Project SAFE

Feasibility of an Agency-wide Information System to Support the Analysts File Environment

Confidential

October 1974

Copy N⁰ 74

Project SAFE

October 1974

CONTENTS

LIST OF FIGURES

LIST OF TABLES

APPENDICES*

*Appendices I – X are located in Room 1E4808, CRS/Systems Analysis Staff, and will be made available for reference.

I. INTRODUCTION AND SUMMARY

INTRODUCTION

Over the years CIA has made a wide array of intelligence resources available to its analysts. Indeed, the Intelligence Community spends a large sum each year to provide these resources and to find new ones. They are made available by such a variety of processing systems and procedures that the individual analyst may have difficulty in finding all the items he needs—particularly if he has a short deadline.

Production offices have continually sought to better exploit intelligence resources by creating their own data bases and files, sharing files of common interest, or introducing new analytical methods or automation. For the most part, these efforts are made at the office level and, at best, answer only office needs.

This report describes CRS efforts to design an Agency-wide, all-source intelligence resource system that would offer all Agency analysts the best support today's technology can provide. It suggests how such a system might be cheaper in the long run than the sum of all the individual systems currently being developed or proposed. The design that emerges is called the SAFE (Support for the Analysts File Environment) Information System.

SUMMARY

CRS began work on Project SAFE in response to a June 1972 directive by Mr. Colby, then Executive Director Comptroller. It said that CRS should "work with the analysts and production offices within the Agency . . . to develop the most effective mix of central bibliographic and document retrieval files and special purpose document retrieval files for individual customer offices, (and) analysts. . . ."

Preliminary development work with the production analysts soon showed what characteristics a SAFE system should have. The concept that emerged was that of a multipurpose Agency-wide information processing system operating through on-line terminals widely distributed among the production offices. SAFE will permit the individual analyst to view his daily mail on-line, route particular items to other analysts, build machine files for himself or his office, and to maintain on-line files. The on-line file building capability will allow the analyst to store a complete text, an extract from it, or an indexed representation of it and to include his own comments on such items. The system will allow the analyst to search the files he creates and, because he has multiple access points to any item, to search them more thoroughly and more specifically than he could normally search a conventional paper copy file. Where document representations are stored in files, SAFE will provide the necessary full text back-up, either by digital storage of text or, more commonly, microforms.

In addition to its role in dissemination and in the support of analyst or office files, SAFE will give the analyst access, through his on-line terminal, to a wide range of resources, including the major CRS data base and several files of the complete texts of intelligence messages. Eventually the analyst may also be able to use the same terminal to reach "external" data bases, including those within the community as well as such commercially available files as the *New York Times* Information Bank. The analyst thus will have, at his fingertips, a wide array of information resources needed in the production of finished intelligence.

1

CRS implemented a model of a SAFE system and made it available to a small number of production offices over an 8-month period in 1973-74. This was defined as the data-gathering phase of the project. Its objectives were threefold: to determine the general feasibility of SAFE; to learn the user's reaction; and to gather data from which to develop more detailed specifications for an Agency-wide system. The SAFE model was modest in that it used inexpensive and relatively unsophisticated software, existing computer resources, a small number of terminals and a selected sample of users. It nevertheless demonstrated all of the major components of the proposed system.

Close cooperation between CRS and the analysts in the production offices has been an important feature of the data-gathering phase. Those analysts played a key role in the design of the pilot system. Indeed, CRS assumed from the beginning that if an Agency-wide system is to succeed, its real users must be involved in its actual design. The pilot branches cooperated fully, and the large amount of data collected has enabled us to define much more clearly the requirements of an Agency-wide system.

Conclusions

The overall reaction of participants in the SAFE pilot operation has been extremely positive. Our evaluation (described in detail in Chapter V) of the pilot system indicates that SAFE is potentially a very powerful tool, faster and more efficient than the resources we presently have. Most analysts who have used the pilot system are enthusiastic about its present capabilities and its potential. Indeed, there is a strong feeling that this is the direction the Agency must take in information processing. All the proposed features of the system have proven valuable, but the handling of text files and the building of analyst files will probably be the most important.

Two of the most significant values of SAFE will be its ability to get incoming material to analysts rapidly and its ability to provide fast access to a wide array of information. It appears to have great potential utility, therefore, in the handling of crisis situations, as reported by one of the pilot branch users: "During the Cyprus crisis, SEC was able to receive relevant reports hours before the reports were available in hard copy. This capability allowed us to stay well ahead of possibly threatening developments and, in fact, alerted us to potentially interesting developments in the Balkans before reports were available thru regular channels . . . I believe the SAFE system has enormous potential for crisis management."

The SAFE concepts were examined by five companies involved in the design of large computer-operated data systems. They believe most of the concepts, with one major exception, are within the state-of-the-art. The exception refers to the part of the original concept that called for scanning paper copy, digitizing it and entering it into the system. In their opinion this is not currently feasible. Because parts of the SAFE concept are close to the outer limits of the state-of-the-art, implementation of SAFE will present major challenges in systems design, software production, and the coordination of much hardware. A similarly large and complex system is not known to exist elsewhere. The individual parts do exist, however, and the contractors agree that SAFE can be built.

Our experiment has persuaded us that the Agency should move toward the implementation of a system of this kind, having the general configuration described in Chapter VI, and that we should immediately begin work on a detailed system design.

Cost

To support 1000 analysts, the proposed system will require a substantial investment over a number of years. Some of this investment will be compensated by a more efficient and integrated use of Agency computer resources; by the assimilation of

2

certain existing systems and operations; and by a considerable reduction in the generation, movement, storage and disposal of paper copy. The system must be justified on the grounds of benefits to the Intelligence Community not on the grounds of economy. We consider these benefits to be improved intelligence products, generated by analysts who are informed more rapidly, more completely and more precisely than ever before.

The estimated cost of SAFE is about 41 million dollars. This sum would cover the software design and development and the purchase of hardware in 1974 dollars. It does not include past costs, personnel costs of CIA employees involved in the project, logistic costs (which may be high), or OJCS costs for continued support of the pilot program. Our estimated cost would be less if the software could be developed in-house (which is highly desirable) and if much of our existing equipment could be used. We have deliberately used the high figure of our cost range to make sure that approval of Project SAFE carries a realistic recognition of the potential financial impact (excluding logistic costs). Development of the SAFE effort is a commitment of up to 41 million dollars and a development period of at least 5 years. It would also represent a major effort—not yet defined—for logistics as well as an undetermined communications investment.

These dollar and time costs are as firm as we can determine from current experience. Both could increase, however, during SAFE's development and implementation. Because we have used the higher cost figure, such increases should not have a major impact on the overall cost of the system.

Finally, the SAFE Information System faces three major problems. First, there are important security considerations involved in the development of a computerized file environment which have not been addressed in this report.

Second, it was noted earlier that, although the concepts of SAFE are within the state-of-the-art, there is no system in existence of comparable size and complexity. There is a related risk. SAFE will become an integral part of the analyst's working environment; if it fails him, he is out of business. Therefore reliability and backup are critical. The Agency has limited experience in building and operating applications where the computer is so intimately tied to an Agency function. What experience we do have tells us that, in addition to high equipment reliability, extraordinary developmental and operational discipline is required even for simple applications of this kind. SAFE will represent a challenge different from any that our computer systems people have ever encountered.

Third, the project need not necessarily be completed by FY 1980; but prolonging the work would probably increase both the cost and risk. The funding need not be so heavily concentrated in the first years as we have proposed; but spreading the funds evenly across all the years will delay implementation and probably increase the risk. Most importantly, SAFE must rationally be a complete intelligence processing system. Because of the cost, we expect to hear proposals to create one-half or two-thirds of the system—to handle some sources of information, but not all; or to serve some production offices, but not all; or to perform some of the functions that are technically possible, but not all. We oppose all such proposals.

II. HISTORY OF PROJECT SAFE

CRS INITIATIVES

In December 1971 the Director of CRS created a task team to write a detailed plan for upgrading the 1,300,000-record, computer-based CRS reference file (AEGIS).[1] The general plan was to convert an off-line batch mode of operation to an on-line interactive mode. This would improve service by allowing interactive searches to be made at remote computer terminals as search requests were received. The ability to enter search requests from remote computer terminals would also theoretically allow Agency production analysts to bypass CRS analysts, who presently serve as intermediaries.

The task team was also to consider methods by which production analysts could add keywords, codes, and documents to the basic reference file. It had long been recognized that many of the analysts' special interests could not be adequately handled by the more general indexing performed by CRS.

In March of 1972 the task team began discussions with representatives from OCI, OER, OSI, DDO (then DDP), OSR, and OBGI in order to inform them of the CRS objective, to learn the extent of their interest as potential input or output users of such a system, and to determine whether any of their requirements should be considered in the proposed upgrading of AEGIS.

OCI and OBGI immediately expressed interest in a system that would give them a computer search capability for their manual office files. OCI was especially interested in reducing the size of its paper files by using a computer control system.

As a result of this interest, the task team conducted an OCI/CRS and OBGI/CRS 2-week experiment, which simulated production analyst input to the CRS AEGIS file. The results were encouraging, and in May 1972 OCI asked if CRS could implement interim measures to allow continued OCI input prior to the upgrading of AEGIS.

AGENCY DIRECTIVE

In June of 1972 the Director of CIA, Mr. Richard Helms, approved a series of recommendations by Mr. Colby, then Executive Director Comptroller. The series included a directive that CRS "work with the analysts and production offices within the Agency, and with such other Intelligence Community agencies as may be feasible, to develop the most effective mix of central bibliographic and document retrieval files and special purpose document retrieval files for individual customer offices, analysts, or other requesters."[2]

CRS RESPONSE

Responding to this directive, CRS first critically reviewed its major file building and information processing capabilities:

    1. The MAD system, an Agency-wide Machine-Assisted Dissemination system developed by CRS for SI electricals;

---

[1]Already Existing General Information System—this reference file is often referred to by the acronym AEGIS, which is also the name of the computer data management program for this file. Other programs could also "manage" the reference file. In fact, later in this paper the RECON program is introduced as one such alternative.

[2]MEMORANDUM FOR THE DIRECTOR, SUBJECT: Automatic Dissemination, June 1972. (Confidential)

    2. The AEGIS system and an on-line version of AEGIS (which, although not considered a candidate for the upgraded AEGIS system as discussed above, allowed for searching from remote computer terminals);

    3. The OLDE computer program, an On-Line Data Entry program by which computer files are created and maintained at remote computer terminals (OLDE was developed as part of the task team's AEGIS follow-on activity);

    4. The OCI and OBGI experiments, which gave some evidence that the analysts were willing to switch from their manual document files to a computer/microfilm system;

    5. The CRS computer center, a center developed to maintain systems like MAD, AEGIS and OLDE.

These five capabilities were the building blocks upon which two related proposals were based:

    1. "Proposal for a Demonstration of an On-line System to Provide Production Analysts with Access to Personal, Office and General Bibliographic Files."

    This work was written in August 1972 by Professor F. Wilfrid Lancaster of the University of Illinois. Its purpose was "to demonstrate a concept, with the object of generating interest and support within the various production offices . . . As the capabilities are demonstrated, user reaction will be observed and gauged . . . We can learn much more about user needs and attitudes from such a working model than we can possibly learn by a paper model and more conventional interviews or questionnaire surveys."

    This working model would attempt to simulate the ultimate system ". . . (which) will give the individual production analyst on-line, interactive access to his personal document file, his parent office files, specially prepared extract files, and a wide range of CRS bibliographic files."

    2. "Prototype of a CRS Production Analysts File Support System as an Interim Step Toward an Operational CRS On-Line System."

    This work was written in August 1972 by Jean Skillman of the CRS Systems Analysis Staff in response to OCI's request for an interim capability. It proposed that OCI analysts would mark the terms by which their documents should be indexed; CRS would input the index records for those documents into a special AEGIS file created solely for OCI. CRS would also microfilm the documents for permanent retention and have computer listings printed regularly, to give OCI analysts an index to their microfilm file holdings. The use of microfilm in this remote system would significantly decrease the volume of OCI holdings, and the printed indexes would give OCI analysts improved access to their documents. This experiment with OCI was the origin of the SAFE concept (later called Module 1) that production analysts would create their own document index files.

A Project SAFE paper based on these two proposals was published in October 1972. The paper (See Appendix 1) described a set of concepts that, taken together, postulated and partly defined a new Agency-wide information processing system for intelligence materials.

The paper also proposed a data collection period during which production analysts would evaluate the utility (not the cost-benefits per se) and practicality of the concepts. First, the concepts would be partly implemented through test systems (called "modules" in the SAFE paper) set up with existing or easily developed computer/microfilm techniques; and then a representative sample of analysts would work with and evaluate the test systems.

III. SAFE CONCEPTS

Many of the concepts discussed below are not new. In 1968 McCracken[3], then in OER, presented his version of a system in which the intelligence analyst could read his mail via a CRT device, disposing of some items immediately, reserving others for later reading, and transferring some into his personal files. The concept of the integrated system—in which the analyst can extend a search from his own files into Agency-wide (or Community-wide) files from the same terminal, using only one query language, or at least similar languages—does appear to be new, however.

REMOTE COMPUTER POWER

The basic SAFE concept is that of making the computer, via a remote terminal, directly available to the production analyst—literally putting it at his "finger tips." The analyst need have no knowledge of computers or computer programs; indeed, another SAFE tenet is to keep the man-machine interaction simple.

The computer is used to conrol index files and other document surrogates and/or whole documents. Indexes and documents now available to the analysts have the drawback of limited access because they must be controlled manually by file drawer headings, index card entries, etc. Material that is available externally, as from CRS, offers improved access because of computer control, but places an intermediary between the analyst and his information. This intermediary offers certain advantages to the analyst by being a specialist trained in the techniques of information retrieval. However, this "delegated" mode of searching for information also has significant disadvantages: the individual analyst has no way of browsing in the machine file, the intermediary may misinterpret the real needs of the analyst, and system response will be delayed rather than immediate. The intermediary may offer additional resources and assistance, but experience tells us the analyst does not fully use them. This experience was brought home in the results of a user study by Professor F. Wilfrid Lancaster[4] in 1969.

Lancaster used 17 finished intelligence documents written by analysts who had not requested an AEGIS Subject Search in support of their research. He conducted searches and took the results of those searches to the 17 analysts for evaluation. On the average, the analysts considered close to 50% of the retrieved citations to be relevant. Four of the 17 felt an AEGIS search would have been of major value to them in the preparation of their reports; one analyst found two documents he did not know existed, which would have been of major value to him; and six were unaware of the existence of the AEGIS system. Although we feel that increased publicity has made Agency analysts more aware of the AEGIS system today, we still question whether they are fully exploiting this and other intelligence resources.

A SPECTRUM OF RESOURCES

SAFE envisions the analyst having access, via his remote computer terminal, to a wide array of intelligence resources. These include the bulk of the "mail" he now

[3]McCracken, M.C., Computers in Economic Intelligence, December 1968. (Secret)

[4]Lancaster, F.W., "Intelligence Analyst Appraisal of AEGIS Subject Searches, "Memorandum to D/CRS, dated 17 December 1969. (Secret)
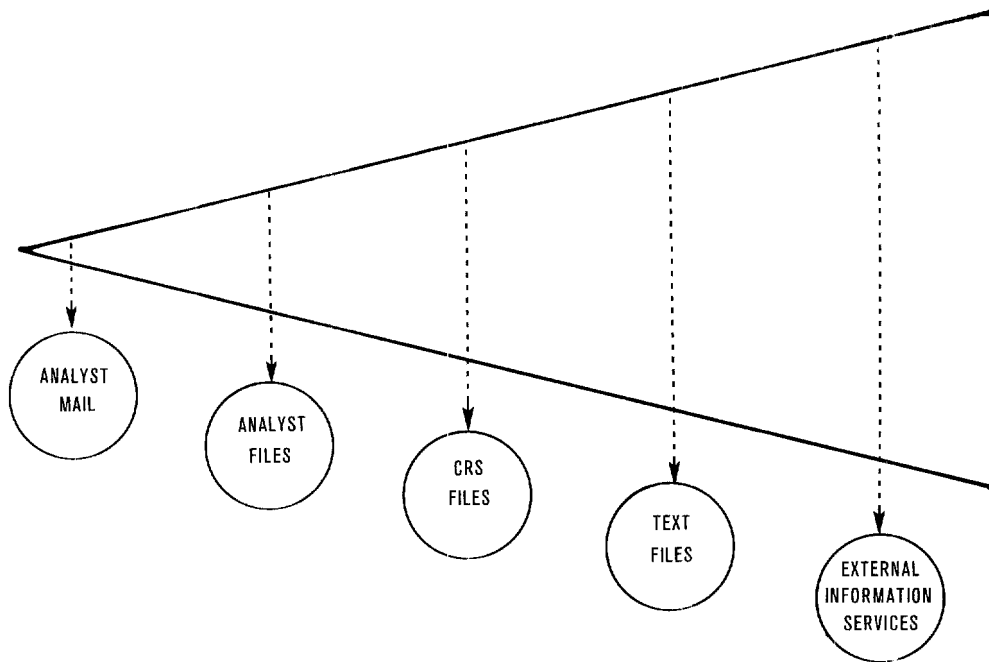
Figure 1. Spectrum of Information Resources Available to Analysts

receives in paper copy form; his personal files, whether they are whole document, index, or data files; office files; CRS files, including AEGIS and certain biographic directories; text files, e.g., State cables and FBIS field traffic; and other sources, which might include indexes of open literature, such as the *New York Times* Information Bank (see Figure 1). For example, an analyst could query his mail file (after a security/identification check) by pressing the mail key on his remote terminal. Within seconds, his latest receipts will begin appearing on the screen. If he notes any item of special significance, he could add his comments and/or index terms and file the entire document or a portion of it under the appropriate file heading. Or he could route the item to another analyst.

If the item suggests a question to the analyst—verification of some fact new to him—he could switch operational modes by pressing the analyst file key. Within seconds, the screen displays a "menu" of available files. The analyst chooses one, and the system displays the proper query instructions. The instructions help the analyst formulate the best search statement for his requirement. If he wishes, he can also check CRS files for pertinent information. He presses the CRS file key and a "menu" of available CRS files is displayed. He proceeds as he would with his personal file. He may also wish to extend his search to the "text" or to the "other" files. If the analyst finds an item of interest in any of these searches, he can view it on his display unit, print it on his terminal printer, or add it to his personal file.

We see many other possible uses for analyst-to-SAFE terminals. For example, the analyst could compose an intelligence article at the terminal and, as he wrote, add contributions from a wide array of files. He could retrieve numerical data from his files, then call up a special module which could make statistical computations. He could also create certain computer files by calling-up and completing a form, which would appear on the display screen, and purge his files automatically by a pre-set purge indicator. In addition to all of this, we foresee an "alerting" capability: an item that met pre-set requirements would set off an audio and visual indicator to catch the analyst's attention.
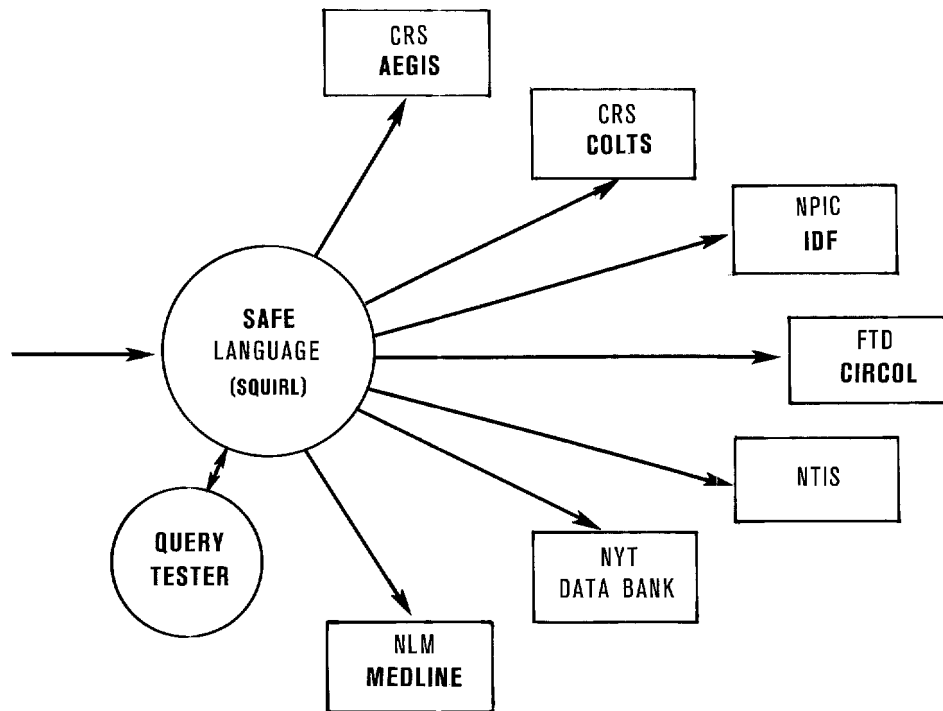
Figure 2. Single SAFE Language

SINGLE SAFE LANGUAGE

Another SAFE concept is that the production analysts need not know computer programing language; simplicity in remote computer terminal operation is a fundamental requirement. SAFE envisions a single "language" with one set of conventions for interrogating any of the various files associated with the SAFE system (Figure 2). Currently the various systems (e.g., CRS's AEGIS and COLTS, NSA's SOLIS, NTIS tapes, *New York Times* Information Bank, NPIC's IDF, FTD's CIRCOL) have their own languages, conventions, aids and hardware. It is questionable how many of these can be mastered even by experts who work with these systems full time, let alone by analysts who might sometimes use one or more systems.

The SAFE language concept has been named SQUIRL (SAFE QUery and Information Retrieval Language). It will give analysts access to any SAFE system file. SQUIRL also embodies a "query tester." This tester would allow the computer to analyze a query in terms of the contents of existing files. The tester would print on the screen the file names most likely to contain the required information.

