PN-ACB-491

ATTACHMENT V

DEVELOPMENT INFORMATION SYSTEM

Background Information

In support of other efforts to improve program planning, project design and program/project evaluation, AID is endeavoring to improve its information storage and retrieval system. A special task force on library and information retrieval has recommended the establishment of a Development Information Service along the lines of the attached report, "Establishing a Development Information Service (DIS). While decisions have not yet been taken on all the recommendations in the report, it is possible to conclude at this time that the three proposed new PBAR automated systems discussed in separate papers (CPDB, ESDB, PPT) will encompass a good portion of the data discussed in the DIS report, pp. 17 to 21, and that the data will be accessible to the new Development Information Service. Furthermore, it is anticipated that the DIS will have an automated file(s) for state of art data. The content of these files is being elaborated at this time. (See attached Scope of Work).

Attachments:

A. Establishing a Development Information Service (DIS)

B. Scope of Work (Functional Requirements)

9/12/75

Attachment: A

PN-ACB-491

BEST AVAILABLE COPY

ESTABLISHING A

DEVELOPMENT INFORMATION SERVICE (DIS)

Submitted to:

Agency for International Development
Task Force on Library and Information Retrieval
Washington, D. C. 20523

By:

Practical Concepts Incorporated
1730 Rhode Island Avenue, N.W.
Suite 200
Washington, D. C. 20036

August 8, 1975

BEST AVAILABLE COPY

PREFACE

This report is submitted to the Ide Committee (The AID Library and Information Retrieval Task Force) by Practical Concepts Incorporated (PCI). This report presents the analysis and recommendations resulting from PCI's support to that committee. Because of the complexity of the issues addressed by this analysis, and its interrelationship to other efforts PCI is undertaking for AID, we owe a vote of thanks to many individuals, including the following members of PCI's own staff:

PCI Contributors to This Report

L. J. Rosenberg	T. Schmidt
R. Formichella	A. Ritchie
R. J. Turner	D. Sachs
M. Hageboeck	M. Marx
A. Feiner	R. Popper

PCI is solely responsible for deficiencies or errors in this report. Any credit is due the members of the Ide Committee:

Members of the AID Committee on Library and Information Retrieval

W. C. Ide	S. Applegate	D. Myren
J. Roush	J. Chamberlain	L. Rhodes
J. Hafenrichter	J. Donnelly	H. Turner
D. Steinberg		

TABLE OF CONTENTS

		PAGE
PREFACE		
I.	SUMMARY	I-1
	A. Super Summary	I-1
	B. Summary of Conclusions	I-3
	C. Summary of Recommendations	I-4
II.	DIS OBJECTIVES AND ALTERNATIVES	II-1
	A. Introduction and Summary of Analysis	II-1
	B. Development of a Functional Specification for the Development Information Service (DIS)	II-6
	C. DIS Files and Functions	II-12
	D. Cost Sensitivity: Theoretical Considerations	II-25
	E. Development Information Service (DIS) Functions and Organization	II-33
	F. Basis for Calculating DIS Costs	II-38
	G. Personnel and Contracting	II-41
III.	SCOPE OF WORK FOR DESIGN & IMPLEMENTATION OF A DEVELOPMENT INFORMATION CENTER	III-1
	A. Basic Objectives and Approach	III-2
	B. Overall (PPT) Schedule for the DIS System Design and Implementation Activity	III-4
	C. Activities and Schedules for the DIS Design Phase	III-8
	D. Manpower Requirements for the System Design Phase (Phase IIB)	III-18
IV.	MANAGEMENT PLAN	IV-1
	A. Assumptions	IV-1
	B. AID's Current Expenditures for Information Services	IV-2
	C. Costs of Developing and Operating DIS	IV-7
	D. Organization and Staffing	IV-15

SECTION ONE

SUMMARY

A. SUPER SUMMARY

AID currently spends more than $2,000,000 per year on information services, yet AID Project Designers are not being provided the information they need. We propose that AID establish a Development Information Service (DIS) that will provide the information needed by Project Designers and integrate existing information activities.

We further recommend that the Development Information Service, after fully meeting the information needs of AID Project Designers, have the potential for becoming an information analysis center that defines the development profession.

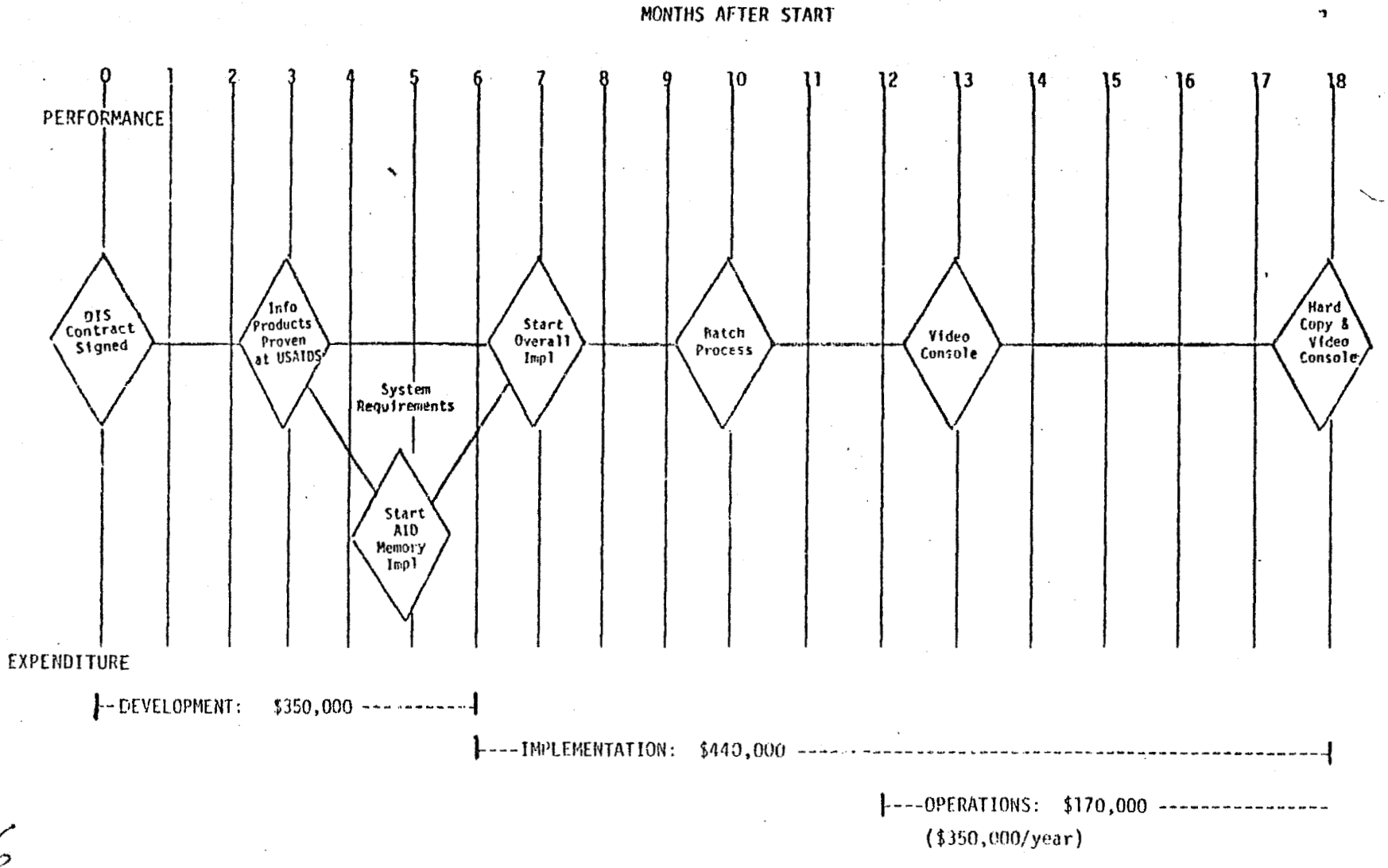
To develop and implement the DIS will cost about $800,000, spent over an 18-month period and during which time information products will be provided to USAIDs. Annual operating budget for a full-scale information service should be approximately $340,000 per year.

Initial DIS design and implementation activities will require contractor support. However, it is recommended that AID top management identify and select a Director of the DIS who will work closely with the contractor team during the development phase and assume operational responsibilities during implementation. The DIS Director should be available for this assignment for a period of 18 months. He should be a senior grade official with experience in information analysis and project management. Experience with AID or development programs is not necessary.

A performance and expenditure schedule for DIS development and implementation is summarized graphically in Figure 1-1.

FIGURE 1-1

SUMMARY PERFORMANCE & EXPENDITURE SCHEDULE FOR DIS



6

B. SUMMARY OF CONCLUSIONS

The Agency should establish a Development Information Service (DIS) as an information analysis center, per the following definition of that term:

> "An information analysis center is a formally structured organizational unit specifically (but not necessarily exclusively) established for the purpose of acquiring, selecting, storing, retrieving, evaluating, analyzing, and synthesizing a body of information in a clearly defined specialized field or pertaining to a specified mission with the intent of compiling, digesting, repackaging or otherwise organizing and presenting pertinent information in a form most authoritative, timely and useful to a society of peers and management."

It may be appropriate for DIS to establish the definitive collection for the development profession. Just as the medical profession is in one sense defined by the National Library of Medicine, the "development profession" can be defined by DIS.

- By taking advantage of related efforts in automation -- particularly the PBAR initiatives to establish the AID MIS -- DIS can have automated capability relatively quickly and at relatively low cost.
- It is feasible that DIS provide direct user service at the USAIDs, through use of interactive remote access consoles.
- The USAIDs have indicated enthusiasm for the DIS concept and skepticism as to whether Washington will in fact live up to the DIS promise.
- DIS capacity and capability can be developed in evolutionary manner -- one small step at a time to achieve immediate but modest levels of capability without compromising the long-term potential.

C. SUMMARY OF RECOMMENDATIONS

- DIS be immediately established as a service bureau within AID.
- DIS implementation adhere to the following:
  - Immediately (30 days after start-up) DIS provides information, based on ARC and other AID collections, to USAID project designers regarding a frequently encountered class of project;
  - DIS provides representative "in-depth" analyses within 3 months after start-up;
  - DIS operational capability established, and the operational system fully defined, at the end of 6 months;
  - Batch processing capability be established nine months after start-up;
  - If full-scale automation is chosen, provide real-time remote access capability (video only) at selected locations within 12 months;
  - Within 18 months real-time remote access, including hard copy print-out, available in the selected USAIDs;
  - DIS be jointly staffed by AID and contractor personnel.
- That ARC and all of its related functions and budget be put within DIS.
- That PPC or PBAR appoint the technical monitor for the contract to operate DIS.
- That a DIS Board of Directors be formed, consisting of representatives from PPC, PBAR, TAB, management services, and the contractor chosen to initiate DIS implementation.
- That the "Ide Committee" be disbanded in favor of the Board of Directors, and the PPC/PBAR monitorship, which shall supervise DIS until such time as DIS achieves full-scale operations.
- When DIS achieves full-scale operations, the Board of Directors select a permanent organizational home for the DIS.
- That a Director of the DIS be chosen to serve for an 18-month term, with the clear understanding that he will step down in favor of a new Director after DIS is fully operational.

SECTION TWO

DIS OBJECTIVES AND ALTERNATIVES

A. INTRODUCTION AND SUMMARY OF ANALYSIS

The Agency needs to improve the way in which information is acquired, stored and disseminated. A library/information science orientation needs to be brought to bear on these issues.

Discussions among the Ide Task Force members centered around information priorities; whether the greatest need was for better technical information, better capture of AID's own experience with related projects and programs, etc. The issue of information priorities was not fully resolved. Rather the Task Force decided to let information priorities be determined according to user priorities -- and the highest priority user is the USAID project design team.

The Task Force concluded that:

- USAID Project Designers must be provided the information they need to create first-rate projects;
- AID should make the investment in personnel and facilities as necessary to provide that information.

Specific recommendations and an implementation approach are discussed in the following three paragraphs:

1. Key Concepts;
2. Mechanism for Implementation: A "Development Information Service";
3. An Action Program.

1. Key Concepts

- Information needs of USAID project designers should be given the highest priority.

- If USAID project design information needs are fully met, then all other needs can be met with low incremental investment.

- Any information "system" must have human intermediation to ensure proper sensitivity to both user needs and information availability.

- Information needs can be characterised in three basic categories:

  (1) Information regarding experience with similar projects, programs, and problems (not necessarily limited to AID's activities);

  (2) Technical data, specifying the state-of-the-art in project-related technologies;

  (3) Context data describing social, anthropoligical, economic, and political variables bearing on LDC problems/programs.

- AID must aggressively outreach to obtain information regarding all three types of information.

- AID must immediately develop mechanisms for sharing its own experience and experience of other donors, with project designers. Thus, among the information sets, first priority must be given to "experience data".

- The flow of information to project designers must be oriented to decision making needs.

