CONFIDENTIAL

3 NOV 1978

MEMORANDUM FOR: Deputy Director for Processing
                  Chief, Special Projects Staff

25X1 FROM   :   [   ]
                         Deputy Director of Data Processing

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

REFERENCE   :   OCR memo, same subject, dated 19 October 1978

1. Attached is a copy of a proposal, submitted by the Director of Central Reference, that CIA offer to the Intelligence Community a centralized bibliographic and document retrieval system based on AEGIS. The proposal would demonstrate to the Information Handling Committee and the Community at large that CIA is not as unwilling as many believe to participate in COINS and other efforts to create a true community data base. Also attached is a statistical trend report showing among other things the number of terminal queries handled by OCR's Central Libraries Division in Fiscal 1978. Accompanying the statistical report is a copy of a note from D/OCR to D/ODP explaining a few of the entries on the report.

2. Before submitting the proposal to the IHC, we need to evaluate carefully what impact this proposal would have on ODP. We also need to satisfy ourselves that OCR's estimates regarding cost and time needed for implementation are reasonable and defensable.

3. I have been asked to put together the IHC presentation. Before I do so I would appreciate receiving your comments on the proposal's impact and on its time and cost estimates. I would also like your views on a question raised by D/ODP about the possibility that a third option may be available to us. This option, falling somewhere between OCR's options 1 and 2, would involve replacing
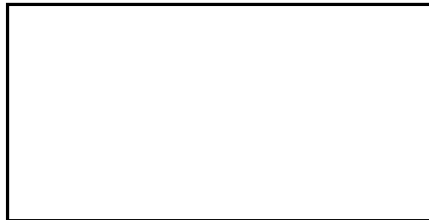
25X1

CONFIDENTIAL

written or telephonic requests with electrical (voice or narrative) two way communications between the requestor and the intermediary (the OCR area reference analyst), permitting the results of the query to be transmitted to the requestor electrically.

4. I will be sharing the OCR memorandum with the Office of Communications and eliciting comments from them about the communications implications of such a proposal.

5. It would be helpful to me if I could have your comments by 17 November.

25X1

Atts: a/s

Distribution:
   1 - ea. Addressee
   2 - O/D/ODP

25X1   O/D/ODP/[   ]:aj/3 November 1978

CONFIDENTIAL

14 November 1978

MEMORANDUM FOR: Deputy Director of Data Processing

STATINTL FROM   :

Deputy Director for Processing

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

REFERENCE   : OCR Memo, dtd 19 Oct. 1978, Same Subject

STATINTL

1.   The attached memorandum speaks to the OCR proposal
outlined in the referenced document on a paragraph by para-
graph basis. [   ] comments are valid and they
highlight some of the real problems.

2.   As a general statement, however, the OCR proposal
does not fully address the problems that would be faced by
ODP in providing such a service.

3.   Any expansion in RECON services would require a
major redesign of the data base, since it is too large for
the present RECON system. This will be expensive. Option 1
assumes no software cost and very low hardware costs. A
better guess at this point would be about $250,000 to remove
all I/O bottlenecks from the existing system so as to make
better utilization of the existing 370/168 CPU, and an
annual maintenance cost of $100,000.

4.   Option 2 will require dedicated support from
Engineering Division, Systems Programming Division and
Operations Division. This is a CAMS-like approach and would
be expensive. Ten positions (two systems programmers and
eight operators) would be needed to support this operation
in a stand-alone environment that is electrically isolated
from our present facilities. Approximately 2,500 square
feet of computer-grade floor space would also be required.

5.   Without any estimate of the transaction loading
and the number of terminals to be supported, we must be
cautious in selecting the appropriate CPU. It may well be
that a 158 size machine is not adequate for a dedicated
environment. In the case of CAMS, a 360/65 was determined
to be drastic overkill initially. Once users got some
experience, CAMS requirements exploded. The cost for a
dedicated CPU in option two is probably closer to 4 million
dollars that the estimate of 2.7 million.

STATINTL

Att: a/s

STATINTL cc: [   ] w/att.

8 November 1978

MEMORANDUM FOR: Deputy Director for Processing, ODP

THROUGH   : Chief, Engineering Division, F/ODP

FROM   :
                 Chief, SEB/ED/F/ODP

SUBJECT   : OCR Proposal for a Community-wide
                 RECON/AEGIS System

        1.   This memorandum documents my thoughts on the
implications, for ODP, of the OCR proposal to offer a
RECON/AEGIS type service to all of the intelligence
community. I found the proposal itself to be quite readable
and well organized. The level of detail is sufficient to
make the case for a feasibility study without inundating the
reader with more complexity than is necessary. In general
the ideas and concerns that occured to me in reviewing this
paper are as follows:

    o I think that some of the communications problems that
       may arise are not emphasized sufficiently. (See
       paragraphs 6 and 7 of this memo.)

    o The 158-Blue CPU might be a candidate for this
       application if bibliographiic option 2 is adopted. (See
       paragraph 6 of this memo.)

    o From the Processing standpoint, the implications of
       bibliographic option 2 are considerable. Floor space,
       peripheral equipment funding and operational staffing
       will all present problems even if a CPU is available.

