(4)

PHASE II FINAL REPORT

PROJECT
CHIVE

Volume IV

SYSTEM REQUIREMENTS

DESIGN APPROACH

CHIVE/R–3–65
1 March 1965

DIRECTORATE OF SCIENCE AND TECHNOLOGY
OFFICE OF COMPUTER SERVICES

WARNING

This material contains information affecting
the National Defense of the United States
within the meaning of the espionage laws,
Title 18, USC, Secs. 793 and 794, the trans-
mission or revelation of which in any manner
to an unauthorized person is prohibited by law.

U.S. Central Intelligence
Office of Computer Services

Project CHIVE:

Phase II Final Report

Volume IV

SYSTEM REQUIREMENTS

DESIGN APPROACH

CHIVE/R-3-65

1 March 1965

010365

25X1

TABLE OF CONTENTS

TABLES ·

Page

FIGURES

Chapter 4.1

PROBLEM DEFINITION

The rationale for Project CHIVE has its roots in a request from the Assistant Director, Central Reference for a study of ADP needs in the Agency and in numerous investigations, conducted by in-house as well as contractor personnel since approximately 1959, all of which generally agreed both as to the increasing complexity of information retrieval within the Agency and the advisability of introducing new hardware and techniques reflecting the present state-of-the-art of the information-handling technology. More than five years ago, in a report issued by the AD/CR under the auspices of the Central Reference Advisory Group (CRAG), it was stated that in one DD/I office (OCR):

> "The volume of incoming information exceeds processing capabilities based on current manual or electrical accounting machine (EAM) techniques;
>
> The proportion of receipts which can be fully processed is declining;
>
> Service from existing facilities is becoming slower as the size of the several indexes increases;
>
> Quality of service in terms of listing, subject correlation, up-dating, and display is declining or not offered because of the limitations of current staff and equipment."

PROBLEM DEFINITION
4.1.

-1-

CONFIDENTIAL

These findings with respect to OCR were complemented by indicated applications for computers in ORR, OSI, and other DD/I offices

4.1.1. PROJECT PHASES

In its total context, the CHIVE Project historically included three major tasks: Task I - To establish a DD/I computer center; Task II - To implement problem-, rather than system-oriented applications ("special projects") on computers; Task III - To study and design a new document/ information retrieval system for the DD/I. Only the third task, however, continues to carry the original project title, and it is that element--broadened in scope to include the intelligence information retrieval requirements of the Agency as a whole--with which this report is concerned.

In brief, the project has been defined as a four phase effort to:

- Study Agency needs for acquisition, dissemination, processing, and control of information.

- Recommend a data processing system to solve Agency information handling problems.

- Implement an operational system in two stages.

Because the magnitude and complexity of the task equals or exceeds any previous design effort in the storage and retrieval area, the search for problem

PROBLEM DEFINITION
Project Phases
4.1.1.

-2-

solutions has been (and will continue to be) charac-
terized by an attitude of relative caution.

Phase I--fact-finding and formulation of the overall
system concept--extended from September 1962 to
September 1963 and included an evaluation period in which
the broad findings and recommendations were reviewed by
a group of representatives from DD/I and DD/S&T offices
and others concerned with Agency information processing
activities. The principal Phase I findings are reviewed
in the next section.

Phase II--detailed system design--began in
October 1963 and is terminated with the submission of
this report. The activity during this period included
several iterative steps to bring the total problem
definition within manageable proportions, the selection
of indexing and retrieval techniques from the broad
spectrum that was available, the detailed tasks needed
to give substance to a system philosophy, and the
testing of techniques and procedures.

Phase III--implementation of an initial segment
of the system--will extend for eighteen months to
October 1966. During this period, an OCR personnel
organization will be formed and trained, EDP equipment
will be acquired, computer programs will be written, the

PROBLEM DEFINITION
Project Phases
4.1.1.

-3-

necessary initial files will be built to support the indexing operation, and a pre-operational test will be undertaken. (See Volume III.)

When a satisfactory level of confidence in the new techniques and procedures is attained, the system will assume operational responsibility (Phase IV) for processing a defined segment of the input flow and a corresponding segment of service to research analysts.

4.1.2. PHASE I FINDINGS

The Phase I effort was directed primarily to determining the Agency's information processing needs through an extensive "grass-roots" survey of analytic DD/I components rather than concentrating on the methods currently used by OCR in performing its document retrieval mission, which were the subject of most of the earlier studies of Agency information processing activities.

In essence, the Phase I findings verified the assumptions which initially motivated the project undertaking:

- Information resources extend beyond the normal (and heavy) flow of intelligence material into most areas of the world's published literature.

> PROBLEM DEFINITION
> Phase I Findings
> 4.1.2.

-4-

CONFIDENTIAL

- <u>Missions and interests</u> cover the spectrum of human activity.

- <u>Formats</u> of the useful material vary widely and include maps, photos, cables, information reports, etc.