- Information activities should be integrated with AID's evaluation system and studies, and with PBAR initiatives to establish a "country program data bank" and a "performance tracking system".

2. Mechanism for Implementation: The Development Information Service (DIS)

The Task Force recommended that information and library activities of the Agency be functionally integrated within a single organization -- the Development Information Service. The charter for the Development Information Service (DIS) would be to immediately meet the information needs of project designers and to eventually evolve into a full service information facility. Salient characteristics of the DIS include:

- The DIS is oriented towards obtaining, maintaining, and providing decision-driven information. That is, the information of interest derives from the real needs of project designers to make decisions concerning alternative project strategies.

- The DIS provides analytic, not descriptive, information. Rather than giving a user the full body of information related to his projects, DIS performs the analytic task of culling the literature and synthesizing an information package explicitly tailored to his needs.

- The key point is access to, not physical possession of the data needed. The Agency need not and should not commit itself a priori to collecting and maintaining a comprehensive body of data, but should instead mount an aggressive program to obtain data from existing sources external to AID.

- The DIS is oriented to information users rather than information sources. Its mission is to determine what information project designers need and get it for them, in the form they need it and when they need it.

- The DIS will bring to the Agency a new orientation toward information and library sciences, with staff who will work directly with information users to ensure that their needs are fully met.

- The DIS will use both conventional and unconventional storage and indexing techniques as suited to the job at hand.

- The DIS will use the current AID reference center and other libraries on an as needed basis, and will improve techniques for indexing and retrieving project, program, and project specific information.

- The DIS will maintain a comprehensive file of "AID-unique" data, concentrating on current information.

- AID memory will emphasize retention of evaluation data, and of the dissenting points of view presented in pre-project documentation.

- The DIS will not attempt to duplicate information collections where those collections are already adequate for AID needs. It will provide access to those collections and will codify them as required to provide AID users with:

  a) Condensed statements of the state-of-the-art relevant to their needs; and,

  b) References and access to further documentation as may be required.

- Based on interaction with the USAID project designers, DIS will produce information packages for the different project design stages in response to USAID needs and requests.

- It is envisioned that such information packages would include:

  1) Prior and current experience with similar projects and programs, considering both AID and other donor activities;

  2) Codified statements of the state-of-the-art regarding relevant technical information, and relevant social, cultural and economic "context" data.

The relationship of DIS to its users and information sources is defined in Figure 2-1.

3. Action Program

The DIS priority must be a dynamic institution, continually evolving in response to information needs and availability. We can get started now, solving immediate priority information needs at the same time that we further define DIS's long-term options. Therefore, the Task Force recommends a six month period of prototype operations, during which time the DIS will:

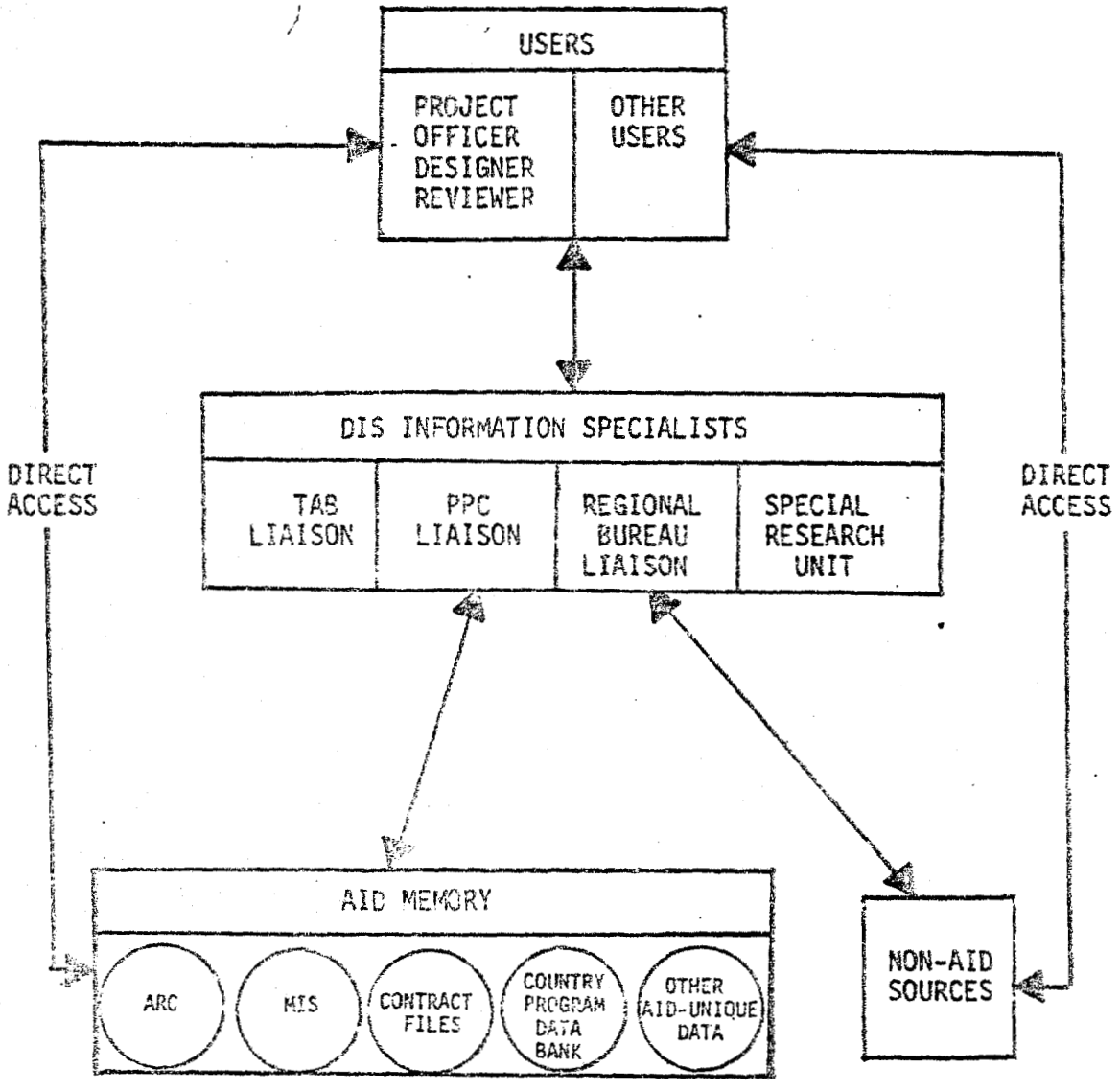


Figure 2-1: Functional Relationship of DIS to Users and Information Sources.

DIS Information Specialists will provide a service to those requesting information but will not preclude their direct access to AID or other information collections. Both bibliographic and summary data will be provided.

1) Define and meet AID information needs for a priority topic or topics;

2) Fully develop a plan for implementing an "AID Memory";

3) Based on its prototype activities, define whether and how DIS should identify, store, index, and codify information available from outside of AID:

4) Collect, analyze and disseminate "experience data" for use by project designers.

B. DEVELOPMENT OF A FUNCTIONAL SPECIFICATION FOR THE DEVELOPMENT INFORMATION SERVICE (DIS)

In order to develop a design specification from which we can deduce system hardware and software, we first develop a functional specification describing the operating characteristics of the system as seen by the user. To allow the "Ide Committee" active participation in the development of functional specifications, PCI offered the following discussion:

1. DIS Modes of Operation;
2. Answers to Basic Queries Regarding Each DIS Mode;
3. Parameters of a DIS Functional Specification.

1. DIS Modes of Operation

There are three plausible modes of DIS operation. For each of these modes we must consider how the system looks from the user's point of view:

a. Automatic Responses -- information forwarded from DIS to the user without explicit triggering events;

b. Semi-Automatic Responses -- information forwarded to users based upon triggering events that are not DIS specific (e.g., routine program and project submissions);

c. Query Responses -- answers to specific information requests.

It is useful to indicate the "purposes" to which the outputs in each of the DIS modes are addressed:

a. The automatic mode defines and characterizes the DIS data base and capabilities so that user requests will be appropriate to DIS capabilities.

b. The semi-automatic mode provides project summaries and related technical data as they have been garnered from existing AID documentation and sources in order to inventory the state-of-the-art as it is known to AID. Although such semi-automatic responses will include references to collections other than DIS, no attempt would normally be made to access or summarize collections outside of active "AID memory".

c. Responses to specific queries inventory any and all data sources in order to answer specific questions posed by the DIS interrogator.

2. Answers to Basic Queries Regarding Each DIS Mode (From the User's Point of View)

For each of the three above noted modes of operation, we answer basic questions characterizing DIS operations from the user's point of view.

1. How does the user gain access to the DIS file?

A. Automatic Mode: Basic materials are sent to the Mission. It is a Mission option from that point on as to the nature and extent of access to this file. (One option would be to maintain, in a single location, looseleaf summaries of DIS data.)

B. Semi-Automatic Mode: Triggered by normal document submissions (e.g., DAPs and PIDs). Responses are cabled, airgrammed or pouched to preparers/signators as appropriate.

C. Issue and Problem Oriented Responses: Mission formally requests problems or topics for DIS consideration. The ability to respond depends upon the information priorities as set by policy makers (including the evaluation community) as well as by the number of information requests and the resources allocated to DIS. (Note that the driving variable for DIS costs will be the number and nature of special requests that DIS responds to.)

2. What information does the user receive?

A. Automatic Mode: Descriptions of the basic materials contained in DIS memory -- with periodic updates, as well as episodic circulation of key information products. Routine descriptive data might include lists of new publications and projects, evaluation reports, etc. Episodic information products, circulated on an as completed basis, might include special information responses to topics of general interest (e.g., if DIS does an inventory of water resource activities, on demand for project designers in the Middle East, that information product might be circulated to all USAIDs).

B. Semi-Automatic Mode: In response to specific types of project documentation. Triggering documents to which DIS should respond might include the DAP, the PID, and the PRP.

In response to the DAP, the DIS will identify programs having similar goals, alternative purposes related to those goals, as well as sector studies, context data, and bibliographies as appropriate to issues raised in the DAP. Project summaries would be forwarded for those projects where good evaluation data are available.

In response to the PID, project summaries would be forwarded for those projects having similar purposes, helping the project designers identify alternative outputs for achieving similar purposes as well as inputs for achieving each type of output. Key assumptions -- with emphasis upon those assumptions that have not proven true in the past -- will also be highlighted. It is in response to the PID that the DIS will focus its responses, as it is at this point (post-PID) that the details of project design should be set forth.

It is possible that in response to the PRP, additional data will be forthcoming from DIS with emphasis upon contractor assessments to allow for selection of alternative implementing agents, detailed schedules, etc. Informational product from the Agency's MIS as well as from the contractor evaluation reports would normally be forthcoming from this point.

Note: The information products defined above are representative rather than comprehensive and information sets are not mutually exclusive.

C. Issue and Problem Oriented Responses: The information product will include results of analyses (performed by ad hoc groups, information sepcialists and analysts) as well as alternative solutions and identification of plausible data sources not covered by the DIS investigation. Context, technical, and project and program data will be forwarded as appropriate. Supporting data of such analyses will include data normally provided as part of the semi-automatic response, but in this mode of operation, DIS will normally outreach to and summarize various non-AID sources.

3. How soon will the user receive a response from DIS?

A. Automatic Mode: This question is not really relevant to the automatic mode. However, it is worth noting that quarterly or semi-annual updates are plausible. It is probable that in no event would submissions be less frequent than annual.

B. Semi-Automatic Mode: With a fixed (say, 30-day) period following receipt of a DAP, PID, PRP, etc.

C. Issue and Problem Oriented Responses: The response time to special requests will be determined on a user-by-user basis. The period required for analysis will depend in large part on the complexity of the problem addressed and the extensiveness and intensiveness of the analysis to be undertaken. It seems reasonable to expect that the DIS should at least acknowledge receipt of the user's information request within (say) 7 working days of receipt. Thus, another output of the DIS would include an acknowledgement of an information request and a probable date at which the information would be provided the requestor. (Special requests to DIS should include "must have" dates -- that is, the requestor should indicate the latest date on which he can accept the inforamtion requested. As a minimum, DIS should quickly note: (a) the plausibility of responding within the time available, and (b) the quality of the search that can be accomplished within that period.)

4. In what form will DIS products be provided the user?

A. Automatic Mode: Probably in the form of looseleaf books and periodic updating sheets for those looseleaf books, as well as bound copies of special reports, etc.

B. Semi-Automatic Mode: Existing abstracts, project summaries, reports and bibliographies. Non-AID sources not included in AID supplied bibliographies will be identified to allow for direct access to such sources.

C. Issue and Problem Oriented Responses: Early submissions and/or supporting submissions may be the same as in the semi-automatic response modes but targeted at the specific problem or question. These responses will frequently include references to individuals and/or data collections not accessed by the DIS information analyst.

5. By what means will DIS products be forwarded to the users?

A. Automatic Mode: Hard copy normally submitted via pouch.

B. Semi-Automatic Mode: Hard copy submitted via pouch and/or hand carried by technical staff participating in related projects. Airgram or cable responses will be used for summaries and to identify the nature of responses that will be forthcoming. Where comprehensive bibliographies are provided it is probable that microfiche or similar condensation techniques will be used.

