

SAF-E-258-78
17 November 1978

MEMORANDUM FOR: DD/ODP

STATINTL

FROM : [REDACTED]
Director, Consolidated SAFE Project Office/ODP

SUBJECT : Comments on Centralized Community Bibliographic
and Document Retrieval System

REFERENCE : Memorandum for CIA Member, Intelligence
Information Handling Committee, from
H.C. Eisenbeiss, Director of Central
Reference, dated 19 October 1978

1. These comments relate to Mr. Eisenbeiss' memorandum on this subject. It is my understanding that this concept is to be presented to the Intelligence Information Handling Committee to determine whether there is interest in pursuing it further. At this relatively tentative state of discussion, I believe the reference memorandum is adequate.

2. The staffing and cost estimates appear to be rather gross, and I would be concerned that RECON would handle a vastly expanded work load even as modified. This would require further investigation based on projected volume and usage which is not identified in this proposal.

3. It would appear that a conservative approach to this problem would involve retrieval through intermediaries as a first step to determining the relative worth of providing on-line service. As the on-line retrieval program is developed, I am concerned that no identification of site preparation costs has been made. If this proposal were tied to the SAFE program as an expanded function, it could take advantage of the SAFE site preparation and development activity. If this proposal is pursued, I would be interested in discussing this relationship in depth.

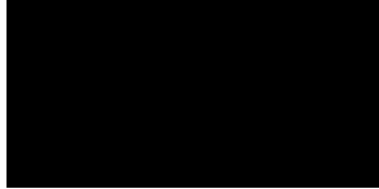
4. Retrieval through ADSTAR of hard copy documents would impose a significant but not identified additional load on the ADSTAR storage system. It would undoubtedly involve some re-design, as well as additional equipment.

Subject: Comments on Centralized Community Bibliographic
and Document Retrieval System

5. If this proposal is pursued, it is imperative that projected usage of the facility be obtained from the interested agencies in order to size the required facility and development adequately.

6. Again, if this program is pursued, I believe that its relationship to SAFE is such that the CSPO and ADSTAR projects should be brought into the planning process.

STATINTL



cc: C/PPAC/CSPO
C/SA/CSPO
COTR/ADSTAR
File

File: Bibliographic
ODP # 8-1919

19 OCT 1978

MEMORANDUM FOR: CIA Member, Intelligence Information Handling Committee

FROM : H. C. Eisenbeiss
Director of Central Reference

SUBJECT : Proposal for a Centralized Community Bibliographic and Document Retrieval System Operated by CIA

1. This memorandum discusses the advantages of adapting CIA's RECON^{1/} retrieval system for intelligence documents to serve as the basis for a centralized bibliographic and document retrieval system to serve all NFIB^{2/} agencies. The memorandum also addresses how such a system could be configured, what services could be provided, how long it would take to implement the system, some tentative estimates as to the possible costs involved and various methods of funding its development and operation. The proposal at this stage is purposefully conceptual and brief, and the cost estimates are extremely conjectural. If you and the other IHC members feel the idea is worth further exploration, additional work by an interagency task force will be required to flesh out exactly how such a system might be brought to reality.

2. The proposed system would be composed of two rather distinct subsystems, namely: a) a bibliographic retrieval subsystem wherein document citations dealing with specific search criteria would be provided to the intelligence analyst, and b) a document retrieval subsystem which would provide the analyst with copies of the relevant document images themselves in either soft copy, paper or microfiche. The system's total cost to the government would be mitigated by the savings it would achieve by making unnecessary certain duplicate and redundant systems in the Intelligence Community.

^{1/} RECON is the on-line version of what is generally referred to as the AEGIS system. AEGIS operates primarily in the batch mode but RECON uses an inverted file technique enabling faster access to the data.

^{2/} Defined as CIA, State/INR, DIA, the Military Service's Intelligence Branches, NSA, Treasury Department, DOE and FBI.

SUBJECT: Proposal for a Centralized Community Bibliographic and Document Retrieval System Operated by CIA

3. Various possible options and means of configuring this system exist, including arrangements involving centralized file creation of both bibliographic and microfilm records combined with decentralized retrieval service (wherein copies of magnetic tapes and the filmed documents would be transmitted on a regular basis to individual agencies for their own use). A number of these options are explored in this paper but not the "centralized/decentralized" approach. Such an arrangement, though technically feasible, is believed to present too many disadvantages in its implementation and operation to warrant further examination.

Why Use RECON?