- <u>Refinement</u> of the data ranges from raw fragmentary data to finished intelligence.

- <u>Use</u> of the data includes background study, corroboration, and current awareness.

- <u>Point of view</u> ranges from the bizarre (e.g., spider reactions at high altitudes) to the mundane.

- <u>Time requirements</u> for processing material include the stringent requirements of current intelligence, the predictable needs of programmed research, and the slow but demanding cadence of basic intelligence.

One primary concern in Phase I was the role played by the analyst files in contrast to those of the central system in OCR. Table 4-1 summarizes this contrast from the analyst's point of view.

The basic Phase I conclusions were (a) that there is a need for greater speed, depth, and breadth of access to the total Agency information resources and (b) that more efficiency is needed in information processing to counterbalance the costs of increased depth and breadth of information coverage and access. As corollaries of these basic conclusions:

PROBLEM DEFINITION
Phase I Findings
4.1.2.

-5-

- both central and analyst files are needed.

- central files and processes should be integrated.

- computers can play a significant, but not a dominant role in intelligence information processing.

- Project CHIVE should focus on the design of a central system.

In September 1963, the CHIVE Evaluation Group sustained these conclusions and agreed that the project should move into the design phase with emphasis on testing as a vital part of the design methodology.

In retrospect--particularly as the project moved into detailed design--the methodology used in Phase I was lacking in several areas. Among other things the interview technique produced results which could be interpreted adequately at a gross level, but were inadequate to provide direction in specific design areas-- such as what type of processing organization the user should see and the balance between named object and subject/concept retrieval which would produce the best pay-off to the user. Consequently, the need for more data gathering is ever present and will continue as long as the project is in a dynamic state. To some extent the CODIB-SCIPS work has been useful and will continue to be consulted.

PROBLEM DEFINITION
Phase I Findings
4.1.2.

-6-

Table 4-1

PHASE I FINDINGS

	Analyst Files	Central Files
VALUE	Primary retrieval mechanism in terms of: - use rate - response time - analyst specifications	- Agency document repository - fulfills community obligations - provides basic reference services to all Agency components
USE	- check validity of new data - determine effect of new data on what is known - handle immediate queries - research	- retrieve data not in analyst files - routine, long lead-time requests - specific requests for biographic and installation data
STRENGTHS	- readily accessible - contain filtered data - tailored to needs and habits - control of concepts and ephemeral topics	- historical depth - broad subject and area base - files organized from several points-of-view - backstops analyst files
WEAKNESSES	- limited indexing depth, access points - duplicative, costly - limited to current interest - inaccessible to other analysts	- slow response - insufficient depth and breadth - lacks single point retrieval - conflicting pro-cedures, duplicative processing

-7-

CONFIDENTIAL

The product of the Phase I effort was, therefore, an admittedly general but sufficiently comprehensive understanding of the needs of analysts (and the environment in which they operate) to enable one to recommend the requisite design goals of a future system, and to suggest a system configuration which would meet these needs in an optimum (if not ideal) fashion.

4.1.3. CHIVE DESIGN BOUNDARIES

A basic restriction has been assumed in the design of the CHIVE system: the project is concerned with that part of the intelligence cycle where pre-processing (such as SIGINT data reduction, photo interpretation, and language translation) leaves off and production and evaluation begins. The functions in between--such as document dissemination, file building and retrieval--are those that provide information processing services for the effective execution of the others in the intelligence cycle and have been considered proper areas for study and design.

The succeeding sections further refine this definition to a manageable design level.

PROBLEM DEFINITION
CHIVE Design Boundaries
4.1.3.

-8-

4.1.3.1. Excluded Activities

The CHIVE system will not and should not be equated with any extant Agency office or collection of offices. Although it will be operated by, and involves most of the present components of OCR, several OCR functions, as indicated below, will remain outside CHIVE. Similarly, CHIVE cannot be regarded as simply a machine configuration consisting of computers, document and index storage devices, communications media, and other hardware--that is, a super OCR Machine Division. Rather it is a complex of people (including managers, input/output analysts, machine operators, clericals, etc.), as well as machines and computer programs, organized in a system context to perform a complex of activities specifically related to the centralized storage and retrieval of information in behalf of a significant population of Agency customers.

The following is a list of functions--categorized from the OCR point of view--which we presently believe should remain external to CHIVE. This is not to say that some of these functions will remain unaffected by the CHIVE system design. Further, the fact that certain operations will be kept out of CHIVE is not meant to suggest that these activities could not be benefitted

PROBLEM DEFINITION
CHIVE Design Boundaries
4.1.3.1.

-9-

by the application of EDP hardware and related techniques.
On the contrary, some (for example, inventory control
problems) are well-established application areas for
automatic data processing while others (library circula-
tion service, etc.), are at least being examined in the
outside world as potentially suitable for automation.
We are suggesting, however, that if such operations are
to be upgraded to EDP, the responsibility for the design
effort whould not rest with CHIVE but must devolve on
OCR (for the necessary problem definition) and on OCS
as required (for technical assistance on hardware and
programming implementation.