SINGLE DOCUMENT STORAGE

About 1,800,000 individual documents are disseminated annually by CRS and the Cable Secretariat, and in the process each one is reproduced in an average of 16 copies.[5] Thus almost 25 million copies are disseminated annually. Some of them are later copied again for filing under two or more subject headings or dossier names. A significant number of these copies will be found in the CRS files. The SAFE concept

---

[5]This figure is based on: a) Morfit, J.C. and Roepe, H.B., Dissemination and Filing Survey, Project ASPIN, July 1970 (Secret); and a July 1974 survey of Cable Secretariat dissemination.
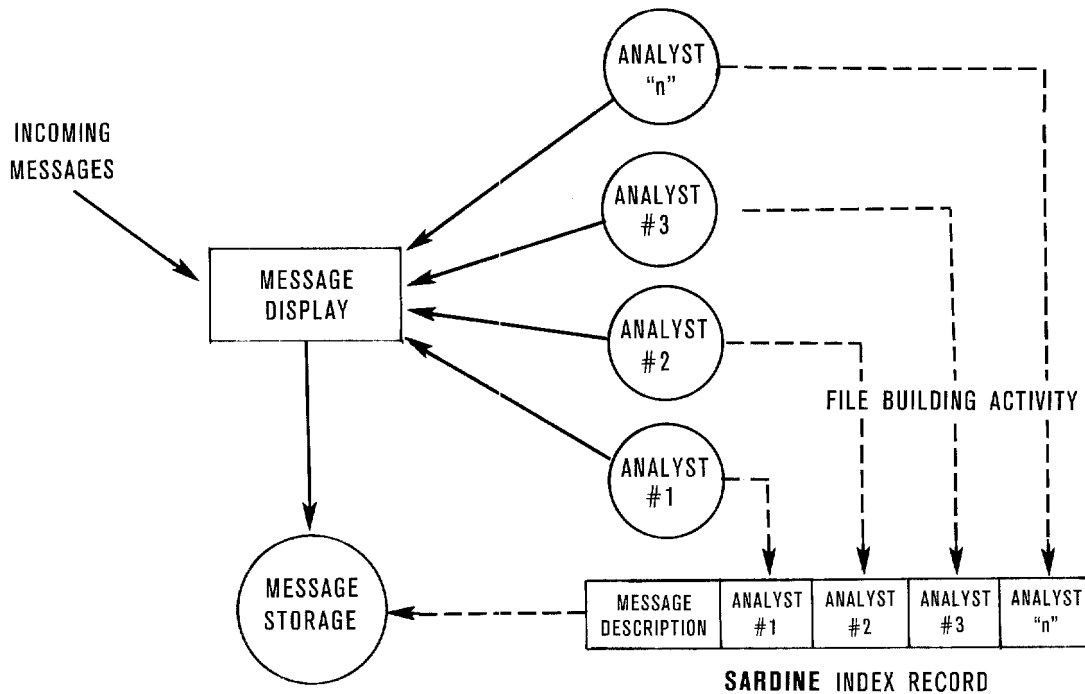
Figure 3. Single Storage Concept

suggests that a document would be stored once; an analyst could electronically view it as often and "file" it in as many places as he liked. The document itself remains stored in one place, but an entry is made in the SARDINE index record[6] for each analyst who has "filed" the document. This SARDINE record structure controls the storage collection, permitting both filing and retrieval (see Figure 3).

A SYMBIOSIS BETWEEN PERSONAL AND CENTRAL FILES

According to SAFE concepts, the production analysts would be able to create, maintain, search and retrieve records from their computer-based personal document files, as well as from other files. SAFE would regard these personal and office files as complementary to the CRS Subject Index files rather than as substitutes. Personal and office files have a number of obvious advantages over the central files. They contain material of significant value to the individual analyst because they include only documents he has considered worth retaining. Their organization—i.e., file headings, index terms, and comments—reflects his special requirements.

The disadvantages of personal files are equally obvious, however. They are not necessarily complete, and analysts may index by terms too specific to serve anyone but themselves. Further, one analyst generally will not be able to use another's file except by special agreement.

One major disadvantage of personal files in paper form is that they offer only extremely limited retrieval of information. A document is usually filed under one heading, and if more than one is required, a duplicate must be filed in another place. Space and unwieldiness prevent truly multiple access to personal files in paper form.

---

[6]SARDINE (System for Analysis and Retrieval of Documents Idiosyncratically Named and Evaluated) is the file management system which, conceptually at least, is the desired replacement for AEGIS.

But a computer-based system can provide a virtually unlimited number of access points to each document.

CRS indexes a wide array of intelligence documents, using terms that allow for general retrieval on topics as wide ranging as the intelligence problem itself. The CRS Subject files are available to all, within the bounds of security requirements. The SAFE concept is that both central and personal files could exist side by side in an on-going system, but that a document need only be stored once. We envision a computer record created for a document whenever even a single analyst (including CRS analysts) decided to "file" it. The document would remain stored in one place, but an entry would be made in the SARDINE record structure that controls the storage collection. Analysts throughout the Agency would have access to the document—through their own entries on the SARDINE record if they created such a record as they read their mail, or through the CRS entry if they did not.

We envision an alerting process to insure that CRS makes an entry on a SARDINE record for every document of interest. If a document normally excluded by CRS selection criteria is considered worth keeping by two analysts—i.e., when an index record contained two or more entries—CRS would then index the document by adding an entry to the document record, so that it would become available to all.[7]

THE "PAPERLESS OFFICE" CONCEPT

We have previously stated that approximately 25 million copies of intelligence materials are disseminated annually throughout the Agency. During a single year, an individual analyst may accumulate several hundred new documents, and a single division of a production office may file tens of thousands. The size of files thus becomes substantial: In August 1972 the Middle East/Africa Division of OCI was estimated to have about 372,000 items on file.

The distribution, reproduction, filing, storage, retrieval and refiling of these documents are expensive. SAFE proposes a gradual shift away from such handling of intelligence materials. Paper copies would be the exception rather than the rule, and intelligence materials would be handled electronically or on micromedia.

[7]Document security practices would, of course, still be in effect.

IV. DATA COLLECTION PLAN

INTRODUCTION

Clearly, the concepts discussed in the previous chapter required an exhaustive, near real-life evaluation before a commitment to an Agency-wide system could be proposed. An evaluation period of this kind is generally called the "data collection" phase of a system design effort (Figure 4).

The general plan for the SAFE data collection phase, as proposed in the October 1972 paper, was designed around a set of test "modules" (Table 1), which were demonstrated to the Agency production analysts who had expressed interest in various SAFE concepts—and, more importantly, a willingness to volunteer the use of their files and their time. Each of the 12 proposed modules represented a specific information processing capability and a specific intelligence resource. For example, Module 3 allowed an analyst to search and retrieve on his remote terminal the full text of State Department cables. Module 6 allowed an analyst to search and file SI "mail" that had been selected for his office by the MAD dissemination system.

Each of the modules either existed or could be developed with a minimum of resources. The first working demonstration would be made at a CRS Computer Center remote terminal. At the same time, arrangements would be made to install CRT display devices, printers, and microfilm equipment within those offices that had volunteered for the program. This equipment would be a prototype SAFE Console Station (SCS). It was expected that the analysts would suggest improvements as they worked with the various modules during the test period. If possible, some of the improvements would be added to the modules during the test period, and the rest incorporated into the final system design.

By January 1973 further details of the data collection plan had been worked out. Pilot branches were selected and mail and file surveys were planned for them. Surveys

CONCEPTUALIZATION –

DATA COLLECTION –

SPECIFICATIONS –

IMPLEMENTATION –

TEST OPERATION and –
OPERATIONAL CAPABILITY

TIME FRAME
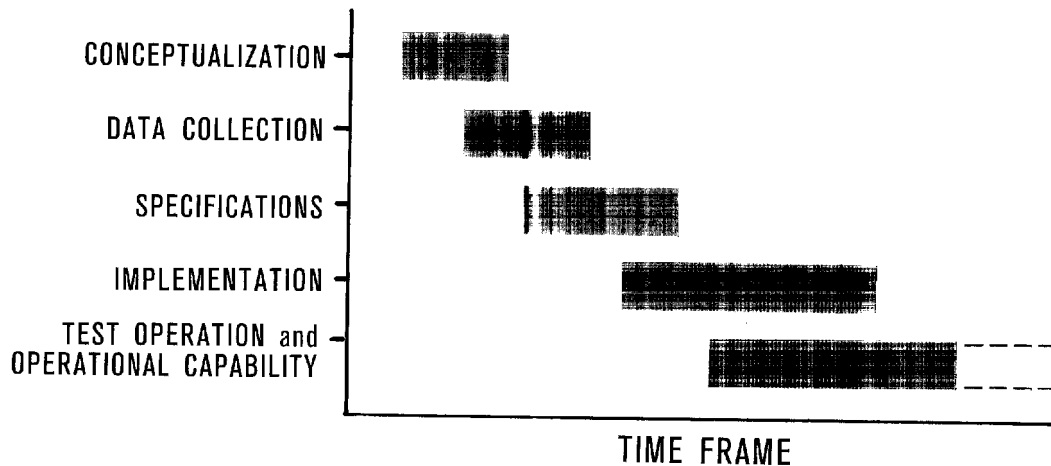
Figure 4. SAFE System Design Phases

13

Table 1

SAFE Modules

Module 1 — User Index/Information Files
Module 2 —Search and Retrieve—SI Messages
Module 2.1—Search and Retrieve—OAK Materials
Module 3 —Search and Retrieve—State Cables
Module 4 — Search and Retrieve—FBIS Field Traffic
Module 5 —Search and Retrieve—Military Cables/IR
Module 6 —Scan, MAD Mail
Module 7 — Search and Retrieve, Module 1
Module 8 —Search and Retrieve, CRS Index (AEGIS)
Module 9 · Scan—State Cable Segments
Module 10 — Edit—Modules 6 and 9
Module 11 — Search and Retrieve, CRS Index (RECON)
Module 12 —Search and Retrieve, External Data Bases
Module 13 — Compute

and test query runs would document the kinds of information the analysts needed during a working day and would determine, in a preliminary way, the ability of the proposed SAFE modules to satisfy those needs.

THE COMPUTER MERGER

In the spring of 1973 the Management Committee recommended, with DCI approval, the closing of the CRS Computer Center and the absorption of its work by OJCS. The effect of this merger on SAFE included a delay of about a year, while OJCS and CRS tried to fit the operating programs and data sets of CRS to OJCS procedures and to a far more complex machine environment. A large number of people in both CRS and OJCS made an extraordinary effort to provide the machine facilities that made the testing of the SAFE modules possible during this period.

By May 1973 a revised Project SAFE paper (see Appendix 1) was published,[8] which explained in greater detail the data collection plan, the participants, and the computer software requirements. It also attempted to adjust the schedules. By November 1973 an additional SAFE paper was published,[9] which explained the final details of the data collection plan and again adjusted schedules to conform to the realities of the computer merger problem. Schedules were revised again in March 1974.

Essentially, Project SAFE activities during CY 1973 were limited to the planning of surveys in the pilot branches, the preparation of software and the introduction of Module 1 (user index files). A serious problem was sustaining the enthusiasm of the production analysts for the promised, but now delayed, implementation of the various modules. Actual implementation began during CY 1974.

ACTIVITIES DURING THE MERGER—CY 1973

Planning

Early in 1973 four pilot branches were selected:

OCI/Middle East Africa Div./Greece Turkey Iran Br. (now called Arab States-Mediterranean Br.);
OSR/Program Analysis Div./Strategic (now called Strategic Evaluation Center);
OER/USSR/East Europe/Strategic Impact Br;
OSI/Nuclear Energy Div/Nuclear Technology Br (now called Sino-Soviet Br);

---

[a]Data Collection Plan: Summary of Pilot Branch Operation and Computer Support Requirements.
[9]Project SAFE, Questions and Answers, November 1973.

By the end of June, however, an OER substitution was made:

OER/Near East-South Asia Div/Trade and Aid Br (now called Developing Nations Div/Trade and Aid Br).

The biggest change made during this planning period was in the method of introducing the modules to the pilot branches. According to the new plan (November 1973), SAFE modules would be introduced singly in the pilot branches without a preliminary CRS demonstration.

Further, the modules would be loosely held together by a special computer program (SQUIRL-I) that would give SAFE some characteristics of a coherent information system, rather than a set of unrelated modules. ·

Pilot Branch Surveys

Beginning in February 1973, members of the SAFE project team conducted surveys of the pilot branches' operating environment. They studied (a) mission, organization, analysts' functions and work flow, (b) mail receipts, and (c) file input. The mail survey was to determine the quantity and source of all unique incoming items in each of the pilot branches. Among other things, it told us how much of the analysts' mail was electrically transmitted to the Agency and how much was received as paper copy. Such information has a direct bearing on the extent to which computer processing can process and store this mail. The file input survey was to determine the number and source of documents selected for filing from the total incoming mail. It would tell us how many of them were also filed and indexed by CRS. Such information has a direct bearing on the extent to which CRS supports analysts' files. These two surveys were conducted between February and July 1973.

Beginning in April 1973, the production analysts were given tape recorders and asked to comment on their information needs as they arose. They were asked to state:

1. Information (What was needed)
2. Purpose
3. Time span of information required
4. Required turn-around-time (How soon needed)
5. Prompt (What prompted the need)
6. Known files to be addressed

This survey (conducted between April and September 1973) was to determine the direction and scope of the team's subsequent data collection efforts. Samples of the analysts' information needs were formulated into search statements, and queries were conducted on the appropriate available modules—for example, a text search of State cables (Module 3) and/or an index search of the *New York Times* Information Bank (Module 12). These searches allowed us to make a preliminary evaluation of the usefulness of the modules. The results of the mail, file input and information needs surveys are contained in Appendix IX.

Software Preparation

The testing of the various modules required the use of several computer programs either in existence or in development. Both MAD and AEGIS were already operational. MAD permits the building of special files of State and military cables, SI messages, DoD IRs and FBIS field traffic for the pilot branches; AEGIS allows analysts to tap directly into a subset of the large CRS index files; and OLDE can be used to build special personal index files.

OLDE would allow analysts to build computer searchable files by entering data at their consoles. Although the OLDE program already existed, it had to be tested in an operational environment. It was introduced into the CRS/USSR area division in

15

February 1973, and within a few months it took the place of OCR (Optical Character Reader) typing and processing of AEGIS Subject Index File records.

In early 1973, work started on COLTS (CRW On-Line Text Search), which was to be an on-line version of the Chase, Rosen and Wallace (CRW) batch text searching system. This system would allow analysts to search, on their SAFE terminals, the text of special computer files of SI messages, State and military cables, DoD IRs, and FBIS field traffic. The program was to have been completed by the end of May 1973, but by then it was evident that COLTS would not work in the OJCS environment. It was redesigned, and final testing began in December 1973.

By April 1973, a contract had been awarded to Operating Systems, Inc. (OSI), for the development of the DEMON (Data Extract MONitor) program. This would enable analysts to search the full text of messages (SI, State cables, etc.) for specific words and, when the words were located, to extract a specific segment of the message. For example, an analyst could specify that each time a certain post name was found in a State cable, he wanted to see the title of that cable, with the option of seeing the whole cable if necessary. Operating Systems, Inc. began installing this program in December 1973.

In October 1973 a contract was awarded to Chase, Rosen and Wallace to develop the first of the OLTA (On-Line Text Analysis) series of programs. The first version of OLTA would allow an analyst to scan and "file" his SI electrical "mail" (those SI items disseminated by the CRS MAD system) on his SAFE terminal before he received the paper copy. The second version would allow the segments found by DEMON to be viewed as mail and would allow the whole message to be viewed on demand. The final version would allow analysts to interact more completely with their SI mail—they could add index terms and comments, extract portions of a message for their files and edit messages before filing them. OLTA I, II, and III became operational during May, June and July 1974 respectively.

Introduction of Module 1

The design of the first of the Project SAFE test modules was based on the results of our early experiments with OBGI and OCI (mentioned in Chapter II), in which production analysts indicated the terms by which their documents should be indexed. CRS then created computer index records, microfilmed their documents, and printed computer listings of their index terms. These listings served as references to the document collection. A continuing Module 1 experiment was in progress in OCI/MEA/GTI as Project SAFE became a reality in late 1972.

The purpose of Module 1 was to give analysts an alternative to their single access, manual document and data files and to then determine whether they would accept a computer file that, while giving them greater access to their documents and data, required them to do the actual indexing. A secondary purpose was to see if microfilm was a suitable alternative to paper copy files. Figure 5A shows a document marked for indexing by an OCI analyst and the microfilm aperture card that represents the document. Figure 5B shows sample entries of the computer listings that serve as a reference to the document collection. Note that the listings have been ordered in different ways, giving the analyst multiple access to his file—by keyword (index term), subject heading, document number and classification. OCI chose its Cyprus and Turkey files for the extended experiments.

In May 1973, discussions started in OSI/NED/NTB. This branch decided to create a computer index file for all its holdings on Worldwide Nuclear Technology. A month later, the first Module 1 listings were made for this branch.

Next 2 Page(s) In Document Exempt

Self-Help Branches

The Project SAFE pilot branches were chosen primarily for the interest of the branch personnel and their willingness to participate. However, other analysts and offices began asking to take part in SAFE activities. In November 1972 an analyst from DCI/IC/SAS had asked for help in setting up a computer-based index to his special collection. The Module 1 system offered a solution to his problem, and by March 1973 the first group of documents from that staff was being processed.

In April 1973 the OER/S division chief asked CRS for help in setting up a file control system for the collection of documents related to Indochina. Here, too, the Module 1 system would have been useful, but CRS resources could not handle the extra load. In August 1973 OER/S and CRS agreed on an alternative. CRS helped the division set up its file structure, including AEGIS file building and microfilming procedures; trained the division analysts in the use of the OLDE program, which allowed them to build their computer file from a remote CRT location; and gave them a procedure manual.

Since this program was devised, CRS has assisted other branches with similar requests. They were called self-help branches, and before the data collection period ended in July 1974 there were 26 of them.

ACTIVITIES AFTER THE MERGER—CY 1974

Introduction of SAFE Test Modules

In November 1973 SAS published the Project SAFE Data Collection Plan, which centered around the introduction, over an 8-month period, of the various SAFE modules in four pilot branches:

        1. OCI/MEA/ARM
        2. OSI/NED/SSB
        3. OSR/SEC
        4. OER/D/TA

The modules were introduced in the order shown in Figure 7. Dated 15 March 1974, it was the final revision to the November 1973 schedule.

SAFE Report Period
SEARCH & RETRIEVE:text files:FBIS field traffic
SCAN: MAD mail, SI *(OLTA-III)
SCAN: message segments: State, (DEMON/OLTA-II)
SEARCH & RETRIEVE: CRS subject indexes (RECON)*
SCAN: MAD mail: SI (OLTA-I)
SEARCH & RETRIEVE: text files: OAKS *(COLTS)
SEARCH & RETRIEVE: text files: State, SI, MIL. (COLTS)
SEARCH & RETRIEVE: SAFE language introduced (SQUIRL-I)
DATA ENTRY: user indexes (OLDE - Production)
SEARCH & RETRIEVE: user indexes (AEGIS on-line)
SEARCH & RETRIEVE: CRS subject indexes (AEGIS on-line)
*secondary priorities

| DEC-JAN | FEB-MAR | APRIL | MAY | JUNE | JULY | AUG |

Figure 7. Data Collection Plan

According to the plan, SAFE modules would be implemented in the four branches at about the same time. In mid-October, however, OCI/MEA/ARM withdrew temporarily because of the Arab-Israeli war. It rejoined in late January 1974, but by late February 1974 the branch decided to discontinue its Module 1 file, because of the termination of the NIS program the file was designed to support.

Search and Retrieve—User Indexes (Module 7)

The pilot branch Module 1 files generated computer listings that served as a reference to the indexed document collection. In December 1973 Module 7, which uses the AEGIS on-line system, was introduced. Using this module (an on-line version of Module 1), analysts can search their index files via their CRTs, and within minutes any "hits" are displayed on the screen. Through the displayed references they can retrieve the documents from their microfilm collection and view them on the microfilm reader/printer. Where the computer reference listings are large or the search questions complex, Module 7 is especially useful. For analysts who have Module 1 "information files," the hits themselves may be answers to the search requests.

Search and Retrieve—CRS Indexes (Module 8)

In January 1974 extracts of the CRS Subject Index (AEGIS) were made available to the pilot branches via the AEGIS on-line computer program. The analyst can search portions of the CRS Subject Index by formulating search questions at his SAFE terminal. The search is made as he waits, and document references appear on the display screen. If the document reference(s) appears to answer the question, the analyst can order the microfilm version from the CRS document library.

Search and Retrieve—SQUIRL I

The SQUIRL I computer program, available to analysts in late January 1974, gives the various SAFE modules the character of a coherent information system. When a user dials into the SAFE system, he sees first the Daily Notes (Figure 8) and next the

```
                        ****************                          15 MAY 74
                            WELCOME
                               TO
                          PROJECT SAFE
DAILY NOTES:                ****************

    ---- ALL USERS ---- PLEASE NOTE ----
         SAFE HOURS HAVE NOW BEEN EXTENDED FROM 0700-1800,
         MONDAY THRU FRIDAY. FOR PROBLEMS OCCURRING 0800-1630
         CALL SAFE 'HELP' NUMBER, EXT 7870. TO OBTAIN ASSISTANCE
         BEFORE 0800 OR AFTER 1630, CALL OJCS ON EXT 6816.

    -PRELIMINARY PLANNING FOR DEMON (MESSAGE SEGMENTS) AND ULTA
     (MAIL FILES) IS IN PROGRESS. THE MAIL FILE WILL BE THE NEXT
     SAFE MODULE IMPLEMENTED ,WITH DEMON FOLLOWING IN ABOUT A
     MONTH.

    -TEXT FILE REORGANIZATION IS ANTICIPATED BY 20 MAY. USERS
     WILL BE NOTIFIED AS SOON AS AN EXACT DATE CAN BE DETERMINED.

    -THE CRS EXTRACTS WERE UPDATED 13 MAY.

DATA COLLECTION SCOREBOARD - DATA COLLECTION FORMS RECD TO DATE:
              OCI/ARM - 85                  USR/SEC - 30
              OSI/SSB - 88                  OER/D/TA - 30

***** TO ADVANCE - DEPRESS ENTER KEY.
```

Figure 8. Daily Notes

21

```
                              **********
                              *  SAFE  *
                              **********
                                  *
              *******************************************************
              *                   *                    *
        ***********        **************        ***************************
        *  SCAN   *        *   SEARCH    *        *       DATA ENTRY       *
        *  (OLT)  *        *      &      *        * OSI/SSB(SCI)  OSR(SC3) *
        ***********        *  RETRIEVE   *        * OSI/FER(SC2)  OER(SC4) *
                          ***************         *      OSR/SFC(SC5)      *
                                  *               **************************
                  *******************************
              *                 *                *
        ***********     ******************    ***********
        *  TEXT   *     *     USER       *    *  CRS    *
        *         *     *  INDEX/INFO    *    *  INDEX  *
        *  (STS)  *     *     (SAG)      *    *  (REC)  *
        ***********     ******************    ***********
```

**** TO SELECT AN OPERATING MODE TYPE THE SWITCHING CHARACTERS SHOWN
     IN PARENS.

Figure 9. SAFE System Outline

SAFE System Outline (Figure 9). This is a summary of the SAFE system and lists the "switching characters" that give the user access to any available part of the system. When he dials into a search and retrieve part of the system, a "file menu" appears showing the file names available for searching (Figure 10). SQUIRL I also allows an analyst to switch to other files.