C. Issue and Problem Oriented Responses: Same as for semi-automatic modes but frequently results will be handcarried to USAID Missions by those technical design staff from AID/W who work with DIS analysts to develop the information package and subsequently are TDY team members working in the USAID to develop the project design. Telephone or other more timely mechanisms for communication will be used as the situation warrants.

3. Parameters for a DIS Functional Specification

The basic parameters establishing the functional specification for an information system are:

1. Extensiveness of coverage: the degree to which the DIS search exhausts the set of all potentially relevant information.

2. Relevance of the information actually provided: the degree to which the DIS selects and/or edits from among the total information of potential interest to select specifically those items germane to the respondent's interest.

3. Extent of abstraction, summarization and analysis of the data base to respond specifically to the user's interests.

4. Timeliness: the time between receipt of a user request and the forwarding of information meeting that request.

5. Information quality: the confidence that the user has that the information is current, valid and reliable.

6. Number of simultaneous queries that the system can handle.

7. Total number of inquiries that the system processes during the course of a given working period.

8. Accessibility: how hard is it for the user to obtain the system services.

The Committee's deliberations regarding the above specifications were focussed on answering the following questions for each DIS mode:

1. Response time: What is the elapsed time between a respondent's query and the reply to that query?

2. Number of simultaneous inquiries that the system handles -- How many people can use the system at once? What happens to the user when he addresses his query to the system but it is currently saturated?

3. How many inquiries is the system expected to handle on any given day? Over the course of a year? Are there peak loading requirements?

4. Extensiveness of coverage: How many of what kind of data sources must the DIS access in order to develop the information package?

5. Depth of coverage: To what extent must DIS analyze, annotate, abstract, etc., as opposed to simply identify and/or forward information?

6. Accessibility: How hard is it for the user to get to the system? What must he do in order to ensure that his request is received? To ensure that it is honored and met within the time schedule he has set forth?

7. To what extent is DIS to be a broker for other data sources: That is, to what extent should DIS be monitoring FAO, World Bank, etc., information products in or-er to select from among these products and distrubute those deemed of high relevance to AID users?

8. Brevity: How concise should the information product be? To what extent should DIS be dealing with distillations as opposed to documents?

9. Specificity: How much information will the user tolerate that is not directly relevant to his topic of interest?

10. Serendipity: How much deliberately non-relevant information will be forwarded in the hopes of sparking some creative act?

11. Data quality: How valid must the data be? Must findings be certified by a committee of experts? How reliable must the data be -- how bad is it if two independent data analysis teams come to different conclusions?

12. In what form must system products be? Books? Xerox copies? Microfiche? Etc.

The Committee reviewed the above list of questions and developed the functional specification shown in Table 2-1.

C. DIS FILES AND FUNCTIONS

1. DIS Operations Under Conditions of Maximum Capability

It is useful to examine the DIS files to consider what the DIS file structure might be like in the best case operational profile -- that is, for the 'maximum operational capability':

1. DIS defines the development science, maintaining a core literature that is so definitive;

2. The central data base is automated;

3. All non-library data are machine accessible;

4. Interactive consoles (at USAIDs and AID/W) provide access to the automated data base and can provide hard copy graphics and text.

20

TABLE 2-1

LIBRARY & INFORMATION RETRIEVAL SYSTEM TASK FORCE
RESPONSE TO QUESTIONS POSED BY
PRACTICAL CONCEPTS INCORPORATED

QUESTION	AUTOMATIC RESPONSE MODE	SEMI-AUTOMATIC RESPONSE MODE	QUERY RESPONSE MODE
1. Response time.	N.A.	Project Level (PID) - 5 working days after project has been approved in A.I.D./W. Sector Level (PIO/T) - 20 working days after receipt of PIO/T.	Project Level - Up to 10 working days. Sector Level - Up to 25 working days.
2. Number of simultaneous inquiries. How many references at once?	N.A.	It is assumed DIS will operate from an automated data base using at least two terminals to answer request from Mission and A.I.D./W staff. A.I.D./W bureaus will also have access to the data base for their own operational needs; however, DIS will service specific requests for information packages, specialized bibliographies, and reference materials.	
What happens if system is saturated?		Second priority will be given to semi-automatic responses.	First priority will be given to specific queries.
3. How many inquiries.			
Daily		(Can't be estimated at this time.)	)
Annually		Project Level - 250 Sector Level - 25	Project Level - 250 Sector Level - 25
Peak loading		Project Level -July-August Sector Level - NA	Project Level - Preceeding PID submission (April-May) and following semi-automatic response (August-September.) Sector Level - NA

Practical Concepts Incorporated

2

4. Extensiveness of coverage	A.I.D. Sources: Expanded Country Data Base (including Country Data File, Country Economic Profile Data, Financial Data), ARC holdings, Master Project File, etc. Technical Sources: National Medical Library, National Agricultural Library, etc. Other Donors: IBRD, FAO, WHO (Health). ILO (Labor), UNESCO (Education), FPA (Population), UNITAR, (Social Development Research), etc. Bi-lateral Dunors: OECD/DAC, SIDA (Sweeden) CIDA & RISD (Canada), etc.		
5. Depth of Coverage	Organized packages of project type documentation; abstract of phamplets, extensive reports and books and provide listings of other donor agencies materials-include abstracts where readily available.		
6. Accessibility: a. How hard is it for user to get to the system. b. What must be done to be sure his request is received. c. and that is time schedule is honored.	N.A.	N.A.	Extremely easy-telephone, walk in, telegram or airgram Rely on communications system Rely on target response time if exceeded send telegram.
7. To what extent is DIS to be a broker for other data sources?	Should have reference materials available with recommended references selected by DIS, Technical Advisory Board.	DIS Technical Advisory Board should select specific references to include in information packages.	DIS should provide abstracts of pertinent outside documentation selected to answer specific query.

Practical Concepts Incorporated

22

8. Brevity	Listings and abstracts journals-backed-up by on-demand printing service.	Provide summary project data such as Project Performance Tracking History List (Attached) and Log. Frame, some documents and references. Reference documents may be obtained from on-demand printing system.	Provide analysis of documentation and forward abstracts of documents- documents themselves are to be obtained from on-demand printing system.
9. Specificity	Tolerate a fair amount of non-relevant data since users will make own selections from listings	May tolerate some non-relevant data since user will realize that packaged material will not satisfy each problem fully.	Will tolerate little non-relevant material since user will expect careful research to satisfy his specific needs
10. Serendipity- hope to spark creativity.	N. A.	Some - Provide a cover sheet for packaged material giving name of researcher and inviting inquiry along other avenues.	More-Provide a personal note from researcher that will recommend other approaches
11. Data quality	Little screening of materials included in system.	Program documents in information packages will be provided as they appear with a summary as to quality or differences in approach Technical documents will be carefully chosen by a board of technical advisors.	Will provide obvious valid information. Otherwise query will be given to a board of technical advisors.
12. Form of System's product.	Listings and abstract journals-backed-up by on-demand printing system.	Information packages, listings and journals backed up by on-demand printing system.	Analytical summary, specific bibliography-backed-up by on-demand printing system.

Under these conditions we consider the following issues regarding DIS composition:

1. Number and type of files: How many kinds of information are there that we need to access? (What is a taxonomy that is appropriate from the user's point of view?)

2. Content of the data files (What data elements are included?);

3. The form in which those files are maintained (machinable, microfiche, hard-copy, sequencing of data elements, etc.);

4. Indexing and access (How do we find the information elements of interest?);

5. Method and frequency with which the files are updated;

6. Analytical and processing capability beyond basic information retrieval functions.

a. Number and Type of Project Files

We envision five basic types of files:

1. Project File -- Containing all Project Unique Information

2. Country and Program File

3. Evaluation and Special Studies File

4. Context File

5. Technical File

Descriptions of these different files follow. It is worth noting that these files considerably overlap PBAR files. For example, the project file will include and integrate data elements from the PPT system, the Country Program Data Base, and to a lesser extent from the financial management system. More importantly, the country and program file will include all information the P-BAR will have in the Country Program Data Bank. Between the context file and the country and program file, DIS

would also exhaust and supplement the set of information that is in PBAR's Economic and Social Data Base. The DIS technical file is of course not covered by P-BAR initiatives. The relative overlap between DIS and the AID MIS files is characterized graphically in Figure 2-2.

b. Content of the Data Files

Given an expanded definition of the DIS, in which it essentially defines development, it is necessary that the DIS data files exhaust the set of "core information" relevant to that science. That is, DIS must identify potentially relevant information and, based on its relative value, summarize, index and maintain the information or the document, or discard the document. This of course does not mean that the entire set of data would be maintained in a machinable form. Certainly at the base of the DIS information pyramid there will always be a conventional document collection.

(1) The Project File

This file either contains or has deliberately excluded (as not worth maintaining) all project-specific documentation. The basic content and format of an automated project file might be as shown in Figure 2-3. Substantive requirements for the project file are that it respond to the bulk of queries regarding the individual project, be amenable to aggregation in terms of key variables of interest, be conveniently indexable, and provide references to other data files both internal to DIS (e.g., the technical data file) as well as external to DIS (e.g., principal investigators, technical references outside DIS holdings).

FIGURE 2-3

POSSIBLE FORMAT AND CONTENT OF AUTOMATED PROJECT FILE

1. Project ID: Significant coding (e.g., per PBAR)

2. Project Status: Begin and end dates; current percentage disbursed

3. LogFrame: In natural language: shows planned and actuals. MoV's include references to evaluation reports where appropriate.

4. PPT and Status

5. Project Summary: 100 word abstract of approach and results.

6. Key Words: Index terms for key work searches. Natual language; glossary evolves.

7. Contractor ID and Evaluation

8. Costs Per Output: Plan and actual

9. Key Technical References

10. Key Project and Evaluation References

11. Key "Context" References

12. Key People: Project management; technical

(2) The Country and Program Data File

The country program data file contains national level planning data, with emphasis upon goals, programmatic hierarchies, and identification of "work breakdown structures" associated with these goals. Basic AID documentation that would be contained in this file would include the DAP 1 and DAP 2, both in hard copy form and in terms of summaries consistent with indexing and retrieval needs. For example, the automated portion of this data base should include national level goals and baselines, sub-goals and sub-goal level assumptions required to achieve those goals, as well as purposes associated with those individual sub-goals. The country program data file will also contain information from the LDCs themselves and from other donors. Thus DIS would treat World Bank and LDC generated reports in the same fashion as AID reports -- summarizing and abstracting where appropriate, and maintaining hard copy and microfiche within the library facility.

(3) Special Studies and Evaluation Data Base

Because of the importance of the evaluation system, and of related special studies (performed by DIS and other analysts), we allocate a special data file to this storage of this information. Although the data will undoubtedly be of great importance to all users, it will have special significance to DIS, which will be an extensive user of this data file in updating other data files as well as responding to special requests for information. It is also through searching of this data base, in combination with the country program data base, that we would hope to highlight opportunities where investment in information (evaluation and analysis itself) can provide high payback.

27

Although project level evaluation reports will most probably be filed in hard copy form in the project data file, the machinable abstracts and summaries would be maintained in this evaluation file. It is the objective of the special studies and evaluation file to point to new directions in maintaining other files as well as to identify the technical and contextual issues that should be properly considered part of the "development science" body of knowledge.

(4) The Context Data File

The context data file starts with AID (and may be normally "entered" via AID assumptions) and reaches far enough to include baseline social, economic, cultural, and anthropological information of proven and potential importance to development planners. This data bank will supplement the country program data bank by adding models and analyses that are not on a country-specific basis, by providing additional references to macro-level political, social, and economic issues, and, most importantly, by codifying the state of knowledge regarding cultural and anthropological issues that bear upon development. It is in this data base that one would find descriptions of the family and economic activities of tribal cultures, up-to-date information regarding quality of life and other social indicators, etc.

Emphasis in developing the data base will be on the failed assumptions and implicit hypotheses where the source of the difficulty was in our inadequate understanding of the social, cultural, and economic processes.

(5) The Technical Data File

This data file must contain information such that an AID planner may gain ready access to any or all sources of technical expertise and knowledge. In theory, as there are few a priori restrictions on the technical mechanisms that AID may choose to utilize, the technical data file must exhaust the field of human technology. As a practical matter, we have a good empirical basis for knowing the types of

technology that development activities normally encompass and can predict many of the new directions. Thus, the probable organization of the data file will be around the technical fields covered by AID's current project portfolio. Much of the basic work in codifying these technical fields has of course already been done by TAB and the summary data contained in the DIS automated file should be compatible with (and perhaps provide automatic access to) the PARIS system.

In response to a routine query, the DIS technical data file must contain sufficient information that for any development-related technology, it can identify the documents, individuals, or institutions, most appropriate to the technology of interest. In response to special queries, the DIS analyst must in fact consider all information sources -- public and private -- as being functionally part of his information domain.