4.1.3.1.1. Liaison Services

This is a staff function, performed for the entire
Agency, involving liaison contacts with all U. S.
Government departments. It coordinates requests for
CIA intelligence information and action which may be
required by these departments and coordinates the Agency
briefing and debriefing program for officials enroute
to or returning from tours of duty abroad. Presently
housed within OCR, it is functionally unrelated to the
input and retrieval of documentary information.

PROBLEM DEFINITION
CHIVE Design Boundaries
4.1.3.1.1.

-10-

4.1.3.1.2. Administrative Staff/OCR

While it is apparent that this organizational component of OCR will become intimately involved, both as a user of system-generated management data as well as a contributor to planning and decision-making having to do with recruitment policy, training, personnel reassignments, job descriptions, budget planning, etc., it will remain (as now) essentially a support-type operation not substantively a part of the document handling process.

4.1.3.1.3. Acquisition (Publication Procurement) Services

The CIA Library, the Graphics Register, and the Map Library Division are somewhat unique among the various central repositories in that all have additional responsibility for the procurement of intelligence information. The Library is charged with the procurement of foreign and domestic open publications, including books, periodicals, and newpapers. In addition, it obtains requested materials through interlibrary loans. The Graphics Register is similarly involved in a photo and film procurement program, while the Map Library Division is responsible for administering [   ]   25X1A

25X1A [   ] [   ]   25X1

PROBLEM DEFINITION
CHIVE Design Boundaries
4.1.3.1.3.

-11-

CONFIDENTIAL

It seems logical to exclude these collection activities from CHIVE on the ground that collection is clearly distinct from the function of indexing and retrieving the materials collected.

4.1.3.1.4. Dissemination Services

The initial, or first level, dissemination of documents received by the Agency from various collection programs is currently handled as follows: (a) Cable Secretariat is responsible for the dissemination of both CIA and non-CIA cables; (b) the Document Division disseminates intelligence reports, finished collateral intelligence publications, Comint (including both teletype and hard copy), and T/KH material; (c) the Acquisitions Branch of the CIA Library disseminates foreign and domestic books and serials to customer offices based on assumed interest as well as specific requirements; (d) the Foreign Broadcast Information Division (FBID), through the Printing Services Division (PSD), distributes FBIS daily reports, summaries, and abstracts ordinarily in some 400-800 copies each; (e) the Foreign Documents Division (FDD), again through PSD, disseminates Agency-produced translations of foreign documents.

PROBLEM DEFINITION
CHIVE Design Boundaries
4.1.3.1.4.

-12-

For the present, CHIVE does not propose to assume the responsibility for the initial dissemination of any of the above materials. Instead, like the Analysis Branch of Document Division or a specialized OCR register, the CHIVE system will be a recipient of documents disseminated to it by the various distribution organizations named above on the basis of its reading requirements.

4.1.3.1.5. Book Cataloging and Circulation

Books and serials selected for the Library's collection are catalogued according to the Library of Congress classification system. This operation as well as the normal circulation activities are excluded from CHIVE. The project will become involved in the exploitation of books and serials only on a selective basis where deeper indexing appears justified by consumer demand.

4.1.3.1.6. OCR Special Projects

Several specific information processing activities in OCR have some potential as EDP applications but have no direct relationship with CHIVE. These include:

PROBLEM DEFINITION
CHIVE Design Boundaries
4.1.3.1.6.

-13-

CONFIDENTIAL

- Punch card services performed by the EAM components of OCR for other Agency components.

- The production of bibliographies resulting from literature searches.

The generalized nature of the computer programs needed to perform CHIVE functions may prove to be useful in providing EDP support to the above activities, but are not designed specifically to accommodate them.

4.1.3.1.7. Translation Services

CHIVE does not believe that the translation activity per se should be an integral part of the central reference activity--i.e., that the FDD analysts should be dispersed among the input and retrieval organizations. From a management point-of-view there seems to be no more logic for doing this than, say, to co-locate FDD personnel with research analysts. On the other hand, it is fully recognized that the foreign document exploitation activity will be affected by, and will in turn have an influence on, the central storage and retrieval operation.

4.1.3.2. User Boundaries

The question of the CHIVE user population has a direct bearing on many aspects of the system design (e.g., input and retrieval speeds, programming, memory

PROBLEM DEFINITION
CHIVE Design Boundaries
4.1.3.2.

-14-

capacity, document coverage, special file requirements, etc.), of which perhaps the most important is that of the subject coverage of the system.

4.1.3.2.1. Service to Non-CIA users

It is evident that at least in certain subject areas the CHIVE system will be required to support non-CIA customers. These areas are those where, by DCID or other external directive, the responsibility for processing and retrieving (as distinct from producing and collecting) certain intelligence information has been delegated to CIA to perform on behalf of the Intelligence Community.

