

Panel Members:

(spokesman)

I. INTRODUCTION

25X1A9a

A. Purpose of Paper

The purpose of this paper is to suggest areas and ways in which machines can be put to use to assist Classification personnel in the performance of their indexing function.

B. Composition of Paper

The paper is composed of two sections - one major, the other minor. The major section identifies areas in which machines have been used within OCR in support of the classification function and outlines some specific applications in this regard. The second and minor section of the paper treats of the working climate which must exist among classification, machine, and reference components in order that advantage be taken of the full potential of machines in the OCR complex.

C. Apologia

I should like to admit a few things right away about this paper: First, there is throughout a presumption that our topic "Contribution of Machines to the Classification Process" is not an absurdity, a presumption that machines can assist persons engaged in the task of classifying data for input to a machine index system. Second, we are talking here about standard EAM machines; that is, punch-card equipment of the type now available in OCR. We are not talking about EDPM machines or high-speed magnetic tape equipment with computer capabilities, etc. Third, there is what may appear to be an undue accent on the experience and practices of the Special Register (SR) in this paper. SR's entire reference system is heavily oriented towards machines - much more so than is the case with any other OCR reference component. So we have developed a habit of trying to make machines do things for us. And one of the things we have tried to make them do is to assist classification people with their indexing function.



Some of the tasks performed by machines in support of classification personnel could be accomplished by other means, of course. This paper, however, will outline something of what has been done in the prospect that you may find some "transfer" value in these applications.

D. Basic Machine Capabilities

Before getting into the specifics of how machines can assist in the classification process, it may be helpful to list briefly the types of processes EAM machines can perform. Each of these processes or functional capabilities may be of help to classification personnel attempting to use machines in support of their classification task.

First, of course, machines can store information, keeping it "at the ready" in the form of punched card files or indexes.

Machines can cumulate or merge information, thereby updating files.

Machines can compare information and check file sequence in the process.

Machines can arrange or sort information into differing sequences.

Machines can select wanted information from a larger mass of data.

Machines can perform simple arithmetic tasks, such as counting cards, adding and subtracting figures, etc.

Machines can reproduce data and, at the same time, adjust or rearrange the relative left/right positions of data fields in the process.

Lastly, machines can print out information.

EAM equipment is today considered to be slow, through comparison with EDPM equipment. However, even EAM machines perform these data-handling processes with great speeds. One of our print-out machines, for example, could keep pace with a complement of 37 Agency-qualified typists.

Well, these are the basic functional capabilities you have at your disposal. I'm sure most of you are familiar with them.

II. MACHINE SUPPORT TO THE CLASSIFICATION PROCESS

Now let us turn to those areas, together with some specific applications, where machines have been used in OCR to support classification personnel in their indexing activities.

A. Classification Manuals

Of course, machines can be used to store, sort, and print-out your classification manuals or code books. There are several advantages to this:

1. Revision of manuals

Because a punched card file constitutes a kind of "dynamic storage" of data (that is, the storage is extremely flexible and mobile), the recording of your classification manuals in machine cards greatly facilitates posting revisions to your coding scheme. Not all code systems change appreciably. Some change a great deal. The Special Register's code scheme for Soviet organizations, for example, undergoes hundreds of changes each year in response to the changes occurring in the organizational structure in the Soviet Union and in response to our improved understanding of this Soviet structure through new intelligence receipts. The Intelligence Subject Code (ISC) has recently been extensively revised and the planning for processing with Minicard equipment may result in further changes. Such revisions are very easily recorded and controlled when your code book is stored in punched cards.

2. Currency of Manuals

The speed of machine print-out makes feasible more frequent printings of your code books, with the very desirable result that the manuals on the desks of your classification analysts are kept more up to date. In a growing classification system, this is particularly important.

3. Multiple Sequence of Manuals (Index to basic book and others)

Through their sorting capability, machines make it entirely feasible to list your code books in varying sequences. The basic sequence for a structured or classed code is, of course, by code number. This sequence groups the topics of your classification

scheme by major category with generic subordination within each major category. If your code books are recorded in punched cards, however, it is possible to list your books in alphabetical sequence by code title, thereby obtaining alphabetical indexes to your basic (code number) books.

Such alphabetical indexes to the basic code books have proved very helpful to classification analysts. The basic purpose of the alphabetical index, of course, is to provide the classification analyst with a tool for quick access to the code number for a given concept through direct alphabetic look-up rather than obliging him to find his topic within the structured arrangement of his basic classed code.

