C-O-N-F-I-D-E-N-T-I-A-L

UNITED STATES INTELLIGENCE BOARD

COMMITTEE ON DOCUMENTATION

MEMORANDUM FOR: USIB Committee on Documentation

SUBJECT: CIA/OCR MINICARD Test

REFERENCE: CODIB-D-23 (8 January 1959)

1. Attached for your information is a summary case history of the limited test of MINICARD as a substitute for the OCR Intellofax System.

2. Our findings are negative. This conclusion is based only in part on our findings that the MINICARD system would not enable us to give substantially superior reference service over that possible with our present system. Weighing very heavily were the present limitations on staff, on space, and on money; these operating assets have been appreciably reduced since the inception of the MINICARD project. Moreover, this reduction has occured in the face of an increase in demand for OCR information services generally, but a relative decline in the demand for literature searches.

3. The decision not to adopt MINICARD as an operational system in OCR does not affect in any way the application of this system elsewhere in CIA. MINICARD has been selected by the CIA Photographic Intelligence Center as a subsystem of its data handling system. As a consequence of the OCR decision, MINICARD equipment, spares and supplies will be released to PIC to augment their proposed MINICARD installation.

Paul A. Borel
Chairman

25X1

C-O-N-F-I-D-E-N-T-I-A-L

SUMMARY CASE HISTORY:
LIMITED TEST OF THE MINICARD SYSTEM
AS A SUBSTITUTE FOR THE
CIA/OCR INTELLOFAX SYSTEM

CODIB-D-23/1
30 June 1960

Purpose of the MINICARD Test

1. The objective of the OCD /now OCR/ proposal of 25 April 1955 submitted to the Project Review Committee was:

> To conduct in OCD an early and large-scale test of a family of data handling equipment known as MINICARD, which is believed capable of substantially improving CIA's Intellofax System as a principal instrument in support of intelligence research. MINICARD promises to contribute improved means for collation of intelligence data, greater speed and flexibility in the conduct of document searches and economies in operation, notably spacewise. 1/

Events Leading to the Test Proposal (1955)

2. From its inception in 1947 to the time of the MINICARD test proposal, the storage and retrieval capability of the Intellofax System was increasingly strained by the flow of information until, by 1955, storage, retrieval and cost problems were considered urgent. The increased growth of the file had been accomplished by multiplication of IBM equipment rentals, storage units and personnel. Analysts' requests at that time for total searches of the seven-year file amounted to 60% of the requests received. Compliance with those requests in categories numbering tens of thousands of cards lengthened search time, multiplied overlap problems, and overloaded requesters with insufficiently refined answers. The possibilities of additional space and personnel ceased, and the alternative to an improved system was reduction in range, speed, and quality of Intellofax service.

3. The proposed MINICARD system held prospects of being such an improved system. MINICARD was said to combine discrete item control, multiple access, flexibility of electronic searching techniques, and inviolate film storage. It could combine coded information and document images which were handled separately by Intellofax. It could ease the critical storage problems with cards and hard-copy. Much faster retrieval rate was expected. In addition it was anticipated that MINICARD would "hasten and expand adoption of common data handling procedures throughout the intelligence community." Six people would comprise the test group; regular search requests after mid-1954 would be submitted to both MINICARD and Intellofax and results compared; decision was expected about 1 July 1957 assuming delivery of equipment about 1 July 1956. Cost reductions would be substantial if other agencies used MINICARD and exchanged Minicards after processing on a common basis; savings in machine rentals would be substantial; the major economy would be space.

- 2 -

4.   Anticipated results, then, were for (a) a community program for compre-
hensive one-time processing; (b) common community storage and retrieval using a
common code, identical equipment and procedures, and inviolate code record and
document storage; (c) improved reference service, particularly with retrieval
according to subject associations, prompt access (50-75% faster), and essentially
simultaneous processing of overlapping requests; (d) economies in operation.
Estimated equipment cost was ⬚   25X1

Developments from 1955 to the Test Period (1959)