        2.   Succeeding sections of this memorandum provide
more detail on specific parts of the proposal. Section
headings correspond to those in the OCR proposal itself. My
thoughts on Mr. May's suggestion for a third option appear
in paragraph 8.

Why Use RECON?

3.   Paragraphs 4, 5 and 6 of the OCR proposal make a good case for the adoption of RECON for this application. My concern is that we do not underestimate the cost of re-design to make this program capable of responding to large on-line system requirements. In addition to making the code re-entrant so that it can handle larger terminal loads, care will have to be taken if data base bottlenecks are to be avoided.

4.   A possibility we might consider, if a considerable re-write is involved, is to recode RECON to run under CICS. The advantages of this approach include:

o The problem of terminal communication interface will be solved with a generalized system with which ODP already has some expertise.

o Tuning, performance evaluation and debugging tools will not have to be developed since they already exist for CICS.

o Program development should be facilitated since CICS will provide many system services and will enforce a relatively standardized coding format on the developer.

o Finally, this will be a step in what should be an ODP standardization goal for all on-line systems. Ideally ODP would support only three on-line systems, VM to meet all time-sharing requirements, GIMS for large scale data base management applications and a CICS environment to support those customers who need special purpose code to access their data bases. ODP could support the CICS environment and the customer would be responsible for the application dependent code run under it.

The Bibliographic System (Option 1)

5.   Although no additional hardware requirement is anticipated for this alternative beyond the one extra disk drive for data base storage, we can expect some performance problems if the RECON workload is significantly increased. Current OCR utilization of the 370/168-1 is about 30% during peak periods. Utilization much beyond this level cannot be realized without removing current I/O bottlenecks. Removing these bottlenecks may require additional channel and disk controller capacity as well as more disk space for this service.

The_Bibliographic_System_(Option_2)

6.   This option, involving non-agency, on-line access
to ODP computer facilities, would present the same problems
as our present CAMS service. The most significant problems
from Processing's point of view are:

o Where do we house an electrically isolated CPU and how
  do we back it up?

STATINTL

o What   kind   of   communication facility do   we   use? A
  conversation with [   ] leads me to believe that the
  COINS network may not be the best option. COINS is also
  considering an ARPA-like follow-on network which would
  also impact our plans.

o Assuming that we could use the 158-Blue for this
  application, after the CAMS backup CPU replaces it in
  FY-81, are we inviting another CAMS-like problem of
  escalating demand?

7.   My first-cut estimate of the best way for us to
implement an option 2 type service would be to put the CPU
in the Ruffing Center with an isolated COMTEN front-end.
Backup would be limited ·to Agency users by moving the
application to another Ruffing Center CPU and switching CIA
user terminals to the internal Agency front-end network.

The_Bibliographic_System_(Option_3)

8.   From Processing's point of view, Mr. May's
suggested alternative has the advantage of simplifying our
communication and security problems since all on-line access
will be within the Agency. It also provides end-user access
to expert assistance and represents a considerable saving
vis-a-vis option 2.

9.   The primary disadvantage that I can see is the
potential for a number of human engineering problems. Many
people do not work well in a telephone interface environment
and   this system   would   undoubtedly require   lengthly
conversational or   teletype sessions. The   problems of
transmitting relatively large amounts of data or operating
in a slow response mode could make this approach unfeasible.

The Document Retrieval System

10.   I'm not familiar enough with the ADSTAR project to make any significant comment on this part of the proposal. The arguments associated with the off-line and on-line options seem reasonable to me. The difficulties associated with the on-line communication volume should, I would think, make this option prohibitively expensive.