4. The RECON subject file, from which the proposed Community data base would be derived, has several advantages over other computer-based document indexing systems currently used by NFIB agencies. Initiated in 1968, the RECON file is the largest and most comprehensive subject index to intelligence reports in the Community. As of September 1978 the file contained 3,000,000 index records. RECON offers access to virtually all substantive intelligence documents originated (given general distribution) by the CIA, DoD, DIA, Air Force, Army, Navy, NSA, State, and NPIC, and some documents from other government agencies of the United States [REDACTED]. The data base contains both raw and finished intelligence reports, includes both collateral intelligence and Sensitive Compartmented Information (SCI), and the area coverage is world-wide. Subjects indexed include government, politics, society, culture, science and technology, transportation, communications, business, commerce, industry, finance, commodities (both strategic and non-strategic), products (civilian and military), resources (including labor and military manpower), and the armed forces. In brief, no area of interest to intelligence is overlooked. Open literature, non-CIA cables, and [REDACTED] reporting are included on a selective basis.

5. The full RECON data base is stored in machine-readable form and is searchable by computer via any one or a combination of the elements used to describe each document. These include the bibliographic description (title, issuing agency, post or origin, date, report number, security classification and dissemination restrictions); area codes (China and the Soviet Union are subdivided to the province and oblast level, respectively); specific place names where appropriate; subject codes; and keywords. The 320 subject codes are standardized broad subdivisions, more than one of which can be assigned to any single document by the indexers in CIA's Office of Central Reference (OCR). The keywords are non-standardized terms added by the indexer based on

STATINTL

STATINTL

STATSPEC

SUBJECT: Proposal for a Centralized Community Bibliographic
and Document Retrieval System Operated by CIA

review of the title and document text; these individual keywords supplement the broader subject codes and thus refine the retrievability of each individual document. The flexibility of such an indexing system allows it to easily accommodate new subject indexing requirements.

6. RECON has an historical depth of 10 years and is the most up-to-date general purpose subject index to intelligence documents available. Approximately 85-90 percent of incoming documents are available for computer search of the index records within eight days after receipt, and by July 1979 this figure will be reduced to three days. Portions of the RECON data base are now available to the Community via COINS, and the total data base itself has been queried on a limited basis by OCR analysts for all NFIB agencies continually since its development. When CIA's earlier bibliographic retrieval system, known as "Intellofax," was in operation, then non-CIA use of the CIA index to intelligence reports was about 45 percent of total queries. With the initiation of the AEGIS/RECON system in 1967-68, however, CIA management placed severe limits on other agency access to these bibliographic records because of substantial reductions imposed on CIA resources. Even under this restriction, however, non-CIA use of the data base has crept upward, and during the first half of CY 1978 the entire data base was queried over 800 times by non-CIA NFIB agencies (approximately 26% of total queries during this period). During the same period, the finished intelligence portion of the RECON data base, which is part of the COINS system, was queried via COINS by non-CIA NFIB agencies over 1,200 times.

The Bibliographic Subsystem--Alternative Configurations And Cost Estimates

Option One: Retrieval Through Intermediaries

7. The least costly approach of providing RECON bibliographic records to the Community would simply entail offering increased service from the system in its present configuration to other NFIB members. Under this arrangement, a non-CIA analyst presents his research request in writing or over the phone to an OCR area reference analyst, who queries the RECON data base and then mails the printed listing of records to the original requester.

8. The primary disadvantages of this system are the delays involved in having to mail the request and the document listing. The existence of an intermediary (the OCR area reference analyst) between the end user of the data and the data base itself can also be a disadvantage, but not without some positive aspects. Among the disadvantages, the requester may have no way of knowing how large or small a document

SUBJECT: Proposal for a Centralized Community Bibliographic
and Document Retrieval System Operated by CIA

listing he will be getting until he receives it from the area reference analyst. Any revision of his query to make his request either more inclusive, more selective, or otherwise more appropriate for retrieving precisely what he needs can only be made after the query has been run and the complete document listing is received through the mail. On the positive side, the intermediary reference analyst usually has a better knowledge than the requester of the subject indexing codes and keywords (including how they have been used), and he can often translate the requester's needs into a more effectively worded query than if the requester is left to his own devices.

