ODP-8-2184
25 JAN 1979

MEMORANDUM FOR: Chairman, DCI Intelligence Information Handling Committee

FROM : Clifford D. May, Jr.
CIA Member, IHC

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

1. Proposal: This memorandum proposes that Intelligence Information Handling Committee study the feasibility and desirability of adopting CIA's RECON bibliographic index and ADSTAR micrographic document storage and retrieval system as a Centralized Intelligence Community Bibliographic and Document Retrieval System, managed and operated for the Community by CIA.

2. Background: a. 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] STATINTL
The data base contains both raw and finished intelligence reports, includes both collateral intelligence and Sensitive Compartmented Information (SCI), and the area coverage is worldwide. 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 STATSPEC included on a selective basis.

b. 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 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.

c. 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.

d. Bibliographic services must be supplemented by document retrieval capabilities. To ensure speedy and efficient retrieval, CIA is building an Automated Document Storage and Retrieval (ADSTAR) System, which is scheduled to enter operation in November 1979. Designed to operate either in batch or online mode, ADSTAR will store documents on microfilm but digitize these images for transmission over broad-band communications links to remote display terminals and printers.

3. Community Options for Bibliographic Service:

a. Offline Service

(1) 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.

(2) The primary disadvantages of this system are the delays involved in having to mail the request and 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 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.

b. Direct Online Service

(1) 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 online access to the data base (stored at CIA Headquarters) via remote visual display terminals (VDTs) in other agencies. Such access could be made available 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.

(2) 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.

(3) If the online 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.

---

[1]This is the precursor of the ultimate SAFE system, designed to assist in all aspects of intelligence production.

c. Online Service through Intermediaries

(1) Somewhere between options a. and b. above would be a system in which community customers would be linked to OCR's area reference analysts in a network of computer terminals. Queries would be presented telephonically or via the computer terminal, and the results of the analysts' online search could be displayed on the requester's terminal.

(2) The advantages of this blend of services are clear and have to do with effective, real-time communications between the area reference analyst and his customer. Questions about individual bibliographic references can be answered and the document listing tailored to the customer's needs. The refined listing could then be printed at the customer's printer as in option b.

4. Community Options for Document Retrieval Service:

a. Batch Mode

Under this configuration the CIA ADSTAR system would produce copies of documents after receiving requests either in writing or by computer terminal command, depending upon which form of bibliographic service has been adopted. The documents would be mailed to the requester.

b. Direct Online Retrieval

(1) In its most sophisticated configuration, remote ADSTAR terminals located throughout the Intelligence Community would allow non-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.

(2) Such an online 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 consists 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.

5. Costs:

a. Any expansion of RECON services will require a major redesign of the data base. This redesign, to remove Input/Output bottlenecks and to render RECON capable of responding efficiently to larger online system requirements, would cost an estimated $250,000, plus annual maintenance of $100,000. These costs are basic and will be incurred if any major increase in the use of RECON is planned, whichever options are adopted.

b. If option 3.a. is adopted, about ten 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.2 At least two more clerks would be needed to address and package listings for mailing and to prepare document and courier receipts. Two additional direct access storage units (one primary and one backup) and one channel address unit would have to be purchased at a cost of $175,000 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. Operating expenses would probably approximate $600,000 per year.

c. If option 3.b. is adopted (and existing communications systems are used), about half of the operating expenses cited in para. 5.b. above would be avoided, for the 15 area reference analysts would not be needed. A large, dedicated host computer would have to be installed, however, at a cost close to $4 million. System software would have to be modified to make the computer program "reentrant," an arrangement enabling the central processing unit to handle

[2]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 purpose 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 customers before controls were imposed on their use of the RECON data base.)

up to 50 online 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. 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 an increased number of users. Also, some combination of software modifications and human intervention may be required to resolve security release problems. Total cost for this effort would approximate $500,000.

d. To house the host computer approximately 2,500 square feet of computer-grade floor space would be required, and ten positions would be needed for the personnel to operate the computer in a stand-alone environment that is electrically isolated from CIA's other computer facilities. The annual operating costs would include an additional computer programmer, and a computer technician, plus higher equipment maintenance costs. The total of these operating costs is estimated to be about $220,000 per year for personnel and $120,000 for maintenance.