5.   When the MINICARD project was approved and an order placed in June 1955,
the equipment was in the blueprint stage. Delivery, scheduled for completion in
December 1956, actually occurred in November 1958, with installation completed
in February 1959. The test period, beginning with document selection, ran from
15 January 1959 to March 1960. During the period between the order-date and
delivery, several major modifications were made in the equipment; we were aware
of them but had no legitimate basis for objection since ours was a program appended
to the Eastman Kodak/Air Force development program. The changes did, however,
invalidate the earlier space and personnel estimates. Operating speeds on the
duplicator, sorter, selector and processor were substantially reduced. Also in
the interim, improvements were made in the IBM equipment in the Intellofax System,
particularly with regard to operating speeds (see para. 28). Moreover, techno-
logical developments in the U.S. and elsewhere with various applications to
storage and retrieval of information advanced very rapidly from the time of the
original MINICARD proposal to the beginning of the test.

6.   By the time the test phase arrived, the earlier expectations from
MINICARD had been somewhat modulated. There was some feeling that extreme
miniaturization which eliminated manual access might prove inferior to our
16mm aperture card system. The combination of codes and images in the same
card was being questioned, as was the loss of the bibliographic Intellofax
tape. Finally the changed equipment specification stimulated the belief that
additional purchases would be required at an estimated cost of 3-5 times the
costs in the original order, with an estimated equipment delivery lag of 1-2
years.

7.   Despite the increased pessimism, the problem remained, requiring
solution.2/ 3/ 4/ On the one hand, Intellofax was a going system and not a
first mechanization approach (hence any conversion must include minor service
disruption, preservation of proven features of the existing system, plus thorough
proof of the replacements); on the other hand, there were no developments in
sight to enable Intellofax to manage 10-20 year indexes with the staff and space
allocated to servicing a 5-year file. MINICARD, though untested, was the only
alternative to EAM (Electrical Accounting Machines) claiming the capability to
handle our task. Not to be overlooked was the fact that at this point, approxi-
mately ⬚   had been invested in equipment, supplies and training; another   25X1
⬚   was required during the last half of FY 59 for maintenance, repair and   25X1
stocking of spare parts; an over-run claim of ⬚   above the original contract 25X1
was under review, not to mention the cost of a one year test.   25X1

C-O-N-F-I-D-E-N-T-I-A-L

8. Intellofax at this time, as today, allowed for 1-2 day dissemination of all documents received - a fact often overlooked in considering the total system; it provided an inviolate (aperture or 35 mm reel) file of documents and enclosures; it allowed for manual access by clerical personnel; it included an index record of subject content of documents. The criticisms of Intellofax were generally: (a) slowness of dissemination of documents (considering the total reporting cycle from inception with the collector); (b) incompleteness of document file (in that poor copy documents, being non-reproducible, were excluded); (c) slowness in retrieval and lack of precision of results. Since Intellofax was backed by IBM cards, the additional complaints touched on limited capacity to record long index terms; slow input and unreliable output, inaccessibility of the cards to the analysts.

9. Still prior to the test, the thought was expressed that subject coding, the slowest phase of Intellofax, would be even slower for MINICARD because of the increased coding complexities; a dual MINICARD scheme (codes and document images on separate film cards) would reduce equipment processing capability by 50%, double the storage space, and create early requirements for additional equipment; the 60-1 reduction ratio was extreme and would heighten the problems with poor quality originals.3/

10. So went the pros and cons. The final recommendation, which was acted upon affirmatively, was that having already made a substantial investment in MINICARD and in spite of the reservations which had developed, OCR should make its own direct evaluation and proceed with the minimum test as described below.4/

Test Period: Input Plan

11. The test program was separately staffed and operated to minimize interference with the regular Intellofax operations. A MINICARD working group consisting of some of our most experienced people from several OCR divisions was formed to prepare for input a sample of 20-25,000 raw information reports containing a normal mixture of document categories by source, format, enclosure variety and with every kind of problem in subject coding and photography, including papertypes, inks, size differences, color differences and varying legibility; we utilized, on the average, 18 people (average salary⬚) for   25X1
over a year. The test corpus was to be fully processed into (a) Intellofax; (b) MINICARD, with codes and document images stored together in a single card; (c) MINICARD, with codes and documentimages stored in separate but related cards. Input was to be accomplished in 8-9 months, depending on Minicoding techniques then to be perfected.