To our knowledge, there are three such specific delegations of reference responsibility (not including those assigned to the DD/P or NPIC), namely:
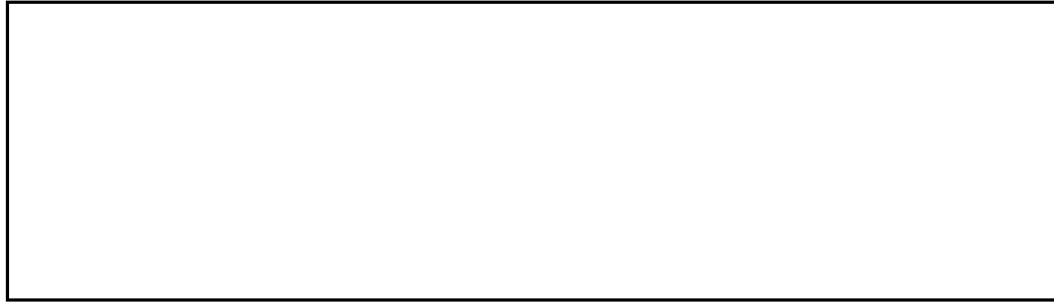
- Maintenance of a file on the exploitation and translation of foreign language publications (DCID 2/4).

There appears to be only two CIA Headquarters Regulations which charge some Agency component (other than NPIC and the DD/P) with providing reference support to non-CIA customers:

PROBLEM DEFINITION
CHIVE Design Boundaries
4.1.3.2.1.

-15-

25X1

As a courtesy to other organizations, and in light of the historical precedent for CIA to perform services for the benefit of the Intelligence Community, the CHIVE system will, at the minimum, respond to requests for information from non-CIA customers (even where the Agency is not specifically charged with servicing outside requests in this area) if the volume and character of such requests would not unduly burden the system. This commitment corresponds to the informal policy practiced by OCR today.

4.1.3.2.2. Service to CIA Users

We would seriously doubt that management would want any distinction made between service to DD/I and non-DD/I components. OCR, certainly, has never restricted its service to the DD/I. Indeed,

25X1

states that the AD/CR shall provide

25X1A

25X1A

PROBLEM DEFINITION
CHIVE Design Boundaries
4.1.3.2.2.

-16-

CONFIDENTIAL

4.1.3.3. Input Boundaries

4.1.3.3.1 Document Coverage

The results of the Phase I study implied that the data base must encompass all documents in use by the analytic offices which the system will serve. If interpreted literally, this requirement alone would make system implementation impossible. Defining document selection criteria which will produce the best pay-off and which can be applied consistently has been one of the most difficult tasks for the project.

This problem is treated in more detail in Volume V of this report, but it still remains unsolved to a great extent. However, certain principles are described here for the purpose of refining the definition of the CHIVE problem.

Individual series within a given document category may be excluded entirely from the system but only after such contemplated action is communicated to, and reviewed by, the CIA customer community.

Specific documents within a series may be excluded from processing without customer authorization.

Foreign documents (the most heterogeneous and unrestricted of all document categories) will be

PROBLEM DEFINITION
CHIVE Design Boundaries
4.1.3.3.1.

-17-

CONFIDENTIAL

largely limited to: (a) that material translated by
FDD and other intelligence organizations in response
to analyst demand; (b) translated items supplied the
system on an ad hoc basis by the research analysts;
(c) Sino/Sov Bloc scientific and technical titles
remaining after analyst review of the Library of
Congress MIRA program. The philosophy of the system
will be to select from that universe of potentially
available foreign documents only those in which analysts
have evidenced an interest as demonstrated by their
translation requirements levied on FDD (and other
organizations) as well as their own personal exploita-
tion efforts. Domestic open source material will be
selected in much the same way.

4.1.3.3.2. Topical Coverage

Assuming that workable criteria can be established
for selecting documents to be processed by the system,
another problem of the same magnitude is encountered
at processing time--what topics in the documents should
be exploited and how? These two aspects of topical
criteria could be restated as breadth (exhaustivity)
and depth (specificity) variables, respectively.
Figure 4-1 is an attempt to show graphically the