C. Form of the Data Files

The form of each data file will be essentially the same. It will consist of:

1. An Automated File, which can be viewed at remote consoles and provide hard-copy print-out of both graphics and text on demand.
2. An Active Collection, of actual documents for which high demand is anticipated. There would be hard copies in the library and, where copyrights allow, the collection would be available in microfiche form for review and distribution. Collections and bibliographies can be sent to the Missions on microfiche.
3. An Archive: documents that are obsolete or not important enough to maintain in active memory. These are stored in microfiche form only; documents are destroyed.

The structure of the automated files should be such as to allow variable size records to be inserted. (Some project records will be much longer than others, and it would be inefficient to arbitrarily allocate the maximum storage space required to each record.)

29

Within the hard copy and microfiche libraries, the basic structure of the five data files would be maintained. However, there would also be a master index (automatically updated) using conventional library indexing techniques. This will simplify inter-library transactions. This will allow us to insert references to documents that are not maintained within the DIS collection. The 'master index' might include index cards identical to those used for the DIS collection with the exception that the actual document identified is within another library - e.g., the State Department library. (Different colored index cards might simplify such inter-library indexing.) Note also that just as adding references to one of the five DIS files would automatically generate a master index, so could generation of a master index card automatically generate index cards (and where appropriate machine references) in each of the five DIS data files.

d. Indexing and Access

The basic indexing technique to be used by DIS will be Logical Framework elements. Since the project files, the country and program data files, the special studies and evaluation file, and the context file, all can be based on the Logical Framework approach, it becomes appropriate to demand as a data file specification that both external users and DIS analysts be able to access information based upon natural language, Logical Framework entries. Thus, if an individual wants a reading on alternative technologies, he should be able to simply specify the output targets of interest to him.

Similarly, if he's interested in learning more about contextual issues, specification of any two or more elemtns of the Logical Framework (e.g., inputs and outputs; outputs and purpose; assumptions and purpose; etc.) should be sufficient to gain entry to the contextual information that is relevant to his needs.

e. File Updates

In the fully automated system the project file could be updated on a real time basis as new information is forwarded from the field and is fed into other systems (e.g., the PPT). The special studies and evaluation file will be updated on a continuing basis, as new information becomes available. However, for the other data files, it may be appropriate to update them (a) periodically and (b) when special DIS studies indicate the need for update. For the country and program data file, the appearance of a new DAP would also stimulate an update. Archives would periodically be destroyed.

f. Analytical Capability of the DIS

The capability of the DIS -- shared between the interactive user console (which may well be a smart console) and the centrally located main frame and peripherals -- could let the user perform analytical as well as aggregating and search functions. A realistic target is that the DIS be able to run any of the models that it offers its users from automated memory. It should also be able to draw graphs from DIS data. Although it is practical and perhaps advisable to give DIS the capability to perform such functions, it is not necessarily true that the users will be best served by pre-programming DIS to run specialized models. Extensive user need would have to be established before dedicating machine space and time to pre-programmed question and answer formats (as for example, those used by the PLATO population activities).

2. DIS Operations for Minimum Capability

If a minimum capability alternative is chosen for DIS there still should be an extensive machine search capability available to it through proper utilization of other Agency systems. Specifically, we havein mind the MIS files (PPT, FINMIS) the country program data bank, the economic and

social data bank as planned by P-BAR initiatives, as well as the PARIS data bank available through TAB. Hence, the real cost of implementing DIS's automated capability is simply the cost of integrating those files with a single set of users -- project and program designers -- in mind. It is moot as to whether DIS performance of that integrating function should be considered a DIS cost. The files should, in any event, be integrated, and DIS use should benefit development of these files. The issue is made still more complex by forthcoming changes in the evaluation systems. Because of the immediate pertinence of the evaluation system to the project and program design functions and its less immediate relevance to the improvements in implementation management envisioned by PBAR, it is probable that de facto DIS will be responsible for integrating the evaluation component into the automated data base.

If DIS has no internal automated capability, DIS would consist of an expanded library and microfiche service -- with the collection screened and abstracted as before, but in this case with DIS maintaining only hard copy, supplemented by DIS information analysts. The DIS information analysts are then themselves users of the various automated data files as well as of the DIS and other library, to respond to inquiries. (See Figure 2-4.)

D. COST SENSITIVITY: THEORETICAL CONSIDERATIONS

The four parameters with which system costs vary most directly are:

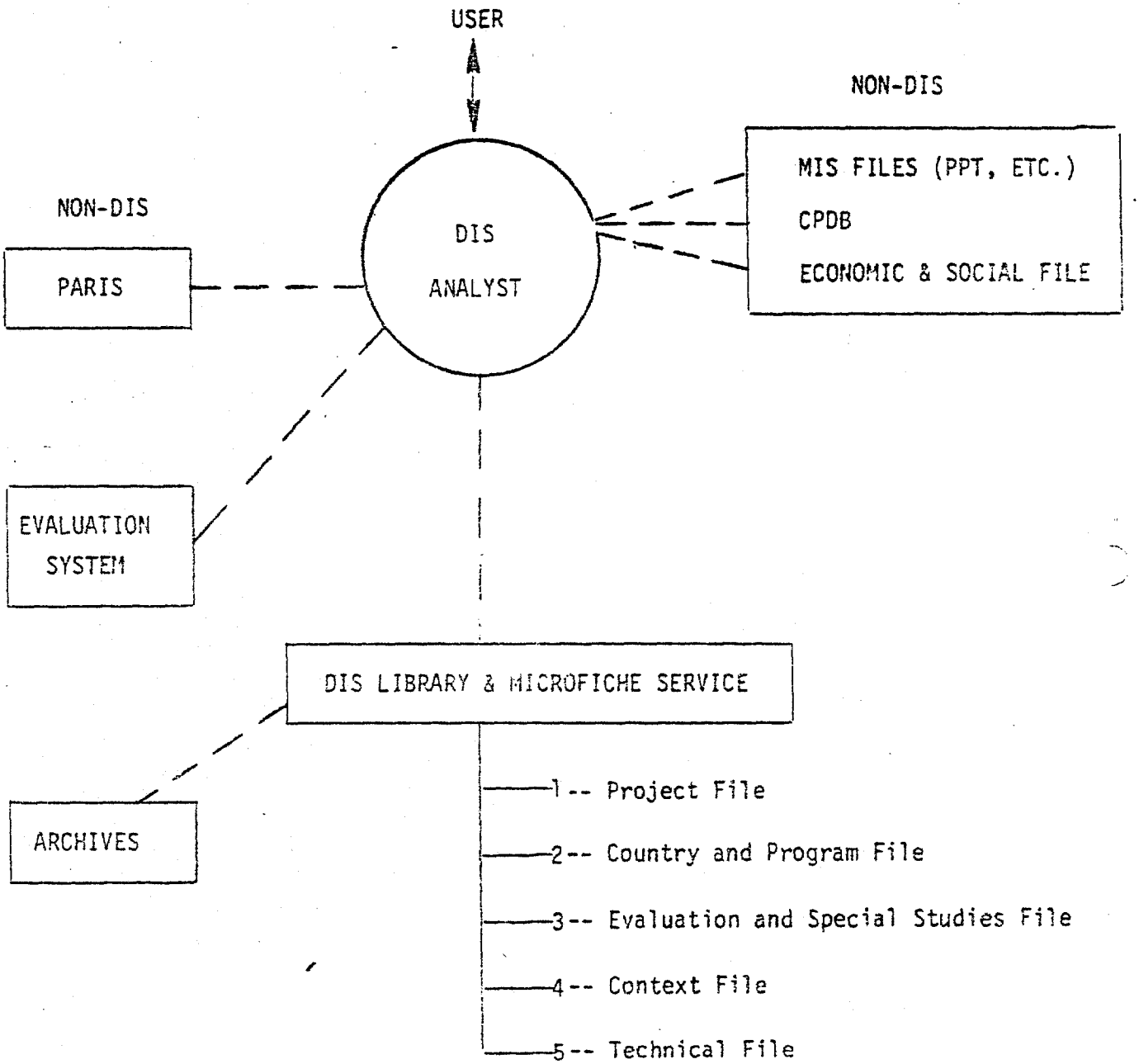
1. Extensiveness of retrieval;

2. Relevance -- the degree to which information provided explicitly responds to the interrogator's question;

3. Timeliness;

4. Validity -- the degree of confidence the user has or should have in the information actually provided.

FIGURE 2-4



Minimum Automation of DIS has DIS Providing Hard-Copy Reference Documents that Support Other Automated Systems. DIS analysts are users of the existing automated capability being developed under PBAR, PPC, and TAB sponsorship.

As a practical matter, timeliness can be considered separately. Our primary concern must be with extensiveness or completeness of coverage, and with quality of information as indicated by both relevance and validity.

1. Extensiveness of Coverage

The concept of extensiveness is graphically illustrated in Figure 2-5. We have two sets of information -- (1) the set of information that is relevant to the user's needs and (2) the set of information that is actually retrived by the system.*

A typical library search might be represented as in Figure 2-5A. Here there is substantial overlap between the information retrived and the relevant information, but a substantial amount of information retrieved is not relevant. The user must both screen more information than he uses and it is problemmatic as to whether he can move from the set of relevant information that is also retrieved to the full set of all relevant information.

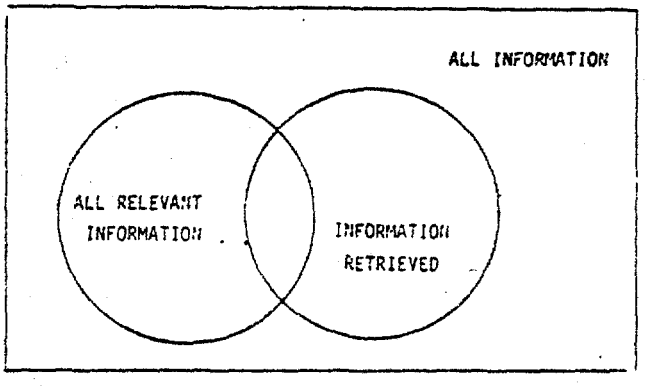
A retrieval concept such as that shown in Figure 2-5A should be relatively inexpensive to operate.

Where it is important that the user exhaust the set of all relevant information, the retrieval concept is that shown in Figure 2-5B. Here we must overretrieve to ensure that we exhaust the set of relevant information, and both the system and the user must pay a price for this. Human analysis is a requisite if you are to ensure that all relevant information is retrieved and, given the size of most data files and information sets, it is probable that machine assistance would also be

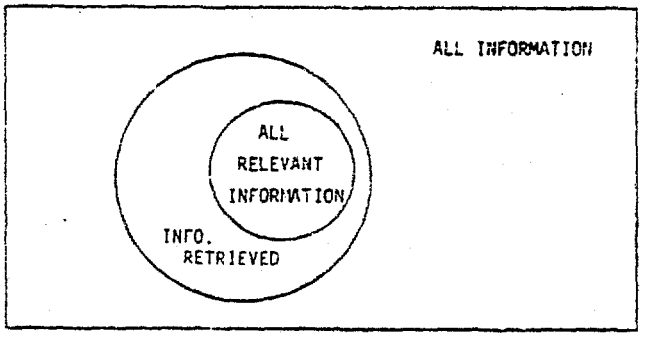
* We ignore the extent to which relevance is a function of quality, which is itself a function of why the user is making his interrogation. All information is relevant to every topic if you carry your information substance far enough. This will be discussed later in terms of relevance, and we will simply assume here that there is a defined and discrete set of "relevant information".

FIGURE 2-5

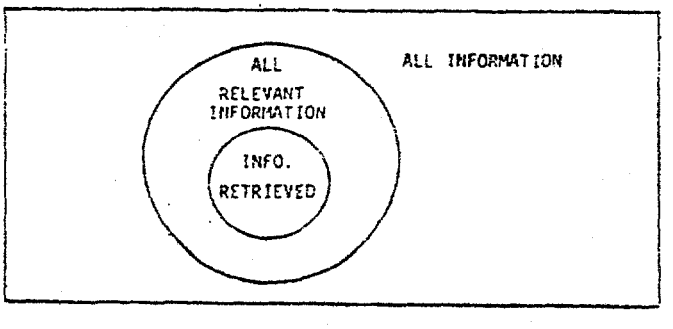
(A) Within the set of "all information", there is a subset of information relevant to a given need. Typically, an information system retrieves some (less than 100%) of the relevant information and some information that is not relevant. The former can be considered a signal and the latter noise. Information specialists are fond of such terms as "trash rate" (percent of non-relevant information), and "hit/miss ratio" (where "hits" are relevant data, misses irrelevant).



(B) If comprehensive retrieval is desired then some "trash" must be retrieved.



(C) A 100% hit rate implies non-comprehensive retrieval.



required. Both design and operating costs for such a system can be expected to be high. Further, the requirements for collection of all relevant information suggests that the size of the centrally maintained file will be ever increasing. Hence, file maintenance costs can become an extremely important factor in affecting costs.