The alphabetical index, however, serves two auxiliary purposes which seem worthy of note. Such an index to the basic code often facilitates more complete and accurate use of the code scheme by the classification analyst in that it serves to alert him to the full range of coverage within the overall coding system of a given term and the multiple meanings this term may possess. For example, the analyst sees a document reference to the term plates, without further specification. In undertaking to assign a code number to this topic he may think of some of the following possibilities but probably would not think of all of them: anchor plates, armor plates, battery plates, boiler plates, cathode plates, cutting plates, dental plates, dinner plates, etc... to pursue only the first four letters of the alphabet.

It may be possible (it often is) in view of the larger context of the document and with the aid of such an alphabetical index, to determine the specific nature of the reference at hand. If not, the analyst at least has the range of possibilities at his finger tips and can code the item in accordance with classification procedures established for such contingencies.

In addition, the alphabetical index provides to classification personnel a tool which may be used to improve the code scheme itself. Not infrequently, confusion creeps into a growing classification scheme through entry into the system of code titles which, although ideationally unrelated, contain similar or even identical title words. The inexperienced or careless analyst in his search

for a code number may find the desired word in the basic classification manual but not the desired concept - with the result that incorrect classification occurs. With an alphabetical list-out of all code titles contained in the system at his disposal, the classification officer can readily locate for survey such potential trouble areas in his coding scheme and, through appropriate re-wording and cross referencing of code titles, minimize the confusion caused by similarity of title words and consequently the potential misuse of the classification system.

The basic code number sequence and the alphabetical sequence by code title (the index) may not be the only sequences it will prove profitable to have your classification scheme listed by. For example, SR's classification scheme for Soviet organizations is listed, believe it or not, in ten sequences: (1) by code number; (2) by title of organization; (3) by city of location; (4) by city of location within Oblast; (5) by SOVNARKHOZ or Regional Economic Council; (6) by Soviet plant number; and by four different echelon levels, viz., (7) by Chief Directorate or Main Administration; (8) by Directorate or Department; (9) by Plant, Trust, or Combine, and (10) by Laboratory, Office, Base, etc., within Plant. It is very unlikely that any classification manual not controlled by machine would be listed in so many sequences. Yet each of these sequences is important to SR's classification effort.

## B. Control of Problem Topics

Another area in which machines can be used in support of the classification process is in the control of "problem topics"-that is, topics which cannot be clearly identified or which do not, for one reason or another, fit precisely into the classification scheme.

In SR, we have developed two types of controls for handling these problem topics, both of which enlist the support of machines.

### 1. The Authority File

The first of these controls is the so-called Authority File. By Authority File, we mean here a file or index of topics which have proved troublesome to code and for which, accordingly, codes have been established by supervisory direction as those

to be used by all input analysts. These code numbers, although perhaps largely arbitrary, by such action become the "official" or "authorized" codes for the topics in question. Thus, the term Authority File, Machines, of course, can be used to store, sort, and print-out the Authority File, and with all the advantages concomitant to machine-card maintenance of the Classification Manuals themselves.

## 2. The Snag File

The second control for "problem topics" is what is called, in the terminology of this paper, the Snag File. The Snag File in SR is a special auxiliary file of those problem topics which are not controlled by the Authority File. The Snag File is maintained in sequence (alphabetically and numerically) by the problem topics themselves rather than by the classification code numbers under which they were indexed, thereby providing direct reference access to problem topics irrespective of codes used.

The Authority File and the Snag File both record coding actions that have been taken. The difference is that the Authority File contains those coding actions which have been thoroughly thought through and constitute authoritative and lasting decisions, whereas the Snag File is composed of coding actions which are not likely to recur and are not considered worthy of long deliberation. Actions recorded in the Snag File might be termed spur-of-the-moment decisions. Strict uniformity in this type of coding action is not necessary because data retrieval is guaranteed by inclusion in the Snag File of all such decisions. Pragmatically, the Authority File tells how to classify a problem topic (providing a single, authorized code for each) while the Snag File tells how a problem topic has been classified (providing a record of the several codes which may have been used in ad hoc coding actions taken). The Authority File, then, is primarily an aid to classification input; the Snag File is an aid to reference recovery.

In SR, machines have been used to prepare a separate deck of cards for problem topics caught in the daily work flow. This culling by machine from our daily work flow of questionable classification entries is accomplished by a simple overpunch in the machine card ordered by the classification analyst whenever he feels he has

not fully resolved all reasonable doubt in assigning to the topic in question the code he has chosen. This is an extremely low-cost method of compiling these data.