Data Entry—User Indexes

Until January 1974 the SAFE task team did the on-line input for the pilot branch user files (Module 1), using the OLDE computer program. This responsibility was turned over to the branches in January, but SAS continued to do the editing until June, when it too was turned over to the branches. Figure 11 shows a completed OLDE form as it appears on the screen.

Search and Retrieve—Text Files

In early April 1974 the pilot branches were given COLTS software that enabled them to search specially selected text files of SI messages, State cables, and military cables and DoD IRs (Modules 2, 3, and 5 respectively). COLTS allows analysts to do full text searches on those documents and to view any message that contains the specified search terms. The analyst can also print a message at his on-site printer or

```
          TEXT FILES ARE UPDATED EVERY 24 HOURS AT APPROXIMATELY 0400.

     INDIVIDUAL SOURCE FILES CONTAIN MOST RECENT 14-21 DAYS TRAFFIC.

     THE NUCLEAR FILE INCLUDES ALL SOURCES EXCEPT FBIS, FOR THE MOST RECENT 60
     DAYS & IS BUILT ON SELECTED TOPICS CHOSEN BY OSI/SSB.

     THE MBFR FILE INCLUDES ALL SOURCES EXCEPT FBIS, FOR THE MOST RECENT 6 MOS.



     THE 24-HOUR SI FILE CONTAINS SI ELECTRICALS FOR THE PRECEDING DAY ONLY
     (0400-0400 HOURS). MONDAY THE FILE CONTAINS TRAFFIC FROM FRIDAY 0400-
     MONDAY 0400.

          1.SI                        2.OAK TARGETS -USSR
          3.24-HOUR SI                4.MBFR
          5.STATE CABLES              6.SAVE FILE SI1
          7.DOD/IR                    8.SAVE FILE SI2
          9.MILITARY CABLES          10.SAVE FILE SI3
         11.FBIS FIELD TRAFFIC       12.SAVE FILE SI4
         13.SELECTED NUCLEAR SUBJ.   14.TEMP SAVE FILE SI5


          **** ENTER THE NUMBER OF THE FILE TO BE SEARCHED (ONE ONLY) ****
```

Figure 10. File Menu

```
74P829152924P3        (K)5(       0PP1              HEADER RECORD           BOYDG                    LINE NO.        1
SEQ. NO P:272107  DOLTYPE:I   CL/DISS:C==   CW:__                                                    LINE NO.        2
SAFE NO:QEB:ZOTB/MX:LBLD_OO189J:19_   ITE--:___  PAGE:____                                           LINE NO.        3
SOURCE/SERIES:_____                                          LINE NO.        4
PUB. DATE:74O7     EX CAT:____   TO PATE:____                                                        LINE NO.        5
TITLE:>BID*1974*mu*ITRBI*BID-:_IOOIB_To_PROVIDE_EGYPT_WITH_WIG:II_SPARE_PARTS,_I                     LINE NO.        6
RODS,_AND_SPARE_PARTS_:==>_CBIBD*5579_//_____                   LINE NO.        7
------------------------------------------------------------------                                  LINE NO.        8
-------------:(7)_                                                                                   LINE NO.        9
SUBJ. CODE:000000  AREA 1:DEG_   AREA 2:____  CMNTS:_____               LINE NO.       10
SUBJ. CODE:_____  AREA 1:____   AREA 2:____  CMNTS:_____               LINE NO.       11
SUBJ. CODE:_____  AREA 1:____   AREA 2:____  CMNTS:_____               LINE NO.       12
SUBJ. CODE:_____  AREA 1:____   AREA 2:____  CMNTS:_____               LINE NO.       13
```

Figure 11. Completed OLDE Form

have a listing of messages printed in the OJCS center. He can save those of interest to him by placing them in one of the five "save" files (personal computer files) alloted to his branch. Using COLTS, he can search and retrieve from these files.

At first the text files were organized by pre-defined file requirements so as to contain only messages pertinent to a pilot branch's area of interest. As analysts gained experience, however, they disliked this limitation and preferred to have access to all State cables, SI messages, etc., and to search against the total receipts.

Scan MAD Mail

During May, computer mail files containing the most recent 5 days of SI message traffic (Module 6) were established; the program used was OLTA I. This module permits the user to scan, on his SAFE terminal, all messages disseminated to him by MAD. He can ignore messages of no interest; have his on-site printer make a copy; enter a message into a "save" file; or route a message to other analysts in the SAFE network. Many software problems with OLTA I arose during the first few weeks, but they were solved by the middle of June.

Search and Retrieve—OAKS

During June 1974 selected OAK files (Module 2.1) were placed on-line for use by the pilot branches. Module 2.1 contains OAK data dating from June 1971. Using the COLTS software, analysts can make full text searches against these files.

Search and Retrieve—RECON

During early July extracts of the main CRS Subject Index File were converted from the AEGIS On-Line system to the RECON system (Module 11). This module was to give production analysts access to a search and retrieval system specifically designed for on-line bibliographic searching. RECON is faster than AEGIS On-Line and offers several computer-generated aids designed to help the analyst in making his searches and to improve the quality of the results. The RECON system is much closer to the "interactive" system we are working toward than AEGIS On-Line. Obtained originally from NASA, RECON is being modified to meet CRS needs.

Scan Message Segments

Module 9, which was the second version of the mail file (OLTA II), was also installed in July. OLTA II allows analysts to scan message segments selected by the DEMON program and to call up the entire message if they desire. (DEMON matches character strings in messages with character strings in "dictionaries" provided by analysts. When it finds a match, DEMON extracts part of the message, as specified by the analyst in his DEMON dictionary). All options available to the analyst in OLTA I are also available in OLTA II.

Scan MAD Mail

OLTA III (Module 10)—the third and final version of the mail file—was installed during the latter half of July. Besides the capabilities of versions 1 and 2, this version

allows analysts to modify messages before storing them in the "save" file. Analysts can add terms (keywords) to messages, add comments of any length, extract selected passages and source documentation, and compose new items for this storage.

Search and Retrieve—FBIS

A text file containing FBIS unedited field traffic (Module 4) was also constructed during July. It contains the most recent 14-20 days of all FBIS field traffic except administrative messages. All capabilities available in Modules 2, 3, and 5 were also available for Module 4.

Search and Retrieve—Library Ready Reference

During July pilot branches were given access to the CRS Library's Ready Reference File, which had recently been automated as a SAFE Module 1 file. It primarily contains references to *Washington Post* articles on topics of intelligence interest.

Self-Help Branch Activities

The self-help activities continued to expand during this data collection period. Table 2 lists the pilot branches and the other organizations that took part in the program or inquired about taking part (the same units are listed by major components in Table 3). Table 2 also shows which modules of SAFE they have used: for example, it shows that most of the self-help branches wanted analyst files, for controlling either a special file or the entire branch holdings.

1-4 Activities of the pilot branches have been discussed above.

5-8 Activities of DCI/IC/SAS, OSR/SF/C, OSI/PSTD/EB, and OER/S have been discussed in the section on FY 1973 activities.

9. OER/CD/IN expressed interest in managing a Steel Plant File through SAFE, if data in that file could also be used in a computational module. This requirement was noted for future SAFE planning.

10. OER/ST/P analysts now have OER finished intelligence index records (extracts from the CRS data base) available for search and retrieval.

11. OSR/TF/W maintains the entire branch file in the SAFE system. Analysts create the index records, and CRS microfilms their documents.

12. OSR/SF/N uses Modules 1 and 7 to access its branch files.

13. OSI/LSD requested a tie-in between the SAFE file building system for cable traffic and the SHOEBOX system, which supports the VIP Medical Program. The OLTA III mail module can provide such a tie-in.

14. OSI/GTB (now OSI/PSTD/SSPB) personnel asked for OJCS help in developing a computer-based system for monitoring hydrographic ships. OJCS expressed interest in SAFE's ability to extract message segments and build files automatically. Planning continues.

15. OSI/NED/NWB has asked that its branch files be added to SAFE. Branch analysts currently are using SAFE Text Files.

16. OSI/NED/NPB has access to the SAFE Text Files.

17. OSI/NED/SA has access to all SAFE modules through the OSI/NED/SSB pilot branch, and uses them to index documents for its own Module 1 file.

18. OSI/PSTD/SSPB has one analyst who requested and now uses the SAFE Text Files.

19. OBGI and ORD discontinued a joint SAFE project after a trial period.

20. OBGI/GD is using Module 1 for a specialized file to determine if SAFE can be adapted to general geographic data base management needs.

21. OCI/MEA/PGI requested and now has access to SAFE Text Files and the OCI pilot branch mail files.

Table 2

List of Organizations Participating in Project SAFE

Pilot Branches	Analyst File Choice (AEGIS)	On-line Data Entry (OLDE)	On-line Search of Analyst Files (AEGIS)	On-line Search of Text Files (COLTS)	On-line Search CRS Subj. Index (RECON)	On-line SCAN/ File SI-MAD Mail (OLTA-I)	On-line SCAN/File State Segments (OLTA-II; DEMON)	On-line "Interaction" SI-MAD Mail (OLTA-III)
1. OSI/NED/SSB	Branch	O	O	O	O	O	O	O
2. OCI/MEA/ARM	....	....	....	O	O	O	O	O
3. OER/D/TA	Special	O	O	O	O	O	O	O
4. OSR/SEC	Special	O	O	O	O	O	O	O
Self-help Branches								
5. DCI/IC/SAS	Special	Inactive	Inactive	....	....	....	....	....
6. OSR/SF/C	Branch	O	O	P	O	....	....	....
7. OSI/PSTD/EB	Branch	O	O	O	O	....	....	....
8. OER/S	Special	I	I	....	....	....	....	....
9. OER/CD/IN	Special	P	P	....	....	....	....	....
10. OER/St/P	....	....	....	....	O	....	....	....
11. OSR/TF/W	Branch	O	O	P	O	....	....	....
12. OSR/SF/N	Branch	P	P	....	P	....	....	....
13. OSI/LSD	Special	....	....	P	....	P	....	....
14. OSI/GTB	....	....	....	....	....	....	P	....
15. OSI/NED/NWB	Branch	P	P	O	P	....	....	....
16. OSI/NED/NPB	....	....	....	O	....	....	....	....
17. OSI/NED/SA	Branch	O	O	O	O	O	O	O
18. OSI/PSTD/SSPB	....	....	....	O	....	....	....	....
19. OBGI & ORD	Special	Discont'd	Discont'd	....	....	....	....	....
20. OBGI/GD	Special	O	O	P	....	....	....	....
21. OCI/MEA/PGI	....	....	....	O	O	O	....	....
22. OCI/MEA/SOA	....	....	....	I	O	....	....	....
23. OCI/WE/SE	....	....	....	O	....	....	....	....
24. OCI/NID	....	....	....	O	O	O	....	....
25. FBIS	Special	P	P	O	P	P	P	P
26. OPR	....	....	....	....	P	....	....	....
27. DDA/ISAS	Special	O	O	....	....	....	....	....
28. CRS/CLD/LY	Special	O	I	....	....	....	....	....
29. CRS/NEA/AB	Special	P	P	P	P	P	P	P
30. CRS/FEPAC/CB	....	....	....	O	....	....	....	....

KEY:
O = Operational.
I = Implementation stage.
P = Planning stage.
Special = A special category file; not the total branch file.

22. OCI/MEA/SOA has one analyst who requested but does not yet have access to SAFE Text Files; this analyst now uses the CRS Subject Index Files.
23. OCI/WE/SE asked and received access to SAFE Text Files.
24. OCI/NID has one analyst who uses SAFE Text Files, the CRS Subject Index Files, and the "mail files" of the OCI pilot branch.
25. FBIS and the Project SAFE task force are studying the applicability of SAFE modules to FBIS needs.
26. OPR was briefed on the SAFE concept and has requested access to Module 8, CRS Subject Extracts.

25

Table 3

List of Participating Organizations by Office

DCI/IC/SAS..................	Intelligence Community Staff, Systems Analysis Staff; now Management, Planning & Resource Review Div., Research & Analysis Br. (MPRRD/R&AB)
OCI/MEA/ARM..............	Middle East Africa Div., Arab States/Mediterranean Br.
OCI/MEA/PGI...............	Middle East Africa Div., Persian Gulf/Indian Ocean Br.
OCI/MEA/SOA..............	Middle East Africa Div., South Asia Br.
OCI/NID....................	National Intelligence Daily
OCI/WE/SE.................	Western Europe Div., South Europe Br.
OER/CD/IN.................	China Div., Industries Br.
OER/D/TA..................	Developing Nations Div., Trade & Aid Br.
OER/S.....................	Southeast Asia Div.
OER/St/P..................	Production Staff
OSI/NED/NPB..............	Nuclear Energy Div., Nuclear Programs Br.
OSI/NED/NWB.............	Nuclear Energy Div., Nuclear Weapons Br.
OSI/NED/SA...............	Nuclear Energy Div., Special Asst.
OSI/NED/SSB..............	Nuclear Energy Div., Sino-Soviet Br.
OSI/PSTD/SSPB...........	Physical Sciences & Technology Div., Science & Science Policy Br.
OSI/PSTD/EB..............	Physical Sciences & Technology Div., Electronics Br.
OSI/LSD...................	Life Sciences Div.
OSI/PSTD/GTB............	General Technology Br. of OSI/PSTD; now Science and Policy Br. (SSPB) of PSTD.
OSR/SEC..................	Strategic Evaluation Center
OSR/SF/C.................	Soviet Strategic Forces Div., Command Analysis Br.
OSR/TF/W.................	Theater Forces Div., Western Forces Br.
OSR/SF/N.................	Soviet Strategic Forces Div., Naval Forces Br.
FBIS......................	Foreign Broadcast Information Service
OPR.......................	Office of Political Research
DDA/ISAS.................	Information Systems Analysis Staff
CRS/CLD/LY..............	Central Libraries Div., Library Br.
CRS/FEPAC/CB...........	Far East Pacific Div., China Br.
CRS/NEA..................	Near East Africa Div.
OBGI/GD..................	Geographic Div.
OBGI & ORD..............	(Engaged in a joint project that has since been discontinued)

25X1

27. DDA/ISAS has, within the SAFE system, a special file of Agency declassification actions. Module 1 computer listings from this file can replace specially prepared reports for Agency circulation.

28. CRS Library's Ready Reference File was converted to Modules 1 and 7. It is queried via a SAFE terminal, and information can be retrieved either on-line or through computer listings.

29. CRS/NEA Area Division joined the SAFE effort to determine the utility of SAFE modules in reference and biographic production work.

30. CRS/FEPAC/CB joined the SAFE effort primarily to study the value of Module 4 (FBIS field traffic) in reference and biographic production work.

Other Contacts

The SAFE Project task force has worked with many other groups, both inside and outside the Agency, concerning aspects of the project. These contacts were important for suggesting further applications of the present SAFE concepts and for adding to our understanding of the concepts and the philosophy of their implementation. Among these groups:

The White House Situation Room has a system similar to the SAFE "mail files."

- State Department is interested in a system that will automatically select and file Soviet visa data. SAFE will have such a capability in the future. This capability will be based on our DEMON program which can extract message segments.
- NSA representatives received a detailed briefing and demonstration of the SAFE system and are interested in setting up a "SAFE system" at NSA.
- NOSIC representatives received a SAFE briefing and discussed the possibility of sending specialized transmissions to CIA to help build SAFE files.
- DIA personnel received numerous briefings and demonstrations on general computer text processing capabilities. DIA is currently implementing a new version of the CIA MAD software.
- DDO/ISG is interested in the text search and data entry computer programs used by SAFE.
- DDO/PS/EVAC is interested in a system that would measure comparative utilization of intelligence sources.
- OCI/WH/LA, OSI/PSTD/FMSC, OWI, IRS, and DDO/OPS/NARCOG have had SAFE briefings and demonstrations, and have expressed interest in using various aspects of Project SAFE.

SAFE DATA COLLECTION TECHNIQUES

During the data collection phase, SAS used several techniques to gather information on module use, software/hardware problems, analyst reaction to the various modules, and general SAFE development:

1. Search Forms—As analysts in the pilot and certain self-help branches used the installed modules, they were asked to fill in a data collection form (see sample, Figure 12). Some 423 forms—indicating at least that many uses of the pilot system—were collected. The information derived from them (summarized in Appendix II) will serve as one of the major criteria for determining the design of a final SAFE system.

2. Finished Intelligence Citations—Analysts were asked to complete a short form (Figure 13) whenever the SAFE system was used in support of a finished intelligence report or briefing. Because this was introduced toward the end of the data collection period, few were actually completed.

3. Help Log—When analysts had problems with SAFE modules, they called a special Project SAFE telephone extension for help. The nature of the problems and their solutions (when known) were recorded in the Help Log. Figure 14 is a sample entry; the log itself is Appendix IV.

4. Mail Log—Analysts were also asked to complete a log sheet to record the frequency and duration of their on-line mail reading sessions. Figure 15 is a sample log sheet; the log itself is Appendix X.

5. RED Book—The basic Project SAFE documentation record can be found in the CRS/SAS RED (Read Each Day) Book. This record—with a subject index—was begun 17 October 1972 and will continue indefinitely. The RED Book (through August 1974) is contained in Appendix VII.

6. Production Analyst Interview—At the end of the data collection period, Project SAFE task team members interviewed 51 people from the pilot and self-help branches. They used a structured interview form, and the average interview took one hour (Appendix III).

7. Production Analyst Reports—At the end of the data collection period, the participating offices were asked to write a critique of the SAFE system, mentioning their experience, impressions, and suggestions for improvement. (Appendix III).

Next 3 Page(s) In Document Exempt

PRELIMINARY SYSTEM DESIGN

As a result of some early comments by analysts on their experience with the various modules and after lengthy discussions with several contractors, the task team wrote a Preliminary System Design paper outlining an Agency-wide SAFE Information System. (Appendix V). It defined a SAFE system in terms of eight subsystems (See Figure 16).

The incoming subsystem described ways of handling each kind of intelligence material as it entered the system. The distribution subsystem showed how the incoming materials might be distributed to 1,000-plus Agency analysts. The current awareness subsystem discussed how analysts would be able to view their "mail" on the SAFE terminal. The file building subsystem outlined how new files would be built, including some automatic file building and a sophisticated file reorganization subsystem. The search and retrieval subsystem indicated the options an analyst would have as he queried the various agency and external files. The data processing subsystem outlined a method whereby data, when retrieved, could also be manipulated mathematically by computational modules (not yet defined). And the intelligence production subsystem discussed the way in which an analyst could compose finished intelligence at his SAFE terminal.

Contractor Reports

Five computer firms were each asked to provide a one-man-month study and review of the SAFE Preliminary Design Report. The contractors were:

        Mitre Corporation
        RLG Associates, Incorporated
        Computer Sciences Corporation
        Operating Systems, Incorporated
        Chase, Rosen, and Wallace, Incorporated

On 29 May 1974 they were briefed in detail on the SAFE concepts, the data collection plan, and the preliminary design report. Each was given a copy of the report and asked to complete its study by 15 July 1974. (The contractors' reports are listed in Appendix VI.) Both RLG Associates, Inc., and Computer Sciences Corporation elected to give CRS a briefing on their major findings.
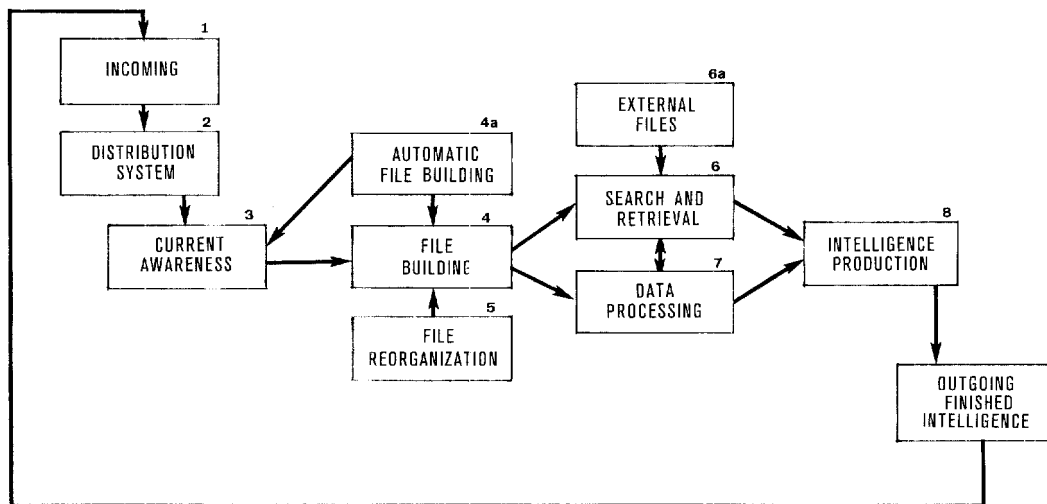

Figure 16. Major Subsystems of the SAFE Information System

V. DATA COLLECTION PHASE EVALUATION

INTRODUCTION

The results of the data collection phase of Project SAFE will be presented in three parts. The first part is a report by Professor F. Wilfrid Lancaster, University of Illinois. Professor Lancaster, who has authored numerous works on the design and evaluation of information systems, is considered a leading expert in the evaluation field. He has been associated with the project since its beginning. (See Chapter II). This report is based on his analysis of the data collection forms, the analysts' critiques of the pilot system, and the structured interview results. All of these have been described in the previous chapters.