e. 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.

f. Option 3.c. would avoid the costs related to the installation and operation of a host computer and the attendant software development costs referred to in para. c. above, but the use of computer terminals to deliver bibliographic information would entail careful systems design and probably the acquisition of a number of "smart" terminals for use by OCR's analysts, terminals with the

ability to store information received from RECON and to deliver it on command to the remote customer terminal, which, in this configuration, would not have direct access to the CIA computer housing the RECON data above. Cost figures for such a system cannot be developed without a major study, but the costs should be significantly lower than those associated with the stand-alone host computer.

g. The costs of Document Retrieval Service Option 4.a. can also be separated into investment and operating expenses. An ADSTAR system augmented to provide Community-wide service would require approximately eight more storage modules to accommodate the assumed 25 percent increase in the number of documents five years old or less that are to be stored in that portion of the system designed to provide immediate retrieval. (These need not be added all at once; two per year could probably take care of the expected annual ADSTAR file growth.) Larger central processing units would be needed to accommodate the greater number of index records and associated support files. For the same reasons more disk packs and disk drives would be needed, the buffer capacity would have to be doubled and at least one other high-speed printer would have to be acquired. If this new centralized document service were to result in a demand for more documents in microfiche, the microfiche output capability would have to be greatly enhanced. Finally, software modifications to the ADSTAR system would be needed. These would all be one-time investment costs, and, while extremely conjectural, would probably total over $1,000,000.

h. The increased operating costs anticipated for an expanded ADSTAR system would include two additional personnel to intervene in the ADSTAR process to resolve document release questions. Two extra clericals would be needed for packaging, mailing, and preparing document and courier receipts for batch requests for documents. Maintaining the various expanded support files (e.g., MIS and Security Access) would require another full-time employee. For preventive maintenance of the additional equipment, the maintenance contract would cost more. These operating costs would probably come to about $150,000 per year.

i. Direct Online Retrieval, as in Option 4.b., would require additional outlays of $750,000 for a central processing unit of greater capacity and associated support equipment, plus $750,000 for more software, and (most importantly) the communications system hardware; the latter would include the communication lines themselves as well as the interface equipment, cryptographic systems, and remote access and display stations. Also, as with the online 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 system study, but undoubtedly the costs would be substantial.

j. It must be emphasized that the various costs described above are only preliminary estimates, subject to change. They are summarized in the tables attached to this memorandum.

6. Funding: There are no resources in the CIA Program for enhancement of our bibliographic index and document storage and retrieval capabilities beyond our immediate needs. If, after its study, the IHC validates a requirement to provide RECON and/or ADSTAR capabilities to other Community agencies and tasks CIA with the development, implementation, operation, and/or maintenance of these enhancements, then the IHC and the Resource Management Staff will have to identify the necessary resources. The resources required to expand and upgrade the existing system to serve the needs of other Community agencies should be provided by those agencies.

7. Time Required for Implementation: a. 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.

b. Off-line bibliographic service (option 3.a.) could be implemented 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.

c. The more advanced approach of providing online bibliographic access (option 3.b.) 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. About the same time would be required to implement a system of online service through Intermediaries using a network of computer terminals (option 3.c.).

d. Centralized document retrieval would be impossible for the CIA until after the ADSTAR system had been imple mented 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 (option 4.a.).

e. An online ADSTAR system that serviced non-CIA agencies via remote work stations (option 4.b.) would take at least two more years for programming user-agency budgets, and acquiring and installing the necessary additional equip-ment. FY 1982 would be a conservative target date.

8. Recommendation: a. We recommend that the IHC sponsor a study in depth of the Community's bibliographic and document retrieval needs to determine whether centralized services of the kinds described above would serve the Communi-ty's interests. The study should emphasize user requirements, system architecture (including communications), and precise investment and operating costs, together with offsetting savings to be made by reducing on-going activities or planned new ventures for which substantial expenditures are planned. Other aspects of the proposal which need research are the security restrictions to be imposed, and floor space requirements for machines and people.

b. If this study demonstrates that centralized services are desireable and economical, we recommend the adoption of RECON and ADSTAR in whichever of the configurations described above most effectively meets the needs of the Community, provided a suitable answer can be found to the questions of manning and funding the Community support.

