CONFIDENTIAL

*A new DCID makes timely this
critical review of CIA's refer-
ence facilities with recommen-
dations for improvements in an
eventual federal system.*

TOWARD A FEDERAL INTELLIGENCE MEMORY

George W. Wright

The problem of storing an ever mounting accumulation of raw intelligence information and maintaining ready access to assorted needles in this haystack is one of the most baffling in the whole field of intelligence management. It is particularly difficult in CIA, where it is necessary to provide community-wide reference services and where no categories of data are excluded from the collection. The problem has been attacked manfully and partial solutions have been achieved; but these solutions have not kept pace with the growing mountain of documents and the sharpened requirements of intelligence analysts. CIA analysts still fall more or less frustrated between the impossibility of keeping adequate personal files and the deficiencies of the central reference service.

It is the purpose of this article to examine the central reference problem critically from the substantive end-user point of view, keeping in mind the intellectual processes and the methodological problems involved in the production of finished intelligence. This is an opportune time for such an examination in view of the new DCID 1/4 creating a permanent IAC Committee on Documentation. The new directive enlarges somewhat upon past community-wide approaches to this problem, and looks toward an integrated community system of compatible individual agency reference services—toward a unified federal intelligence Memory. The Committee will seek to develop appropriate relationships within such an integrated system, so that individual structures may function harmoniously and usefully within the framework of the whole.

The framework to be provided for the federal intelligence Memory would seem to have five theoretical functions:

Function I: *It should integrate the information handling capabilities of all intelligence agencies and other special collections as sub-sets of a federal reference system.* As a corollary of this function, and specifically to facilitate interchange and wide use of all raw intelligence information, the central framework must insure compatibility in the development of information handling systems and equipment within individual agencies.

Function II: *It must insure that raw data from all sources, both open and classified, can easily be brought to bear selectively upon any given substantive problem.* This is the basic requirement from which derive such procedural problems as how to deal with the flow of information from any particular agency or source. The function presents difficult problems in the development of adequate techniques for dealing with current unclassified literature.

Function III: *It should insure that the central reference service is responsive to intelligence priorities, not just to frequency of demand.* Factors underlying such responsiveness include the form of document storage and directness of access thereto; the techniques of search in indirect access and the resulting speed, completeness, and relevance of document retrieval; and the provision of special collections. As a corollary the central framework must provide for placement of document collections and indexing devices within IAC agencies in accordance with needs deriving from their assigned responsibilities and for the maintenance of a central all-source collection in CIA for internal and community use.

Function IV: *It should seek to solve the problem of dealing simultaneously with several high priority requests which require the same files, equipment, or personnel.*

Function V: *It should provide for continuing consultation with users as a basis for improving procedures and should furnish oral and written guidance to users to enable them to employ the facility as effectively as possible.*

In moving from theoretical functions at this level toward more specific management problems within the Memory, there is always a danger of losing the orientation to research and substantive services in favor of procedures and approaches which facilitate internal housekeeping. The discussion which follows will attempt to retain the end-user point of view, and

above all to keep in mind the reasoning and discriminate judgments which go into the development of intelligence products. The present analysis, however, does not get into the important related problems of formulating information requirements or of collecting and evaluating information in the field. Rather it deals with the general problem of facilitating the use of that information which has been sent to the central Memory.

Basic Problems of Information Storage and Retrieval [1]

It is generally recognized that intelligence draws heavily on open-source information, and that the unique element in intelligence research is the careful assimilation of open-source and classified information into a timely, all-source analysis. This requires the systematic treatment of myriad incoming documents, periodicals, and books.

The intelligence reference function is differentiated from ordinary reference services primarily through its servicing of needs for classified documents. Although calls on these documents have in CIA experience constituted considerably less than half the reference requirements, the importance of classified documents to the intelligence officer and to the policy maker is inestimably greater than this proportion would indicate. Their importance derives from a substantive content not available in open sources, from their timeliness, and from the reliability of controlled sources.