The second part of this chapter consists of excerpts from the analysts' reports (copies of the reports will be found in Appendix III).

The third part consists of a comparative analysis of the five contractor reports (found in Appendix VI) on the feasibility of Project SAFE.

EVALUATION OF THE SAFE PILOT OPERATION

Summary and Conclusions

The major purpose of the SAFE experiment was to determine if computers could assist the intelligence production analysts by providing faster dissemination of intelligence materials, giving them greater access to personal and community files, and enabling them to produce more timely and thorough intelligence reports. To conduct the experiment a rather imperfect representation of the proposed system was created. To quote the OSI/SSB report, a "low cost package of hardware and software (was) assembled for test purposes." The entire data gathering phase was plagued, particularly in its early days, with severe problems of system availability and reliability. Moreover, the system was very unstable in this phase of its development in that changes were frequently being made (including major changes related to query language) and new capabilities were constantly being added.

In view of the imperfection of the system, the fact that it was in a constant state of flux, and its general lack of reliability, the results of the experiment seem very promising indeed. Although not all analysts in the participating branches used SAFE extensively, those who did were generally extremely positive in their reactions. There appears to be much less resistance to a "paperless" operation than we might have expected when the experiment began. Indeed, the whole concept has been received with considerable enthusiasm. It is clear that several branches have already become heavily dependent upon SAFE, that even in its present imperfect form it has been able to make a contribution to intelligence production, and that several of the participating branches now feel sufficiently dependent on SAFE as to be seriously handicapped were SAFE now withdrawn.

There is some evidence that all the present features of SAFE have definite utility. As might be expected, different analysts regard different features as being of greatest importance. It seems clear, however, that the really key features are those giving rapid and in-depth access to the complete text of messages or message extracts, together with the features that give a branch the capability of organizing and searching their own files at a level of specificity or complexity that has never previously been possible. In its

experimental phase the system was sometimes used in just the way the designers hoped: a "widening horizons" approach in which a search for information was conducted over several available files. A particularly significant finding, noted by several analyst users, is that SAFE has definite potential value in the management of crisis situations.

It is obvious that a fully operating SAFE system must be considerably different from the experimental system. It must be completely reliable, have greater accessibility, use a common query language, provide a wide variety of tutorial and searching aids, and give access to a much wider range of files and to files with a much greater time span. Each of the existing modules needs considerable improvement, and new modules are needed. Terminals must be widely available, and the "work stations", of which the terminal forms one part, need to be designed to take careful account of the "human factor." The various modules need to be available for longer hours of operation. The mail feature must be completely "real time," and all other files must be kept as current as possible. In times of crisis, it must be possible to call up SAFE modules at unscheduled hours, including weekends.

The data gathering phase of the SAFE project must be judged extremely successful for a number of reasons. First, the feasibility of a system of this type has been clearly demonstrated in an operational environment. Second, user reaction has been quite positive and frequently enthusiastic. Third, the potential value of such a system in intelligence activities has been proved, at least in a preliminary way. Fourth, and in some ways most important of all, we now have a fairly clear picture (from all the accumulated data and surveys of user reaction) of what an "ultimate" SAFE system must look like. The implementation of such a system on a wide scale will be expensive. On the other hand, there is already considerable evidence to suggest that the system could make a very significant contribution to the work of the Agency. By having access to intelligence materials more rapidly than ever before, by having immediate access to an extremely wide range of such materials, and by having the capability of searching these materials at levels of specificity and complexity never possible before, the conscientious analyst will be better informed than he has been in the past. Improved intelligence production must inevitably result.

Evaluation

SAFE has been made available in a "pilot" or "data gathering" mode, as described elsewhere in this report, with the following objectives in mind:

1. To determine the attitudes of analysts towards SAFE in general and to the various separate features it provides;
2. To determine how useful a system of this type would be to production analysts; and
3. To collect the data needed in order to move the system beyond the experimental and into a more fully operational mode.

For this last purpose a vast quantity of data have been collected on how the system has been used, how frequently, with what degree of success, what its failures and limitations are, what features analysts would need in an expanded system, volumes of documents, of searches, times involved, and so on. This section is concerned more with the first and second objectives noted above, i.e., with the evaluation of the SAFE pilot operation in terms of its value to intelligence analysts, and the reaction of those analysts to its use.

To evaluate the SAFE pilot operation, a number of data gathering procedures were used. The most important for the purposes of this section were the basic Data Collection Forms (Search forms), Finished Intelligence Citation forms, interviews with

analysts, and the analysts' reports. The data presented are based largely on 423 completed Search forms (a cutoff date of 26 July was imposed), 51 completed interviews with SAFE users, reports from participating branches, and forms recording use of SAFE in intelligence production.

Results from "Record of Search" Forms

Data were summarized from 423 forms completed by 26 July 1974. Six branches are represented in this aspect of the data gathering, as shown in Table 4. The volume of use of various files, as reported in these forms, is shown in Table 5,[10] with a breakdown by participating branch. Note that the 423 forms accounted for over 500 individual file uses. This is because some searches were conducted over several files although the results were reported on a single form. As a category, the full text files were used more extensively than the CRS (AEGIS) index extracts or the branch files. The most heavily used file of all was that of State cables, with 109 uses reported on 423 forms.

The most complete set of results from the Record of Search forms is presented in Table 6. This Table shows, for each of nine files, how frequently the file was used, by

Table 4

Return of Data Forms by Participating Branches[1]

Branch	Number of Forms
OCI/ARM	139
OER/D/TA	40
OSI/EB	9
OSI/SSB	166
OSR/SEC	47
OSR/SF/C	22
Total	423

[1] All forms containing some useful data that were received before the cutoff date of 26 July 1974.

Table 5

Volume of Use of Various Files

Branch	Branch File	CRS Index Extracts	State Cables	SI	IR	Mil Cable	Nuclear	Mail	MBFR	OAK
						File (N = 538)				
OCI/ARM	....	22	63	28	26	11	1	1	....	....
OER/D/TA	3	8	7	4	5	....	....	....	....	....
OSI/EB	2	7	....	....	....	....	....	....	....	....
OSI/SSB	18	87	38	33	9	10	23	12	....	3
OSR/SEC	38	43	1	3	2	1	....	....	1	....
OSR/SF/C	28	....	....	....	....	....	....	....	....	....
Totals	89	167	109	68	42	22	24	13	1	3

NOTE: The number of file uses exceeds the total number of data collection forms (Table 4) because some searches, recorded on a single form, were conducted over several files.

[10] Because the data collection period ended before the FBIS field traffic file was made available, this file is not included in Tables 5-7 and 11.

Table 6

Search Results Presented by Data Base Used

Data Base	Number of Searches	Number of Branches	Value of Search				Were Citations Viewed On-line?		Were Hits Printed at Terminal?		Were Hits Printed Off-site?	
			Major	Consid-erable	Minor	None	Y	N	Y	N	Y	N
Military cables............	22	3	....	5	9	6	12	5	3	12	0	16
DoD IRs................	42	4	4	17	18	5	39	1	21	17	0	38
SI.....................	62	4	6	19	18	9	44	6	23	25	2	40
State cables..............	102	4	5	42	21	17	70	19	46	39	10	71
CRS index extract file.....	78	5	10	42	12	12	59	16	31	42	20	55
Branch files..............	61	5	8	35	12	8	54	7	27	34	6	54
Mail files................	13	2	1	....	4	3	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.
OAKs...................	3	1	....	1	1	....	2	0	0	2	0	2
Nuclear file..............	21	2	2	6	6	6	15	5	8	10	1	18
Total................	404 [1]	....	36	167	101	66	295	59	159	181	39	294

[1] "Search," in this table, refers to the use of a data base by an analyst in order to satisfy a particular need for information, thus, any one "search" may include several separate sub-searches (i.e., individual strategies). This explains the discrepancy between this table and Table 5, in which each separate strategy is regarded as a single search. Note also that some data are missing from this table (data that were not supplied by the analyst).

how many branches, the analysts' judgment of the value of the search results, whether citations were viewed on-line, and whether the relevant items (hits) were printed at the terminal or off-line. The data in Table 6 are from 404 separate searches, each search representing a single information need.

Overall, somewhat more than half (203/370) of all searches for which "value" data were supplied by analysts were judged to have been of either major or considerable value. Note that these value judgments vary with the individual file used. Searches in the CRS index extract files were, on the whole, judged of high value, 10/76 (13%) of major value, and 52/76 (68%) of either major or considerable value. Searches in the file of State cables also scored well on this value scale, 47/85 (55%) being judged of either major or considerable value. In contrast, only 25% (5/20) of the searches in the military cables were judged of considerable value and none was judged of major value. The sample in this case was quite small, however. It is also worth emphasizing that a particular search may be judged of no value for many different reasons. For example, an analyst may be looking for information that does not in fact exist, or for information that does not exist in the particular file he is consulting. If he finds nothing, he will judge his search of no value. However, it must be recognized that the system behaves perfectly if it retrieves no references when no relevant references exist. In other words, some of the searches shown to be of no value in Table 6 are of no value for reasons that have nothing directly to do with SAFE as an information retrieval system.

Also from Table 6, it can be seen that SAFE appears to be mostly used for relatively short searches that can frequently be satisfied simply by viewing citations at the terminal. In less than half of all searches were citations printed out at the terminal and only a small number of searches (39/294) resulted in a request for an off-line printout.

Table 7 presents further data on the value of the SAFE system. In 63% of all searches for which these data were reported, the SAFE files were judged to provide information that would have been difficult or impossible to locate in other ways within

the required time. In most cases the analyst indicated that the search would have been difficult, rather than impossible without SAFE. Note that the CRS Index extract files appear particularly useful in allowing searches that would have been extremely difficult to conduct in other ways. The two cable files were judged of least value in this respect. The broad purposes for which SAFE files were used are shown in Table 8, which is self-explanatory, and the time spent on SAFE searches in Table 9. As anticipated, the majority of on-line searches were of relatively short duration. In fact, the great majority of searches in any on-line retrieval system should be 15 minutes or less. It is likely that, with increased experience and improved system reliability, a higher proportion of all SAFE searches will fall in the "up to 15 minutes" category.

Table 7

Did SAFE Files Provide Information that Would be Difficult or Impossible to Locate in Other Ways?
(By File Searched [1])

	Yes [2]		No		
File	No.	%	No.	%	Total
CRS index extract.......	51	76	16	24	67
Branch files (includes mail files)...........	45	66	23	34	68
State cable............	54	53	47	47	101
SI....................	37	66	19	34	56
DoD IR...............	25	66	13	34	38
Military cable..........	10	48	11	52	21
Totals...............	222	63	129	37	351

[1] A few files for which very few data exist have been omitted. Data from six branches are included in this table.

[2] For eighty-five searches only, the analysts indicated whether the search would have been (a) difficult or (b) impossible. "Impossible" searches were in the ratio of 11/85, or 13%.

Table 8

Purpose for Which SAFE Files Were Used [1]

	Branch					
Purpose	OCI/ARM	OSI/SSB	OSR/SEC	OER/D/TA	OSR/SF/C	OSI/EB
"Mail check" (current awareness of new receipts)................	32	40	....	8	....	....
Topical substantive information need........................	85	75	34	30	14	7
To locate specific "known message".	21	32	2	5	2	1
To compile a bibliography........	....	2	....	2	....	....
To prepare an intelligence report ...	....	2	....	....	....	....
To check completeness of branch/ individual files................	....	4	....	....	....	....
To support writing of collection requirements..................	....	4	....	....	....	....
Other uses.....................	5	6	3	....	3	....

[1] These categories are not mutually exclusive. That is, a particular search might be conducted for more than one reason.

CONFIDENTIAL

Table 9

Time Spent in On-line Search (N = 304)

Time (Minutes)

Branch	Up to 15	16–30	31–60	Over 60
OCI/ARM......	50	43	27	....
OER/D/TA.....	11	4	7	....
OSI/EB........	1	1	2	....
OSI/SSB.......	48	35	22	3
OSR/SEC......	11	12	8	1
OSR/SF/C.....	10	6	1	1
Totals.......	131 (43%)	101 (33%)	67 (22%)	5 (2%)

(76%)

Table 10

Response Deadline By Branch [1] (N = 329)

Information Needed Within:

Branch	Minutes	Hours	Days	Weeks
OCI/ARM......	57	60	14	2
OER/D/TA.....	8	12	5	1
OSI/EB........	0	1	3	4
OSR/SEC......	5	6	2	6
OSI/SSB.......	17	30	41	36
OSR/SF/C.....	8	6	3	2
Totals.......	95 (29%)	115 (35%)	68 (21%)	51 (15%)

[1] Numbers under each time category represent number of searches.

Tables 10 and 11 present data indicating how quickly information was needed for the various uses of the SAFE system. Note the wide range of response deadlines. While most information needs have a deadline of minutes or hours, some are longer-term and can be satisfied in days or even weeks. It is in the rapid response situation (information needed in minutes or hours) that SAFE is likely to be of greatest value. As expected, there is a variation in the required response time by branch (Table 10); there is a less pronounced variation by type of file (Table 11).

Finally, Table 12 presents data on analyst reaction to the speed of SAFE searches. An overwhelming majority of respondents felt that finding information through SAFE was faster than finding the same information through search of manual files.

Table 11

Response Deadline by File
(Number of Searches in Each Category)

Information Needed Within:

File	Minutes	Hours	Days	Weeks
CRS index extract.....	5	19	21	14
State cable..........	33	33	13	20
SI..................	14	25	12	11
Military cable........	7	7	4	1
DoD IR.............	15	12	5	2
Nuclear.............	5	2	9	4
Branch files..........	19	20	9	4
Mail (including SAVE) .	2	1	5	1
OAKs..............	....	2	1	....
Totals [1]...........	100	121	79	57

[1] The totals in this table are greater than those in Table 10 because a search, as identified in Table 10, could involve more than one file.

Table 12

**Speed of Searching SAFE Files as Compared with the Search
of Manual Files for the Same Information**

Number of Searches for Which Data are Available	Number of Branches Represented	SAFE Faster	Speed Same	SAFE Slower
87..............	6	79	3	5

Results of Interviews with SAFE Users

Between 22 July and 16 August 1974, interviews were held with 51 SAFE users, ranging from branch chiefs to secretaries, having experience with the system. Both pilot branches and self-help offices were included in the interviews. Most of those interviewed were production analysts. The interviews were highly structured, following a carefully prepared interview form. These forms may be found in Appendix III of this report. The results of these interviews are summarized as follows:

Potential Value of SAFE as Demonstrated in Data Collection Phase

No Value				Major Value
1	2	3	4	5
0	1 (3%)	6 (15%)	14 (36%)	18 (46%) (n = 39)

Close to 50% of all respondents judged SAFE as potentially of major value to their work and a very large majority (32/39) gave the system a high value on the five-point scale.

Ranking of Importance of SAFE Features

Users were asked to rank six features of SAFE in order of importance. The final ranking was obtained by taking the rank positions (on a six-point scale) assigned to each feature by individual respondents and averaging them. The lower the final figure the higher the overall ranking.

	Number of Times Ranked in this Position						
Feature	Highest (1)	(2)	(3)	(4)	(5)	Lowest (6)	Overall Rank
Access to Branch (incl. SAVE) files	14	6	5	3	2	1	1
Access to full text message files	7	9	9	2	4	2	2
Viewing full text mail	7	1	7	8	7	1	3
Viewing segment mail	3	6	7	7	6	2	4
Access to CRS index extracts	5	6	4	6	5	6	5
On-line data entry	3	7	3	3	2	14	6

39

Use of Branch Files

Respondents using branch files:		Yes	27
		No	21
Frequency of use:		Daily	5
		Weekly	13
		Monthly	9
Does SAFE allow searches that would be difficult		Yes	26
or impossible to conduct in manual files?		No	1
Does SAFE allow searches that, while theoretically		Yes	24
possible in manual files, might not be conducted because of the time involved?		No	3
Can SAFE files be searched faster than the manual		Yes	23
files they replace?		No	4
Are there any additional advantages offered by SAFE		Yes	24
in handling branch files?		No	3
Are there any disadvantages associated with the		Yes	22
use of SAFE with branch files.		No	5

In discussing the type of search that can be handled easily by SAFE but only with difficulty in manual files, most analysts referred to multi-aspect searches (e.g., searches involving a relationship between two or more countries) or really comprehensive searches on a particular subject. Others mentioned time-related searches, and some pointed out that SAFE is really useful when searching for topics that may be of secondary importance in a document and thus not reflected in the arrangement of manual files. Greater depth of indexing and specificity of retrieval were also mentioned as advantages of SAFE.

SAFE was generally judged much faster than manual files for most types of searches, especially for comprehensive searches. For very simple searches SAFE may offer no advantage over manual files as far as speed is concerned. It was also mentioned that the value of the mechanized system increases greatly as the size of the files increases.

Several users were able to identify additional advantages of SAFE in the handling of branch files: reduced storage space, greater accessibility (of all materials to all analysts), a single convenient source for searching, the discipline imposed upon the analysts (who must read and understand a document before they can index it), simultaneous file access by multiple users, less likelihood of "losing" documents, and the ability to produce printouts of document citations.

Several disadvantages of SAFE were also mentioned: delay in updating the files, the complete reliance on microform which is judged less convenient to read than paper copy, the reliability and availability of the system (including terminal availability), the time involved in indexing, coding and other input operations, the need to construct a formal Boolean search strategy, the time involved in learning how to use the system, the present restricted scope of the system in terms of the number of sources included, the necessity of going to a second location to obtain full copy of many documents (i.e., where the text is not available digitally), and the cumbersome quality of the present log-on procedures.

Use of CRS Index Extracts

Number of respondents who have made use of the CRS index extract files:		28
Frequency of use:	Daily	2
	Weekly	11
	Monthly	15

40

Value of this SAFE feature:

No Value				Major Value
1	2	3	4	5
1	6	5	10	6 (n = 28)
(4%)	(21%)	(18%)	(36%)	(21%)

Does respondent expect to use this feature frequently
in the future?

Never				Very Frequently
1	2	3	4	5
0	9	10	6	3 (n = 28)
	(32%)	(36%)	(21%)	(11%)

In using CRS extract files does respondent ever
find documents he was not previously aware of?

Never				Very Frequently
1	2	3	4	5
1	3	10	10	4 (n = 28)
(4%)	(11%)	(36%)	(36%)	(13%)

Has respondent encountered problems in the use of CRS extract files?	Yes	20
	No	8
Does respondent prefer to search this file himself, on-line, or to delegate the responsibility to a member of the CRS staff?	Self:	20
	Delegate to CRS:	8

Various problems in the use of CRS extract files were identified by respondents. Lack of familiarity with CRS indexing policy, lack of consistency and general "shallowness" or generality of indexing were frequently mentioned. Other problems mentioned were delays in getting material into the system, the need to obtain full text rather than an indexed record only, the complexity of the required search logic, the general availability and reliability of the system, slowness of the search (AEGIS rather than RECON software), the time required to learn how to use the system, and the need to input a search strategy several times for different years of the file (because of the present file organization).

By a wide majority (20/28) SAFE users prefer the ability to search CRS files themselves, rather than delegating this task to a CRS analyst. The reasons given span the full range of the advantages of on-line systems: response is faster, system is more convenient because it is immediately at hand, the search can be interactive and will permit browsing, and the analyst is spared the necessity of trying to convey his information need to someone else. The few who do prefer to delegate the search to CRS cite as reasons the saving of their own time and their feeling that CRS analysts know the data base better and are thus better able to conduct a comprehensive search.

Use of Full Text Files

Number of respondents who have made use of the
full text search capability: 29
Frequency of use: Daily   10
                                                            Weekly   12
                                                            Monthly   7
Text files by frequency of use: State cables   27 users
                                       SI messages   21 users
                                       DoD IRs   21 users
                                       Military cables   18 users
                                       FBIS   11 users
                                       Nuclear   9 users
                                       OAKs   3 users
                                       MBFR   1 user

Ranking of text files by value to respondents:

File	Rank[11]
State cables	1
SI messages	2
DoD IRs	3
Military cables	4
FBIS	5
Nuclear	6
OAKs	7
MBFR	8

Value of the text search feature of SAFE:

No Value				Great Value
1	2	3	4	5
0	1	6	10	12 (n = 29)
	(3%)	(3%)	(38%)	(56%)

Expected frequency of use of text search feature:

				Very
Never				Frequently
1	2	3	4	5
0	1	6	10	12 (n = 29)
	(3%)	(21%)	(35%)	(41%)

Success in the use of full text files:

Very				Very
Unsuccessful				Successful
1	2	3	4	5
0	4	9	8	7 (n = 28)
	(14%)	(32%)	(29%)	(25%)

---

[11] The ranking was achieved by adding, for each file, the rank positions assigned to it by individual analysts and averaging these. The lower the resulting score, the higher the rank of the file.

Has respondent encountered problems in searching text files?	Yes	14
	No	14

Major problems relate to the comparative slowness of the search, lack of standardization of zones from file to file, the inability to input one strategy for searching all text files, garbles occurring in text, the inability to suppress the lengthy first zone (header data) in messages, the lack of a "window" around hit terms, the inconvenience of the file organization (analysts want to search the latest entries first), system reliability, the lack of complete up-to-dateness in the files which can run from 3-24 hours behind real-time (one respondent indicated that the major value of SAFE lies in "crisis management" and that messages should be available for searching at the same time they reach the Operations Center), the lack of synonym tables, and the search format used.   In the interview, SAFE users were asked if other types/sources of text files would be of value. Twenty-nine respondents indicated an interest in additional text files. The most frequently mentioned source was DDO reports, i.e., 00-B and CS reports, (requested by 28 of the 29 respondents). The *New York Times* Information Bank (not full text) was mentioned six times and "wire services" four times.