9. The following costs are foreseen if the current system of Community access to RECON is simply expanded. About 8-10 more document indexers and dissemination personnel would be needed to process the additional material expected to be added to the data base, in addition to indexing certain categories of documents in greater depth to satisfy the anticipated specific needs of various agencies. An additional typist would be necessary for the added input to the data base. Two additional camera operators would be needed in OCR's Microform Processing Branch to handle the increased volume of incoming documents to be filmed. Fifteen more area reference analysts would be needed to handle the added volume of requests^{1/}. At least two more clerks would be needed to address and package listings for mailing and to prepare document and courier receipts. An additional direct access storage unit would have to be leased in order to store the greater number of document citations in the data base. No additional computer equipment, software, personnel or floor space would be required. These operating expenses would probably total more than \$500,000 per year. (See the attached table for a summary of all cost estimates.)

Option Two: Direct On-Line Retrieval

10. If CIA's RECON data base is to be made available to all other NFIB agencies, there is a preferred alternative to merely expanding the operation described above. This would be to provide on-line access to the data base (stored at CIA Headquarters) via remote visual display terminals (VDTs) in other agencies. Such access could be made available

^{1/} It is extremely difficult to accurately estimate the number of index search requests that would be levied on CIA if RECON were made available to the Community without restriction. However, for the purposes of this memo, it is assumed that the current level of requests would increase five-fold. (This figure is largely a guess, based partly on OCR's experience with non-CIA requesters before controls were imposed on their use of the RECON data base.)

SUBJECT: Proposal for a Centralized Community Bibliographic
and Document Retrieval System Operated by CIA

on a 24-hour/day basis if necessary. Bibliographic references displayed on these remote VDTs could be printed immediately on medium-speed (300 lines/minute) printers co-located at each VDT. In this connection it should be pointed out that since the fall of 1973 a variety of intelligence analysts in CIA have been successfully querying the entire RECON data base directly via the SAFE Interim System^{1/} remote VDTs without OCR intervention. These analysts were formally trained to search the data base and are provided with guidance when necessary.

11. The principal advantages of this arrangement include the significantly faster availability of the document citations to the analyst, plus the capability for the analyst to work directly with the data base. The latter feature would enable the analyst to determine if the subject codes and keywords he had chosen were producing references to the kinds of documents he needed; he could also see how large his document listing would be and modify his query parameters if necessary. All this could be done before ordering a printout from the system. For standing requests for index searches the capability to query the data base via the batch mode would be retained, rather than requiring the analyst to repeatedly compose his query at a terminal.

12. If the on-line arrangement outlined is adopted, existing data communications systems such as the COINS network should be able to handle the transmission of the RECON bibliographic records from CIA Headquarters to requester terminals located at other NFIB agencies. Assuming that the COINS network were used, the following tasks would have to be undertaken. A dedicated host computer would have to be installed and the RECON system software would have to be modified to make the computer program "reentrant," an arrangement enabling the central processing unit to handle up to 50 on-line requesters simultaneously. This would entail a one-time payment to a contractor, and would require approximately three man-years of his work and one calendar-year of time. An extra programmer and technician would each be needed in OCR's computer support unit to work with the contractor during the software modification and later to maintain this software and troubleshoot the system's operation.

13. In addition to making the host computer operational for RECON, a number of other tasks would be required. The software interfaces connecting the computer, the message processor, and the COINS network would have to be developed. Certain additional software and hardware changes would be needed to adapt the RECON system to accommodate

^{1/} This is the precursor of the ultimate SAFE system, designed to assist in all aspects of intelligence production.

SUBJECT: Proposal for a Centralized Community Bibliographic
and Document Retrieval System Operated by CIA

an increased number of users. Also, some combination of software modifications and human intervention may be required to resolve security release problems. If all the necessary equipment were bought outright, the investment expenses are estimated to be about \$2,700,000.

14. If the necessary equipment were rented instead of purchased outright, its cost is estimated at about \$780,000 per year, including maintenance.

15. The annual operating costs would include an additional computer programmer, a computer technician, and three more computer operators, plus higher equipment maintenance costs. The total of these operating costs is estimated to be about \$175,000 per year.

16. In addition to the extra personnel--including indexers and microphotographers--already mentioned, a centralized staff of about three or four people (\$60-80,000/year) would probably be necessary to coordinate new indexing requirements from participating agencies; to train personnel to use the system and to provide on-going guidance once the system enters operation; and to handle trouble calls and transmit questions to appropriate operating personnel.