The approach to handling these documents is consequently of fundamental importance in a system of information storage and retrieval, but the same logic of approach extends to incoming unclassified materials as well. In both types there are

---

[1] Technical note: Although this article, for simplicity's sake, frequently refers somewhat loosely to "information" storage, it actually means "document" storage. Information storage and retrieval in the technical sense applied to modern electronic computers is in use in some areas of intelligence—SAGE and some aspects of war gaming are examples—and feasible for certain others, but information storage and retrieval in this sense can never fully replace the basic raw intelligence document collections. The reasons are very complex. Suffice it for our purposes here to say that the processes for producing finished intelligence must continue to challenge the sources, to apply consistency tests to fragmentary information in the basic documents, and to apply other varying criteria in order to assess the credibility of the information and to arrange the information into ever more meaningful patterns.

three primary substantive dimensions by which they can be organized—namely, *time, country* (or area), and *functional content* (politics, economics, military subjects, science, etc.) In both types also there are two general difficulties which cannot be overcome completely or adequately by any simple over-all system. One is that the *full* significance of all the content elements cannot be recognized or understood, even under optimum conditions, immediately upon receipt of the document for interpretation or processing. We shall call this the Limited Immediate Recognition Problem; it is discussed further in examining indirect access techniques below. The other difficulty is that any given document may refer to numerous countries or to numerous functional fields of knowledge. This difficulty, which we shall call the Multi-Country/Multi-Function Problem, precludes the development of special collections of all relevant materials for each possible subsequent research project.

It is largely these two general problems—Limited Immediate Recognition and Multi-Country/Multi-Function—which make exclusive reliance on analyst files impossible as an over-all system and make multiple access to a central document file necessary. The Limited Immediate Recognition Problem makes *direct* substantive access to intelligently organized central files of the raw intelligence reports necessary in a system designed to insure maximum utilization of available information. With these considerations in mind let us turn to the maintenance of the federal intelligence Memory and to the three main determinants of its capability: document storage—form and logic; indirect access techniques; and supplementary capabilities and special collections.

Document Storage — Form and Logic

The central reference system, with its huge and ever increasing volume of material, is forced to use some kind of photographic reduction of the hard copy documents it receives,[2] one important device now in use being the aperture card which holds up to ten frames of microfilm. But photoreduction brings the user immediate and immense disadvantages: he

---

[1] The files of individual research offices, in the form of hard copy, must, because of the space they occupy if for no other reason, remain "gem" collections, rather than complete documentation over any long period. The two general problems mentioned above obviously join the conspiracy against completeness in an analyst's file.

often cannot get a recent document from the federal Memory until it has been microfilmed, mounted and coded; under some storage systems he must resort to some index device to identify and locate appropriate documents; he can look at a document only through a viewer or in re-enlargement; he cannot easily compare it with other documents. For these reasons it is clear that exclusive reliance on photoreduction in the Memory tends to restrict utilization of raw intelligence documents. An obvious way out of this difficulty is to adopt parallel hard copy and photoreduced files for the current year, while the documents are most in demand, and discard the hard copy only when it is say one year old.

The logic of the filing arrangement for raw intelligence documents is one of the most critical determinants of the federal intelligence Memory's capability, affecting as it does the amount of material concentrated for direct access by the analyst. The present arrangement of filing raw documents by central acquisition dates within their respective issuing agency series scatters through the entire document collection the associated reports from a given country. The analyst who wants one particular report and knows the issuing agency and number has immediate access to it, at least in theory; all others, whether area or functional specialists, must resort to search by one of the varieties of indirect access indexing techniques, all of which have significant limitations (see below). A filing arrangement by country of origin,[3] on the other hand, with a second breakdown by issuing agency in chronological sequence, would provide direct and immediate access for country specialists and would similarly serve functional specialists in some measure, to the extent that issuing agencies specialize each in its own functional field. Certainly the current hard copy files urged above should have this arrangement, and its advantages would extend also to the photoreductions of older material unless parallel hard copy files are to be maintained indefinitely. An incidental but important characteristic of this system is

---

[3] There are certain applications problems involving unusual characteristics of the various issuing agencies which must be worked out. The main requirements are that each reporting series be kept homogeneous, and that cables and dispatch series be kept separate. The theoretical problem can be best handled by a centrally-designed prefix numbering system covering at least three variables, agency, country (or post), and means of transmission (cable or dispatch).

that it is open-ended and permits the addition of other classified series such as photointelligence as well as of parallel open-source series such as FBIS (rearranged chronologically by country of origin), newspapers, and foreign affairs material from unclassified wire services.