The information concept currently recommended for the DIS is shown in Figure 2-5C. Here our objective is to retrieve the most salient portion of the relevant information. This is economical from the user's point of view in that all information he obtains is relevant information. Although coverage is not comprehensive, he can reinterrogate the system, based upon knowledge gained in his initial interrogation(s), until he is satisfied that adequate retrieval has been effected. Given proper attention to system design, the initial retrieval can be highly systematized, usually automated. At the same time, central files need not be large. Further, extensive human mediation by the DIS analyst is required only after the user has interacted with the DIS file one or more times. This should both keep the DIS's analysts start-up time at a relatively low level and also means that the analyst's work with the user is both a direct service to that user and a file update and maintenance process. (The analyst may add information to the basic file, provide additional indices or cross-index, or indicate limitations of data sources he contacts during his secondary retrieval activities.)

The cost implications of the information concepts shown in Figure 2-5C are relatively low operating costs, based upon slightly higher costs of implementation. Higher implementation costs are experienced because of the need to carefully define operating processes and procedures, develop relatively sophisticated indices, and perhaps most importantly, to carefully control the quality as well as quantity of information put into file.

The information concept illustrated by Figure 2-5C can be effected at low operating costs because in essence it has low fixed costs and the variable costs associated with responding to user queries are at once a direct service and simultaneously improve the system.

2. INFORMATION QUALITY

We will briefly discuss information quality in terms of its two components of relevance and validity.

a. Relevance

In one sense restricting the information set to only that which is relevant to a given use may be a theoretical as well as a practical impossibility. We assume there is a question in the mind of our user. Perfect relevance is identically only that information required to answer the question. On the otherhand if we gave him only information that answered the question it is doubtful that the user would be satisfied. He would want more information -- to ensure that he formulated his question properly and to know enough about the source of that answer to establish its validity and credibility. A retrieval containing no information extraneous to the user's desires is impossible, because different users will have different criteria for acceptance.

We also know that as long as we restrict our system to existing documentation -- articles, final reports, etc. -- then we again cannot have perfect relevance. Books, articles and reports are not written to answer questions posed by system users. Thus they contain much information that is irrelevant. A given report may contain only a few words or numbers of importance to our users -- although those few words or numbers may be very important.

The trade-off here is between the total size of a file (the amount of information and number of documents yielded by a given search) and the number of users of that particular file. Clearly if there are only one or two journal articles on a given subject it is perfectly appropriate for the system to maintain those articles in their "natural" form, unless the number of users becomes truely exceptional. (As the number of requestors increases, it might become cost effective for the system to summarize these

required. Both design and operating costs for such a system can be expected to be high. Further, the requirements for collection of all relevant information suggests that the size of the centrally maintained file will be ever increasing. Hence, file maintenance costs can become an extremely important factor in affecting costs.

The information concept currently recommended for the DIS is shown in Figure 2-5C. Here our objective is to retrieve the most salient portion of the relevant information. This is economical from the user's point of view in that all information he obtains is relevant information. Although coverage is not comprehensive, he can reinterrogate the system, based upon knowledge gained in his initial interrogation(s), until he is satisfied that adequate retrieval has been effected. Given proper attention to system design, the initial retrieval can be highly systematized, usually automated. At the same time, central files need not be large. Further, extensive human mediation by the DIS analyst is required only after the user has interacted with the DIS file one or more times. This should both keep the DIS's analysts start-up time at a relatively low level and also means that the analyst's work with the user is <u>both</u> a direct service to that user and a file update and maintenance process. (The analyst may add information to the basic file, provide additional indices or cross-index, or indicate limitations of data sources he contacts during his secondary retrieval activities.)

The cost implications of the information concepts shown in Figure 2-5C are relatively low operating costs, based upon slightly higher costs of implementation. Higher implementation costs are experienced because of the need to carefully define operating processes and procedures, develop relatively sophisticated indices, and perhaps most importantly, to carefully control the quality as well as quantity of information put into file.

The information concept illustrated by Figure 2-5C can be effected at low operating costs because in essence it has low fixed costs and the variable costs associated with responding to user queries are at once a direct service and simultaneously improve the system.

2. INFORMATION QUALITY

We will briefly discuss information quality in terms of its two components of relevance and validity.

a. Relevance

In one sense restricting the information set to only that which is relevant to a given use may be a theoretical as well as a practical impossibility. We assume there is a question in the mind of our user. Perfect relevance is identically only that information required to answer the question. On the otherhand if we gave him only information that answered the question it is doubtful that the user would be satisfied. He would want more information -- to ensure that he formulated his question properly and to know enough about the source of that answer to establish its validity and credibility. A retrieval containing no information extraneous to the user's desires is impossible, because different users will have different criteria for acceptance.

We also know that as long as we restrict our system to existing documentation -- articles, final reports, etc. -- then we again cannot have perfect relevance. Books, articles and reports are not written to answer questions posed by system users. Thus they contain much information that is irrelevant. A given report may contain only a few words or numbers of importance to our users -- although those few words or numbers may be very important.

The trade-off here is between the total size of a file (the amount of information and number of documents yielded by a given search) and the number of users of that particular file. Clearly if there are only one or two journal articles on a given subject it is perfectly appropriate for the system to maintain those articles in their "natural" form, unless the number of users becomes truely exceptional. (As the number of requestors increases, it might become cost effective for the system to summarize these

articles. In theory, if I can abstract a journal article in 8 hours, I am justified to do so if I can save 16 users 1/2 hour each. As a practical matter the cost effectiveness calculation is much more complex. There are much subtler pros and cons involved in abstraction and summarization.)

On the other hand, if I have thousands of documents in my collection, all of which bear on a given topic, then clearly I must do something to manage user access to the collection. The first thing I do is develop taxonomies such that I direct my user to smaller and smaller subsets of the file. However, what happens when the finest grain subset still yields scores, hundreds, or thousands of documents? This can occur because when there is an extensive literature on a relatively fine grain subject or, equally important to us, when the user resists getting "too fine grained" about his information needs because he is still in the process of formulating his questions. In such a case we have no alternative but to summarize or abstract.

If we anticipate very few users, then it is appropriate for the system to present the entire collection -- letting the user do his own screening and selection. However, where we anticipate multiple users or where the data topic is extremely important, we are forced to abstract and summarize. As a practical matter we must recognize that very few users, when faced with a task of digesting several hundred references, will in fact do so. Thus, <u>we must assume that without our assistance in predigesting the data file users will miss relevant information</u>.

The other trade-off in building up a selective file is of course the extent to which the burden of indexing and abstracting is put on the DIS versus the original developer of the information. For AID data, this trade-off is already headed in a satisfactory direction.

The Logical Framework approach is an extremely useful, perfectly standardized project taxonomy -- dividing a project into 16 different elements, each of which is indexable and retrievable. The PAR, although never intended as a

document for transferring experience, provides a modicum of standardization. A revised PAR should take DIS needs into account. Development of a standardized format for in-depth evaluations should both improve the in-depth evaluations and provide a conveniently indexable data that will be highly relevant to wide ranges of users.

b. Validity

The essential test of the validity of information provided by the system is whether or not the user succeeds in the enterprise he undertakes as a result of that information. That "success or fail" information will be available through the evaluation system, thus again underscoring the importance of the evaluation system to the DIS.

The important cost trade-off here is the extent to which system responses must be validated by various technical or peerages -- certifying the validity ofthe information provided.

As with most of the other issues surrounding the DIS, the issue of validity is less complex for AID generated, project-specific, information then it is for "technical" information. For most purposes, it seems probable that identification of the projects from which experience has been gleaned will be an adequate indicator of information validity. On the other hand, certification as to the state-of-the-art in growing soy beans will be quite a different issue. Certainly an important connection must be made to such sources of expertise and certification as the TAB.

descriptive, information. Rather than giving a user the full body of information related to his projects, DIS performs the analytic task of culling the literature and synthesizing an information package explicitly tailored to his needs.

There are many sources of information and referral services outside of AID. The key point is access to, not physical possession of the data needed. The Agency need not and should not commit itself a priori to collecting and maintaining a comprehensive body of data, but should instead mount an agressive but eclectic information access and utilization program.

The DIS will maintain a comprehensive file of "AID-unique" data -- establishing a "selective" AID memory. The memory is selective in that, consistent with DIS's orientation, it would continually cull its files to retain only those data worth maintaining. It is possible but not necessary to integrate AID archives with AID memory. Such integration could facilitate maintenance of contracting files.

2. DIS STRUCTURE

Figure 2-6 illustrates the organizational components of DIS. Personnel are drawn from operational and service bureaus and bring their specialized expertise together in a single unit, drawing on the resources of their "parent" organizations.

The overall structure of the DIS consists of four basic elements:

1. The Office of the Director of the DIS;
2. Information service standards, a staff function;
3. A utilization activity;
4. An accession activity.

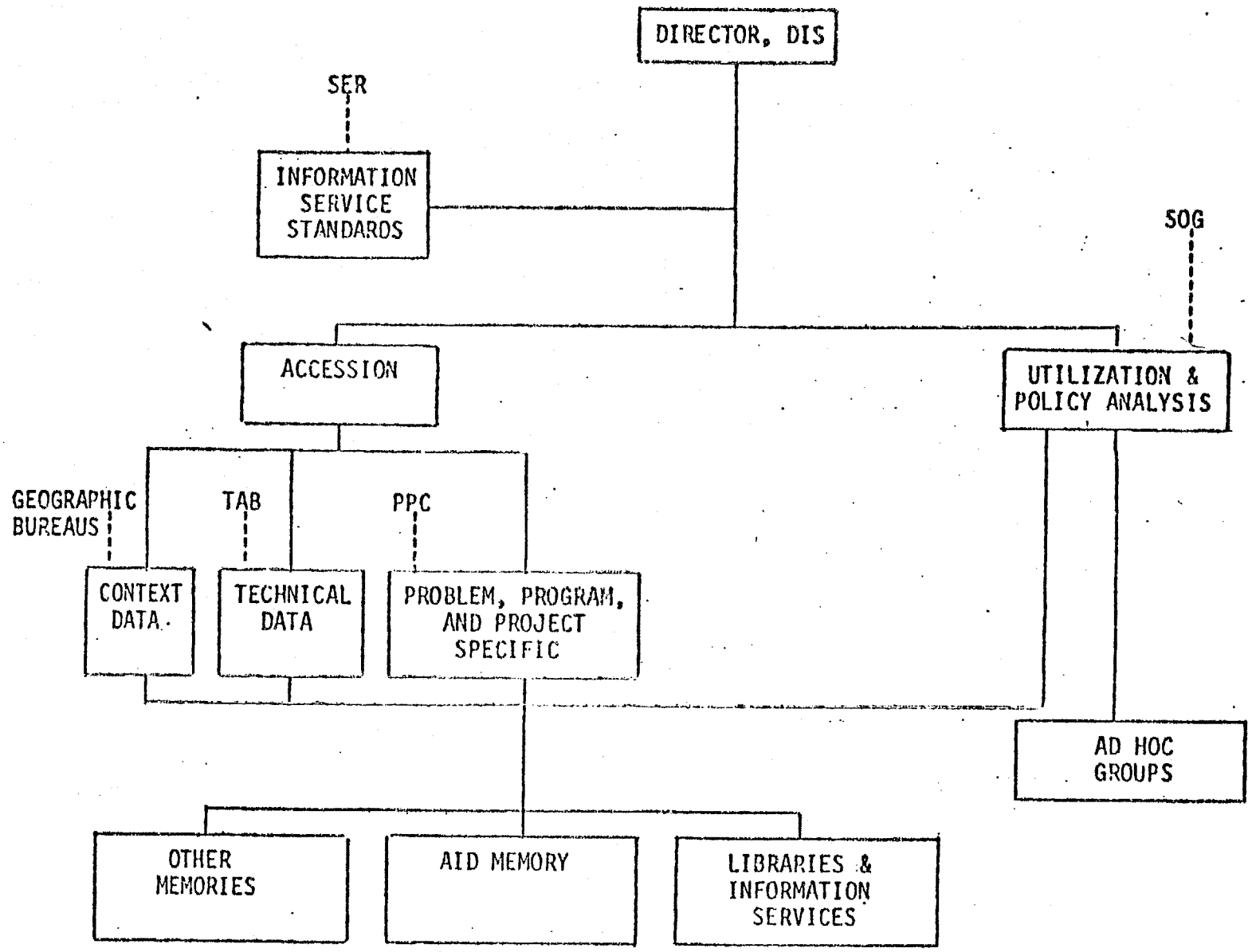


FIGURE 2-6: ORGANIZATIONAL COMPONENTS AND RESOURCES OF DIS

The utilization element is key. The utilization function will focus as required around individual user needs to ensure that the full information resources of the Agency and elsewhere are brought to bear.

The utilization function as well as other users of the DIS are supported by the accession function. The accession function acquires and provides three categories of information:

(1) "Context" and Baseline Information;
(2) Technical Information;
(3) Problem, Project, Program, and Management Information.