Use of Mail Files

Number of respondents who have made use of
mail files: 25

Value of this feature of SAFE:

No Value				Great Value
1	2	3	4	5
2	2	3	9	9 (n = 25)
(8%)	(8%)	(8%)	(12%)	(36%)

Has on-line access to mail any advantages over present receipt in paper copy?	Yes	20
	No	5

Advantages of on-line access:

Faster receipt	17
Faster internal disposition	12
Easier disposition	11
Reduction of paper handling	12
Other	9

Degree of success in the use of SAFE
mail files:

Very Unsuccessful				Very Successful
1	2	3	4	5
0	6	6	6	7 (n = 25)
	(24%)	(24%)	(24%)	(28%)

Problems encountered in the use of SAFE mail files:

Legibility	3
Screen-fill speed	17
Scanning time	16
Other	15

In addition, a number of respondents suggested other advantages of handling mail in the SAFE on-line mode, including: greater assurance of seeing a relevant message (not otherwise disseminated to him, or lost in internal routing), analyst can work at his own comfortable pace (knowing that he is not withholding the document from someone else), ease in filing, and tendency to view much more mail of potential interest than is likely with manual dissemination. It was also pointed out that an important point in SAFE's favor is the availability of weekend mail on Monday morning. The segment display feature of the mail file (OLTA II—DEMON) was generally received with enthusiasm.

Several disadvantages of SAFE were also mentioned by various users of the mail files, including: lack of an upper/lower case capability, the commands to route and view messages take too long to execute, zone 1 of messages is too long, dissemination profiles should be at branch or analyst levels rather than at division levels, too much typing is required of users (there is need for more function keys and/or use of a light pen), it should be possible to skip "empty hours" automatically and to abort a particular screen display at any point in the screen filling cycle (to move to the next message), and the capability for producing hard copy printout, when needed, should be improved. One analyst expressed some doubt about the integrity of SAVE files; he felt that some messages he had asked to save had, in fact, been lost.

On-Line Data Entry

Twenty-six of those interviewed have had experience with this aspect of the system, either directly or through supervising others.

Assessment of the Value of this Feature

No Value				Great Value
1	2	3	4	5
0	0	1	9	16 (n = 26)
		(4%)	(35%)	(61%)

Advantages claimed for on-line data entry are: items get into the file more quickly, paper handling is reduced, secretaries prefer it to the filing of paper copy and some feel it is faster, accuracy of input, close analyst control of input leads to better and more efficient retrieval, error correction is easy, and file maintenance is fast.

System reliability and the availability of terminals were the major problems cited. Some analysts and secretaries feel that the process is more time-consuming than input to manual files (there is some difference of opinion here). Others complained of the inability to call back a record immediately once it has been entered ("paging backwards").

Use of OLTA III (input)

Only eighteen of those interviewed claimed to have had any experience with OLTA III (viewing of mail combined with on-line input). Most of these, together with a few who claimed no experience, answered several of the specific questions on the basis of seeing a demonstration of this capability (See Table 13).

Overall Rating of OLTA III

No Value				Great Value
1	2	3	4	5
0	2	3	5	8 (n = 18)
	(11%)	(17%)	(28%)	(44%)

Table 13

Rating of Various Characteristics of OLTA III

(All on a Five-point Scale)

	Little Value	Scale of Importance			Great Value
	1	2	3	4	5
Indexing a record by adding codes (n = 17).	3 (18%)	....	1 (6%)	3 (18%)	10 (58%)
Indexing a record by extracting keywords (n = 17).............................	....	2 (12%)	3 (18%)	3 (18%)	9 (52%)
Extracting passages for storage (n = 17)..	2 (12%)	1 (6%)	2 (12%)	3 (18%)	9 (52%)
Entering comments (n = 17).............	2 (12%)	1 (6%)	2 (12%)	4 (23%)	8 (47%)
Originating file records (n = 17)..........	1 (6%)	3 (18%)	2 (12%)	5 (29%)	6 (35%)
Editing messages before filing (n = 17)....	7 (41%)	1 (6%)	3 (18%)	....	6 (35%)

General Reaction to SAFE and Problems Encountered In Its Use

General Attitude Towards Move in the Direction of a "Paperless" Operation:

Very Negative				Very Positive
1	2	3	4	5
0	2 (4%)	10 (22%)	16 (35%)	18 (39%))n = 46)

Only two respondents expressed a negative attitude toward the paperless environment. Some of the reasons given were rather nebulous, relating to the greater "confidence" engendered by paper, and a feeling that information transmitted by paper copy is easier to absorb and to remember. Another objection relates to the desirability of having paper copy to annotate (although this can also be done at the SAFE terminal).

Problems Encountered with Various Facets of the SAFE Operation:

Problems	Yes	No
1. System availability	46	5
2. Use of terminal	15	35
3. Log-in and log-out procedures	14	37
4. Query formulation		
AEGIS	14	23
RECON	6	17
COLTS	6	25

The major problem in the entire experimental phase has been the general reliability of the system, especially in the early months of the project. It is clear that, when fully operational, the system must be available for longer hours and must, through appropriate backup facilities, be fully reliable in operation. These problems are discussed elsewhere in this report and need not be elaborated here. Comparatively little trouble was encountered with the terminal as such, although terminal availability was a problem in many cases (i.e, not enough terminals were available when needed to carry out all the functions that SAFE intended to provide).

Some specific problems were reported in relation to the terminal, and several of these have been mentioned already. They include the slow screen fill, the need for more function keys, the need for a light pen to facilitate selection and extraction, the feeling that the keyboard is too "cluttered", lack of synchronization between terminal and printer, the difficulty of taking notes at the terminal (because no facility was provided), and eye strain caused by prolonged use (a complaint from only a very small minority of users).

Some problems with the log-in and log-out procedures were also encountered. The following problems were mentioned or suggestions made: each step of the process should be separately validated (rather than waiting until the end of the operation), the log-on and security check takes too long, there should be a common sign-on procedure for all SAFE programs, it should be possible to switch from one mode of use to another without repeating the log-in procedure, and the "file menu" should be capable of being suppressed.

Some dissatisfaction was also expressed with the variety of query languages in use. There is a widespread feeling that a common query language should be applicable to all types of files. In general, after users had had some experience with it, RECON was considered a great improvement over AEGIS because of its interactive character. On the other hand, some users preferred AEGIS because, once the search statement had been entered the user could leave the terminal. In this case, however, the on-line facilities are not being used to their best advantage, i.e., in an interactive heuristic, manner. The COLTS language was also readily accepted by most users. Indeed, very few problems were reported with this full text mode of searching. Many users feel that the AEGIS language is too complex and that it is too easy to make errors when using it. An error usually means that the user must enter the entire strategy again. The major criticism of RECON is that the user manual supplied is inadequate and, indeed, there seems to be some general feeling that SAFE user aids of this type could be improved. Few specific objections were made about the COLTS language although one user suggested the need for a "greater than" and "less than" search capability, and some concern was also expressed about the speed of searching, especially if larger text files are used. Again, it was suggested that the "zone 1" of messages should be suppressed. The use of the various message zones created problems for some users. Some other users felt it should be possible to modify a COLTS query without the use of so many commands and that multi-character commands should be replaced with function keys. The need for a "don't care" character was also mentioned.

Learning factors associated with SAFE use: all 49 users who responded to this question indicated that they found SAFE easier to use as they gained increased experience.

Other Problem Areas and Suggestions for Improvement

Two open-ended questions in the interview were used to identify additional problems encountered in the use of SAFE and to gather further suggestions as to how the system might be improved. Some of the more important of these are listed below:

1. The ability to enter a search strategy once against multiple files is needed. Alternatively, multi-source files must be created.
2. Files are not updated frequently enough.
3. The printer is too slow and has poor legibility.
4. Terminals should have greater buffer capacities.
5. More warning is needed when the system is going down.
6. On-line access to a central microform document store is needed.
7. A single common query language is essential.
8. Faster search capability for text must be provided.
9. The file coverage needs to be improved in terms of the number of files included and the time span of these files.
10. Files should be available for longer periods each day and should be capable of being called up at other times (e.g., weekends) during a crisis situation.
11. Standardization of log-on and log-off procedures is needed.
12. It should be possible to order a document (e.g., in microform) at the terminal.
13. It should be possible to store search strategies for later use.
14. There is need for inclusion of a thesaurus, or at least a table of synonyms, within SAFE.
15. It would be desirable to be able to select certain messages to be printed from a larger set of items retrieved in a search.
16. Files should be organized so that the latest messages are searched first.
17. For text search, it would be desirable to be able to enter more than one search at a time and have two or more running at once.
18. Each work station must be designed to take human factors into account. Space for writing must be provided.
19. Improved tutorial capabilities are needed.
20. For crisis management, there is a need for "real time" receipt of messages; messages should be available in SAFE files the same time they are received in the Operations Center.

SAFE Support of Intelligence Production

Toward the end of the data gathering, a form (Figure 13) was introduced to record documented cases of use made of the SAFE system in support of intelligence production. Unfortunately, in the time available only twelve documented cases were recorded, three from OCI/ARM, three from OSI/SSB, three from OSI/NED, and three from OSR/SEC. Intelligence products resulting from these uses of SAFE included items in the Daily Surveyor, NID and SID items, a working paper, a briefing, and support provided to the Cyprus Task Force. In the case of five of these intelligence products it was judged, by the analyst responsible, that SAFE was the only source of information (in the time available), and in six additional cases it was judged that SAFE provided the information more quickly than any other channel. The full text files were the major sources used to support production of these intelligence items, with State cables contributing to five products and SI text to four.

Reports From Branches Participating in the SAFE Data Gathering

Parts of these reports are included in the next section and in full as Appendix III to this study. They stand on their own merits and do not need detailed commentary here. Nevertheless, it is worth noting that these reports reinforce the data reported from other aspects of the study. In general, they indicate a strong commitment to the SAFE concept and a desire to see the pilot system develop into a fully operational system of greater scope, sophistication and, obviously, reliability. It is interesting to note that the advantages claimed for SAFE are the advantages, presented earlier in this report, that were anticipated by the designers of the system before it was ever implemented, namely: material available more rapidly, a level of access to material that has never previously been possible, saving of space and paper handling, and the ability of an analyst to extend a search for information over many different files so that he can "bring more evidence to bear on a given problem." Particularly significant in the branch reports is the recognition that SAFE could have an important role to play in "crisis management." A hint of its potential in this respect is given in the report from OSR/SEC which indicates that, even in its present somewhat primitive form, SAFE was able to contribute to keeping OSR analysts abreast of current developments on the Cyprus crisis. In fact, it was reported that in some cases relevant NSA and State cables were available through SAFE before they were delivered to the Cyprus Task Force. The mail handling, message saving and full text search capabilities of SAFE make it potentially of great value during crisis situations.

ANALYSTS REPORTS

For the most part the reports received from the pilot and major self-help branches were divided into two sections: a general critique of the SAFE concept and its usefulness to their operation, and a section dealing with the specific improvements they thought should be added to any final system. Excerpts from their reports are included here; the full reports are found in Appendix III.

From OSI/SSB Pilot Branch:

> "We are basically quite enthusiastic about SAFE and the potential of such a system. We consider the approaches which Project SAFE has taken to be sound and sensible. We recognize that what we have seen and used thus far represents a low-cost package of hardware and software assembled for test purposes. Nevertheless, even in this highly imperfect state the system has been useful to us. We have already become very dependent on AEGIS indexing for branch records; it would be a real hardship for us to lose such a capability.

> "If the attached report at times seems to be critical of SAFE, this was not our intent. Any criticisms in the report are against the test system and are only suggestions to upgrade the test system into a really useful system.

> "It has been a real pleasure being involved in the test phase of Project SAFE and being allowed to make an input to the design of an ultimate system. It has been a special pleasure to work with your staff. Rarely have we met people who were so enthusiastic about a project and were so pleasant and cooperative. They have made an outstanding effort to meet our needs both during the test phase and in the design of an ultimate system."

From OER/D/TA Pilot Branch:

> "The year-long experiment in OER/D/TA/MILAID with Project SAFE has proven that (1) analysts can readily adapt to the elimination of hard copy receipt and processing of mail and (2) analytic capabilities can be enhanced.

The savings of time and space afforded by the system plus the rapid search capability represent a highly desirable electronic package."

From Various OCI/MEA Pilot Branch Analysts:

"It took a lot of effort to sell SAFE to analysts wedded to their paper files—which are replenished at least four times daily and on weekends, and which are unfailingly available when needed. SAFE never quite achieved a reliability factor high enough to convince doubters or satisfy converts. Few analysts yet believe that SAFE techniques will do a better job of filing and retrieving their own current material than they can do themselves with paper. But they now recognize that it can retrieve from a variety of files materials on unusual or unfamiliar combinations more surely and more quickly than they can by hand.

"We will continue to use the same features we have found most useful as long as they are available, though without the goad of logs and reports, we will do so less frequently. SAFE is generally the fastest means of retrieving a document when only a number—filing time, embassy cable number, NSA code number—is known. For the branch chief or someone filling in for a regular analyst it can also be less tedious than searching another analyst's idiosyncratic file. And in the middle of the day SAFE is worth using in the expectation that it may have items more current than our hand-delivered mail.

"My experience with the SAFE Pilot System, although limited in scope, has been positive. The system sometimes seems rather cumbersome and even frustrating, but I believe EDP definitely has a role to play in our current intelligence production. Whether the benefits will warrant the cost is for others to determine.

"The Mail file, although not directly beneficial to the PGI Branch except in a tangential way, significantly shortened the time required for dissemination of documents processed by the 'Mail Run'. (Several OCI/MEA/PGI analysts used OCI/MEA/ARM's SAFE files.)

"COLTS is of particular value as it permits the recovery of text rather than merely documentation as in the case of AEGIS. On one search, a DOD IR was recovered long before it was received in hard copy through normal dissemination.

"The AEGIS system, although limited to documentary retrieval, was of value in developing a list of sources on both the Euphrates River Dam projects involving Turkey, Syria, and Iraq and the current opium situation in Turkey. The rapid printer in the MEA Division was particularly useful. Unfortunately, sometimes acquisition of the documents themselves, after acquiring the registry numbers, largely cancels out the advantages of the rapid search capability of the machines."

From OSR/SEC Pilot Branch:

"Before becoming immersed in the minutia of evaluating SAFE hardware and software, we should step back and review what SAFE was intended to do, why and how well it has performed so far. The series of experiments which came to be called Project SAFE were first described to us in a memorandum written by Mr. Eisenbeiss, Director, CRS. Mr. Eisenbeiss stressed three main points: (1) The necessity to make improved data storage and retrieval more readily accessible to analysts had been recognized and approved at the highest levels of the Agency; (2) the experiment would be conducted in the

49

offices of real users; and (3) the users would determine, by their behavior and attitude toward them, which, if any, of the SAFE modules deserved incorporation into a new DDI-wide data handling system.

"Taking account of SAFE's experimental nature, the incremental way in which new modules were introduced and the work load imposed on SEC by world developments over the past 18 months or so, SAFE has been a solid—if not unqualified—success. Not all members of SEC used SAFE. Among those who did, the frequency varied greatly. Some modules proved much more useful than others. But there was a clear progression from the reluctance of most to learn how to use the machine to recognition by most of the great potential benefits of a SAFE-type system implemented on a wide scale. The more deeply one became involved in using SAFE, the more clearly one saw the potential, and the more willing one became to face a future in which most data would be handled electronically.

"SAFE, then, has succeeded on two fronts. It has demonstrated the feasibility of acquiring, storing and retrieving large quantities of data electronically. And, it has demonstrated that analysts will use the computer as their facility and familiarity increase and as their confidence in the reliability of the system grows. The clear opinion of all SEC analysts who have used the system is that the demonstrated utility of the SAFE modules now available— even in their relatively crude experimental form—more than justifies their retention. To give them up entirely is, by now, unthinkable.

"The benefits gained from SAFE are of two basic kinds—the ability to do more or better the same kinds of things and the ability to do new things. The most immediately evident one is the ability to store and search vastly more information than previously possible. But this uses the computer to do no more than extend the paper files. A more fundamental consequence is that with masses of data more easily available, an analyst can bring more evidence to bear on a given problem. Further, the analyst feels more inclined to check his files before writing because he knows it can be done quickly and comprehensively. Still, this is using the computer only to do what files have always done.

"An interesting effect of having files available on the computer is being able to do searches or use data in ways not previously possible. For example, by making OSR requirement numbers a searchable keyword, it is possible to use one of our branch files to easily answer such questions as what kind of information is being received in response to our requirements, for what countries, how quickly, from what kind of source and from what collector. In short, the SAFE branch file can be used to manage the collection effort on that subject in a new way.

"The SAFE experience has driven home the necessity for thorough preparation of training classes, instructors, and manuals. We are sympathetic with the problem SAS had in preparing experimental modules and simultaneously writing manuals and training analysts. But in implementing a larger system, the importance of good and continuing training on all modules cannot be overemphasized. The training program, moreover, will have to cope with the reality that most analysts will have little familiarity with the computer, will be at best ambiguous in their response to a highly automated data handling system and will be highly individual in which modules they learn to use first and which they ignore.

"A large scale SAFE-type system will represent a very substantial change in the working environment of nearly all analysts. It will force changes in the many ways analysts have maintained their files and organized their time. If the DDI-wide system is to be implemented efficiently these human factors and training requirements deserve full consideration and advance planning. Both problems may be better handled if computer-assisted learning techniques are used to allow analysts to approach the new system more nearly on their own terms and pace."

From an OSR/SEC Analyst:

"In response to your request for an evaluation of the SAFE system, I offer the following personal observations. During the Cyprus crisis and more recently in relation to events in the Balkans, I had an opportunity to use the SAFE system in a crisis management mode. The system proved to be an extraordinarily useful device in this respect. The mail distribution system (OLTA) and COLTS were of particular importance. As you know, the ultimate objective for crisis management systems is a real-time capability for information processing and distribution. While SAFE does not have this capability, it brings us much closer to it than in the past. During the Cyprus crisis, SEC was able to receive relevant reports thru the OLTA system many hours before the reports were available in hard copy. This capability allowed us to stay well ahead of possibly threatening developments and, in fact, alerted us to potentially interesting developments in the Balkans before reports of this were available thru regular channels.

"I believe that the SAFE system has an enormous potential for crisis management. Our experience in SEC has been very positive in this regard. As improvements in the system bring it closer to a real-time capability, its importance as a crisis management tool will continue to increase."

From OSR/TF/W Self-Help Branch:

"As a branch we wish to express our satisfaction with the SAFE system in what it presently does for the branch and in what we anticipate it doing in the future. One of the most attractive features of the system is that it can be tailored to fit the user's needs and, therefore, be something more than a product of some systems analyst's intellectual exercise. This branch also has been heartened by the hard work and serious commitment of those in CRS who have assisted us in getting our SAFE program underway.

"Although this branch is a relatively late entrant to the SAFE system, the branch files—through filming—already have been reduced in size by about one-third. The branch analysts look forward to testing file accessibility and hope that it will be improved through direct on-line input and retrieval. The SAFE system holds the promise of being able to make the ever increasing flow of information a manageable phenomenon, and to help stave off the accumulation of innumerable safes with unmanageable files.

"The branch has not yet received its own computer terminal, but expects one before the end of this month. Even though this terminal will be shared with another branch, its close proximity to the branch should facilitate the input and retrieval duties of our secretary, as well as make the computer directly available to the branch analysts—presently not the case.

"Direct availability of documents and information is important to an analyst. In that respect, SAFE not only provides the promise of efficiently organized, and hopefully speedy access to, branch material, but will permit

direct access by analysts to information from other files such as the Full Text Files, CRS Index Extracts, and the Library Ready Reference File. Although the branch has not yet used these files as part of SAFE, we anticipate that direct access will eliminate the time-consuming red tape now part of retrieving documents from the files of other departments. The direct availability of information to analysts would be further assisted by continued expansion of the Full Text Files to include such material as Clandestine Services reporting and State EXDIS cables.

"To a large extent, many of the elements of the SAFE system remain to be tested by this branch. Nevertheless, our interest is high, and SAFE certainly qualifies as the best potential answer that we have seen to the problem of information storage and retrieval. In the last analysis, however, as with any such system as SAFE, the individual analyst is reliant on machines, the reliability of which leaves room for improvement. Considerable emphasis needs to be placed on improving system reliability as a first step in convincing potential users that automated information storage and retrieval systems will be able to truly satisfy their needs."

From OSR/SF/C Self-Help Branch:

"SF/C is a Self-Help user of the SAFE system. As such, we have not enjoyed the fringe benefits associated with the different modules made available to the pilot branches. Even so, I would venture to say that SF/C is more dependent upon SAFE and possibly more convinced of SAFE's indispensability than any other branch.

"SF/C's SAFE system does not merely supplement the branch research files; it is the branch research file. Briefly, here is how it works. The only documents filed in hard copy form in SF/C are those which contain maps or photos that become unreadable when apertured or microfiched. All other documents, whether they relate immediately to research under way or only might some day be of interest in our research, are converted to aperture cards or microfiche, and the original paper documents are destroyed or routed on to other components. A standard coding sheet which SF/C designed is attached to each document and routed around the branch so that analysts may enter on it the keywords found in an index of keywords created in the branch which describe the document's contents. The coding sheet and document are then routed to the branch Intelligence Assistant, who assigns an SF/C accession number to the document, inputs into the computer the reference data from the coding sheet, and files the document whether as hard copy, aperture, or microfiche. Some documents are not filed at all but merely referenced in the computer. SF/C has processed nearly 2,000 documents in this way. The file grows at an average rate of about 75 per week, and this rate will rise as the branch's research activities expand. (We are only one year old.)

"SF/C cannot operate effectively without the SAFE system. The quality of our research is solely dependent on the quality of our files. SAFE is our file system. We are striving to establish in SF/C the finest, most comprehensive, most usable repository of all-source information on command and control subjects in the intelligence community. We could not aspire to so ambitious a goal without SAFE.