This simple alteration of the primary filing arrangement would thus provide some very important substantive advantages while offering no substantive disadvantages and few if any internal housekeeping disadvantages within the Memory. The new logic largely copes with the Limited Immediate Recognition Problem by permitting, when appropriate, immediate and direct recourse to primary document files for indefinite periods by country of origin and by issuing agency. The Multi-Country/Multi-Function Problem is still with us in a large measure, however, and we shall need some other device unless we wish to rely on the analysts' experience to suggest what other country files should be looked at for a particular research project.

It is primarily to solve the Multi-Country/Multi-Function Problem, and especially to go into the myriad details of certain functional fields such as economics and the military, that indirect access techniques have been devised. It must be recognized at the outset, however, that all such efforts are impaired by the Limited Immediate Recognition Problem in a manner which cannot be fully compensated for in these indirect methods. The indirect methods, nevertheless, are necessary elements of any over-all system designed to overcome the deficiencies which cannot be removed by improving the organization of the primary document file.

Indirect Access Techniques

There are two general types of devices for indirect access: the abstract,[4] which has its conventional meaning of a brief gen-

---

[4] The abstract as a form of indirect access appears to have in isolation rather limited substantive value when applied to raw intelligence documents, which frequently are sketchy, fragmentary, and disjointed. Its most important application is to unified, coherent, journal-type articles as in the Chemical Abstracts series and to finished intelligence studies. In raw intelligence report applications the abstract adds but little to the code form if the latter is applied satisfactorily. Intelligence reporting, however, probably should be standardized to provide an abstract or summary in the first paragraph.

eral summary, and what we shall call the "code form," which indicates the specific categories of information contained in the counterpart document. (These devices may be used singly or in combination, and either or both may be combined with the basic document under some sophisticated systems involving photoreduction.) Both devices can handle the Multi-Country Problem and therefore complement the proposed new file logic for document storage; and the code form is well suited to cope with certain types of detailed functional content as well. Another great advantage of the code form in application to large volumes of material is that it lends itself to machine search.

In theory, machine search rapidly works out the implications of current information-selection instructions on past document classification decisions. Machine search proper enters the process after master code categories have been established and after the content of incoming documents has been matched against these code categories. The over-all system is thus designed to permit the substantive and security classification of incoming documents on a routine basis, so that when an intelligence project is levied the substantive analyst can ideally obtain without delay (Speed Test) a group of documents comprising all those in the system which contain relevant information (Completeness Test) and no document which does not contain relevant information (Relevance Test). Unless the documents are attached to or associated with their counterpart code forms, the research analyst obtains a list of relevant document citations from which he orders retrieval [5] from the document file. There is some tendency toward incompatibility between completeness and relevance—to assure completeness one often must risk some irrelevant documents—and sophisticated systems permit the user to lean in one direction or the other according to his project needs. The greater the number of digits in the classification code, the greater the selectivity for the research analyst and the greater the speed advantage of sorting by machine.

---

[5] The Intellofax system, discussed below, combines the abstract and the code form. After machine search has been completed, a researcher then, on the basis of the counterpart abstracts, has the option of *not* retrieving some documents which machine search found to be relevant. The rationale of inserting this option is not obvious in past applications of the abstract.

Retrospective machine search systems, however, are only as effective as the external human judgments which select the pigeonholes for the incoming documents on the one hand, and as the external judgments which decide what pigeonholes to empty for the analyst's request on the other. All retrospective machine search systems, in fact, have three sensitive points— the master code, the document analysts or coders, and the search instruction writers—which limit the efficiency and reliability of recovery built into the actual searching techniques.

The Intellofax machine search system used by the CIA reference service for handling classified documents has been severely criticized on the ground that it is unreliable, unselective, and costly, and that it is incapable of providing, conveniently if at all, some important services which are desirable in a federal Memory. The unreliability and lack of selectivity stem in a large measure from lack of progress in the initial coding of incoming documents, the notable exception being the adoption of the principle of using one code throughout the intelligence community. This code, the Intelligence Subject Code (ISC), however, lacks a fundamental unifying logic, and has not been adequate to cope with the many new demands levied upon it. It is difficult, if not impossible, to apply the code consistently and accurately because categories have not been defined properly and given items appear in numerous places without adequate cross referencing.