12. A detailed procedural manual6/ was compiled, to be corrected as experience dictated. It was distributed to Eastman Kodak for information and comment.7/ It included sections on the equipment, system work flow, machine operations, MINICARD coding procedures, phrase coding and retrieval procedure. After 9-12 months of input and retrieval-testing of each of the three approaches mentioned above, the plan called for projection of performance rates and costs of each to full-scale operations, utilizing a minimum five year document collection. Coding of the corpus was carried out throughout the test for Intellofax and the separate code/document image MINICARD approaches. A coding variable which was

not originally proposed was the use of the revised Intelligence Subject Code for the MINICARD retrieval test and the old form of the ISC for the Intellofax retrieval phase.

13. In addition to the basic purpose of equipment comparisons and the secondary operational testing of the revised ISC, it was thought that by-product information could be obtained on: (a) the role and importance of a source card file; (b) the utility of a hard-copy document collection; (c) the effect of OCR policy on "nodexing", i.e., not indexing, certain information.

Output Plan

14. The retrieval phase was to begin in late January 1960 (it began in February) after completion of the coding, completion of machine input, consultation with other Agency offices, and selection and preparation of the test questions.[8] This test phase was to run a month, the first two weeks of which would involve submittal of 200 questions at the rate of 20 per day. Since reference service is the justification for all the documentation activity, including machine operations, its goals were the continually prominent yardstick of the investigation. The determinants of customer satisfaction to which the yardstick was applied involved: (a) comprehensiveness, to ensure that all pertinent items are retrieved; (b) precision, to ensure that all items retrieved are pertinent; (c) speed, in terms of gross elapsed time. The 200 test questions included action (live) requests and simulated requests, suggested by retrieval test, other OCR, other CIA and non-Agency personnel. Also involved would be requester interviews to define and refine questions and to check customer satisfaction with content and format.

15. An outline of evaluations to be considered[9], looked something like the following:

    a. Comparison of Intellofax and MINICARD for the test corpus and projection for five-year file
       (1) Personnel and training requirements
       (2) Machine and supplies requirements
       (3) Monetary costs
       (4) Quality of reference product
       (5) Capacity for normal and crash requests
    b. Usefulness of clear text for retrieval
    c. Reliability of phrase structure vs. sentence links for retrieval
    d. Free vs. bound modifier practice
    e. Code rules vs. "word" additions to code outlines
    f. Files management procedures
    g. Coding practices and tools (including authority files, tag definitions, coding depth, area code adaptations for subject modifier use, pagination coding, etc.)
    h. Fixed field coding by clerical vs. analysis personnel
    i. Nodex criteria
    j. Retention value and purging.

Statistical Findings[10]/

16. Approximately 15,000* documents were coded for input and 185 test questions were processed. For Intellofax the retrieval end product was the Card-List Camera tape of bibliographic citations; for MINICARD, a first page print was provided for each document retrieved by the selector. Of the 185 runs made, 120 retrieved specific documents in one or both systems; 65 retrieved no specific documents. Both systems combined retrieved 997 specific references of which MINICARD obtained 788 or 79% of the total and Intellofax obtained 649 or 65% of the total. Marginal or rejected references in MINICARD were 42% and in Intellofax, 65%. Returns were analyzed according to three categories: (a) close match, in which 85% or more of the specific documents were retrieved by both systems (these represented 19.2% of the 120 analyzed runs); (b) disparate match, in which 15% or less of the specific documents were retrieved by both systems (36.7% of the 120 runs); (c) divided relevance, in which 16-84% of the specific documents were retrieved by both systems (44.1% of the 120 runs).

17. Of the 120 runs which recovered relevant documents, 73 included clear text in addition to numeric codes; in 53 of these, documents were retrieved by logical expressions requiring clear text. Among the 23 close-match runs, 13 involved documents retrieved by logical expressions requiring clear text; in 44 disparate runs, 28 included equations using clear text -- in 15 of these, documents were retrieved by equation requiring clear text; and in 53 divided relevancy runs, 29 included equations using clear text, in 25 of which documents were retrieved by such equations.

18. Machine operations: Productive, idle and down time percentages representing thirteen month averages for the MINICARD equipment (cameras, processor, sorter, duplicator and selector) were 36.4%, 42.6%, and 24.2% respectively. Detailed breakdown for individual pieces of equipment was as follows: Camera I: 24.6%, 56.7%, and 19.1%; Camera II: 32.7%, 48.3%, and 19.5%; Processor: 37.6%, 45.8%, and 21.6%; Duplicator: 45.5%, 33.5%, and 28.1%; Sorter: 39.9%, 35.4%, and 29.1%; Selector: 39.0%, 36.7%, and 27.7%.