These sets of information overlap. Overlap is desirable because it provides built-in cross-references among the information sets.

The "sub-units" within DIS (noted in Figure 2-6) will be responsible for information relevant to the above information sets and will:

a) Define and provide information consistent with the DIS charter;

b) Distill and codify the state-of-the-art in the information area for which they are responsible, and generate "routine" information packages;

c) Provide bibliographic searches and indexing capability as required to access any and all information services or libraries.

Another unit -- development of information standards -- will set standards for bibliographies and abstracts, establish specifications for information storage and retrieval, develop standards for literature reviews (whether prepared by AID or contractors), define bibliographic and cross-reference requirements for contractor reports, etc.

The "Utilization and Policy Analysis" unit will be charged with utilization per se. This unit works directly with users, setting up ad hoc problem-oriented user groups as required to ensure that DIS serves real needs. Ad hoc user groups would crystallize around special problems and programs for which there is little development experience and develop "non-routine" information. This unit will also evaluate DIS output, perform policy analyses, support responses to Congressional requests, etc. -- aggressively utilizing information just as other DIS components aggressively access information.

3. DIS IMPLEMENTATION

The Agency is lacking in both the skills and the attitudes required to establish a successful DIS. We are faced with the need to effect important change not just in the way AID staff access information, but also in the way we build the utilization of information to our normal programming and management processes. We recommend that these two deficiencies be addressed by:

1. Establishing a DIS with a temporary charter of initiating prototype operations and planning permanent operations;

2. Obtaining contractor support to provide necessary technical skills on an interim basis;

3. Assigning AID staff to the DIS on a temporary basis, with the explicit recognition that the individuals involved in the initial operations, planning, and formulation of the DIS will not be part of its permanent staff.

The model we have in mind is similar to that used by industry in establishing major operating facilities. Planning and construction management is responsible for developing a new oil refinery, but does not become part of operating management. Operating management overlaps the initial "project managers" so that there is an efficient but gradual turnover of responsibilities. The temperament and skills required to establish a new facility are very different from the temperament and skills required to operate such a facility.

F. BASIS FOR CALCULATING DIS COSTS

We will here discuss cost of operating and developing DIS. It should be noted that given the on-going initiatives in the evaluation system, and the development of various PBAR data banks, etc., it is PCI's perception that the automated DIS capability can be achieved at a negative cost of development. That is, it is our hypothesis that by integrating the various data banks from the high volume user's point of view (the project and program designers), the DIS can effect economies in development of these various systems that will more than offset the management and integrating costs required to achieve that effect.

1. Cost of Machine Operations

Given PBAR's plans for development of automated data bases of which DIS would be a primary user, there is no cost to DIS automation. That is, it is already planned to have interactive consoles accessing these various data bases. Those consoles would presumably in any event be available to DIS analysts as one of the multiple users. However, it would be appropriate to specify as a DIS unique cost, the cost of perhaps three such "smart" interactive consoles. Given the technology at the moment, the capital cost of the consoles might be around $30,000. Depreciating those consoles over 5 years, we can consider an annualized equipment cost of perhaps $6,000 per year.

There is, further, no reason to expect that the frequency, method, or nature of file updating procedures should be substantially different given DIS as the coordinating body that it would be were these systems to be operated independently. Therefore, we assume no additional cost for file updating.

There is one cost of automation that must be attributed solely to DIS -- the automation of the technical data base. The cost of operating an automated DIS will range from zero (PARIS provides all automated capability used by DIS) to operating a data file in which there are at least as many records as there are projects plus bibliographies and abstracts -- perhaps 500 records of 10,000 words each.

2. Development of the Automated Data Base

This issue is again clouded by the degree to which DIS costs should be considered independent of costs to develop related systems and data bases. On the assumption that there are in fact on-going activities as defined by PBAR and PPC, (the AID MIS and the evaluation systems), then it would be appropriate to assume a DIS capability supplementing those activities and consisting of perhaps 5 analysts plus clerical and support staff. This estimate is exclusive of the costs of developing the technical information data base, which costs should be borne entirely by DIS. The functions to be performed by such individuals, and from which approximate costs can be developed, would be as follows:

1. System development team leader: An information and data processing specialist;
2. Development of the project file;
3. Development of the country and programming data file;
4. Development of the context file;
5. Development of the evaluation and special studies file.

47

Based on rough time estimates already provided the Committee (e.g., by the PBAR task force) the system development activities should embrace approximately 12 months, thus involving 60 professional man-months exclusive of clerical and logistic support.

The machine time and facilities required to perform this integrating function cannot be separated from those required to implement such systems without DIS oversight. Therefore, no machine costs have been estimated. (Costs of terminals for use exclusively by DIS analysts are estimated at $30,000 as noted earlier.)

3. Costs of Developing Technical Data Files

The costs of automating the technical data file will of course bear extensively on the nature of that file. However, it seems safe to assume that given the availability of in-depth information within such systems as PARIS and encapsulations of that information in the special studies and other data bases, that the essential technical file should be limited to brief written statements summarizing the state-of-the-art in a given technical field, plus annotated bibliographies to capping literature, and references to subordinate literature not maintained in DIS. (It would be expected that the capping literature would in every case be maintained within the DIS working library.)

We assume that 100 technical fields would be summarized, and that approximately 10,000 words will be required for each record.

4. The DIS Document Collection: The Working Library and Microfiche

Even if DIS is given the charter to codify the development science -- the maximum capability considered for this system -- it still seems doubtful that the document collection would exceed in size that of a small college library. By a combination of controlled access and open stacks -- e.g., reading rooms in special technical subjects -- we might have to expand the staffing level. However, such staffing should not require more than one junior level librarian per reading room. It might also be to preserve restricted access, DIS should obtain documents to be maintained in the State Department library. This seems like a particularly plausible alternative for secondary readings from which the context file has been derived.

G. PERSONNEL AND CONTRACTING

The activities that DIS personnel must perform are listed in Table 2-2. The order in which they are listed is intended to approximate the degree of difficulty in delegating DIS functions to a service bureau. In this listing the lower the number, the less comfortable AID would be with delegating responsibilities to a service bureau.

The last two items on the list -- maintenance of equipment and distributing microfiche -- are functions that could relatively be easily performed by a service center. The first -- data analysis -- is inevitably at least partially an AID function.

In Table 2-2 we consider three types of staff alternatives:

1. Functions directly performed by government-hire DIS staff;
2. Functions performed under contract;
3. Functions provided by AID's internal service bureaus.

3. DIS Analysis and Management with Contractor Operation of a Service Center

This option is an attractive one in terms of convenience. Direct hire DIS staff then become primary users and managers of the information facility for which they have contracted.

4. A Hybrid: DIS Shares Analysis Functions with a Contractor and Utilizes Existing AID Equipment

Although this may not be a long-term outcome, it is a probable interim solution. AID direct hire and contractor staff cooperatively undertake analytical functions, with contractor services providing the interface between the analyst and existing AID processing capability. This option is extremely flexible. It should be relatively easy to move from this option to any of the others. For example, DIS central staff could begin to take over shared functions and then the information specialist functions provided by the contractor. Alternatively or simultaneously, AID computer center staff could absorb machine-oriented functions. This hybrid is attractive in that it provides for rapid start-up but can quickly phase out contractor activities. It also allows great flexivility for short-term assignment for AID staff to the DIS.

5. DIS Contractor Run

This is the option that was chosen by such government facilities as the NASA Scientific and Technical Information Facility, a comprehensive information center bearing important resemblences to the DIS functions as delineated by the Committee. An important benefit of this approach is that it allows AID staff to concentrate exclusively on analysis of information rather than the techniques of obtaining and processing information. (That is, the contractor facility could generate responses to all semi-automatic and routine requests from both inside and outside the Agency, limiting the Agency's need to modify its own personnel structure and providing a convenient fee for service mechanism.)

We will consider the two boundary conditions -- (1) all functions performed by DIS direct hire personnel and (2) all functions subcontracted. (We need not consider the condition where all functions are performed by a service center internal to AID. That configuration would simply recreate DIS under another name.)

Clearly the determination as to how DIS functions will be performed and distributed will be made during the design phase. However, a few words on some representative alternatives shown in Table 2 are in order.

1. DIS is Self-Contained

The difficulty in this alternative is that DIS gets into the business of equipment maintenance and operation. Since a portion of the attractiveness of the DIS is the availability of existing computer capability within AID, as well as existing systems, this is not a probable outcome.

2. DIS Analysis and Management Plus Use of an AID Service Center

This represents the probable maximum utilization of an AID service center -- where the service center activities themselves are not reporting directly to DIS staff. It is plausible that a service center activity could be responsible for putting data in machine form and updating files. However, a strong case can be made for including the file updating task as a DIS operational responsibility. (This argues strongly for a natural language file updating procedure -- a capability that will be examined closely in the design phase.)

SECTION III

SCOPE OF WORK FOR DESIGN & IMPLEMENTATION OF A DEVELOPMENT INFORMATION CENTER

The Agency for International Development intends to establish an information analysis center, to serve those who plan and implement development projects. This information analysis center shall be known as the Development Information Service (DIS). It shall maintain a comprehensive record of AID project and program information (although not necessarily a comprehensive file of documents), as well as information, bibliographies, etc. from any and all sources, as required to best serve its users. In general, then, the DIS shall serve development scientists consistent with the following (generally accepted) definition of an information analysis center:

> "An information analysis center is a formally structured organizational unit specifically (but not necessarily exclusively) established for the purpose of acquiring, selecting, storing, retrieving, evaluating, analyzing and synthesizing a body of information in a clearly defined specialized field or pertaining to a specified mission with the intent of compiling, digesting, repackaging or otherwise organizing and presenting pertinent information in a form most authoritative, timely and useful to a society of peers and management." 1/

Over time, the DIS may serve non-AID users. However, its first and primary users will be USAID personnel.

1/ This definition was developed originally by Dr. Edward Brady, for presentation at a National Bureau of Standards Conference in 1967. It has subsequently been published in Science Magazine and is the operational definition for more than 100 federally funded information analysis centers in the U.S.

The purpose of the scope of work set forth here is to so design the DIS that it can be implemented within time and cost constraints that are both reasonable and cost-effective in terms of AID's immediate and long-term future. To achieve this purpose it will be necessary to define alternative configurations for review by Agency management.

The overall schedule of DIS activities will consist of three phases:

Phase I: Design and Development
Phase II: Implementation
Phase III: Operations

The scope of work contained here refers only to the first of these phases -- design and development. These design and development activities shall be undertaken to identify and select from among Phase II and Phase III alternatives. Implementation and operation phases are thus discussed in detail sufficient to highlight these alternatives.

To clarify the nature of the DIS, and its method of evolution, the remainder of this section describes:

A. Basic Objectives and Approach
B. PPT schedule for the overall effort;
C. Activities and schedules for the design phase;
D. Discussion of manpower needs during system design.

A. BASIC OBJECTIVES AND APPROACH

The basic approach to DIS design and development is to actually provide the service to DIS users and then based upon that service provision determine the costs, benefits, and most cost effective approaches to full-scale implementation.

Figure 3-1

PCI 72-0

LOGICAL FRAMEWORK FOR SUMMARIZING PROJECT DESIGN

Est. Project Completion Date _____

Date of this Summary _____

Project Title: PARTIAL LOGFRAME: DESIGN STUDY FOR A USAID DEVELOPMENT INFORMATION SYSTEM

DEVELOPMENT HYPOTHESES	NARRATIVE SUMMARY	OBJECTIVELY VERIFIABLE INDICATORS	MEANS OF VERIFICATION	IMPORTANT ASSUMPTIONS
If Purpose, Then Goal	Program Goal: The broader objective to which this project contributes: AID Project and Program Planners have state-of-the-art information conveniently available.	Measures of Goal Achievement: 1. Semi-automatic responses deemed relevant by 90% of recipients and "very valuable" by 50%. 2. Project reviews discern: • better use of experience; • awareness of technical and conceptual issues (including benefit incidence). 3. Estimated value of 1. and 2. at least double the cost of DIS.	Self monitoring DIS system which counts responses and flags deadlines missed.	Concerning long term value of program/project: USAID planners use the information sent out by DIS in making project and program decisions.
If Outputs, Then Purpose	Project Purpose: Initiate implementation of a cost-effective DIS.	Conditions that will indicate purpose has been achieved: End of project status Implementation Phase initiated with endorsement of key Bureaus and personnel.		Affecting purpose-to-goal link: Basic file automation available through PBAR initiatives.
	Outputs: 1. Full scale information analysis support provided for important project or class of projects. 2. Based on 1 above, standard (semi-automatic response) packages prove value to USAIDs. 3. Procedures and organization established to ensure effectiveness of automatic, semi-automatic, and special response modes. 3.1 Organization and staffing defined based on alternative work loads. 3.2 Space and equipment projections and budgets. 3.3 Procedures for liaison with other information sources, libraries, and other AID systems. 4. DIS utilization demonstrated to field and AID/W planners and managers. 5. Procedures for continuing evaluation and improvement of DIS. 6. Procedures for continued enforcement of utilization.	Magnitude of Outputs necessary and sufficient to achieve purpose. 1. Alternative information products -- based upon differing extent and intensiveness of information support -- assessed for cost and benefit. Most cost effective alternative endorsed by USAIDs. 2. Missions respond favorably to information products. Utilization of information reflected in subsequent PIDs and related project documentation. 3.1 Organizational and personnel requirements make sense in terms of availability of personnel as well as system requirements. 3.2 Realistic in making good use of available AID facilities. Cost minimized through integration of existing functions -- e.g., ARC. 3.3 Two-way exchange procedures established with best sources, all plausible sources identified; DIS represented when planning new AID systems. 4. DIS operations simulated and proven to USAID personnel. Information packages developed following DIS procedures are deemed by recipients to have value substantially in excess of projected costs. 5. User evaluation routinely scheduled. Impact evaluations (are projects better because of DIS packages?) scheduled and include interrogation of all users of special studies. 6. Positive as well as negative incentives provided for utilization.	Simulation. Field and AID/W pilot study. Monitoring by AID managers and future users of DIS.	Affecting output-to-purpose link: Successful completion of a DIS design study.