To make matters worse, the organization and staffing of the document analysis sections lack specialization, balance and adequate procedures for assuring high-level analysis in the various intelligence fields. Furthermore, the search instruction writers also lack specialization, and have not been kept fully informed on the coding decisions which were being made by document analysts. Moreover, their substantive decisions on what categories of data to recall have been made unilaterally, without adequate consultation with the research analyst. As a result of these deficiencies, the really conscientious research analyst, in order to be sure he has all the available information bearing on his problem, should theoretically forego the selectivity of the six-digit code and make broad requests at about a two or three digit level; that is, he should deliberately ignore the capability of the search apparatus and use it like a conventional card file.

Vigorous efforts are now under way to develop a more flexible and better balanced machine search system, of necessity a more costly one, and especially one able to cope with the Limited Immediate Recognition Problem which plagues all indirect access techniques.[6] But there is little point in spending huge sums of money to develop and purchase a high machine search capability if this capability cannot be utilized because of a much lower capability elsewhere in the system, namely, in coding documents or in writing search instructions. Large machine search expenditures are rational only if similar effort is made to get comparable quality in the three sensitive spots involving the concomitant human effort: the code, the coders, and the search instruction writers.

The Code. The main principle to be followed in formulating the master code for indexing document content should be to focus to the greatest extent possible on *general categories of observable data* in a manner which obviates the necessity for the coder to blur the classification process through the introduction of personal assumptions. Within the general categories the code should then go to particular sub-categories and modifiers. (Categories should be defined properly, and given data should be either treated in one place in the code rather than scattered about, or adequately cross referenced.)

The search for any general category of documents should yield, along with its family of sub-categories and modifiers, the documents of the unmodified general category for which specific sub-category identifications could not be made when they entered the system. Under this type of coding, highly selective runs would be made into a particularly relevant sub-category or modifier code for the direct evidence. But by Boolean algebraic manipulation, the research analyst can select from within the general category homogeneous categories of knowns and unknowns which bear indirectly upon a problem concerned with the particular sub-category, and this may result in further

---

[6] The Minicard system, for example, combines or associates the code form with the photoreduced counterpart document. This system has not been fully tested in intelligence applications, but it appears to offer unusual flexibility in use and to facilitate the interchange of documents and code forms. As regards files, Minicard could provide the country files recommended above and still permit machine search for specifics within that logic.

identification of some unmodified general category data. Other portions of the code will have to deal with more abstract categories of data.

Centralized Community Coding. The analysis and coding of incoming documents within CIA is at present carried out in four sections which are organized to specialize on types of documents according to issuing agency and which are staffed by personnel having political science or general education backgrounds. In general, a single person analyzes a given incoming document. The analysis is usually reviewed by one other person, but there is no method for assuring that the implications of the given observable phenomenon are coded completely in all relevant functional specializations. There are no economists, military specialists, and physical scientists to recognize and identify data in these fields.

The coding sections should be regrouped, probably under the general guidance of the IAC Committee on Documentation, to provide both functional and area specialization. It is recommended that groups be organized first by functional specialization, for example a political and social section, an economics section, a military section, a physical science section, and perhaps a geography section. Within functional sections there probably should be area specialization, for example an economist for Bloc economies, one for western European economies, one for non-Bloc Asian economies, etc. Within the military section, other specialties could be introduced, for example, experts able to identify information bearing on Soviet missiles and possible missile sites. Briefings should be arranged on various subject problems, particularly those having high intelligence priority. Finally, estimates and gaps-in-intelligence reports from all major IAC research groups should be routed to and discussed within these sections.