19. Statistics on documents in the corpus: Total documents amounted to 24,633 of which 14,963 were coded. The average size document contained five pages, six file expansions and 16 code words. Totals and averages for numbers of tagged words, images, documents requiring more than one MINICARD, cards used, first and second generation cards produced, etc. were kept and extrapolated for a full-scale MINICARD operation based on the following assumptions:

   a. 1000 documents** and 20 requests would be processed each day.

---

*See para. 19: overall total includes Nodex material and reports Minicoded by Air Force.

**Present daily input in Intellofax is about 700 documents per day. The projection to 1000 was made to allow for anticipated increased volume and to assess a systems capability greater than that which we now have.

b. File expanded Minicards representing 70 documents would be received from Air Force each day.

c. Minicards mounted in aperture cards and filed by document number would be used to service requests for specific documents.

d. Equipment in use (or planned) for reproduction of the present aperture cards would be modified to reproduce the Minicards mounted in aperture cards.

e. Vital records would receive a summary listing of subject and area codes with document numbers for each code and Minicards mounted in aperture cards.

f. Tabulating cards with an electrostatic first page image of the document would be used as a source file.

g. Any new MINICARD equipment procured would have the same operating capabilities as that on hand.

h. Service personnel would work a staggered shift and check the equipment before each day's operation.

25X1

Interpretive Findings 10/, 11/, 12/, 13/

21. The test demonstrated clearly that subject control of information and the procedures employed thereto are the principal areas to consider in the development of a successful system.

22. The test revealed several advantages of MINICARD over Intellofax. In the main, these were of a nature which could, with additional cost in manpower or money, be incorporated into the Intellofax system. The over-all retrieval test showed an appreciable qualitative advantage for MINICARD. This advantage was attributable chiefly to coding and procedural techniques developed by the

Working Group during the test period. These would be largely transferable to a revised Intellofax system.

23. The validity of comparisons between MINICARD and Intellofax could have been affected by the following factors: (a) the revised ISC and new Area Code were used for MINICARD input and the old ISC was used for Intellofax; (b) document coding was reviewed for MINICARD input at the rate of one in three documents but such review occurred rarely for Intellofax; (c) document coding transcript sheets were available for cross check when MINICARD retrieval was at variance with Intellofax; (d) phrase coding, unlimited boundaries and logical relationships were available in MINICARD but not Intellofax; (e) clear text was used in MINICARD only; Intellofax used bibliographic citations. In the judgement of the Working Group, the revised ISC provided improved structure, form and subject unity which was compensated for in part by Intellofax in this test situation by normative procedures developed over a long period, and familiarity with the application of the old ISC. The potential advantage of MINICARD input-review to provide more standardized and higher quality input was compensated for by the fact that the Intellofax team had had supervised training, standard interpretation and published coding reminders in their previous work with the old ISC. Although document coding sheets were available only to MINICARD for cross-reference check, the main file of the Intellofax system provided a source of control against the test deck of punched cards. Phrase coding, unlimited boundaries and logical relationships were not available to Intellofax. The extent to which this can be added to present equipment is limited by the capacity of EAM for determing logical relationships by using matching techniques and separate files for direct entry to major retrieval problems. Clear text was an advantage to MINICARD; clear text could be used if we re-designed the Intellofax card.

24. Information controlled by both MINICARD and Intellofax includes: subjects/commodities/organizations/personnel types; area/related area; modifiers for subjects; security classification; locator number; control number; publication date. Information controlled by MINICARD but not by Intellofax includes: names of persons, organizations or geographic locations; subject and commodity specifications by clear text entry; modifiers based on Area Code (direction indicators, nationality and comments codes); control of format (maps, charts, bibliography, etc.) by clear text entry. We have not found information-date control to be of sufficient value to justify allocation of space in a re-designed Intellofax system; the Group recommended that the other categories be studied for possible incorporation in a revised Intellofax system, and a preliminary study has now been made. 14/