3. Authority File and Index to Classification Manual Combined

It may be worthy of note that in SR we have combined "authority file" entries with the regular code book entries, thereby producing a single alphabetical listing of these combined decks of cards which serves the classification analyst both as an index to the basic code book and as an authority listing for problem topics. That both the authority file and the code book are stored in punched cards makes merging and sequencing of these data a simple matter.

C. Input Quality Control

A third major area in which machines can be of assistance to the classification function is that of input quality control. There are several techniques which may be employed here - the objective being to catch errors in the index cards before they are merged into the standing indexes of the service system.

1. Daily Work Listed for Survey

One technique is to prepare listings in code number order of new index cards as generated in the daily work flow. In the hands of the classification analyst, such listings make it feasible for him to catch (1) impossible codes, (2) alphabetical entries which are incorrectly spelled, and (3) entries which do not conform to established procedures governing the manner of entry and fielding of data. This technique has been used in the Special Register as a check point in our quality control efforts for all our basic indexes.

2. Daily Work Matched Against the Classification Manual

Another technique aimed at minimizing input error consists of matching by machine new index cards against the official classification manual codes in order to isolate all impossible codes. This technique is currently used by the Machine Division in connection with the Intellofax system.

This same technique is used by SR in the case of our Soviet place name index. This matching of new index cards against our area classification manual validates the accuracy of both area code numbers and area place names.

3. Automatic Authentication of Data Entry Patterns

Another technique for the control of input quality consists of screening by machine new detail cards to assure that prescribed patterns of data recording are being maintained. Conformity to prescribed patterns can be checked at multiple points in the index cards by a single pass through the machines. In SR, all the major fields of new subject and commodity index cards are checked by this technique before being merged into the standing indexes.

D. Correction of Index Cards

A fourth major area in which machines can be used in support of the classification function relates to the correction process. Every index system has its errors and it is one of our less pleasant tasks to try to get them out. Machines can help.

1. Correction as Step in Service Processing

The following technique, now used in SR, possesses the particular virtue of limiting or restricting the correction effort to those portions of the index file actively used in the servicing of requests. As a standard step in the processing of each search of our machine indexes, a listing is prepared of all index cards recovered. This listing shows all data contained in every index card selected in the search. Documents referenced in these index cards are then pulled from the document file and are scanned for pertinency, possible follow-up runs, etc., prior to release to the consumer. If, through this scan, a document is found which does NOT relate to the requested topic, it is known that an error exists in the index card which produced this particular reference. By turning to the machine listing, the person effecting correction can easily determine which portion of the index card is in error. The listing also provides him with all data required for the deletion card employed in the correction process. Data for the correction card, of course, must usually come from re-analysis of the document - a

step which is taken before the document is returned to file.

2. Detail File Conversion: Old Codes to New

Another facet of the correction process with which machines can help is the conversion of old codes to new codes in the index file following a change in the classification scheme itself. When the classification scheme has been altered, the index file must be correspondingly updated or Reference personnel are forced to work with multiple recovery systems, which is, of course, undesirable. Machines can be of great assistance in this regard through automatic conversion of the index cards from codes in the discarded system to their equivalent codes in the new system.

Of course, this type of conversion can be effected on punched data only. All data in SR's system is punched. In the Intellofax card, however, a substantial portion of the data carried is not punched but is entered as reproduced typewriter text. This textual data would, with present OCR equipment, be lost in the type of conversion depicted above (except when correction is effected by "under punching"). There are indications, however, that new equipment may be on the market before long, which could circumvent this difficulty, permitting Intellofax card reproduction without loss of unpunched or textual data.

3. Suspect Portions of Detail File Listed for Survey

There is yet another way in which machines can facilitate the correction process. It is not unusual that in the course of operating a large index system, certain sections of the system, for one reason or another, become suspect. In such case, it has often proved helpful to use machines to list out for study by the classification officer all cards in the suspect sections of the system. Analysis of these listings may lead to compensatory actions such as (1) card corrections, (2) reprocessing of batches of materials, (3) altering or tightening of input procedures, and (4) revisions to the classification scheme.



III. THE TRINITY OF MACHINE REFERENCE SYSTEMS

Now, in conclusion here, I'd like to moralize a little about the working climate necessary to the mechanized reference system.

We have been talking here about ways in which machines can be put to work in support of our classification people. I think it will be evident to you that support activities such as those outlined in this paper suggest a working atmosphere of mutual understanding and cooperation among all elements of the system.

The point I'd now like to stress is that a mechanized reference or data-handling system of any appreciable complexity or scope can function properly only when a high degree of operational integration or synthesis exists among the three principal components of the system; i.e., Classification, Machine, and Reference.