Procedurally, every incoming raw intelligence document should be routed to each functional section for analysis, to assure competent examination for implications in all intelligence aspects. This innovation assumes that current batch handling be replaced by discrete handling of individual incoming documents. Its functional orientation could, and I think should, lead to a centralized and highly sensitive coding for the entire IAC to replace the several duplicative operations which individually have limited competence in some fields. The

coding slots could be staffed jointly by CIA and the IAC community in accordance with assigned primary responsibilities. In any case it would be profitable to have select advisory personnel from IAC agencies assigned to the functional sections on a temporary or rotational basis. All these measures serve to restrict significantly the scope in which the Limited Immediate Recognition Problem operates, but clearly they do not eliminate the problem. And, in view of this recognition problem, the general decisions not to code or to photoreduce certain types of documents—the so-called "NODEX" guides—should be carefully reviewed by community users.

Search Instruction Writing. The central reference service has also underestimated the importance of the search instruction writer. This person, usually a trained librarian but understandably insensitive to the *indirect* evidence which bears on specific research problems, is nevertheless making substantive judgments on each such problem which requires reference material, in that he determines what categories of coded data are relevant to it. If he makes this selection unilaterally, his inexpert substantive determination removes responsibility from the research analyst for further data probes. Present Intellofax procedures call for "another look" if *no* documents are recovered on the basis of the first instructions or if a *known* document is not turned up, but in the more typical cases short of these extremes there is no way of assuring that the instruction writer has ordered all or even most of the categories which the research analyst should study.

There should be a reconsideration of the question whether the formulation of the master code used by the document analyst is really adequate in the search instruction phase. Document analysis is primarily the matching problem, resembling inductive reasoning, of subsuming the document content to the master code. In search instruction writing there is primarily the deductive problem of calculating what data bear upon a given research problem. It is therefore possible that two sets of code books would be more effective, a basic one for document analysis and a cross-referenced one for search instruction. The latter might bring together code categories which usually bear upon certain typical and frequent research problems.

Search instruction writers should specialize more than at present and should undergo special training on research methods. Daily current intelligence briefings, as well as reading finished intelligence within their specializations, might be helpful to them. Procedures for keeping them informed of the coding sections' decisions on particular coding problems should receive continuing review. Above all, instruction writers should never make unilateral decisions on what categories of data to search for. The research analyst must be made more familiar with the problems of coding and should participate actively in the formulation of search instructions.

Ideally, for optimum functioning of an indirect access reference system, the research analyst himself should have coded all documents and should write the search instructions for material relevant to his immediate problem. It is only by approaching this ideal more closely, through procedures based on an improved understanding of the formidable communication and comprehension problems involved, that the cost of machine search can be justified. These considerations apply both to Intellofax and to the more complex machine systems under experimental development. (See, for example, footnote 6 above.)

Problems of Political and Military Dynamics. Machine search has its greatest potential value for those documents whose content aspects contain easily defined and recognized logical categories. Economic activities, physical country descriptions (including missile site characteristics), target information, military hard goods, order of battle, biographic information and other broad categories of data can be handled conveniently and with great rapidity by Intellofax or by some other retrospective machine search. (In line with Function II above, machine search can conveniently be extended to include unclassified material relating to selected high priority National Intelligence Objectives.) But these machine systems are inconvenient, if useful at all, for certain other information retrieval requirements.

Especially for political and military dynamics—the delicate tasks of inferring strategies, objectives, and motivations and of identifying and weighing opposing forces—there usually is no substitute for intact chronological files by country and issuing agency. In these pursuits the relevant categories of data are

not fully known, and in addition they can change frequently, perhaps with the demise of a political leader. Moreover, purely economic or purely military data sometimes later acquire critical political meaning, even if only through an implicit threat. Furthermore, there may be very indirect shreds of evidence in the raw documents which suggest new lines of inquiry or which contribute to the testing of hypotheses on the possible strategies of various factions or interests, shreds which seldom can be identified *a priori* for coding purposes but acquire meaning gradually with successive study of preceding and subsequent events. Finally, the machine search system is incomplete; certain types of documents such as FBIS, cables and Weekas are not coded.

In the field of political and military dynamics perhaps more than in others, a further deficiency of the present central reference system is a serious one—delays and gaps in the actual retrieval of documents. If it requires several days or weeks to retrieve or re-enlarge an eight- or ten-month country file, if it requires even two days to furnish prints of a hundred or so documents, if documents received in recent weeks are not made available because they are in photoreduction process, then the area analyst with an immediate need cannot be serviced by machine search, regardless of how well the material may be coded or how wisely the search instructions are written. Intelligence officers with important policy briefing functions simply cannot afford to be kept waiting while the slow, painstaking process of assembling country files takes place. The responsible country analysts must have direct and immediate access to the intact files by country, preferably in hard copy, for which this article pleads.