25. The advantages, then, of MINICARD are: clear text and phrase coding capabilities and file space savings.

26. MINICARD disadvantages noted include the following:

a. Attempts to establish code-unit boundaries and to provide linkage between given MINICARD words within a phrase proved difficult for machine processing. To accomodate the MINICARD work, the notation of the revised ISC

has been restricted to six digits; likewise natural language as identification of a given item has been entered in artificial form as one MINICARD word. This limited the value of natural language entries and necessitated rigid control against loss of relationship or meaning within the MINICARD Record.

b. Indication of pagination within MINICARD was found impractical. Each word within a given MINICARD phrase may be taken from a separate page of a report processed. To paginate would require linkage and code-unit boundaries.

c. Although digital information in MINICARD can be supplemented, deleted or altered by making a modified duplicate, this is a slow and laborious process, amounting to the reprocessing of all information in the document. All corrections must be photographic; item by item standarization via the gang-punch technique is not possible. Conversion of a MINICARD file to reflect code changes is difficult, and the use of systems change indicators, as suggested by Eastman Kodak, is not the answer, according to our input people.

d. Manual access to the detail file is a valuable asset in the reference facility. MINICARD does not provide this facility.

e. Punch card equipment lends itself readily to dictionary building; MINICARD equipment does not.

f. The building of special purpose files as a by-product of the detail indexes is not feasible.

g. Identification of coding error is possible through mechanical matching of detail entries against an established deck of approved codes. This technique is possible with Intellofax but not with MINICARD.

27. **Equipment:** We have mentioned operating speed reductions engendered by equipment modifications. In addition, the block sorting operation (in which the Minicards are placed in the 100 magazine file blocks) which was to have been performed mechanically, failed because of the rigid tolerances needed when positioning the block over the transport belt.10/ The overall assessment from the equipment standpoint was that (a) it did not perform in accordance with the specifications established at the outset of the project; (b) the modifications of operating specifications would necessitate substantial maintenance and supplies (and hence increase the requirement for standby equipment); (c) MINICARD equipment does not easily lend itself to fit in as a subsystem of OCR's overall machine system, much of which would have to be continued to process materials not suitable for control by MINICARD.

Conclusions Concerning the Original Project Proposal

28. The expectation of economies in the Agency from a common community program for one-time processing, common code, and identical equipment and procedures hasn't come about and from all indications would not come about through MINICARD, whether this Agency adopted it or not. The Air Force is using

MINICARD and does exchange information with us -- involving 70 documents per day, as compared with our present daily input of about 700 documents. Technological advances since 1955 have been such that alternative systems have been presented and no other agency in the USIB plans to use MINICARD. One of the aims of the USIB Committee on Documentation (CODIB) has been the stimulation of compatibility of systems considered, and this goal has by no means been reached; to hope for identical systems is just not realistic. As to the common code, the ISC was approved in principle as the USIB community code and, hopefully, will be adopted by all some day, but to date only Air Force and CIA use it. Moreover, the code including the revision, was developed apart from MINICARD considerations and applies regardless of the system.

29. As for MINICARD providing an inviolate storage file, the same is provided by the aperture card/microfilm reel facilities of Intellofax -- and we feel we must keep this manual access capability. Retrieval according to association and other capability for more pertinent retrieval can be built into Intellofax with the experience garnered from the test. As for speed: input is slower for MINICARD because of its increased complexity; this fact was noted both during coding and preparation of machine logic. Access time did not prove to be faster than Intellofax. It should be noted that the outlook is for improved EAM equipment: collator operating speeds of 1300 cpm vs present 480 cpm; selectors at 1000 cpm; sorters at 1000 cpm with further promise of doubling. By the same token, second generation MINICARD equipment promises much improved capability in speeds and other economies, such as the single machine combining selector and sorter. Our test did not go into the relative merits of these improved capabilities.14/ Simultaneous processing was not really tested. Retrieval quality (legibility) was not impressive.

30. The major economy mentioned in the original project proposal was space. Owing to the changed specifications, additional equipment hence additional space would be required. There is no question but that the MINICARD files would occupy less space than IBM cards; the ratio of Intellofax work and file space to that of a two-card MINICARD system might approach 5:1 or more. This is a problem which must be tackled but it is not, in our view, the paramount problem. We assign higher priority to quality of input and relevance in retrieval. Steps have been taken to reduce the need for hard copy document files and a procedure for further culling material from the active and inactive files in OCR and at Records Center is under development.