If there was ever a case for the left hand's knowing what the right hand is doing, it is the mechanized reference system. A data-handling machine is a very precise and exacting piece of equipment. It imposes upon its users demands of formidable rigidity. It is essential that data symbols and their inter-relationships in the machine system preserve constancy of meaning from the initial point of classification input, through all the processes of machine manipulation, to the terminal point of reference output. The machine will accommodate no ambiguity ... no difference of interpretation. The burden of constancy lies with the personnel operating the system.

The mechanized reference system, then, has a "need-to-know" principle all of its own, a principle quite the opposite in its effect from the need-to-know principle of security doctrine. Instead of restricting knowledge and communication, the "need-to-know" principle of the mechanized reference system transcends the barriers of organizational compartmentation and proclaims that all components of the mechanized system need to know a great deal about one another.

What are some of the specifics of this "need-to-know" principle?

The Classification component, in order to do its job, needs to know the nature and capacity of the basic machine record; needs to know the function, design, and maintenance sequence of all data files in the system; needs to know the functional capabilities of the machines and something of their speeds; needs to know the nature of the end product desired by the Reference component; needs to know the avenues of approach to the data files and the search techniques which will be relied upon in fulfilling requirements; needs to know the scope and accent

of requests to be placed, including shifts in the emphasis of consumer interest; and needs to know of success and failure in the operation of the system as guide posts to classification development.

The Machine component, in order to do its job, needs to know the nature of the classification schemes; needs to know the data categories requiring discrete maintenance sequences; needs to know the substantive inter-relationships of data recorded in the system; needs to know the nature of the products the Reference component will require; needs to know the recovery techniques or access routes the Reference component will wish to employ; needs to know the scope and number of requests to be serviced and the time limits imposed on servicing them; and needs to know the nature of new types of consumer needs so that new applications of the machine potential may not be neglected.

And the Reference component, in order to do its job? Well, the Reference component, ideally, needs to know everything about its sister components. It needs to know the substance of source materials and the classification coverage of these materials; it needs to know all classification schemes and techniques and procedures; it needs to know the design, data coverage, and maintenance sequence of all mechanized files; and it needs to know machine capabilities in searching and otherwise manipulating these files so that reference recovery approaches may be efficient and fruitful and so that new methods of exploiting the potential of the system may be conceived and activated.

Our insistence on this point may seem like a tempest in a tea pot, but, in view of experience already gained and in view of the demands in this respect which the new data-handling equipments now on the horizon will make upon its users, we feel our point to be both timely and of substance.

Now all this stress on the "need-to-know" about one another is not, of course, to suggest that you cannot have internal structure within your mechanized reference system. You've all seen organization charts which fracture our Office, Divisions, Branches, etc., into neat little black-walled cells or boxes with very straight and very narrow paths between for communicating upward and downward (but not laterally). Well, this cellulation or compartmentalization is admittedly necessary for numerous administrative reasons and is all well and good for the purposes intended. There are, in fact, lots of very splendid benefits from compartmentalization. It is a great help in the Hearts and Flowers department; it is the only element of order in Time and Attendance reporting; it

intersperses enough supervisors among us so that employees often get to know the faces and sometimes even the names of their bosses - and vice versa; it has not been neglected as a mechanism for justifying promotions; it is absolutely indispensable when it comes to "orientation briefings"; and it can be a great comfort to hard-living employees by giving virtual assurance they can safely expect not to have to speak to a living soul before the morning coffee break!

So, compartmentalization has its justifications and none of us would know how to live without it. But, there is the grave risk here, nonetheless, of which we are warning - the risk that organizational form take precedence over function - that delineation of elements within your system effect a separation of the properly inseparable. The Classification, Machine, and Reference components of a mechanized reference system either work together or they do not work at all. They are not permitted unilateral license. They are highly interdependent elements of a single entity -- the Reference System. They are, if you will, a trinity, composed of three, but constituting one.

Unless this interdependence is recognized, unless the "need-to-know" principle is practiced, unless close inter-unit work relations are established and sustained, your mechanized reference system will not only suffer a deficiency of imaginative service applications, it may well even fail to fulfill the basic functions it was established to perform.

PANEL IV

DISCUSSION

QUESTION: Does any structural organization or other mechanism exist to foster understanding and cooperation among the three major components (Classification, Machine, and Reference) of the mechanized reference system?

25X1A9a [REDACTED] The "need-to-know" more about one another is a matter of great concern. The present OCR seminar is the first Office-wide effort to break down organizational barriers and I hope to have similar meetings at least semi-annually, each to deal with some aspect of the information problem confronting OCR.