                                    PROBLEM DEFINITION
                                    CHIVE Design Boundaries
                                    4.1.3.3.2.
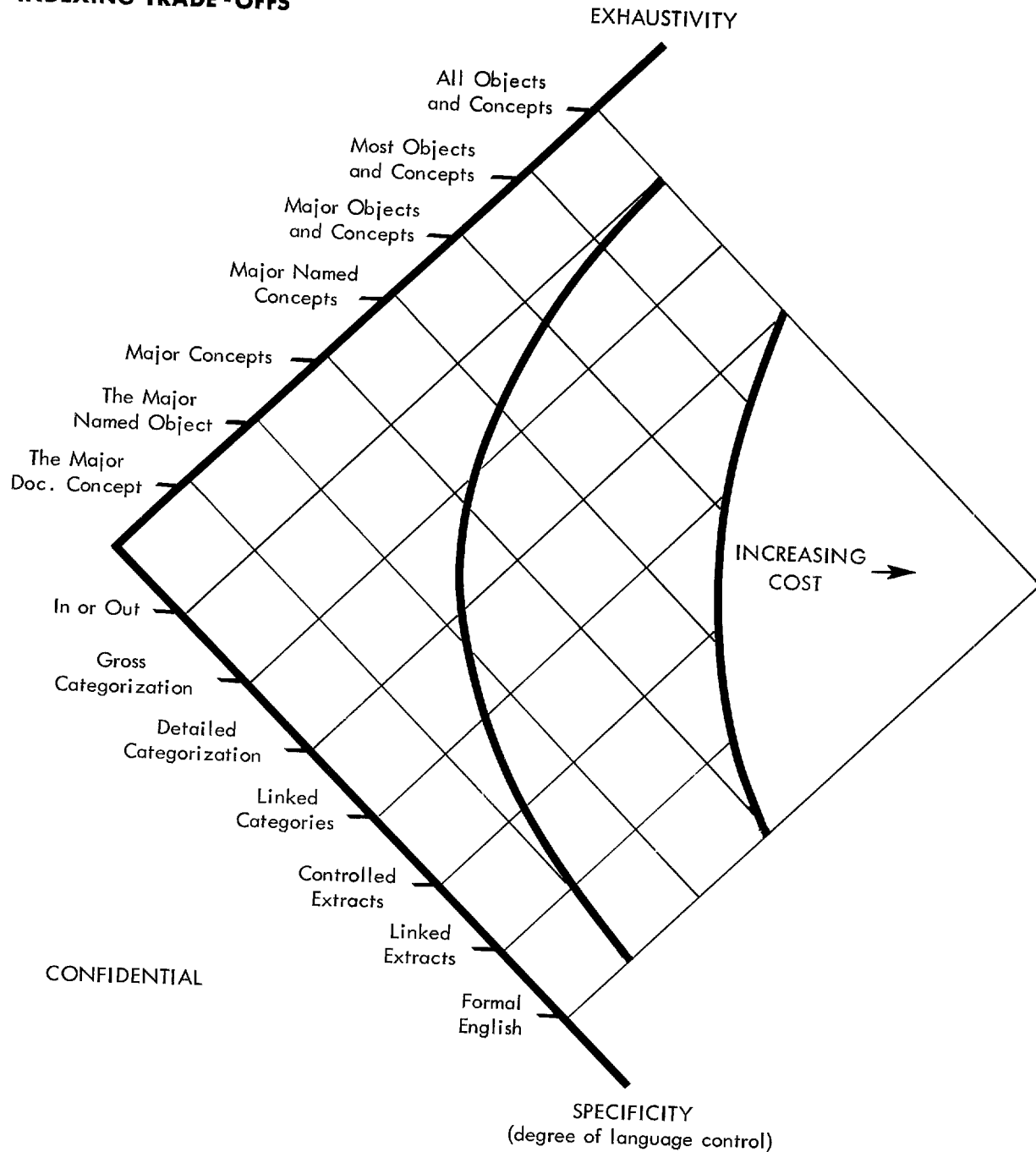   -18-

interaction of these variables when cost is considered.
Stated briefly--for a given investment, depth of
exploitation must be sacrificed if greater breadth
of topical coverage is desired (and vice versa).

Each of the existing OCR systems could be placed
(at least approximately) on this graph. CHIVE's
intent is to increase both the breadth of topical
coverage and the depth to which the topics will be
exploited in such a way that the investment required
is no more than the aggregate investment in the
present OCR systems. From one point of view, at
least, this has been the essence of the design effort.

The notion of "exploiting" input material goes
beyond the need to retrieve it on demand. Moving
out on both axes on Figure 4-1 should permit a
capability for large-scale manipulation of the data
base to assist the analyst in correlating material
and generating inferences. This, indeed, is one of
the primary justifications for using a computer--
which, unlike simpler machines, has both the speed
and logic capabilities to assist in these functions.

PROBLEM DEFINITION
CHIVE Design Boundaries
4.1.3.3.2.

-19-

INDEXING TRADE-OFFS

EXHAUSTIVITY

All Objects
and Concepts

Most Objects
and Concepts

Major Objects
and Concepts

Major Named
Concepts

Major Concepts

The Major
Named Object

The Major
Doc. Concept

INCREASING
COST

In or Out

Gross
Categorization

Detailed
Categorization

Linked
Categories

Controlled
Extracts

Linked
Extracts

CONFIDENTIAL

Formal
English

SPECIFICITY
(degree of language control)

4.1.3.4. Service Boundaries

Two basic assumptions which have influenced CHIVE development are (a) that the CHIVE system and the analyst files will complement one another, and (b) that the system's primary customer is the Agency desk analyst. The system is to provide a variety of services in support of analytic activity, ranging from extensive research service on a project basis, through documents related to an analyst's current requirements, to fragmentary information in answer to a factual request.