Supplementary Facilities

The central reference system should be a house of many mansions. It should include, in addition to its reorganized complete file of classified documents, photoreduced and coded for machine search for functional analysts, and its hard-copy file by country of current classified and open-source material for broad political analysts at a country level, a number of supplementary facilities. Some of these are represented in CIA by existing registers and special libraries. For example, the important Industrial Register provides direct access to reports

on numerous Bloc industrial installations. There should be added an improved reference assistance service with substantive competence (area and functional), somewhat analogous to the Legislative Reference Service of the Library of Congress; a complete collection, by country, of the speeches and communiques of political leaders; area source registers; a file of FBIS dailies by country of broadcast origin; and arrangements for making revealed US policy positions on a given country available for quick reference. Finally, it is possible that for selected high priority intelligence objectives, selected unclassified material should be coded for the purpose of achieving the rapid all-source objective cited in theoretical Function II above.

Speeches, communiques, and other position papers by major political leaders theoretically are available in central reference, but access to them requires a tedious search of *NY Times*, FBIS dailies, State and CIA reports, and foreign newspapers. These materials are of such usefulness to national intelligence in showing the evolution of political leaders' public positions that special efforts should be made to make complete files by country available within CIA on a moment's notice. This service, involving routine search through relevant incoming source documents plus nominations by substantive area analysts, would result in a file similar to the present Bloc economic plans collection.

Area source registers should maintain a listing of the publications within or relating to each country, with data on the usual subjects covered in each, its orientation, apparent backers, etc. This file can borrow as appropriate from *The Political Handbook of the World* and from Library of Congress reference facilities and publications. Such a device has considerable potential for filling important data gaps, and would be useful in liaison work with other libraries.

FBIS material can be systematically included within the central reference system by a simple, inexpensive device. Existing FBIS regional dailies could be split up into countries to form new reference volumes containing the accumulation of individual country output over some months. Each new reference volume would comprise two parts, the index and the broadcasts, and each part could be set up on a day-to-day chronological basis. In this form, FBIS would parallel the proposed primary document file according to country of origin and

within that by issuing agency.   Alternatively it could be bound and indexed as a book, but indexers should have an intelligence orientation.

In short, the central reference system should thus develop a combination of machine search, country files, and other features with a goal of achieving balance and flexibility.   The criteria for balance and flexibility are two: the attainment of speeds of reaction which are generally consistent with the intelligence priorities of existing and foreseeable types of projects; and the maintenance of a capability of filling effectively all reasonable requests and needs which are now experienced and those which are likely to have a significant bearing on national intelligence and security within the next five to ten years.   Consideration should be given to the problem of simultaneous high priority requests which make use of the same raw intelligence documents, reference personnel, or other capabilities and to the problem of making the entire community's assets available when appropriate to researchers in any of the IAC agencies.

CIA now has primary responsibility for studies looking toward the assignment of more specific and differentiated responsibilities among IAC agencies for maintaining information storage with rapid search and retrieval capabilities.   It must take the lead in developing a master system to integrate the compatible assigned capabilities of other IAC agencies, as well as those of the Library of Congress and other special collections, as chambers of the federal intelligence Memory.   Especial emphasis should be given to the provision and placement of information handling capabilities—realistically conceived in the perspective of the data and intellectual processes involved—to facilitate the analysis and weighing of factors which tend to upset political equilibria in countries of the Free World or to alter the strategic balance in the world situation.   These capabilities certainly should have the highest information handling priorities in the intelligence community.

This review has been very critical in tone.   The underlying point, however, is not that there are better reference systems elsewhere, that the existing facilities are not of considerable value, or that no progress has been made in the past few years. Rather, the point is that the international situation is moving

into a subtle phase in which the time required to recognize new strategic and tactical developments and assess their implications will become increasingly important. The existing reference facilities are not yet good enough to meet this need.