25X1A9a [REDACTED] One inter-unit mechanism is already in being - the "OCR Composite Group" established to strengthen the servicing of Intellofax requests. This group is composed of representatives from the Reference Branch of the Library and the Analysis Branch of the Document Division.


The Document Division periodically schedules members of its Analysis Branch to work with the Intellofax retrieval component of the Library in order to acquaint themselves better with retrieval activities and problems encountered. The Machine Division is also represented when appropriate through its standby member of the Group.

There are many ways to circumvent organizational compartmentation. In the Special Register, the barriers of compartmentation are diminished by circulating all consumer requests as written up by Reference to the classification analysts as a double check on validity of retrieval coding; by establishing and exercising direct working-level channels among those actually carrying out assigned tasks, irrespective of Branch of assignment or supervisory lines of communication; by staffing the servicing component of Reference only with persons who have served at least two years as classification analysts; by periodic re-training of Reference personnel in Analysis operations; by exchanging written operational procedures among Reference, Analysis, and Machine units; and by inter-unit staff meetings called by any unit when the need for same arises.

Other possibilities include: (1) additional training programs established by each component for the benefit of selected personnel from sister components and (2) the repetition, after 4 to 6 months, of the familiarization tours now given new EOD's, with follow-up tours thereafter every year or two.

QUESTION: Will the effect on inter-unit coordination be negative if OCR machine operations are consolidated into a single machine center such as is currently under consideration?

25X1A9a

 The purpose of consolidation is largely economic and I do not believe consolidation of machines need have a harmful effect on efforts to coordinate activities of Classification, Reference, and Machine components. As more complex and costly machines are acquired, there is a natural and concomitant tendency to consolidate machine facilities because of the prohibitive costs of multiple installations and the technical specialization required to operate such equipment. The increased need for coordinating input and output activities when working with such equipment may counter-balance any separation of input and reference personnel from machine personnel resulting from the physical consolidation of machine installations.

Summary of Final General Discussion

Most of the discussion and comment in the final session of the Conference centered on consumer requests and appropriate OCR reactions thereto. [REDACTED] 25X1A9a pointed out that our response to most information requests must, of necessity, be a collective or "team" response because of the complexity of our retrieval problem and the substantive interrelationships of the materials received and classified by the various components of OCR.

[REDACTED] 25X1A9a reflecting on "requesters he has known," stated that they generally fall into one of two categories: those who have no knowledge of how to retrieve the data pertinent to their request, and those who feel they know better than the information specialist how to perform the search. Personally, he said, he prefers the former, although this usually means that a reference analyst must spend considerable time with a requester in order to determine exactly what it is that he seeks. Requests from within the Agency, he stated, usually reflect considerable understanding of the retrieval problem, while extra-Agency requests do not. As for the advantages to classification of better titling of reports, he agreed that increased training of collection officers in report preparation and titling would be useful as long as such titles as "Waltz me around again, Mohammed" continue to be received.

The session ended with some brief observations on the oft-discussed possibility of establishing a central point in OCR where customers could obtain coordinated reference service. [REDACTED] 25X1A9a indicated that centralization and coordination of our service activities would be materially aided by the physical accommodations planned for OCR in the new Agency building. In the more immediate future, he added, it is likely that there will be a greater number of intra-OCR briefings and increased interchange of personnel between the various OCR divisions.

Approved For Release 1999/09/24 : CIA-RDP84-00951R000400070003-0

CONFERENCE ON PHILOSOPHY OF DOCUMENT CLASSIFICATION IN OCR

21 November 1959

Chairman - Paul W. Howerton

25X1A9a

0900 Introduction to the Nature of Classification . . .

0930 Panel I - The Intelligence Subject Code . . . . .

1030 Break

1045 Discussion

1100 Panel II - Classification Tools . . . . .

1200 Lunch

1300 Discussion (Panel II)

1320 Panel III - Supplements to the Main Classified  
File . . . . .

1400 Discussion

1430 Break

1445 Panel IV - Contribution of Machines to Development  
of the Classification Process . . . . .

1515 Discussion

1545 General Discussion . . . . .

25X1A9a

Program Assistants . . .

Logistical Support . . .

Approved For Release 1999/09/24 : CIA-RDP84-00951R000400070003-0  
**FOR OFFICIAL USE ONLY**

**FOR OFFICIAL USE ONLY**  
Approved For Release 1999/09/24 : CIA-RDP84-00951R000400070003-0