It is extremely difficult to account for the wide spectrum of working habits of analysts in our design. Nor is it possible to estimate the influence which the proposed system will have on the size and character of analyst files and his analytical techniques. Our basic approach to this problem has been to design the system in spite of analyst file activity and organization, applying our best judgment to the notion of a "service of common concern" in conjunction with a (hopefully) rational approach to analyst requirements. It is intended that our initial implementation will include an intensive program of user familiarization and training so that analyst working techniques will

PROBLEM DEFINITION
CHIVE Design Boundaries
4.1.3.4.

-21-

evolve to the point where his files eventually supplement the central system, and are restricted to current data of interest and basic reference aids which he must access on a daily basis.

Our objective is to provide three levels of service (with respect to processing sophistication). These are described below as they relate to the manipulation of system inputs.

4.1.3.4.1. Selected, unaltered inputs

This refers to the system's ability to retrieve input data for the most part in the same form in which it was entered--documents, their index records, dictionary data or whatever.

4.1.3.4.2. Processed, collected, linked inputs

This refers to the ability to (a) reorganize input data in any manner desired for output, (b) link together pieces of data from different index records (principally extracted information referring to named objects), and (c) maintain collections of such data when necessary in physically discrete packages (records and files) which have been gathered together from the complete collection of input documents. No exhaustive analysis or evaluation

PROBLEM DEFINITION
CHIVE Design Boundaries
4.1.3.4.2.

-22-

is implied in such organizations or collections of data,

except to the extent that they will support the finished

25X1 [ ] responsibilities of OCR. This capability will

permit the maintenance of indexer reference aids and the

storage of data relationships discovered in the review

of machine responses to requests so that these relationships

are available for future reference.

4.1.3.4.3. Evaluated, correlated inputs

This refers to the system's ability to (a) derive

new information from collections of input data,

(b) generate consistent, concise, and complete summaries,

and (c) store and retrieve such derived data. It is

anticipated that this level of processing service (which

is bordering on intelligence analysis) will be largely

25X1 limited to the finished [ ] responsibilities of

OCR.

4.1.3.5. Security Boundaries

If OCR is to serve in fact as the central repository

of positive intelligence information for all components

of the Agency, it must process and store any information

of continuing intelligence value--whatever its security

classification. For this reason we are placing no

PROBLEM DEFINITION
CHIVE Design Boundaries
4.1.3.5.

-23-

security boundaries on the kinds of documents which the system will handle. In addition, we are assuming that current management philosophy would also support "all-source" clearances (in the traditional, if relative, sense of that term) for CHIVE system personnel to the same extent as it has for its intelligence production analysts. This is not to suggest that compartmentation could not be made to work, but this would seriously aggravate the communication problem--not to mention file maintenance, dictionary development and query processing.

The procedural and mechanical safeguards recommended in Volumes V and VII have been developed to preserve the necessary security requirements but still permit the necessary flexibility to perform system functions.

4.1.3.6. Cost Boundaries

CHIVE system design has proceeded with no Agency management guidelines on costs beyond those that can be implied by the steadily increasing emphasis on cost consciousness at all levels in the Agency.

A general, informal injunction calls for an operational cost burden which will not exceed the budget and manpower

PROBLEM DEFINITION
CHIVE Design Boundaries
4.1.3.6.

-24-

now needed to operate OCR (as the system approaches full capability over a number of years). This statement is made with some reservations; system designers should not be totally constrained by a cost ceiling. Techniques or hardware may prove to be of sufficient importance that an overall increase in capability can be realized which far exceeds the additional costs required. This is an intangible area; the necessary proof of such capability must be determined in a realistic manner if and when such system implementation decisions are made.

4.1.3.7. Technical Boundaries

A casual glance at the literature on information processing technology seems to suggest that revolutionary equipment and computer techniques are just around the corner. The system designer appears to have an impressive array of new methods to solve his problems. In-depth analysis of these new techniques, however, reveals that for the designer who must implement a system within a reasonable time frame, there is no panacea nor even one technique that offers outstanding advances.

Computer equipment is getting faster, cheaper, and smaller, but more or less brute-force methods must

PROBLEM DEFINITION
CHIVE Design Boundaries
4.1.3.7.

-25-

still be used to make them perform useful work in non-numeric applications. The promises of automatic indexing and effective communication with large, complex machine-stored files seem to be as far away as they were five years ago.

Even outside the computer area, no breakthroughs are evident. Advances in the indexing art can be characterized as innovations of questionable worth.

CHIVE has proceeded on the basis that we cannot wait for new technologies to develop, nor can we materially accelerate their development with R&D projects beyond the pace that the normal competitive environment will provide. In short, design decisions have been based on techniques and equipment which are well-proven or which warrant the risk for an ultimate pay-off.

PROBLEM DEFINITION
CHIVE Design Boundaries
4.1.3.7.