31. The MINICARD capability for greater depth of coding deserves mention. The Working Group does not believe that depth of coding to the degree originally considered is required -- such depth just isn't in the documents themselves. In a small test-within-the-test sampling representing about one weeks production (400 documents), 80% of the documents required three subjects or less; 91% required five or less; 95.5% required eight or less. To the degree that this sample is representative (and it is thought to be) if expensive equipment were bought on the basis of providing greater coding depth, it would be applicable to 4-5% of the documents under control. Another way of looking at the time expenditure for input to MINICARD is to consider that over 50% of the coding time was

spent on less than 10% of the documents.15/

Concerning Secondary Test Considerations

32. The revised ISC provides for more precise input than the old code. During the course of the test the Group was able to identify 301 subjects which occurred with sufficient frequency that they should be (and were) added to the revision prior to its issuance. Operational testing of the revised code for over a year afforded excellent experience for its use with Intellofax and developed a cadre of analysts from several OCR components who returned to their own components with the advantage of this experience.

33. No significant conclusions were reached as to by-product information on files management, indexing policies, utility of the hard-copy collection, retention value or purging, or the importance of the source card file. Clear text was a MINICARD plus and will be applied to Intellofax. Phrase structure was not considered necessary by the Group. As to free vs. bound modifier practice, the Group would prefer the free modifier but the revised ISC is bound.* Coding practices and tools are being explored and preliminary recommendations have been made concerning their application to Intellofax. Fixed field coding could probably be done by clerical personnel.

34. Generally it was concluded that the methods of subject control developed during the test should be used in redesigning Intellofax. The MINICARD dictionary, Training Manual, card forms and transcript sheets have application for Document Division use. The value of machine listings as index building aids was demonstrated.

With Reference to the Over-all Office Picture

35. The Intellofax, or any other machine system, role is only a part of our service facility. Requests for machine searches constitute partial use of the Library, difficult to pinpoint accurately but estimated at about 20%; considerable use is also made of the Biographic, Graphics, Industrial and Special Registers. In the overall OCR picture, Intellofax furnishes 8% of the total references provided and represents 0.6% of the total requests received (FY 59 figures).** With increasing pressures of budget and personnel restrictions,

---

*i.e., at present. It may be that it can be unbound.

** In FY 59 there were 2070 requests for Intellofax runs among 313,277 requests levied on the office as a whole. These machine runs furnished 809,581 selected references out of an office total of 10,806,335 references supplied. Other requests levied on OCR almost certainly resulted from references furnished by Intellofax runs but are not identifiable as such. It should also be noted that the term "request" is an ambiguous one which could run the gamut from a telephonic transaction completed while the customer waits to a substantial machine and man-hour expenditure; no better unit of measurement than the term "request" has been developed as yet.

Page Denied

(2) identification of named commodity which is classed in a single code entry having more than one commodity; (3) geographic place names (perhaps in the Bloc only). Provision for clear text entry of the following to be deferred for further study: names of research establishments; names of industrial firms; identification of terms other than commodities; names of geographic locations outside the Bloc; names of persons.

e. Changed IBM Card Format: a proposed card design has a built-in capacity for 32 additional types of tagged information. In a document recovery system, variable types of information within a fixed field identified by tag are as effective in retrieval as that derived from a one purpose field. This form of information control can be readily adapted for input to a magnetic tape system.

Summary and Decision

38. As a result of the test the Working Group concluded that MINICARD did not live up to what had been hoped for in terms of our own problem. There were demonstrable advantages, but important disadvantages were also discovered. As a consequence the Working Group did not recommend a conversion from Intellofax to MINICARD, but recommended instead the modification of Intellofax to incorporate as many of the advantages of MINICARD as were technically and administratively feasible. OCR management took these findings into account, along with the proportional role that machine searches play in the overall OCR service picture, and the present limitations on staff, money, and space. These considerations led to a decision not to adopt the MINICARD system as a substitute for the OCR Intellofax system.

Appendix: Source references and selected bibliography on MINICARD.

25X1

Page Denied

Next 2 Page(s) In Document Denied