The Document Retrieval Subsystem--Alternative Configurations
And Cost Estimates

17. If a centralized document retrieval service in CIA is envisaged to supplement the centralized bibliographic retrieval service, then the CIA's current document retrieval system would have to be significantly enhanced to accommodate the increased work load. The system as it now operates is capable only of handling the present request load. For this reason future requests for copies of documents, whether generated by either of the bibliographic retrieval options discussed above, would have to await implementation of the CIA's Automated Document Storage and Retrieval (ADSTAR) system, scheduled to enter operation within CIA in November 1979. Like the bibliographic retrieval system discussed above, the ADSTAR document retrieval system could operate in either a batch or on-line mode. In either mode, ADSTAR employs digitized images in its document retrieval and display processing, and present plans call for transmitting such document images directly to CIA user analysts at their remote locations over an upgraded communications network implemented as part of the SAFE system.

Option One: Batch Mode

18. Under this configuration the ADSTAR system within CIA would produce copies of documents after receiving a request for them either

SUBJECT: Proposal for a Centralized Community Bibliographic
and Document Retrieval System Operated by CIA

requesters to query the CIA's central ADSTAR library and display the text and print hard copies of whichever documents the NFIB analyst selected from his RECON listing.

22. Such an on-line document retrieval system, however, could not be developed on the basis of existing data communications systems, such as the COINS network. This is because the bandwidth capacity to handle ADSTAR document image transmissions, which consist of approximately four million bytes per page image, is not available in existing Community networks. The data transmission problem could be eased somewhat by using advanced data compression techniques, but even such a compressed data transmission would require an estimated one million bytes per page image.

23. Development of such an on-line document retrieval system, compared to the ADSTAR batch mode, would require additional outlays for a central processing unit of greater capacity, more software, and (most importantly) the communications system hardware; the latter would include the communication lines themselves as well as the interface equipment, encryptors, decryptors, and remote access and display stations. Also, as with the on-line bibliographic retrieval system, appropriate measures would have to be taken to handle security release problems before this system is implemented. We cannot estimate the total of these additional costs without tasking communications specialists to undertake a study of the problem, but undoubtedly the costs would be substantial.

Funding

24. Funding could be accomplished in at least four different ways, each of which has its advantages and disadvantages. One possible method involves user agencies supplying personnel to CIA according to a ratio proportionate to the additional input burdens each agency would impose on the RECON system plus the use each agency made of the system. This method has been used between CIA and NSA for reference support under Project Millstream. Its applicability when a number of agencies are concerned, however, is questionable. There is the problem of allocation of manpower compensation from individual agencies whose costs to the system are fractions of many years. There are also the problems attendant with periodic replacement of personnel and with the loss of control by CIA in applying its own personnel selection procedures and standards to all of the people working in the CIA.

25. A second alternative would be to have user agencies transfer funds to the CIA to pay for their portion of the input and use made of the RECON/ADSTAR system. This would be similar to an arrangement during the 1950's and early 1960's between the State Department and the CIA,

SUBJECT: Proposal for a Centralized Community Bibliographic and Document Retrieval System Operated by CIA

whereby the latter transferred funds to the State Department to pay for the CIA's use of State Department biographic files. This approach is easier to arrange and manage than the transfer of personnel, but is complicated by the situation in which a number of agencies must defend a portion of their budgets that are allocated to a program run by another agency. Furthermore, this alternative does not address the question of personnel, so a situation could arise in which the CIA had enough money, but had not been authorized enough additional slots for the people needed to operate the system.

26. A third way would be to have those developing and operating costs of the system that are associated with Community service (including the additional positions required) made part of the budget of the Intelligence Information Handling Committee (IHC) and to charge the IHC with defending this portion of its budget each year before Congress. A peculiarity associated with this arrangement would be that the investment and operating funds for an essentially integrated system would have to be split between two budgetary sources, and potential complications could develop if differing budgetary priorities ever arose between the IHC and the CIA.

27. The fourth possible method would be to increase CIA/OCR's budget to allow it to finance the development and operation of the system itself. Such a proposal was made by OCR as an "enhanced" option in its FY 1980 program call, but it was rejected. If adopted, however, it would have the advantage of administrative simplicity and would avoid any complications arising from splitting the source of funds for developing and operating the system among different organizations.

Time Required for Implementation

28. Any planned expansion of the CIA's bibliographic and document retrieval system would require a thorough and detailed study of at least six months' duration, plus time to hire whatever additional personnel the study will have called for.

29. The maximum Community-wide service that could then be implemented would be batch bibliographic retrieval via OCR area reference analysts, with document retrieval accomplished through each NFIB agency's own document library. This arrangement could be set up as soon as additional service personnel were hired, possibly as early as six months after completion of the initial six-month preliminary study, assuming that the requisite floor space could be acquired.