"SAFE is simple to use, easy to understand, and (when the hardware and software work, which is most of the time now) instantly available and responsive to our needs. Best of all, SAFE is flexible and ready to serve the

52

analyst in countless ways which neither its designers nor SF/C analysts initially foresaw. For example, SF/C codes defectors as documents, assigning them keywords which describe their knowledgeability, including rank, nationality, branch of service, etc. Thus, an analyst can identify through our SAFE file a defector source with special knowledge on a given question and can levy requirements on him through the Interagency Defector Committee. The same procedure allows SF/C to evaluate how we have used reports from a given defector. Likewise, keywords have been devised for specially compartmented collection systems. These keywords allow SF/C to evaluate our use of these systems. We also have special keywords for retrieving documents which give good overviews of selected topics. Such documents are useful to new analysts or to individuals who drop by the branch to read up on, say, the organization of the Soviet Ministry of Defense. Instead of racking our brains or searching the computer file to identify a good basic reference document on such a topic, we simply query on the keyword for Ministry of Defense and prefix it with 'BR' (Basic Reference). Each week we discover additional ways that SAFE can serve our filing and research needs."

CONTRACTOR REPORTS

This section compares the evaluations of the SAFE system made by five contractors: Chase, Rosen, and Wallace, Inc. (CRW); Mitre Corp.; Operating Systems, Inc. (OSI); Computer Sciences Corp. (CSC); and RLG Associates, Inc. They were asked to analyze the system proposed in "Project SAFE: A Preliminary Design Report" (published 29 May 1974). If they found it feasible, they were to give suggestions for implementation; if not, they were to propose alternative designs. The purpose of the evaluation reports was to solicit ideas on design rather than proposals for building SAFE.

Four of the five contractors believe Project SAFE is feasible. An interesting correlation exists between a company's familiarity with Project SAFE and its views on SAFE's feasibility. Mitre, which has done the least work with the Agency on SAFE, said it "cannot be implemented due to the combination of rapid response time requirements and the large size of the data base." RLG, having a little more experience, said the "timing requirements for search and retrieval are not currently obtainable," but SAFE is "realizable if its development is stringently controlled and staged in incremental modules concerted with the evolving technology." The other three firms have all had much more experience with the SAFE system concept and, with a few reservations, believe that the system can be implemented.

All the contractors found difficulties in the document image storage concept (digitizing paper copy material). They estimated that between 1 and 1.5 million bits of storage per image would be needed. The problems involved in handling this amount of storage and displaying the image make this concept infeasible at the present time. As an alternative, CSC proposed a cost-effective microfilm system that can be implemented with commercially available hardware. Mitre said the CSC microfilm approach "does not seem possible" but "might be considered as a stopgap;" Mitre favored a video tape system, if the query load is not too severe. CRW, professing no expertise in the area, suggested that digitizing techniques might be possible in the future but gave no specifics. OSI and RLG did not discuss document image storage in their papers, but during subsequent discussions with SAFE personnel they concurred with the other contractors.

53

All the contractors mentioned that the SAFE terminal must be specially designed because no currently available terminals meet the system specifications. Because its design and manufacture could take considerable time, CSC suggested that choice of terminal design be one of the first decisions made.

All the contractors stated that the use of minicomputers could provide the fast response times required by SAFE. CSC, OSI, and RLG based their hardware design on minicomputers. CRW based its design on two tightly coupled main processors (computers). Mitre made no detailed hardware design but stated that "the idea of using multiple minicomputers is attractive."

Hardware Design

CRW, CSC, RLG, and OSI were allowed only one month to evaluate the SAFE Design Report and to propose hardware configurations. That time was too short for them to make a detailed study. Consequently, this report discusses their designs in general rather than in detail, and points out only those weaknesses that are serious enough to degrade the systems' functioning.

Similarities in the designs are discussed below in the section Comparative Analysis of Hardware Designs. Figure 17 is a flow chart of the hardware designs in which the computer systems are separated into five general levels: terminals, front end processors, main processors, rear end processors, and data storage. Table 14 describes the operations that occur on each level.

CRW was the only contractor to suggest a hardware system design based on two large computers rather than on many minicomputers. CRW engineers felt this design would increase transmission speed, reduce hardware requirements, and reduce core requirements for I/O (input/output) interface because of the local, direct attachment of terminals. Two large, tightly coupled computers do all the software work for this system. When a request arrives from the terminals, the computer with core space available handles it. Using a hierarchical concept, CRW ties its fastest physical storage devices to the indexes most frequently used. The levels of storage, from fastest to slowest, are: core; fixed head disks; movable head disks (with redundant storage that allows access to data through a variety of devices); movable head disks with single storage; and a destructive write device. The more general indexes are stored on the faster memory. Thus, the master indexes in core have pointers to the master user indexes in fixed head devices, which have pointers to the user SANS index in redundant storage movable head disks. With this storage hierarchy, general files (user ID index) will have a faster access than specific files, which are used less often.

Two of the contractors (CSC and RLG) discussed their opinion that a large main computer design would not work. Any large computer, however fast, must still process data sequentially, and RLG felt that sequential processing of 500 or 1,000 users' work could never produce the fast response required for a SAFE system. CSC saw two important disadvantages to a large main computer system: at some point further expansion becomes almost impossible; and the redundancy needed to insure system reliability would require an expensive duplicate system.

The CRW design (two large computers) could be regarded as a system based upon one large main computer with a backup system that is sharing the load, rather than sitting idle waiting for the main computer to fail or trying the impossible task of paralleling the main operation. In all probability, both processors would share the work equally. This is an interesting technique to get optimum use of a back-up system. It is "tightly coupled" because the computers share certain core and supervisory software. This design overcomes CSC's objections; the total system can handle more than the projected requirements, and both systems are being used.

25X1

Table 14. Hierarchy of Operations

	CRW	CSC	RLG	OSI
Terminal............	1. Terminals strictly I/O device 2. Function keys	1. Terminals strictly display 2. 3 display areas each screen, internal memory. 3. Alarms and indicators 4. Function keys	1. Medium sophisticated terminals	1. Terminal strictly I/O 2. Special function keys 3. Terminal has own internal memory
Front end processor...	None	1. Working storage 2. Primary private storage 3. Interface with analysts 4. Some query programs doing easier searches.	1. Interface with analysts 2. Do *ALL* query processing 3. Possibly interconnect all processors to work in parallel on queries	1. Interface with analysts 2. Local data and program storage 3. Validate and prepare queries for satellite processor 4. Processes data available on this level.
Main processor.......	1. All communication to terminals 2. *ALL* query processing 3. All communication to data base Note: System has two tightly coupled computers. All requests going into single queue and taken first come first serve. Will share some common memory.	1. *ALL* communication 2. Search and retrieval functions on SARDINE data base.	1. Strictly handles communication from rear end processors to front end processor.	1. Performs searches on local SARDINE records at user level. 2. Interface data routing network to terminal support processor. 3. Support file handling of mail, SARDINE records, indexes, and analyst comments for local users.
Rear end processor...	None	None	1. Handles communication to enable disks to be read in parallel and data routed to main processor.	1. Route data from one point to another. Could be from one terminal to another or from data base to terminal. 2. Interface with OJCS 3. Interface with associative file processor. 4. The associative file processor will search disks in parallel on 1000 keywords or do direct access retrieval.
Data storage.........	1. Fixed head direct storage. 56 million bytes—hold pointer files. 2. Movable head direct storage. 7 billion bytes—Hold text messages (one year's worth). 3. Mass storage. 20 billion bytes—Hold entire data base accessible 4. Off line storage—(magnetic tape)	1. Storage technology corp. (STC) five disk systems each holding 6.4 billion bytes. Storage for SARDINE records and primary data storage. 2. TBM (mass storage similar to ORACLE) holding 343 billion bytes—Secondary storage.	1. Special IBM 3330—11 disks each holding 1.5 billion bytes.	1. Disks—Each associative file processing system has 4 billion bytes of storage. 2. ORACLE for secondary storage

The RLG objection still remains, even with two large, fast computers, because each must still work sequentially to a certain extent. During peak use (especially during a worldwide crisis, when many analysts will be making complex queries) the system could slow down significantly. CRW claims to have statistical data showing where the slowdown would theoretically occur.

The CRW design has another potential weakness in its concept of tightly coupling and sharing some of the same core and software. If a problem arises in the shared resources—and these have no backup—the entire system comes down. The shared resources would require constant checking to insure reliability, which could minimize but not remove the possibility of a total system shutdown.

A final weakness, which was not discussed, was having so much storage hardware tied directly to dual computers. Tying four or five fixed head disk units, 30-50 normal disk units and a mass storage system all into one dual-computer complex would probably encounter serious difficulties.

CSC proposed a network system with minicomputers connecting the terminals to a large central computer. CSC said this "strawman" hardware configuration was "developed for discussion purposes," implying that it is incomplete. Each minicomputer has an 88-megabyte disk storage, holding the working storage and primary private files of 16 analysts. Searches on those files would occur on the minicomputer level, while all communication, search, and retrieval functions on the SARDINE data base would be handled in the main computer. (The 88-megabyte disk is made by Storage Technology Corporation, and its specifications are not currently available.) CSC proposed only five disks for holding the entire SARDINE data base and primary storage; there is some doubt that such a system could handle a peak workload, but CSC did not discuss a backup system. In short, this discussion of hardware design was somewhat superficial.

The RLG design stresses hardware rather than software: "Software technology is not currently at a state where it can provide efficient and timely access to data in a large data base that is continually changing." Believing that a single computer cannot handle the problem, RLG advocates multiple minicomputers with 20 analyst terminals tied to each. The minicomputers are all connected to one large computer. Special file and record processors are used between the disk data base and the main computer, allowing parallel transmission of data from all disk surfaces to the main computer. The main computer does nothing but control input and output, and the minicomputers do all the searches and software functions. (RLG provided an alternate configuration that would provide a parallel search capability by duplicating a query in all of the minicomputers.)

The heavy workload placed on the minicomputers in the RLG design raises a major problem: a minicomputer may not have the speed or core to execute a variety of programs simultaneously. Attaching a large disk storage and hardware text searching module (General Electric Co. makes one) to each of the minicomputers would make a more realistic system. The RLG report did not discuss a backup system, but during their briefing they said it would consist of duplication of hardware.

OSI designed the most complex hardware configuration, having a level of minicomputers attached to a level of medium computers attached to another level of minicomputers used for connecting the medium computers. This configuration permits the analyst to get at his files (stored in a disk pack associated with the minicomputer he normally uses) from any other console in the building. There is disk storage at all levels, which greatly reduces the communication required. To search larger files in the entire data base, OSI suggested an associative file processor, which it is developing, that will allow parallel searching of all disks on 1,000 keywords.

58

This system has two areas for potential degradation. The first is its complexity: the more complex a system, the more likely it is to break down, with long repair times. The OSI system is so complicated that it might take much longer to assemble than the SAFE schedule permits or than the contractor anticipates. The other potential problem is the associative file processor, which may be so often interrupted that it becomes useless. Any request for direct access to the main disk data base interrupts any associative search in progress. With 500 terminals, the number of interruptions could be significant. The seriousness of the problem will depend on the data base design and the quantity of data that can be stored at the mini and medium computer levels.

The interconnection of computers gives the OSI system an inherent redundancy and backup. Tasks can be routed around problem equipment, and the routing control processors can find alternative paths for data. The breakdown of one of these processors could interrupt the systems's bypass capabilities, but this could be overcome by additional connections.

Mitre did not present a detailed hardware design, because it judged the SAFE project not currently feasible, but discussed the subject in general terms. The contractors like the idea of using multiple minicomputers and point out that special purpose processors, such as associative processors, may be needed if bottlenecks develop. They believe that competition for communication lines will be one of the primary problems; and they discuss a special piece of hardware they are developing, a Time Division Multiple Access (TDMA) device which would facilitate parallel transmission of data.

A TDMA communication line partitions the information on the line by time intervals. A specific terminal will not read from the communication line except during its time interval. Each terminal uses one channel on the TDMA line, and Mitre can install more channels per line to increase the number of users. A significant problem with TDMA lines is the absolute necessity that time clocks in the terminals be synchronized with clocks in the computers. A slight divergence will cause one terminal to read information being sent to another.

Comparative Analysis of Hardware Designs

Each proposed design has problems that could be serious enough to degrade system implementation, but there is no doubt the contractors, with a little thought, could solve them. The purpose of the evaluation reports, however—to solicit ideas—was satisfied. The variety of designs ranged from an extensive software to an entirely hardware oriented system, and the next task was to look for common ideas among them.

A specially designed SAFE terminal is one such idea. Mitre says the hardware components are available on the shelf and need only be properly put together. CSC says no existing equipment will meet the SAFE requirements, because the terminal design needs extensive human engineering. This is because the analyst will spend long hours at the terminal and must learn to remember what he reads from the terminal as easily as what he reads from hard copy.

On the next level, most of the contractors favored minicomputers, with fifteen to twenty terminals tied to each minicomputer. Trouble in any one of the minicomputers affects only a small part of the network. Disk storage at this level, containing the analyst's personal files or files frequently accessed from that terminal, gives the response time required by SAFE. If problems arise at the main computer level, the analyst still has many capabilities left in the terminal-minicomputer network.

There is less agreement about the main computer level. The main computer serves as either a processor or an I/O controller, or both. After the main level, two designs

proposed an additional processor (rear end computer) to interface with the data base; this is needed if the data base is accessed by associative techniques. The two basic designs are: a main computer that does some processing and all the I/O and is tied directly to the data storage devices; or a main computer that only does processing and is tied to a rear end processor that allows parallel accessing of the data base.

The CRW design (tightly coupled main computers and no minicomputers) was discussed above. Because no other contractor presented this concept, we have no interplay of variants to study. Comparison of a single main computer level versus a network of minicomputers would demand far more information than that presented in the contractors' reports. An in-depth simulation and statistical analysis would be necessary to determine which is more feasible.

The data storage technologies varied with each system design. Each contractor proposed the storage system most familiar to him.

File (SARDINE) Designs

Three contractors (CRW, CSC, and OSI) discussed in detail their file design for referencing documents. Their designs are outlined in Figure 18; the arrows indicate relationships between files and the directions of the information flow.

CRW broke the SARDINE record into a hierarchical set of independent files. A search will cascade from the master index to the specific user file or public file, to the specific file for each keyword, to the master locator index, to the documents that satisfy the query. To save space, keywords are compressed during the indexing of documents. The term equivalence table lists the index terms and a number that represents each term in the users' indexes. This compresses a multi-byte word to two bytes and saves considerable space in the users' indexes. A minor disadvantage is that the system must translate all the index terms for an analyst's query into the equivalent numbers before the system can search for them in a document.

The CSC design contains a file structure (permanent document reference data set) that resembles the SARDINE record more closely than does the CRW hierarchy of files. All dissemination to the analyst index flows through the analyst control data set (which acts as a master index). Comments are not imbedded in the analyst index but are linked to the analyst control data set. The CSC analyst index contains almost the same information as the CRW master user index and user SANS index. The two exceptions are that CSC puts the mail file pointers into the dissemination index and comments into the comment index.

The structure of the OSI basic SARDINE control record resembles fairly closely that discussed in the CRS Preliminary Design Report. For each new item, a basic SARDINE control record is built that processes it into the system and a document delivery record is built that sends it to those files designated as dissemination points. As they read their mail, analysts can store selected items in files referenced by the File Header Record. Records are stored within each of these File Header Records in the Intra-File Pointer Record structure. In this structure pointers link each document to 1) its predecessor, 2) its successor, 3) document catalog information, 4) the SARDINE central record and 5) related comments.

The OSI design stores the pointers from documents to index terms, whereas the CRW design stores the pointers from index terms to documents.

File Reorganization

The large size of the SAFE data base and the requirement for a fast response time dictate a hierarchy of data storage. A more recent or more important document should be in a faster part of the storage hierarchy. To keep the documents reasonably
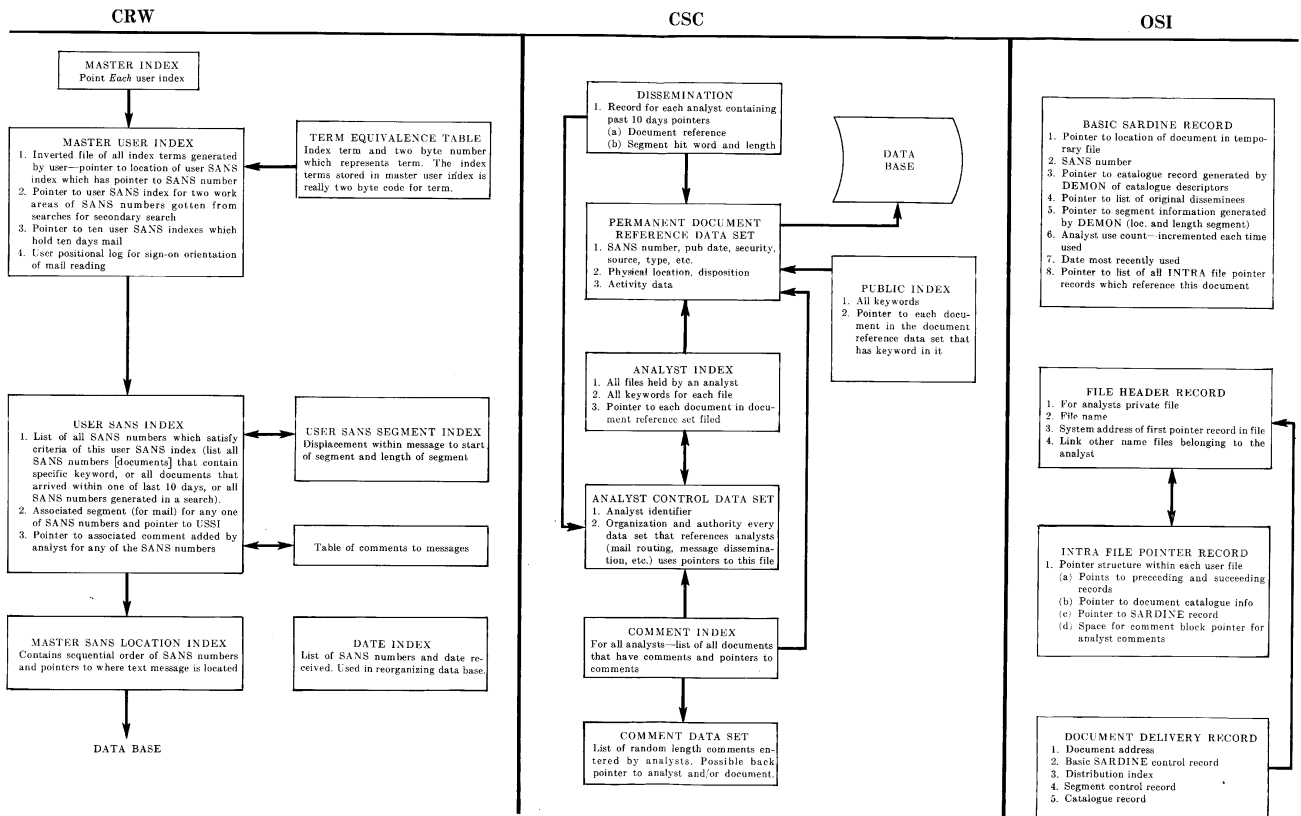
60

CRW

CSC

OSI

MASTER INDEX
Point *Each* user index

MASTER USER INDEX
1. Inverted file of all index terms generated by user—pointer to location of user SANS index which has pointer to SANS number
2. Pointer to user SANS index for two work areas of SANS numbers gotten from searches for secondary search
3. Pointer to ten user SANS indexes which hold ten days mail
4. User positional log for sign-on orientation of mail reading

TERM EQUIVALENCE TABLE
Index term and two byte number which represents term. The index terms stored in master user index is really two byte code for term.

USER SANS INDEX
1. List of all SANS numbers which satisfy criteria of this user SANS index (list all SANS numbers [documents] that contain specific keyword, or all documents that arrived within one of last 10 days, or all SANS numbers generated in a search).
2. Associated segment (for mail) for any one of SANS numbers and pointer to USSI
3. Pointer to associated comment added by analyst for any of the SANS numbers

USER SANS SEGMENT INDEX
Displacement within message to start of segment and length of segment

Table of comments to messages

MASTER SANS LOCATION INDEX
Contains sequential order of SANS numbers and pointers to where text message is located

DATE INDEX
List of SANS numbers and date received. Used in reorganizing data base.

DATA BASE

DISSEMINATION
1. Record for each analyst containing past 10 days pointers
(a) Document reference
(b) Segment hit word and length

DATA BASE

PERMANENT DOCUMENT REFERENCE DATA SET
1. SANS number, pub date, security, source, type, etc.
2. Physical location, disposition
3. Activity data

PUBLIC INDEX
1. All keywords
2. Pointer to each document in the document reference data set that has keyword in it

ANALYST INDEX
1. All files held by an analyst
2. All keywords for each file
3. Pointer to each document in document reference set filed

ANALYST CONTROL DATA SET
1. Analyst identifier
2. Organization and authority every data set that references analysts (mail routing, message dissemination, etc.) uses pointers to this file

COMMENT INDEX
For all analysts—list of all documents that have comments and pointers to comments

COMMENT DATA SET
List of random length comments entered by analysts. Possible back pointer to analyst and/or document.

BASIC SARDINE RECORD
1. Pointer to location of document in temporary file
2. SANS number
3. Pointer to catalogue record generated by DEMON of catalogue descriptors
4. Pointer to list of original disseminees
5. Pointer to segment information generated by DEMON (loc. and length segment)
6. Analyst use count—incremented each time used
7. Date most recently used
8. Pointer to list of all INTRA file pointer records which reference this document

FILE HEADER RECORD
1. For analysts private file
2. File name
3. System address of first pointer record in file
4. Link other name files belonging to the analyst

INTRA FILE POINTER RECORD
1. Pointer structure within each user file
(a) Points to preceeding and succeeding records
(b) Pointer to document catalogue info
(c) Pointer to SARDINE record
(d) Space for comment block pointer for analyst comments

DOCUMENT DELIVERY RECORD
1. Document address
2. Basic SARDINE control record
3. Distribution index
4. Segment control record
5. Catalogue record

Figure 18. File Structure

distributed among the various levels of the hierarchy, SAFE requires a method of dynamic file reorganization.

Most of the contractors said little about this important problem. CSC and OSI suggested reorganizing the file hierarchy in batch mode during the least active periods of SAFE usage. CRW is the only contractor to discuss file reorganization in detail, and the only one to propose a significant hierarchy of storage mediums (the others suggest one or two mediums).