[redacted] STATINTL

Clifford D. May, Jr.

Att: a/s

Distribution:
Original - Addressee, w/att. 1-D/NFAC 1-Comptroller
1 - ODP Registry, w/att. 1-D/OCR 1-DD/P/ODP
2 - O/D/ODP, w/att. 1-C/SG/IMS/DDO
1-[redacted] IHC/IMS Rep.

STATINTL STATINTL

O/D/ODP/BJohnson:caj/5 Dec. 78/retyped: 24 Jan. 79

1-[redacted] 10/16/79

PRELIMINARY ESTIMATES OF COSTS OF COMMUNITY DOCUMENT RETRIEVAL SYSTEM

Requirement	Option 4.a. Positions	Option 4.a. One-Time	Option 4.a. Recurring	Option 4.b. Positions	Option 4.b. One-Time	Option 4.b. Recurring
Hardware (storage modules, CPU, disk drives, buffer, printer and software		1,000,000+			1,000,000+	
Maintenance			150,000			150,000
Document Release Control	2		40,000	2		40,000
Clerical Service	2		25,000			
Files Support	1		20,000	1		20,000
Additional ADSTAR Hardware, Software					1,500,000	100,000
Communications				Unknown	Unknown	Unknown
Sub-Totals	5	1,000,000	235,000	3	2,500,000	310,000
Total Annual Cost Assuming 5-Year System Life			200,000* $435,000			500,000* $810,000

*Annual figures represent 1/5 of the one-time totals shown in preceding column.

PRELIMINARY ESTIMATES OF COSTS OF COMMUNITY BIBLIOGRAPHIC SYSTEM

(RECON)

Requirement	Option 3.a. Positions	Option 3.a. One-Time	Option 3.a. Recurring	Option 3.b. Positions	Option 3.b. One-Time	Option 3.b. Recurring	Option 3.c. Positions	Option 3.c. One-Time	Option 3.c. Recurring
Redesign RECON		250,000	100,000		250,000	100,000		250,000	100,000
Bibliographic Service									
Off-line									
- 13 Index/Dissem/Clerical, 2 Camera Op., 15 Area Reference Analysts	30		600,000	15		300,000	30		600,000
- Add. Direct Access Storage Unit		175,000			175,000			175,000	
On-line (Direct)									
- Host Computer					3,200,000*				
- Software					500,000				
- 10 Operators, 1 Tech, 1 Systems Analyst, 3 Requirements Coord.				15		280,000			
- Operating Costs						120,000			
On-line (Intermediary)									
- Smart Terminals								250,000	
- Software								250,000	
Sub-Totals	30	425,000	700,000	30	4,125,000*	800,000	30	925,000	700,000
Total Annual Cost Assuming 5-Year System Life			85,000** $785,000			825,000** $1,625,000			185,000** $885,000

*Plus 2500 sq. ft. of floor space.

**Annual figures represent 1/5 of the one-time cost in preceding column.

☐ UNCLASSIFIED ☐ INTERNAL USE ONLY ☐ CONFIDENTIAL ☐ SECRET

ROUTING AND RECORD SHEET

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

FROM:
Clifford D. May, Jr.
2D00 HQS

EXTENSION [redacted]

NO.
ODP-8-2184

DATE
25 January 1979

STATINTL

TO: (Officer, designation, room number, and building)	DATE RECEIVED	DATE FORWARDED	OFFICER'S INITIALS	COMMENTS (Number each comment to show from whom to whom. Draw a line across column after each comment.)
1. STATINTL [redacted] 1W02 CHB				George: I finally was able to move the Centralized Community Bibliographic and Document Retrieval proposal. I would like to have time on the next IHC agenda for Harry Eisenbeiss and I to explain our pro-posal. STATINTL [redacted] C. D. May, Jr.
2.				
3.				
4.				
5.				
6.				
7.				
8.				
9.				
10.				
11.				
12.				
13.				
14.				
15.				

FORM 3-62 610 USE PREVIOUS EDITIONS ☐ SECRET ☐ CONFIDENTIAL ☐ INTERNAL USE ONLY ☐ UNCLASSIFIED