SUBJECT: Proposal for a Centralized Community Bibliographic
and Document Retrieval System Operated by CIA

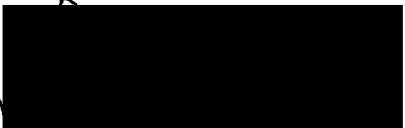
30. The more advanced approach of providing on-line bibliographic access would probably require at least two years after completion of the initial six-month study. During this period, software modifications would have to be accomplished, additional equipment would have to be acquired and installed, and non-CIA agencies would have to program their budgets for the communications equipment and remote terminals they must fund.

31. Centralized document retrieval would be impossible for the CIA in either a batch or on-line configuration until after the ADSTAR system had been implemented and operationally tested for at least six months. This would make ADSTAR available for Community-wide use no earlier than June 1980, and then only for batch retrieval.

32. An on-line ADSTAR system that serviced non-CIA agencies via remote work stations would take at least two years for programming user-agency budgets, and acquiring and installing the necessary additional equipment.

Unexplored Issues

33. The foregoing examines some basic considerations regarding the establishment of a centralized bibliographic and document retrieval system. If the IHC feels this proposal is worth pursuing, then the questions of user requirements, system architecture, and precise investment and operating costs would all have to be thoroughly researched. In addition, other unresolved issues relating to these and other aspects of the system would have to be studied in detail. These include security arrangements, floor space for machines and people, and the cost and funding of communication lines, printers, remote terminals and other equipment at participating agencies. Finally, we would want to examine what savings such a system would provide within the Community, either by reducing on-going activities or planned new ventures necessitating substantial expenditures in labor and hardware for systems now in the design stage.


H. C. Eisenbeiss

STATINTL

Attachment:
As stated

Option 1 - Retrieval Through Intermediaries

	One-time Costs	Annual Costs
Hardware		\$ 24,000
Staffing		500,000
ADSTAR Costs	\$1,000,000	150,000
	<hr/>	<hr/>
		\$ 674,000
	\$1,000,000 ÷ 5 =	200,000*
		<hr/>
TOTAL		\$ 874,000

* Pro rata annual share of initial one-time costs, assuming a system life of five years.

Option 2 - Direct On-Line Retrieval

	<u>Purchase</u>		<u>Lease</u>	
	One-time Costs	Annual Costs	One-time Costs	Annual Costs
Hardware	\$2,700,000			\$ 780,000
Maintenance		\$ 70,000		
Software Modification	500,000		\$ 500,000	
Staffing		755,000		755,000
ADSTAR Costs	1,000,000	150,000	1,000,000	150,000
	<hr/>	<hr/>	<hr/>	<hr/>
TOTAL	\$4,200,000 ÷ 5 =	\$ 975,000 840,000* <hr/> \$1,815,000	\$1,500,000 ÷ 5 =	\$1,685,000 300,000* <hr/> \$1,985,000

* Pro rata annual share of initial one-time costs, assuming a system life of five years.

SENDER WILL CHECK CLASSIFICATION TOP AND BOTTOM

UNCLASSIFIED	CONFIDENTIAL	SECRET
--------------	--------------	--------

OFFICIAL ROUTING SLIP

TO	NAME AND ADDRESS	DATE	INITIALS
1	Mr. Danny May 2000 Hqs		EM
2	EO	2008/28	GD
3	DD		ADJ
4	D/ODP	10/31/78	CM
5			
6			

<input type="checkbox"/> ACTION	<input type="checkbox"/> DIRECT REPLY	<input type="checkbox"/> PREPARE REPLY
<input type="checkbox"/> APPROVAL	<input type="checkbox"/> DISPATCH	<input type="checkbox"/> RECOMMENDATION
<input type="checkbox"/> COMMENT	<input type="checkbox"/> FILE	<input type="checkbox"/> RETURN
<input type="checkbox"/> CONCURRENCE	<input type="checkbox"/> INFORMATION	<input type="checkbox"/> SIGNATURE

Remarks:

Copy to: C/MS } Done 10/20.
 C/SPS } (E)

STATINTL

~~Is this a response to a [redacted] question?~~

In reply to request from D/ODP G.D

Copy to: DD/ODP } for [redacted]
 D/ODP } Done
 11/1/78. (E)

DD/ODP - Action
 Done 1/2/79

FOLD HERE TO RETURN TO SENDER

FROM: NAME, ADDRESS AND PHONE NO.	DATE
OD/OCR, 2E60 Hqs	

UNCLASSIFIED	CONFIDENTIAL	SECRET
--------------	--------------	--------