CRW presents a valuable storage concept: temporary upward mobility. It is more important that documents located on slower storage mediums should be raised to faster mediums when analysts need rapid access to them, than the normal shift from faster to slower mediums (based on age and use) should occur. This means upward motion should be accomplished on-line at any time, whereas downward motion can be a batch function.

The complexity of file update may be reduced by having analysts' files stored locally. A significant portion of new record updating can occur at the minicomputer level. File reorganization of the master location file and the records themselves can occur on the main computer and at lower levels.

The discussion of file reorganization procedures is limited because it would have to be based on a detailed storage design at all levels and on a complete hardware design—however, the contractors did not have time to construct such designs.

Software Designs

All the contractors agreed that software must be specially designed to meet the requirements of SAFE and that altering any existing software system to try to meet them would only degrade its overall performance. The size, complexity and fast response time of SAFE will need specially tailored search algorithms and control monitors.

All the software must be completely "reentrant," that is, must allow any user to begin a search from his console, no matter how many analysts are using the same software. The master copy of each software package must be permanently available. It can be stored at the main computer, from which copies can be sent on request to the minicomputers. All software documentation must be consistent, and therefore standards must be established before any contract work begins. A package like HIPO (Hierarchical Input Process Output) could serve as a basis for such standards.

Universal Language—SQUIRL

The concept of a universal language that a SAFE user could learn and apply to any data base is attractive; but the complexities of translation involved make it impractical. The contractors suggested alternatives that, together, offer a solution. CSC suggested that a universal language be developed strictly for the software built for SAFE. OSI suggested that extensive HELP tutorials on the terminal be made available upon entering a program to guide inexperienced users through the design and submission of a query. (A set of these tutorials could be developed for any data base and new ones developed as additional data bases are added.) The more often an analyst uses a data base, the less he would need the HELP tutorials. The option of not invoking the tutorials at all must exist.

Reliability and Backup

The SAFE system *must* have extremely high reliability. None of the contractors discussed reliability in detail, though CRW devoted a few pages to hardware reliability.

63

All contractors agree that hardware must be duplicated as a backup to any system. Some backup is inherent in the hardware configurations; the CRW backup computer is working all the time. In the minicomputers, backup is achieved by following alternative routes of data flow. We must assume that almost all the hardware (channels, storage devices, connecting devices, etc.) will be duplicated.

Performance monitoring can provide one type of backup. It tabulates all operations performed; then, if the system goes down, the engineers will be able to restart it and know which operations have been completed.

The contractors' discussion of system reliability is limited partly because it (like file reorganization) would have to be based on a detailed system design, which they did not have time to draw up. The final system design must have two levels of reliability. One will assure that there is a backup for any software or hardware component whose failure could bring down the whole system. The second, and harder to achieve, will assure that even in a degraded mode due to hardware failure, the system will appear almost normal to the analyst.

System Cost

The contractors' cost figures vary widely because they are based on different hardware configurations. A comparison of the estimates would be misleading, since the final configuration has yet to be chosen, and cost is one of the criteria or considerations used in the decision. Hardware costs would range from $15 to $30 million and software costs from $8 to $12 million, if all equipment were purchased. An additional $10 to $20 million could be required for a backup system. This gives a total of $33 to $62 million, which is quite a range.

The use of hardware already in the Agency, depending upon the final design, could lower those costs significantly. Any software work done by the Agency could also reduce total costs.

Management Procedures

In addition to developing software and hardware design specifications, the SAFE task team must determine project management procedures. All the contractors agreed that implementing the SAFE system will require a high level of project management. The system must be built in stages (modules) and all stages monitored in detail. Either the Agency or a contractor must manage the entire program and subcontract as needed. Whoever is in charge, the fewer the independent software and hardware contractors building the system, the fewer the problems of coordination.

VI. PROPOSED SAFE INFORMATION SYSTEM OUTLINE

INTRODUCTION

We interpret the analysts' evaluations of the SAFE modules and SAFE concepts as a general endorsement—with qualifications, or reservations. The qualifications, which relate to system reliability, file contents, user aids, response times, etc., are being studied.

We interpret the contractors' evaluations of the technical feasibility of the SAFE concepts as a general endorsement with qualifications. These qualifications relate to the technical difficulties of digitally converting and storing data obtained on paper copy medium; the problems of response time for large files; and the inherent difficulties in the SQUIRL concept. They are being studied and are taken into consideration in the system proposed in this report.

This chapter outlines a proposed SAFE Information System that will satisfy the analysts' two fundamental needs: computer searching of digitally stored message traffic (Text Files) and maintenance of computer-based analyst files.

The proposed system resembles that system hypothesized in the SAFE Concepts chapter of this report and described in the Preliminary Design Report (Appendix V). However, because of current technical and cost restrictions, this design differs from the hypothesis in four important aspects:

1. Material received in paper copy form will be stored in microform rather than in digital form. The conversion to digital form is still an objective.

2. An item received by electrical transmission need only be stored once, regardless of the number of analysts who may have "filed" it; but, as a corollary of item 1, material received in paper copy form will have to be stored in as many microform collections as are required.

3. External files, such as the *New York Times* Information Bank, will not be a part of the present system proposal; their inclusion is still an objective.

4. The system response time (time required to complete an analyst's transaction) will vary depending on the size of the files and the "operation" being performed. The original hypothetical response times now appear impractical.

The first step in a system development program would be to design the system in detail; this design would require 4-6 months to complete. The description that follows is in three parts: System Overview of proposed SAFE capabilities; File Operation, which outlines the relationships among the major files; and Preliminary Hardware Design, which includes an estimate of total costs.

SYSTEM OVERVIEW

The system capability can be summarized by describing the SAFE Console Station (SCS), the files it can access and the processes it can perform. (See Figure 19). The SAFE system should, where practical, be integrated into the general Agency data processing environment; a SAFE terminal should be able to access other Agency data bases in addition to SAFE files.

Figure 19. Overview of the Proposed SAFE Information System
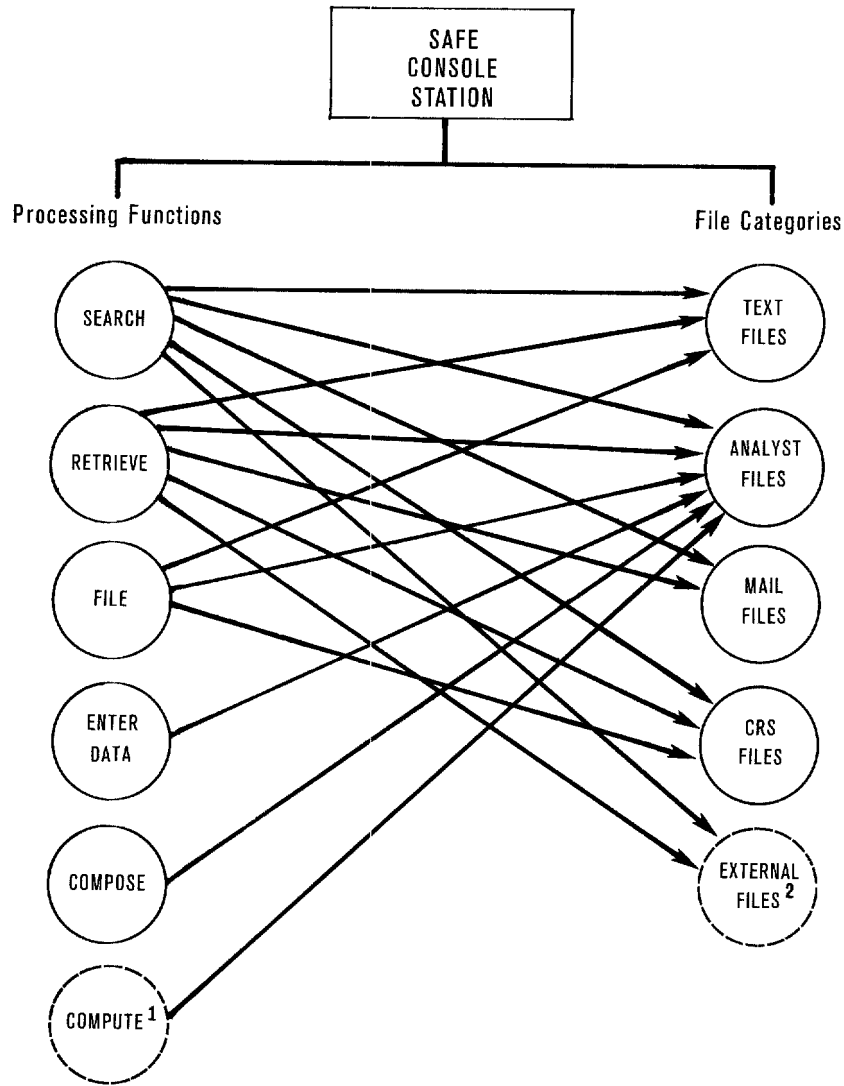
1. Compute functions not to be considered in early SAFE System but is a future objective. Compute would tie the file system outputs into existing (or new) OJCS compute programs.

2. External file not to be considered in early Safe System but is future objective.

SAFE Console Station (SCS)

The production analysts will use the SAFE Information System through an SCS. The SCS is more than a simple cathode ray tube (CRT) device. For example, it may consist of a "local" terminal (digital viewing screen and keyboard) stationed at every few desks; a digital printer reasonably close to the terminal; and a "regional" microfilm viewing screen, film storage device and printer. The keyboards will have

function keys that control the file categories to be accessed and the functions to be performed.

The viewing screens must feature readability and general ease of use consistent with today's state-of-the-art. The SCS will be designed with either two screens or a split screen, so that an analyst can view information on one part while entering data on another. The SCS will have an alerting device which will bring a predetermined "priority" message to the analyst's attention. Analysts will be advised automatically of any operating abnormalities.

File Categories

1. Text Files are those electrically received transmissions that may be processed and stored in digital form. They currently include:
   —FBIS field traffic
   —SI messages
   —DoD IRs
   —State cables
   —OAKS
   —CIA/IAS
   —Military cables
   —Wire services (Reuters, AP, etc.)
   —DDO selected information cables

These items (except for certain sensitive or highly classified items) will be held for 14 days, during which time analysts with the proper clearances can access them for processing and possible inclusion in their own files.

2. Analyst Files are those created and maintained by analysts. They may be document reference files (which contain indexes to specific documents) or information files (which contain data and may or may not refer to the source documents).

3. Mail Files are a subset of the Text Files; each mail file contains a selection of electrically received transmissions that have been processed into it by the Cable Dissemination System. A "distribution index" ties a specific message to a specific set of analysts.

4. CRS Files include the Subject Index File (two million records and growing), a major document reference system. CRS indexers select documents for indexing in this file according to predetermined criteria. Other documents of special merit may be "activated" for the system. SAFE proposes an additional selection criterion, whereby CRS will index any additional document if two or more analysts have "filed" it and if the security classification of the document permits a "public" index record. (The process is described below in the section on Indexing and Filing of Digitally Displayed Items.) CRS files will also probably include certain biographic and installation information files and certain library reference files.

Processing Functions

1. Search—Analysts will be able to perform searches on any of the above files. In the case of Text Files, they may search by specifying any word or combination of words and asking to see the documents in which they appear. The other files will have different search capabilities, but to the extent practical a common language/procedure will guide the analysts through their searching. A search in the Mail File would probably be a simple scan of items

received since the last search. Special aids will be made available to analysts who are unfamiliar with any particular file.

2. Retrieval—Documents or information that match a search parameter can be displayed on the screen and printed at the SCS. The mode of retrieval will vary depending on the file and the file storage medium. Figure 20 shows the retrieval options available.

3. File—Analysts can "file" any document being viewed on the SCS display screen, whether it is a microfilm or digital display. Table 15 shows the file options available. If the document is a paper copy receipt the filing instructions are considered to be in the Data Entry category discussed below.

4. Data Entry—Analysts may create or add to analyst files by calling up the appropriate "form" on the screen and then entering data directly on the displayed form.

5. Compose—Analysts may use the compose function to write and edit. This "document" can then be filed with other intelligence items or in a special project file to which other items can be added.

FILE OPERATION

This section describes briefly how the proposed system will work. For the most part, this description was developed from the outline contained in the more detailed Preliminary Design Report, published in May 1974.

Search and Retrieval—14 Day Temporary Text Files

Figure 21 shows the proposed schema. Digital message traffic is received after being processed through CDS (1) or other OC sources (2). This traffic is processed through the SAFE Automatic Cataloging program (3), which sets up one computer index file record (called the Basic SARDINE record) for each message. The record (4) contains the standard SAFE Number (SANS), classification, date, and file name. Messages in this temporary text file are held for approximately 14 days (5).
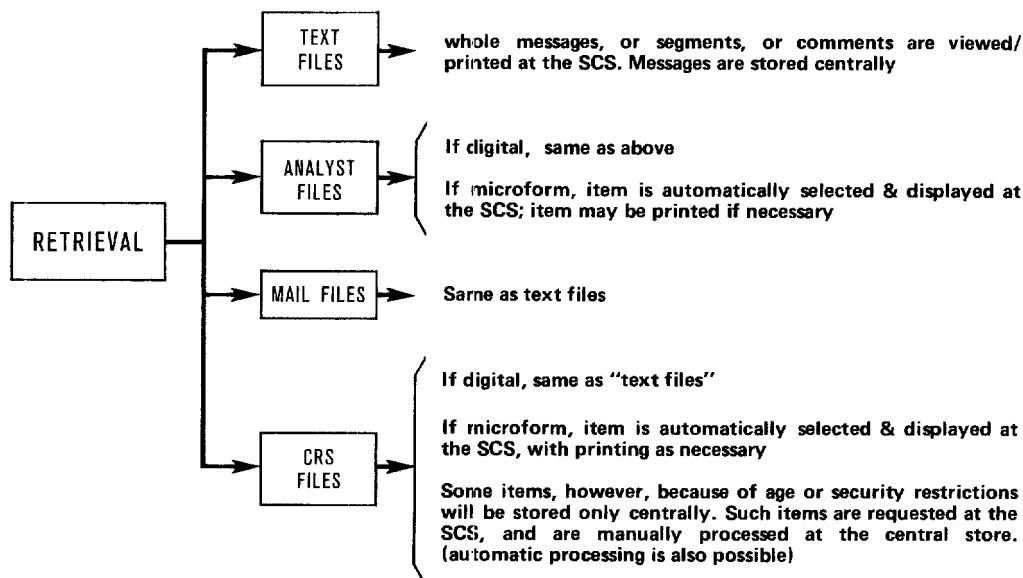
Figure 20. Document Retrieval Options for
the Proposed SAFE Information System

Table 15

Filing Options Available to Analysts as They View Documents at the SCS

Filing Option	Description	Applicability
1. Indicate file name........	Document will appear to be filed under that file name.	Microfilm and digitally displayed documents.
2. Add index terms to the document.	One or more words, or word phrases may be used to further describe the document.	Same as above.
3. Add comments..........	The analyst may add evaluative comments about the document.	Same as above.
4. Extract data...........	Analysts may extract data from the document; whole paragraphs or specific segments.	Digitally displayed data only.

When an analyst searches (6) this file, he may limit his search to any parameter he chooses, e.g., date, post number, security classification, keyword in text, etc. If the number of hits exceeds a certain level, he will have the option of refining his query to reduce the number of hits or having them printed in the OJCS center. Otherwise, he can ask for the whole item to be displayed, or he may ask for only the segment of the item that contains the search terms. He further has the option of printing (7) or filing (8).

Search and Retrieval—Mail File

When a message from CDS is routed into the temporary text file, at the same time (see Figure 22) the list (Distribution Index, DI) of who gets that message is routed to the DI file (2). When an analyst asks to search and retrieve from his mail file, this index determines what messages are sent. The analyst need only ask for "mail" to start scanning the items that have been selected for his office since the last time he viewed his mail file. The analyst can also elect to further route (8) the messages being scanned.
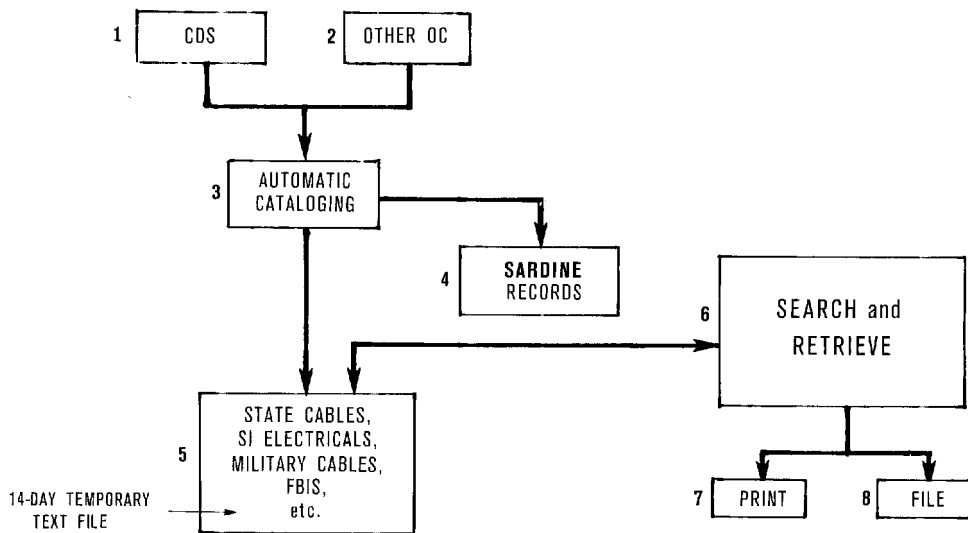
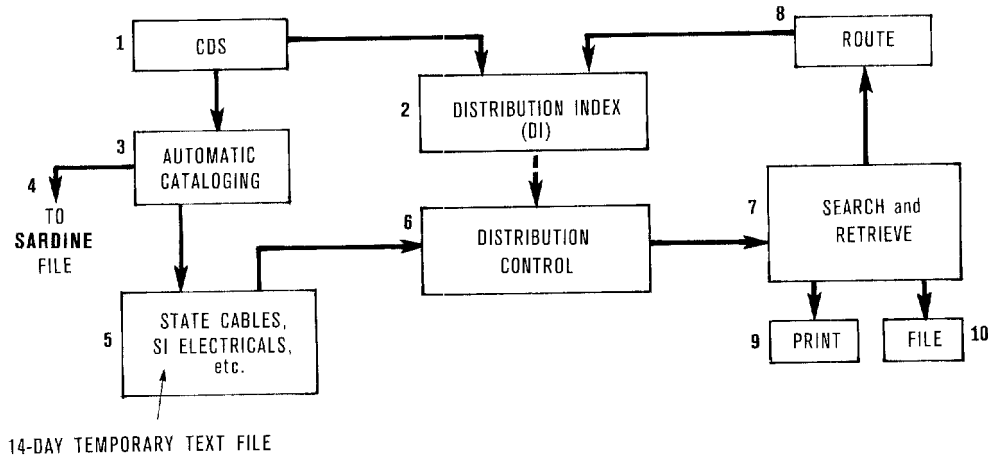Figure 21. Search and Retrieval From 14-Day Temporary Text Files

Figure 22. Search and Retrieval - Mail Files

This routing automatically updates the Distribution Index so that it will be available on some other screen—if that analyst has been cleared for the item. Analysts can also print (9) and file (10).

Indexing and Filing of Digitally Displayed Items

The creation (see Figure 23) of temporary text files (4) from OC (1,2) and the creation of the Basic SARDINE record (5) have been discussed above under Search and Retrieval of 14 Day Temporary Text Files. When an analyst chooses to "file" (6) a digitally displayed text item, what he really does is add his file criteria (be they file names, keywords, or whatever) to a record (7) associated with the SARDINE record already created for that item. He may also use a data entry form to create a comments file (8) for the text of comments he wishes to make on the document. When he next retrieves that document, his own comments (but not those of other analysts) will appear with it.

SARDINE relates the proper comment to the proper user and to the proper text document. The above connections are made as the analyst views the document on his SCS screen, and his data entry form is displayed concurrently with the message. If any analyst has added a file sub-record to the Basic SARDINE, it will affect the file reorganization (9), because after 14 days each item in the temporary text file must be moved to another storage area.

If a given item has not been put into any file, even that of CRS, then it is processed via computer output microfilm (10) to a central microform collection (11) or is processed to the lower order digital storage, the Tera-Bit Memory (TBM) (12), which may be an alternative to microform storage. The SARDINE record continues to exist for that item.

If an item has been entered into one or more files, it will be transferred to the primary text file (13). Analysts will be able to do text searching on all items so stored. Items remain stored in primary text until the next reorganization, when the date and activity of each record are automatically reviewed. If an item has not been retrieved for a given period of time, it too will be routed to microform or TBM storage and out of the more expensive digital primary text.
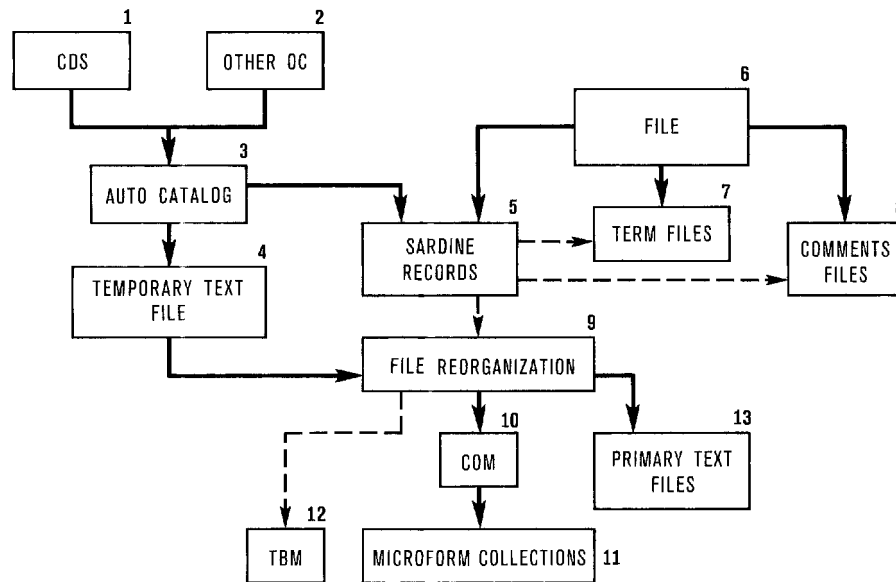
Figure 23. Filing of Digitally Displayed Items From Text Files

Indexing and Filing of Non-Digitally Received Items

In a typical sequence (see Figure 24), an analyst receives a document in paper copy form (1) and reads and marks data (2) that are to be filed. He enters the data on a form that appears as a display on the SCS (3). The particular form is tailored to the kind of file being built. Data so entered goes into term files (4) or comments files (5) as appropriate, and the location is recorded in the SARDINE record (6), which "points" to the CRS microform version of the original document (7). Whenever the SARDINE record is retrieved, it references that document.