© Practical Concepts Incorporated, 1972
Washington, D.C.

54

The logical framework for design of the DIS is shown in Figure 3-1. The key to this logical framework is immediate and extensive involvement of USAID mission staff 30 days after start up. Thus, the assessment of DIS effectiveness will be based upon its proven value in the field.

Inputs are not shown in Figure 3-1. The inputs and activities are spelled out in subsequent paragraphs of this section, and in Section 4 of this report.

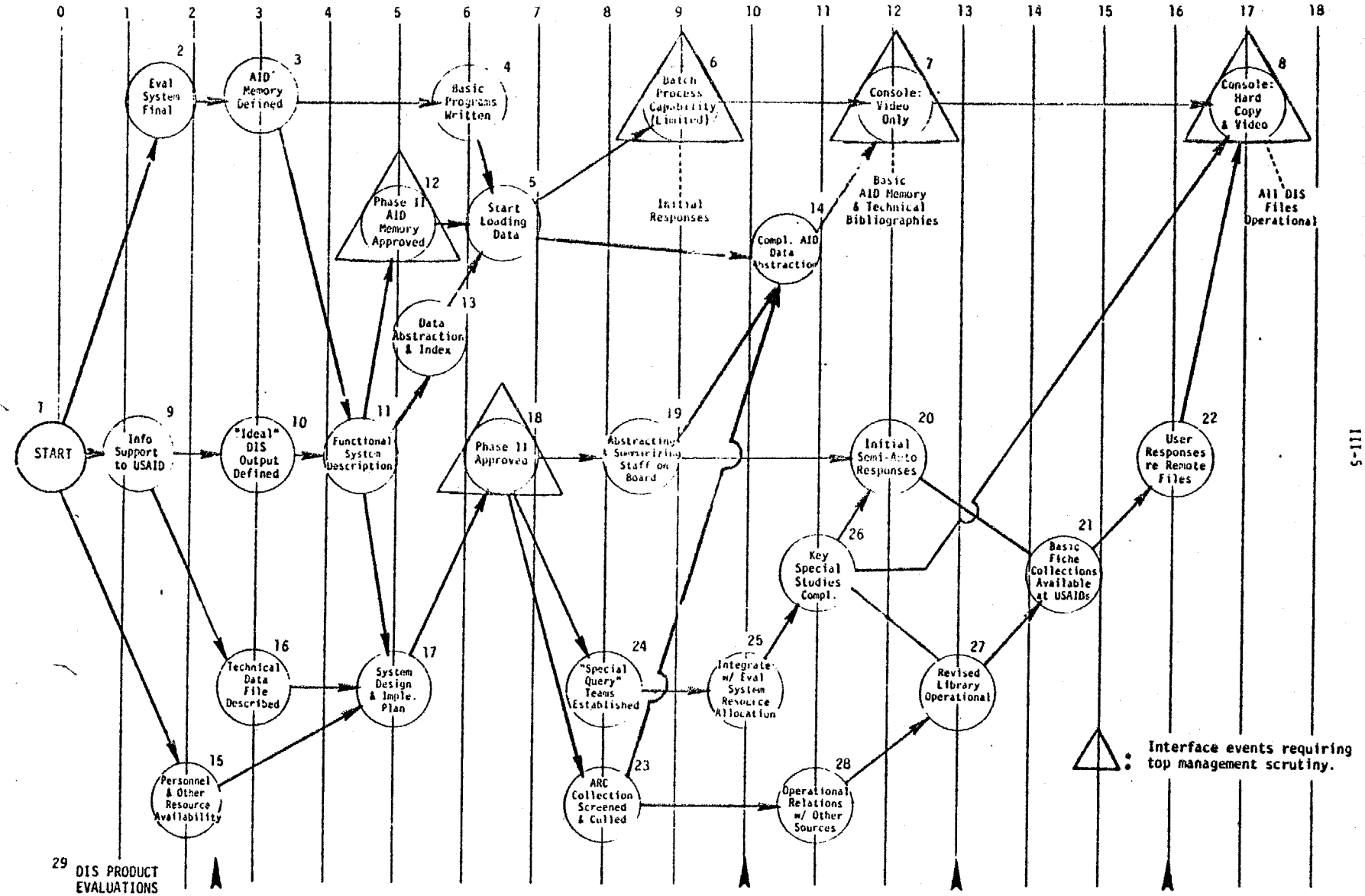
Note that the goal of the design activity: AID project and program planners have state of the art information conveniently available -- becomes the purpose of a subsequent implementation activity and within the manageable interest (therefore the output of) the operational DIS. Thus, the over-riding objective of the design and development activity is to establish a plan by which the goal noted in Figure 3-1 comes within AID's manageable interest.

To be judged successful, the design study must produce a DIS that passes pilot and simulation tests of its ability to provide state of the art information to AID planners. The implementation phase, in order to be judged successful, must actually provide the information. Once the system is operating, then evaluation of the information product must be based solely on the real improvements made in project and program design.

B. OVERALL (PPT) SCHEDULE FOR THE DIS SYSTEM DESIGN AND IMPLEMENTATION ACTIVITY

A PPT schedule for DIS design and implementation is shown in Figure 3-2. PCI estimates that system design and implementation can be completed in approximately 18 months. DIS outputs would be routinely available after the ninth month (batch processing only). Basic (prototype) outputs become available 30 days after start of work.

FIGURE 3-2: OVERALL PPT NET FOR DIS IMPLEMENTATION



1. Start-up should be timed so subsequent activities can provide real value in helping shape the selected project or program. It should also include a resource commitment of a quality and level to cover the transition period between design and implementation. (Availability of such transitional resources will avoid the waste of resources inevitable if we phase down completely before starting implementation.)

2. Definition of the AID memory requires a working definition of the evaluation system and particularly of the revised evaluation reports. Those reports must contain transfer of experience information that will (as a minimum) direct DIS analysts to appropriate projects and evaluation reports. Evaluation reports (and project histories) must be available for a sufficient number of projects that meaningful aggregations can be made.

3. The composition of the AID memory, in terms of number and types of files, data content for each file, etc., should be completed as early as possible during the design study, to allow us to start writing programs for developing the files.

4. By the end of the sixth month, programs should be written for all files where the data elements have been specified. As a minimum there shall be a complete program for at least one file -- probably the project file. Basic file development and updating programs should be available for all files with the possible exception of the technical file.

5. Basic data generated during the design study will be loaded into files immediately after the programs have been developed -- to debug the programs and to get a head start on establishing processing capability.

6. By the end of the ninth month, we should have a capability to batch process such searches as: (1) What projects have purposes similar to this one?; (2) What are the key assumptions that have been made in such projects?; (3) What inputs are normally associated with these outputs?; etc. If our current design approach holds, we will be able to identify projects from any set of Logical Framework elements.

7. As the data base evolves, it will be appropriate to add interactive capability to the batch processing. The minimum level of capability would include responses to such queries as: (1) How many projects have purposes similar to this? (Answer); (2) Show me the first five of these (Answer); (3) Etc. Hard copy output would not be available at the console but would be available on request.

8. At the point of full operation, scheduled for around the 18th month, current plans (based upon the PBAR predictions) call for the addition of hard copy printers at remote locations. This would allow the user to demand and immediately receive hard copy products for any of the information in the automated data base. Such hard copy production would however, be subject to limitations based on the priorities of the user and the availability of reserve capacity. (A plausible compromise would be to honor large-scale hard-copy requests only during non-peak periods of system operations. Thus, hard copy requests might be aggregated and met during the midnight to 3:00 AM period.)

9. This study must simulate information packages that could plausibly be made available by the DIS, assess their value to USAID planners, and project costs of such packages for a fully operational DIS.

10. Based on knowledge of what has been most valued by USAID planners, the design team will identify optimum inputs to and outputs from the DIS -- the most cost-effective transactions -- setting a design goal for full-scale DIS operations.

11. This specifies how the system works, including descriptions of data files, etc., in detail sufficient that it is understandable to lay persons and also immediately useable by the system design team to translate it into an operational implementation plan.

12. To ensure that DIS activities do not delay the other automation activities that it is integrating, we envision a separate but limited approval of the functional system description -- allowing development of the AID memory component. The requirement here is that the AID memory activities be so designed that resources will not be wasted even if DIS analysts make only modest utilization of automated capabilities (beyond those necessary for AID's own memory).

13. Abstraction and indexing of data reviewed during the earlier system activities, as well as complementary data deemed appropriate, must be effected in a quantity and of a quality such that DIS programs can be debugged and run.

14. By the end of the first year of activities, the historical AID project data, abstraction and indexing of prior projects will be essentially complete. By essentially complete we mean that for all projects where it is possible that there will be high user interest, summaries, abstracts, indexes, and machine summaries will have been prepared.

15. Those AID personnel whose qualifications make them appropriate staff members for DIS will be identified. Personnel requirements that cannot be met by existing AID personnel will be "red flagged".

16. The content and format of the technical data file will be established in sufficient detail that system implementation activities can begin. There must be at least some overlap with TAB data systems, (e.g., PARIS) to ensure mutual accessibility.

17. The design must demonstrate its cost-effectiveness in terms of value received vs. incremental costs of operation, and meet all planning requirements imposed by the Agency.

18. Approval should not be just of the implementation, but with full recognition of budgeting for operational costs. Note: Development of the DIS automated capability has not been compressed to the maximum extent, in recognition that inevitably the pacing item will be the availability of good data. Since "good data" will be mostly the result of implementation activities, it must be remembered that every day of delay of Phase II approval will constitute a day of delay in full-scale operations.

19. As abstracting and indexing procedures become debugged and the data file evolves, abstracting and summarizing staff must be brought on board to ensure that the total AID memory is abstracted, indexed, and put into both machine and document files by the end of the 12th month. It is probable that some portion of the abstracting and the summarizing activities will be done under contract, as the initial compilation of the data base will represent a peak load that will probably never again be experienced.

20. By the end of the 12th month, through both batch processing and some interactive console work by DIS analysts, semi-automatic responses should be routinely forwarded for all PIOs. (Earlier responses to PIOs as well as special queries will of course be made. However, at the end of the 12th month, DIS should honor the obligation to respond to each PIO.)

21. High-interest, high-usage portions of the DIS collection will be made available to the USAIDs for analysis of their value. Such collections will be made available only where there is Mission interest, where this is a substantial file of relevance to the Mission needs, and the Mission need can be met in a timely manner.

22. Formal evaluation of the fiche collections will establish cost and benefit for both individual Missions and large-scale dissemination.

23. The ARC documents represent a body of knowledge of indeterminate value. That collection will be screened and culled so that DIS maintains only those documents that are worth keeping and with a clear indication of how and when those documents will be deemed obsolete.

24. Key to the integration of the technical data base with the other data bases is the development of the special query teams. Such teams must include information analysts, TAB subject specialists, subject specialists external to AID, and program personnel.

25. Both evaluation and DIS special queries are investments in information. This event highlights the fact that those investments should be made cooperatively. The key target is that there is a system for sharing information priorities and mutually allocating resources to those priorities.

26. The value of DIS's product must be continually demonstrated. Further, DIS's own special studies will be comprising major portions of the DIS data base. Hence, it is of utmost importance that some special studies be completed and made available and widely disseminated before the start-up of full-scale operations. At least three large-scale special studies should be completed by this point, and at least six in-depth bibliographic responses.

27. This must be a complete collection of all AID documentation deemed appropriate for retrieval by program and project planners and designers, indexed to simplify access, as well as to facilitate inter-library exchanges. Reading and work space should be appropriate to the anticipated demand.

28. Inter-library exchange programs should be established as appropriate. Unless compellingly demonstrated otherwise, this should include as a minimum the World Bank, State Department Library, CIDA, OECD, the Department of Agriculture (e.g., liaison through CRIS to obtain the latest information on agriculture research), Medlars and Medline.