FUNDING

11.   In regard to funding the two basic bibliographic retrieval options, my feelings are as follows:

o The off-line alternatives (option 1 or Mr. May's option 3), using the current OCR service machine, would require more than a one disk drive equipment enhancement at $24,000 per year. Any significant increase in the workload imposed on the RECON program may well require re-coding of the software and additional I/O resources to guarantee an acceptable level of system response. $100,000 per year might be a better estimate of the true cost involved.

o The price for option 2, the on-line alternative, is quoted as 2.7 million dollars for a dedicated CPU. It would seem more reasonable for ODP to use the 158-Blue for this service. Since this machine is already purchased, a considerable saving could be realized. Because the RECON machine would have to be isolated for security reasons, dedicated peripheral and front-end communication equipment would have to be acquired. This equipment would cost approximately 1.5 million dollars to buy. If a new CPU were to be acquired for this service the estimate of 2.7 million dollars for the entire configuration is probably low by about one million dollars.

Implementation Time

12.   The lead time of one year for an off-line retrieval capability (6 months for planning and 6 months for implementation), should be enough time for SPD and ED to insure that a workable system will be available if the manpower can be allocated to this task. Conversion to MVS and testing and tuning under increased loads may be all that is required.

14 November 1978

MEMORANDUM FOR: Deputy Director of Data Processing

STATINTL   FROM   :   ☐

Deputy Director for Processing

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

REFERENCE   :   OCR Memo, dtd 19 Oct. 1978, Same Subject

    1.   The attached memorandum speaks to the OCR proposal
outlined in the referenced document on a paragraph by para-
STATINTL   graph basis. ☐ comments are valid and they
highlight some of the real problems.

    2.   As a general statement, however, the OCR proposal
does not fully address the problems that would be faced by
ODP in providing such a service.

    3.   Any expansion in RECON services would require a
major redesign of the data base, since it is too large for
the present RECON system. This will be expensive. Option 1
assumes no software cost and very low hardware costs. A
better guess at this point would be about $250,000 to remove
all I/O bottlenecks from the existing system so as to make
better utilization of the existing 370/168 CPU, and an
annual maintenance cost of $100,000.

    4.   Option 2 will require dedicated support from
Engineering Division, Systems Programming Division and
Operations Division. This is a CAMS-like approach and would
be expensive. Ten positions (two systems programmers and
eight operators) would be needed to support this operation
in a stand-alone environment that is electrically isolated
from our present facilities. Approximately 2,500 square
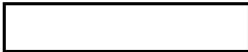feet of computer-grade floor space would also be required.

5. Without any estimate of the transaction loading and the number of terminals to be supported, we must be cautious in selecting the appropriate CPU. It may well be that a 158 size machine is not adequate for a dedicated environment. In the case of CAMS, a 360/65 was determined to be drastic overkill initially. Once users got some experience, CAMS requirements exploded. The cost for a dedicated CPU in option two is probably closer to 4 million dollars that the estimate of 2.7 million.

STATINTL

Att: a/s

STATINTL cc: [   ] w/att.

DD/ODP

ODP # *2-1919*

1 9 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
RECON1/retrieval system for intelligence documents to serve as the basis
for a centralized bibliographic and document retrieval system to serve
all NFIB2/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 [   ] STATINTL

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 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 [   ] reporting are included on a selective basis.

STATSPEC

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

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 dis-
advantage, but not without some positive aspects. Among the disadvantages,
the requester may have no way of knowing how large or small a document

-3-

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 simul-
taneously. 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 inter-
faces 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

-6-

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

via a document listing sent through the mail (Bibliographic Retrieval
Option 1, discussed in paragraph 7) or via a command entered by the
requester on his remote terminal in another NFIB agency (Bibliographic
Retrieval Option 2). These documents would then be mailed to the requester.

19.   The costs[1] of such a document retrieval system can be separated
into investment and operating expenses. An ADSTAR system augmented to
provide Community-wide service would require approximately eight more
storage modules to accomodate 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.

20.   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.

Option Two: Direct On-Line Retrieval

21.   In its most sophisticated configuration, remote ADSTAR terminals
located throughout the Intelligence Community could allow non-CIA

---

[1]   For the purposes of estimating costs, it is assumed that the
number of documents processed into the data base will increase by 25%
above the present level. This figure is based on the current volume of
cables and other material (consisting primarily of finished intelligence
produced by various unified and specified military commands) received by
CIA that is now being processed on a selective basis only into the RECON
data base.

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 manyears. 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,

-8-

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.

-9-

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 invest-
ment 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

-10-

Option 1 - Retrieval Through Intermediaries

	One-time Costs	Annual Costs
Hardware		$ 24,000
Staffing		500,000
ADSTAR Costs	$1,000,000	150,000
		$ 674,000
	$1,000,000 ÷ 5 =	200,000*
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

	Purchase		Lease	
	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
		$ 975,000		$1,685,000
	$4,200,000 ÷ 5 =	840,000*	$1,500,000 ÷ 5 =	300,000*
TOTAL		$1,815,000		$1,985,000

---

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