-26-

Chapter 4.2.

SYSTEM OBJECTIVES

From an analysis of the Phase I effort and the
preceding discussion on the refinement of the CHIVE
problem definition, a list of system objectives can
be summarized from three points of view.

4.2.1. SYSTEM USER OBJECTIVES

- Broader document coverage

- Increased indexing specificity

- More exhaustive indexing

- Capability to answer more complex questions

- Faster service

- Single point service

- All-source, integrated output

4.2.2. SYSTEM OPERATOR OBJECTIVES

- Increased input rates

- Increased file utilization

- Micro-storage media for documents

- Reduction of retrieval time

- Integrated organizational structure

- Reduction of manual labor

- Improved communication with users

- Increased record lengths

- Common system vocabularies

4.2.3. SYSTEM MANAGEMENT OBJECTIVES

- Flexibility to meet changing needs

- Good system evaluation tools

- Efficient use of available manpower

- Improved pay-off/cost ratio

These objectives are given here without narrative elaboration because they are either self-evident or have been justified in previous documentation. They remain objectives--proof of accomplishment, particularly as to improved pay-off cost ratio will be difficult to establish in comparison with present OCR or other systems.

SYSTEM OBJECTIVES
Management Objectives
4.2.3.

-28-

Chapter 4.3.

DESIGN METHODOLOGY

A formal approach to the design of a large information processing system requires that a series of steps--taken in fixed order--be well defined and carried out in an environment that provides good communication between the people involved. These steps might be enumerated as follows:

- Problem definition

- System requirements

- Functional specifications

- Subsystem definitions

- Subsystem design

- Subsystem test

- System integration

- System test

- Acceptance

Theoretically a project should proceed through these steps in serial fashion. Practically, of course, these steps are difficult to separate. Nor can they proceed without some iteration.

DESIGN METHODOLOGY
4.3.

-29-

A review of the methodology used in Phase II of Project CHIVE indicates that it had but a vague resemblance to the textbook approach. Beyond the intangible aspects of human frailty which corrupt the best of plans, there were some significant factors which frustrated any attempts at a formal approach.

A fundamental hardship was our inability to synthesize the analysts' needs into a meaningful and objective set of system requirements that could and would be endorsed by Agency management. Evaluation of project goals by management and the coordination with future operators of the system has been characterized by a "constructively skeptical" attitude. This has been conditioned by the investment in the present system and implications of the proposed change and by the significant absence of positive results from earlier EDP information system developments in other USIB Agencies.

With only the broadest of guidelines to work with, the system requirements had to be largely self-imposed by the system designers. This put management in a responding rather than in an initiating role. Consequently, the system designers concentrated on techniques--some of them rather generalized--in the hope that they would

DESIGN METHODOLOGY
4.3.

-30-

ultimately fit a loosely defined set of requirements.
The unsettling effects of this approach are obvious.

Secondly, the designers worked under several
layers of coordination. This was to be expected, but
each of these layers consisted of service groups.
That is, the CHIVE group could be considered a service
group to OCR, which in turn is dedicated to serving
the production analyst. This group is the core of the
Agency, which in itself is a service organization to the
National Security Council. Since it is always very
difficult to define the scope of activities of a service
organization, the CHIVE designer at the bottom rung of
this ladder has lived in a poorly defined problem area.
Most of the system designers, however, were OCR graduates
whose work was supplemented by present OCR senior officer
involvement. Thus we have enjoyed an advantage of
in-house capability over other design efforts which were
more contractor bound.

While it would be possible to pursue further the
occupational hazards of systems analysis--perhaps even
to the extent of providing material similar to that
given in the SCIPS Stage I Report, Volume 5--it is
believed that such a discussion would serve no useful
purpose in this report.

DESIGN METHODOLOGY
4.3.

-31-

CONFIDENTIAL

4.3.1. EVOLUTIONARY APPROACH

Previous OCR studies have looked to improvements in existing systems, while the CHIVE effort has attempted an overall synthesis and redesign. The nature of the systems problem is such that it should be attacked all at once-- at least in the early concept and thinking stage. Prolonged effort at this level, however, is obviously frustrating and not very productive.

One of the major decisions early in the design was that an evolutionary approach must be taken. There has been no indication that there is a need to institute new procedures, organization, and techniques for the entire system at one time. In the absence of such pressure, a smooth upgrading of the current system is obviously wise. However, a minimal threshold invest-ment in design and operational effort is needed to "bootstrap" the system into existence. Two distinct thresholds were considered:

- in personnel, the threshold is low. A relatively small pilot group can test and adjust procedures and still provide an operational capability.

- in computer equipment (and its operating programs) the threshold must be quite high. The effort needed to provide a minimal retrieval capability for a small file is the same order of magnitude as that for a large file.

DESIGN METHODOLOGY
Evolutionary Approach
4.3.1.