An analyst may see only a microform copy of a document. He can still file it by following steps 2-7.

Search and Retrieval—Analyst and CRS Files

When the analyst searches and retrieves from his own or from the CRS files (see Figure 25), he uses various term files (1) and the SARDINE data structure (2) related to them. When the search is complete, he may view the SARDINE records and the term file entries that satisfy his search statement. These may themselves contain the information that answer his question, or the analyst can retrieve the pertinent documents. Documents in digital form are retrieved from a primary text file (3) or the lower-speed TBM (4) device. Once a set of these digital documents (or analyst comments (5) about them) are retrieved, they are available to the analyst in a special computer file called a "work space" (6). Documents thus retrieved can be further searched by text search techniques (7) or refiled (8). Documents in microform are retrieved from the regional storage facility (9) associated with an analyst's SCS.

Some documents will be beyond a given age limitation or will be of a special security category. Such documents must be retrieved from central storage (10). Requests can be made directly from the analyst's SCS; the documents are processed manually.
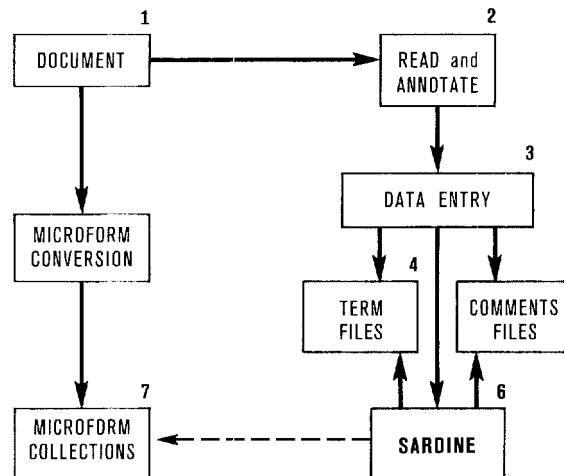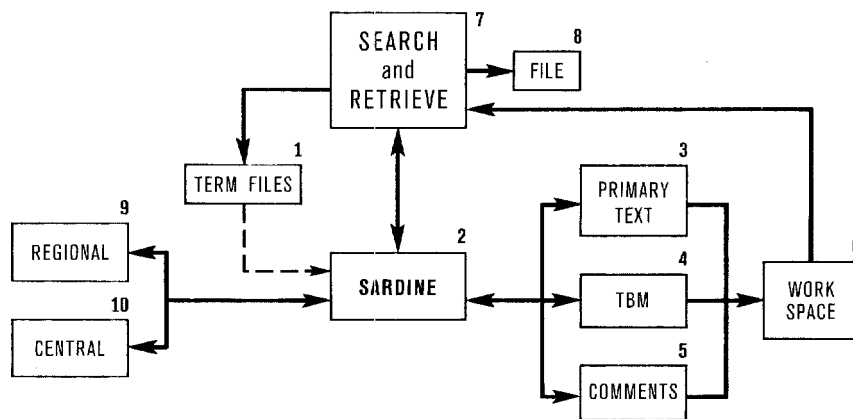
71

Figure 24. Data Entry

Figure 25. Search and Retrieval of Analyst and CRS Files

PRELIMINARY HARDWARE DESIGN

Introduction

The preliminary concepts of the system design were discussed by a joint CRS/OJCS task team, which had been directed to determine the major parameters for an updated SAFE Information System and to consider how those parameters would influence the system design. Once the parameters were established, the team considered various ways of implementing them and discussed the merits of special versus general purpose computers and of distributed versus central processing. The team decided on a distributed network of minicomputers attached to general purpose computers doing central processing.

The following, more detailed hardware design was made by a team of CRS computer specialists, based upon a consensus of the overall system configuration determined by the joint CRS/OJCS task team. This system design indicates the possible magnitude and cost of a SAFE Information System.

Because many of the SAFE requirements are still approximations, the team considered two possible configurations. The larger and more expensive one might be able to do the job; the smaller and less expensive one probably will not be able to handle peak workloads. Because of the large volume of data that will be vulnerable to both hardware and software failures, file backup and alternate routing procedures will be required at all levels of the system. In addition to backup equipment, SAFE will require processing and electronic file storage equipment to restore service after either an external problem (e.g., fire) or an internal problem (e.g., equipment malfunction) destroys some part of the electronic files in the system.

As exact SAFE requirements are derived, the detailed system design phase of the Development Plan (Chapter VIII) will determine the final system configuration, which will probably lie between the minimum and maximum configurations presented.

SAFE Configuration Description

The proposed system requires hardware for four processing levels: the analyst's console, forward processing, central processing and central microfilm storage (see Figure 26).

- Analyst's Console Level: It is proposed to install some 500 consoles, about one for every two analysts. For every five consoles (approximately) there will be a regional microfilm reader and storage device. This device will contain microfilm images of documents (nonelectrical receipts) that were filed by the analysts and a sub-set of the central (CRS) microfilm storage. The contents of this sub-set will be controlled by security and document age.
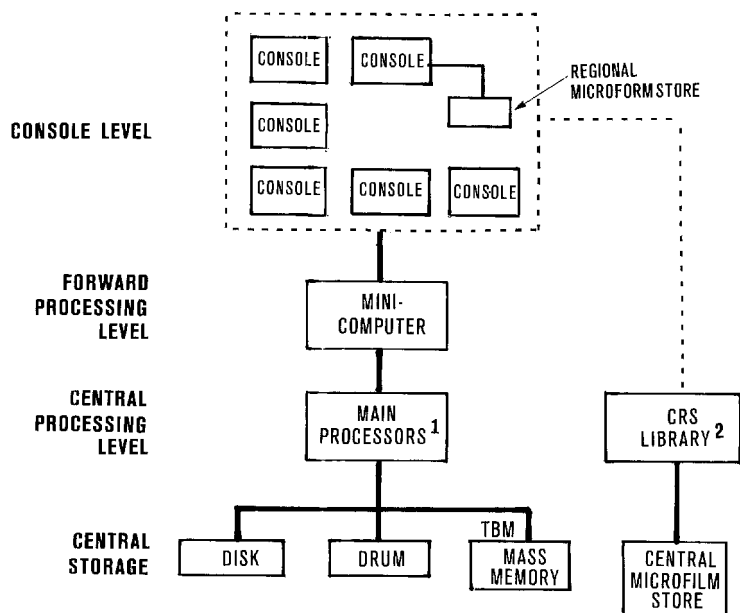
1. may consist of two general purpose main frames (small system); or may consist of four special purpose main frames (large system)

2. central processing may remain manual (low-cost system) or may be automated (hi-cost system)

Figure 26. Proposed Hardware Configuration

73

- Forward Processing Level: It is proposed to station about 50 minicomputers in the Agency, averaging one mini for every 10 consoles. This network of minicomputers allows the SAFE consoles to be less sophisticated and therefore less costly. It also allows the processing of simpler tasks (reading mail, writing and editing reports, and checking syntax of commands for errors) to be accomplished at a level closer to the analyst and relieves some of the work load on the central processors.
- Central Processing Level: The complex computer functions of monitoring the system, text searching, index searching and maintaining the data base will take place at the central processing level. The minimum computer configuration needed is two large (IBM 370/168 size) general purpose computers. All of the functions will be performed in either machine, and each will back up the other. Some members of the task team doubt that this minimum system will have enough computing power to handle the workload, especially during peak periods. The failure of either computer would seriously degrade the entire system. An alternate design uses four large computers (IBM 370/168 size). They are specialized; two maintain the data base and search text files, and the other two search the private and public index files and do text searching of the current 14 day text file. Should any one computer fail, its mate would be able to maintain the function with little or no system degradation. This system is more expensive but guarantees maximum backup and high computing speed.

  In both systems the electrically received data and index files are stored in a two-level storage heirarchy. The primary storage level consists of approximately 75 disk drives (IBM 3330 size) with a couple of fixed head devices used as a buffer. Depending on age and frequency of use, the data will be reassigned to a mass storage TBM system.
- Central Microfilm Storage (CRS): The central storage facility will contain all items processed by CRS as well as some aging items sent back from regional locations because of security restrictions. The minimum system design would continue the present manual system with one additional feature: analysts at their consoles would be able to automatically order those documents not available regionally. The subsequent delivery would be manual. The alternate design calls for automating the central facility so that documents ordered automatically could be delivered automatically. The expense of an automated system might be justified if document requests levied on the central facility were to increase significantly. At present, however, the SAFE plan does not include automating the central microfilm facility.

Hardware Costs

Comparative costs of the two computer systems are shown in Table 16. The price of IBM equipment was used to judge the cost of the main processors and disk/drum storage system. When specifications are better defined, perhaps some other type of equipment of the same computing power could be used. The terminal cost is calculated for 500 terminals. The mini-processor/communication system is based upon 50 mini-processors and the associated computer communication lines. The cost shown for the mass storage (TBM) is not that of a complete system but of an expansion of the system the Agency is currently purchasing. The programming costs include the initial programming of all the software for the system and the maintenance programming needed thereafter. The costs cited do not include the expense of altering existing

facilities to accommodate the new equipment, nor the expense of additional personnel to maintain it. The next chapter will discuss some of the cost savings and benefits associated with the SAFE Information System.

Table 16

System Costs
(In Millions of Dollars)

	2 General Purpose Computers	4 General Purpose Computers
Terminals......................	5.0	5.0
Mini computers and communication lines.......................	2.5	2.5
Main computers................	11.0	18.0
Card reader/punch, printers disk/ tape storage..................	4.0	4.0
TBM—mass storage............	1.0	1.0
Microfilm system...............	1.5	1.5
Software......................	6.0	6.9
Initial rental for main computer, and total system maintenance cost........................	2.5	2.5
Total cost...................	33.5	41.4

VII. COST-BENEFIT CONSIDERATIONS

INTRODUCTION

A cost-benefits analysis of the proposed $30-$40 million SAFE Information System is not possible at this time. We cannot assign a dollar figure to the potential value of the system to the production analysts for whom it would be built. We can, however, cite the arguments of the analysts that the SAFE system would improve the finished intelligence product by offering new analytic techniques, data bases and data base access. Also we can show that the SAFE system could improve the organization and allocation of Agency computer resources. And we can suggest areas where dollar savings may occur that would at least partly offset the cost of SAFE.

IMPROVED INTELLIGENCE PRODUCT

The arguments offered here are those made by the analysts in their critiques of the pilot system. They have already been cited in Chapter V but are quoted here, in part, because of their particular relevance.

"SF/C is a self-help user of the SAFE system. I would venture to say SF/C is more dependent upon SAFE and possibly more convinced of SAFE's indispensability than any other branch . . . SF/C's SAFE system does not merely supplement the branch files; it *is* the branch research file . . . We are striving to establish in SF/C the finest, most comprehensive, most usable repository of all-source information on command and control subjects in the intelligence community. We could not aspire to so ambitious a goal without SAFE . . . Scraps of information of interest to us can be found in all of the file modules being considered for incorporation in SAFE in the future . . . The more files we can dig through, the better chance we have of coming up with meaningful tidbits, and no one can predict where those tidbits will be found. Given the fantastic capabilities of computers, I see no reason to arbitrarily restrict the scope of our search for information by limiting the number of files to which we will have access. We want them all!!! And I promise you that we will learn how to exploit them." (OSR/SF/C comments).

"The most immediately evident one (benefit) is the ability to store and search vastly more information than previously possible . . . A more fundamental consequence is that, with masses of data more easily available, an analyst can bring more evidence to bear on a given problem. Further, the analyst feels more inclined to check his files before writing because he knows it (checking) can be done quickly and comprehensively . . . An interesting effect of having files available on the computer is being able to do searches or use data in ways not previously possible." (OSR/SEC comments.)

"During the Cyprus crisis and more recently in relation to events in the Balkans, I had an opportunity to use the SAFE system in a crisis management mode. The system proved to be an extraordinarily useful device in this respect. The mail distribution system (OLTA) and COLTS were of particular importance . . . SEC was able to receive relevant reports through the OLTA system many hours before the reports were available in hard copy. This capability allowed us to stay well ahead of possible threatening developments

77

and, in fact, alerted us to potentially interesting developments in the Balkans before reports of this were available through regular channels. I believe that the SAFE system has an enormous potential for crisis management." (Comments of one OSR/SEC analyst.)

"The SAFE system holds the promise of being able to make the ever increasing flow of information a manageable phenomenon, and to help stave off the accumulation of innumerable safes with unmanageable files." (OSR/TF comments.)

"The year-long experiment with Project SAFE has proven that . . . analytic capabilities can be enhanced. The savings of time and space afforded by the system, plus the rapid search capability, represent a highly desirable electronic package." (OER/D/TA comments.)

In summary, we believe that the data collection experiment demonstrated that the proposed system will help Agency analysts provide a better intelligence product. A better product may be a piece of incoming intelligence more thoroughly indexed and annotated for later reference; or information routed to users faster and more efficiently; or a more thoroughly researched piece of finished intelligence.

We believe the SAFE system will offer analysts improved techniques for monitoring and manipulating a large amount of incoming intelligence items, for searching files they could not otherwise use in the time before their deadlines, and for scanning incoming mail minutes after it arrives in the Agency.

In acquiring new technology, the Agency has traditionally emphasized the information collection side of the intelligence problem rather than the information analysis side. As this continues, it resembles building an ever larger cone for a funnel while keeping the same sized neck, and expecting the flow to increase. Agency analysts cannot now digest all the information they receive; they often cannot quickly find yesterday's piece of intelligence when it suddenly becomes relevant today. The task force feels that the development of the SAFE Information System represents the required parallel emphasis on the analysis side of the intelligence problem.

IMPROVING COMPUTER RESOURCES ALLOCATION

Computer and microfilm information systems to support production analysts have often been developed on an essentially individual basis. Each office would set out to meet its particular needs without knowledge of or coordination with other offices with similar problems, and the overall development of the Agency's information system has suffered. Proper development requires a unifying concept that would relate, for example:

—a file building requirement in OSR with one in OBGI,

—a text search and edit requirement in OSI with a text indexing requirement in CRS, and

—a text segment extract requirement in OWI with an automatic cataloging requirement in CRS.

A unifying concept would reveal the relationships between such varying requirements, and enable the task force to derive a common denominator.

Lack of a unifying concept has resulted in unnecessary developmental costs and, probably, unnecessary acquisition of computer equipment.

The task force suggests that the SAFE Information System could be such a unifying concept; that it is wide enough to embrace most of the information processing requirements of the production analysts; and, in short, that SAFE could improve the organization and allocation of Agency computer resources.

POTENTIAL SAVINGS

Savings could follow the adoption of the proposed SAFE methods for handling the Agency's electrical and paper receipts and the proposed SAFE text searching system. SAFE would also change the pattern of CRS use of computers and manpower. These changes are discussed below, but no dollar figures are projected.

Handling of Electrical Receipts

Approximately 20 million copies of electrical messages are disseminated yearly at the Agency. The cost of the existing operation is considerable; the existing equipment, supplies, space and manpower will no longer be needed if they are replaced by more efficient equipment and more efficiently used space and manpower.

Handling of Paper Receipts

The SAFE system plans to continue the current routine microfilming of documents that are received only as paper copy. Instead of keeping them all in a central location, however, SAFE would make a large collection of the microform documents available in regional storage devices and thus lighten this load on the central storage facility. This central facility now manually microfilms documents that were received as electrical messages. SAFE will enable the central facility to receive computer output microfilm (COM) processing, reducing the use of manpower.

Text Searching

During the data collection period analysts used the digitally stored text files to obtain messages that they may or may not have expected to receive through the regular delivery of SI messages, State cables, FBIS field traffic, military cables or DoD IR electricals. Analysts used various parameters in their search of those files and could change the parameters as their requirements changed from day to day. They found these searches valuable:

"I've used COLTS (text searching program) primarily to retrieve messages referred to in other cables but nowhere to be found in our mail."

"COLTS produces messages faster than hand delivery."

The proposed SAFE Information System would regularly update the text files as messages are received from OC. Its improved text search capability will allow analysts to repeat a question without having to reformulate it every time, and to view only titles or segments rather than the whole text, whenever they are scanning many messages for relevant items.

The task force anticipates that text searching will at least partially replace the dissemination of messages to user offices; and, possibly that someday intelligence messages will not have to be read and reread before reaching the ultimate customer. To the extent that shuffling, carrying and reading the mail are reduced, the Agency can save money.

Changes within CRS

If project SAFE becomes an operational reality, it would satisfy most of the present CRS requirements for computer support, as well as some other Agency requirements, and release a significant amount of OJCS resources.

Under SAFE, CRS will continue to analyze documents to create the "public" index record. Some increase in indexing may be required, but we feel money would be saved overall because CRS will be able to use the on-line analysis and automatic cataloging functions. Also, CRS will need fewer specialized analysts for routine reference work, because SAFE will permit production analysts to search many of the CRS files for themselves.

VIII. DEVELOPMENT PLAN

INTRODUCTION

Chapter VI of this report outlined the proposed SAFE Information System, its capabilities, possible hardware configuration, and cost estimates. This chapter describes the development plan of the SAFE Information System and projects the number of developmental phases required through FY 1980 and the expenditure required each fiscal year for the same period. These estimates are tentative and will certainly change as a result of the first phase activity (detailed system design).

DETAILED SYSTEM DESIGN PHASE

In the first phase of the SAFE Information System development, the task force must draw up detailed design specifications. It will have to verify that the system hardware configuration suggested in Chapter VI is correct or spell out the new configuration.

Once the hardware configuration is fixed, the task force must draft detailed specifications on individual components. If the minicomputer/main processor configuration remains the preferred one, studies must be performed to determine the optimum mix of the functions performed by the mini and main computers. The task force must spell out the requirements for the SAFE Console Station and decide whether or not to use existing terminal equipment. The task force must also fix the detailed specifications for the computer software, and determine how the overall project is to be managed.

Task Team

Project SAFE will demand a new task team composed of various specialists. Many are already Agency employees; some must be hired. This team would guide the detailed system design phase and the project management plan mentioned above. It would also maintain the interim SAFE system now in use in the various developmental branches. The analysts who are still working with the pilot system—at their own request—will continue to play an important role as SAFE is developed Agency-wide.

The task team would consist of 13 to 15 full-time analysts from the following organizations:

- CRS/SAS—Six or seven analysts engaged in project management, system design, and interim system management.
- CRS/SSD and OJCS—Two analysts studying hardware configuration.
- OJCS—One analyst, engaged in coordination, would keep OJCS informed of SAFE progress and would seek OJCS expertise as required.
- Contractors—Four or five systems analysts from a major software/system firm to analyze the implications of the expected load and queueing through computer simulation and modeling.

It would also need four part-time personnel as follows:

- OC—One person, familiar with the Cable Dissemination System of the Cable Secretariat, who will coordinate the SAFE requirements with those of the Secretariat.
- ORD—One person who would monitor industrial and academic research developments in areas of interest to SAFE.

● Contractors—Professor F. W. Lancaster, of the University of Illinois, as consultant on the design and evaluation of information systems; and Professor A. Meltzer, of George Washington University, as consultant on the design of computer systems.

OTHER DEVELOPMENT PHASES

During the SAFE System Design phase the task force will schedule the developmental tasks, the hardware implementation and the various operational phases. It will also estimate the funds required for each phase and for each calendar year.

Figure 27 is a tentative schedule and cost statement, drawn up to give working data to the planners. This schedule is based on the hardware configuration outlined in Chapter VI.

This Figure also shows the three basic developmental activities (system design, software development, and hardware acquisition) phased across 6 fiscal years, 1975-80.

The system design has been discussed in Chapter VI. Software development might be divided into as many as five distinct phases; for example, the first phase has a system monitor and a text search package. The first estimate of the size of these efforts is also shown in Figure 27 by the length of each expected phase, the number of man years (my), and the estimated expense. The last phase (5) is the continued maintenance of the completed system, here estimated as six man-years (my) each fiscal year.

Hardware acquisition is shown here as large main frame computers (mf), minicomputers (mc), SAFE consoles (sc), and film systems (fs) for the maximum configuration discussed in Chapter VI. The acquisition is spread across 5 fiscal years. By the end of FY 1976 the system will have two main frames (computers), eight minicomputers, 50 SAFE consoles and six film systems. By the end of FY 1980 the full-grown system will consist of four main frames, 50 minicomputers, 500 SAFE consoles and 100 film systems.

The operational phases, shown in Figure 27, are coordinated with software development and hardware acquisition. For example, phase 3 operations begin when phase 3 software is available. Phase 3a depends on the acquisition of phase 3a hardware and extends phase 3 capabilities to more users.

The proposed acquisition and installation schedule is ambitious and meeting it will require extraordinary effort. We feel the effort of giving analysts the earliest possible operational date is justified if we are to sustain their enthusiasm for the SAFE system. The magnitude of the effort will be considered during the detailed design study phase. The acquisition of building space and utilities as well as computer hardware may be exceedingly difficult in the proposed time frame. The earliest possible operational date that is anticipated for phase 1 would be after January 1976.

The estimates of total FY dollars shown in the figure are based on software expenses calculated at $7 million and hardware purchases calculated at $34 million.

The proposed SAFE Information System has not directly addressed all the security implications evident in the transition toward a "paperless" office. Major changes in current security procedures may be required in the new environment. We recognize, of course, that only selected persons will have access to special category items. However, we must assume that analysts who have access to basic SAFE files (e.g. State cables, SI messages) may occasionally see items not within their "need to know." We believe this assumption may be necessary if SAFE is to remain a viable system. The ability to provide strictly enforced "need to know" software could significantly increase the cost of a SAFE system. Specific security requirements will be addressed in the detailed design phase of the SAFE system.

25X1

Confidential

Confidential