-32-

A manifestation of the high threshold investment in EDP is the attendant risk that must be accepted. That is, this portion of the system must be constructed and checked out before any real measure of its value can be determined.

Starting with a solid personnel and EDP base, the system can grow in relatively small increments-- assuming more responsibility, accepting more inputs, servicing a larger segment of the user population-- with little risk beyond that accepted for the initial system.

4.3.2. TESTING

The caveat in the recommendations of the Phase I CHIVE Evaluation Group to consider testing as an integral part of the design effort has been followed as much as possible. Because the expense of planning and performing in-depth experiments is high, testing has been performed at the technique or module level and has been limited to the most critical design areas. For example, the complete cycle of system input, file building, and output has not been simulated. In a sense, putting the system "on-the-air" is the only way

DESIGN METHODOLOGY
Testing
4.3.2.

-33-

to test it adequately. It is hoped that if it fails,
it can be adjusted to make it work rather than
scrapped for a complete restart.

The CHIVE testing experience has, in general,
been very useful. Beyond achieving its basic purpose,
the need for attention to detail has uncovered subtle
problems and has assisted in maintaining a good
design perspective. In a very real and deliberate
sense, Phase III itself will be in large measure a
test of the proposed system.

4.3.3. PROJECT ORGANIZATION

The number of people participating directly in
Phase II of the project has ranged from 13 to 43 over an
eighteen-month period. These people had diverse
backgrounds and came from several organizations.
They brought with them a wide spectrum of points of
view. This in itself sounds foreboding, but their
motivation was such that their parochial concerns
helped rather than hindered the effort. Also, as
noted above the use of considerable in-house talent
was a distinct asset. The role played by the basic
components involved in the project are discussed
below.

> DESIGN METHODOLOGY
> Project Organization
> 4.3.3.

-34-

4.3.3.1. Office of Computer Services

Primary responsibility for the system design resided with the Development Division of OCS. In addition to coordinating the tasks of contractor personnel, personnel in this division were involved in all tasks of the project--from generating basic concepts to evaluating alternative detailed procedures and planning the initial system.

In some areas, personnel from other components of OCS played a significant role either through consultation or assistance in planning for the acquisition of equipment and programs. The CIA Computing Center was used extensively in simulation and testing.

4.3.3.2. DD/I CHIVE Officer; CHIVE Support Staff

At the outset of Phase II, the Executive Assistant, O/AD/CR was designated as the principal representative of the DD/I on the project. In addition to providing liaison with DD/I components (primarily OCR), he added guidance and made decisions where the facts and circumstances warranted. He was assisted by two OCR people (from BR and SR) constituting the CHIVE Support Staff. This staff was accommodated by OCS slots and

DESIGN METHODOLOGY
Project Organization
4.3.3.2.

-35-

resided in its Development Division. They were
intimately involved in the design and experimental
work in addition to their coordination duties.

STAT

4.3.3.3. [   ] Federal Systems Division

The principal contractor on the project was [   ]   25X1
which provided the bulk of the manpower in EDP design
work as well as assisting in non-EDP systems analysis.
Unlike most contractual arrangments, the [   ] group was   25X1
integrated into the total CHIVE team rather than being
assigned isolated tasks. They worked full-time at
Headquarters in Development Division spaces.

4.3.3.4. Stanford Research Institute

25X1A   The services of [   ] of [   ] were used on   25X1
Project CHIVE for consultation as well as for specific
tasks in his special areas of competence. He worked
off-site, assisted on an ad hoc basis by other [   ]   25X1
personnel with liaison provided by an [   ] analyst at   25X1
Headquarters working on the same tasks.

4.3.3.5. OCR Experimental Indexing Group

A significant indexing experiment (see Volume V)
was undertaken in CHIVE which used the services of

DESIGN METHODOLOGY
Project Organization
4.3.3.5.

-36-

18 OCR personnel drawn from several divisions for about 4 months. Their work was planned and supervised by full-time members of the CHIVE design group.

4.3.4. COORDINATION WITH OTHER AGENCY COMPONENTS

Through the DD/I CHIVE Officer, the prospective users of the proposed system were asked to participate in some data gathering and experimental work. In addition, the various offices in all directorates were kept abreast of the design concept and were asked to react to it. It is anticipated that this coordination will gain momentum as we proceed toward implementation.

DESIGN METHODOLOGY
Coordination
4.3.4.

-37-

Chapter 4.4.

SUMMARY

The purpose of the Phase II report is to describe
and recommend a system with arguments restricted to
the pro's and con's of specific design areas. The
basic concept of the system and its gross justifica-
tion are discussed in previous documentation. This
volume was included in the report to provide limited
context for the substantive description of the system.

To emphasize and re-emphasize that the CHIVE
problem is complex would contribute little to
documentation at this stage. Our purpose has been
to keep a steady pace in cutting through the unending
maze of problems rather than belabour them. Without
this momentum the problem would continue to grow at
a faster rate than our ability to cope with it.

SUMMARY
4.4.