29. During the initial period of evaluation of start-up operations, user response should be evaluated after the fact in terms of both acceptability to the user (relevance, quality, timeliness, [illegible]) and impact -- the extent to which projects are actually improved as a result of DIS products.

The schedule shown in Figure 3-2 reflects recognition that:

1. DIS can provide an important source of energy to accelerate PBAR's program of automating the "AID MIS";

2. Data quality is of paramount importance. Hense, all DIS activities involve active collaboration with users and information sources.

The key interface events, for which top management scrutiny is suggested, are shown in Figure 3-2 inscribed within triangles. These key events are:

Phase two AID memory approved: development of AID's own data banks is scheduled for independent approval to reduce the danger that DIS development will delay the agency's overall automation program. AID memory development is scheduled for approval at the end of the fifth month.

Phase two approval: this event marks the end of development and the beginning of implementation. This is the point at which the agency must commit to a scale of implementation and, to a lesser extent, of operational activities.

Batch process capability should be initiated by the ninth month, at which point the agency should assess the desirability of additional automated capability as well as the value of a more extensive data base.

Video console capability should be added by the end of the twelfth month, at which point the basic questions to be presented to agency top management include whether or not hard copy capability should be added to the remote console, as well as the number and location of such consoles. Tentative decisions should be made as to the desirability of allowing direct interrogation of non DIS files - e.g. through remote access link-ups to the National Library of Medicine, existing commercial systems, etc.

Hard copy as well as video capabilities should be added before the eighteenth month, in at least some remote locations. The pacing item may be telecommunications capability, suggesting that a brief telecommunciations survey be added to implementation activities. At this point the agency will have a good idea of the remote interactive capability desired at the USAIDS and should be considering the extension of that capability to other donors, LDC's, etc.

A conceptual issue to be addressed at each of the key decision points is the extent and nature of human buffering between the DIS and its ultimate users. The basic approach is non-exclusive: we will both provide direct access to DIS and other files and also offer analyst-intermediated service. However, as the system evolves, we should consider optimizing in favor of one of the two approaches. (Both points of view - direct natural-language access to the data base, and the necessity for human intermediation - will be represented in the development activities.)

The overall schedule as shown in Figure 3-2 is concerned primarily with an overview of the system evolution and fails to emphasize the key element of PCI's recommended development phase: direct and immediate service to users in the USAIDs. Therefore, even the casual reviewer of this report should refer to the following discussion of Phase 1 activities.

C. ACTIVITIES AND SCHEDULES FOR THE DIS DESIGN PHASE

The design/development phase will "learn by doing". Information products will be provided to representative users, for a representative projects to both test the value of such products and improve procedures for development the products. By properly choosing the projects to be analyzed, a multiplicity of objectives will be served:

(1) A "special querry response" will be used to directly improve a USAID project/program;

(2) A related "semi-automatic response" will be circulated to planners considering similar projects;

(3) A technical data file will be developed upon which additional results can be based.

To respond to USAID interest in the DIS, initial interactions with the USAID project design team should be effected within 30 days and all information products tested within six months.

The design phase for the Development Information Service (DIS) will be described in terms of five types of activities:

1. Work directly with users;
2. Development of the AID Memory;
3. Development of DIS approaches to subject and problem searches;
4. Assessment of AID's resource capabilities for implementation and operations;
5. Analysis and implementation planning.

These activities are inter-related and will be initiated simultaneously. However, it is clear that the latter two are logically and functionally derived from the first, or as a minimum must be confirmed through subsequent user interactions.

We also recognize that the first of these -- interaction with users -- is of particular importance. In the process of defining information required to support project design activities we will to some extent redefine or clarify the nature of the project design process. The interaction between information provider and project designers should thus be mutually rewarding.

The five above activities are discussed in the following. We assume that a prior activity has occurred -- the selection of both a subject area and a country in which project designers wish to formulate new project initiatives. Examples of such selections might be water resources (subject)

and Ethiopia (country where we seek to formulate projects). Other alternatives might be development management in Portugal, population dynamics in Pakistan, etc. Ideally, the selection would involve both a general topic requiring a subject search (water resources) and a project within that technical field (e.g., purpose: irrigate 40,000 acres in the Sinai). Response of the first mode would allow us to develop techniques for developing a technical data file responding to "special queries"; a response to the latter for providing "semi-automatic responses" based on PIDs. The interaction between these two activities should be beneficial to both and should also generate data for comparing value added by the more extensive response.

1. Work Directly With Users

The sequence of activities for work with users is shown in Figure 3-3. As may be noted we have broken this into two simultaneous and interacting activities -- (1) developing the "semi-automatic response" and (2) developing the "special response" and technical data file.

We assume that the activity is occuring in the period between the PID and the PP. The DIS team simulates four different semi-automatic responses for the same project:

a. Based upon documentation available in AID/W;
b. Based upon work in the USAID to help formulate response requirements;
c. A response based upon the full knowledge of the special search;
d. A "perfect response" in the sense that we prepare it after the fact, knowing what has been most highly valued by the Mission staff.

By comparing the quality of the various responses we will assess the value added by individual information gathering activities. The final test of all activities will of course be quality of the USAID project. (We will have available the designs developed during each interaction.)

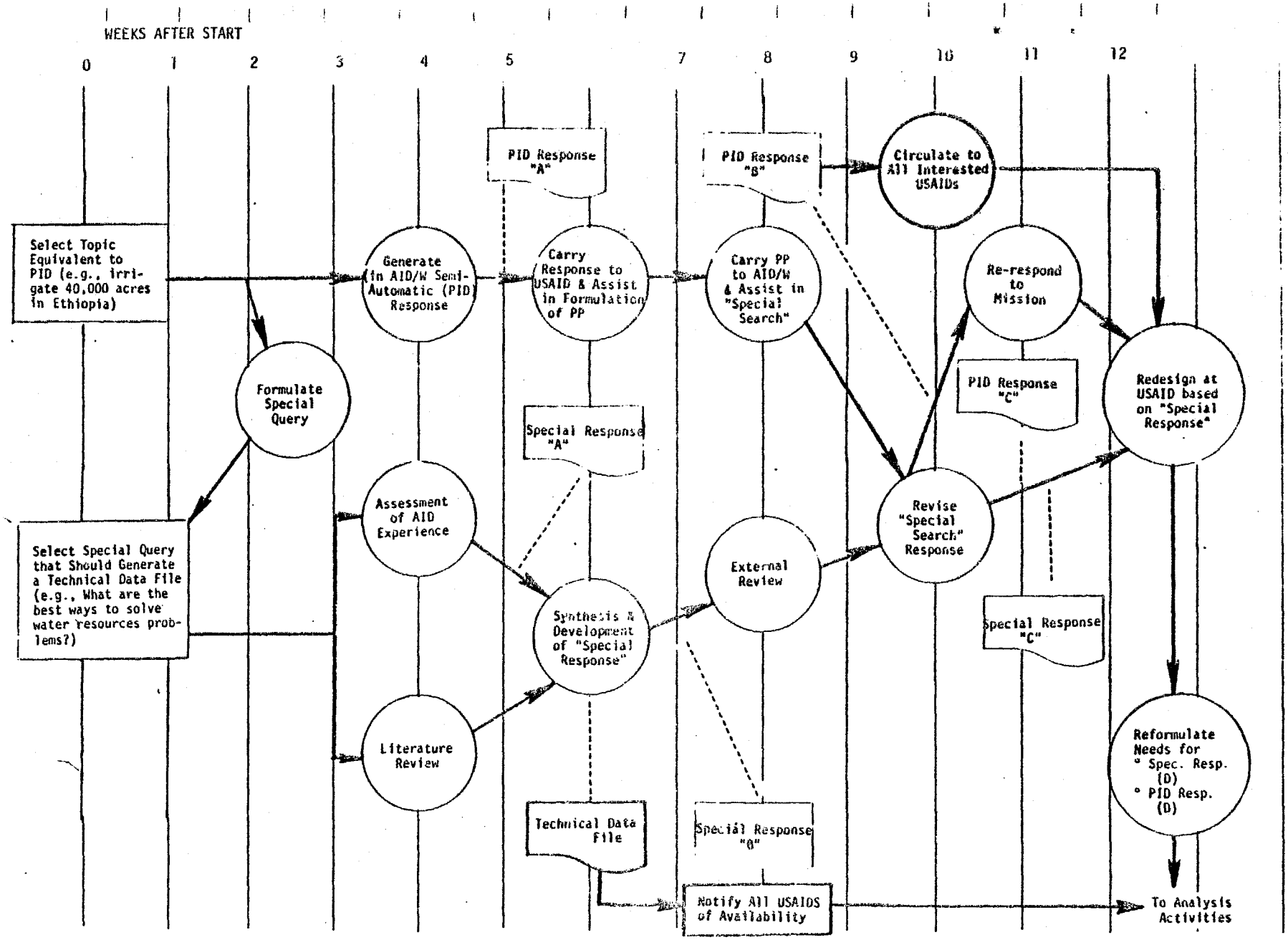


Figure 3-3: The sequence of activities for work with project designers consists of two parallel paths: (1) generation of a semi-automatic (PID) response and (2) generation of a special response in the technical field appropriate to that project. For both activities, the final test of the information product is its real value to the USAID planner. That value is verified and observed through work with USAID staff.

Thus, as a minimum output, the Mission should have a better project. It is realistic to expect that the value added by the study team will exceed study costs -- thus providing our DIS data at negative or at least minimum cost.

In similar fashion, we have four special responses:

A. Based on AID/W assessment of AID experience;

B. Based on AID-managed synthesis of a literature review and its own experience;

C. Based on external technical reviews of the response package, as well as work with the team that had participated in design activities at the USAID;

D. A final formulation, with full knowledge of what has been of greatest value to the Mission.

Simultaneous with the activities to develop optimum responses for Mission needs, the on-site team will also be sharpening their insights into the nature and form of information that USAIDs need to provide DIS analysts in order to improve the quality of the DIS response. Thus, in addition to optimum DIS <u>responses</u>, we will also have optimum DIS requests. The optimization criteria for the information requests would of course be minimizing load upon the requestor while ensuring an effective and efficient DIS response to his needs.

2. <u>Development of the AID Memory</u>

This set of activities will set up the procedures and analytical processes for obtaining and screening all AID information to determine its relevance for DIS. This task is a timely one and it is probable that a hard look from the user's point-of-view -- which is the view that DIS will be taking -- will shed insight into the utility of current documents, and into ways of improving current reporting.

There are two paths to this activitiy. The first will be to directly support the information activities of Task 1 -- obtaining information on past projects through analysis of all available documentation, in an attempt to support the project and program design activities. This will give us some insight into the value as well as availability of AID experience.

The second path for this activity will be an inventory of all AID documentation, development of the mechanisms for screening those documents, indexing them, summarizing them and abstracting them as appropriate in machinable files. DIS is responsible to its colleagues who are planning development projects to maintain a comprehensive record of AID experience. Basic DIS inputs -- to be screened for value -- are listed in Table 3-1.

A network for development of the AID memory is shown in Figure 3-4.

Actual automation of the AID memory should be initiated immediately after the design study.

For further insight into the value of DIS use of AID data, the interested reader is referred to PCI working papers included as Exhibits to this report:

A. Tentative Approach to Integrating the DIS Semi-Automatic Response with AID Planning Processes

B. Representative Information Products Requiring No DIS Analysis

C. Basic Approach to Technical Subject Search, as Used by PCI to Assess State-of-the-Art in Use of Telecommunications to Provide Social Services.

TABLE 3-1

REPRESENTATIVE INPUTS TO BE SCREENED BY DIS FOR RETENTION OR SUMMARIZATION

A. INPUTS

1. AID Inputs

1.1 AID Approval Documents

1.1.1 DAPs
1.1.2 PIDs (including project codes; especially benefit incidence code)
1.1.3 PRPs
1.1.4 PPs

1.2 AID Implementation Documents

1.2.1 PIOs
1.2.2 PPT Reports
1.2.3 Contracts
1.2.4 Interim and Final Reports
1.2.5 All Other Contractually Required Reports

1.3 AID Evaluation Documents

1.3.1 PARs (revised)
1.3.2 Project Completion Reports
1.3.3 Sector Studies and Assessments, Special Evaluations
1.3.4 Contractor Evaluation Reports
1.3.5 Project Histories
1.3.6 DIS Responses to Special Queries
1.3.7 "PPT Histories"

1.4 Other AID

1.4.1 Identification of and access to other systems
- PARIS
- Disaster Relief Data Base
- PPT
- FINMIS
- Indicator Data Bank

1.4.2 Special data needed to form such special data bases as:
- Project matching system
- "Context" data file
- identification of key non-AID references and sources

1.4.3 Special documents requested by:
- "Holdings Committee"
- Special study analysts

2. Non-AID Inputs

2.1 As Established by DIS Special Analyses

2.2 Bibliographies

2.3 TAB Bibliographies by Subject Field

2.4 Bibliographies of "Capping" Data for Social and Economic Context

2.5 As Recommended by Joint DIS/Other Donor Information Symposia

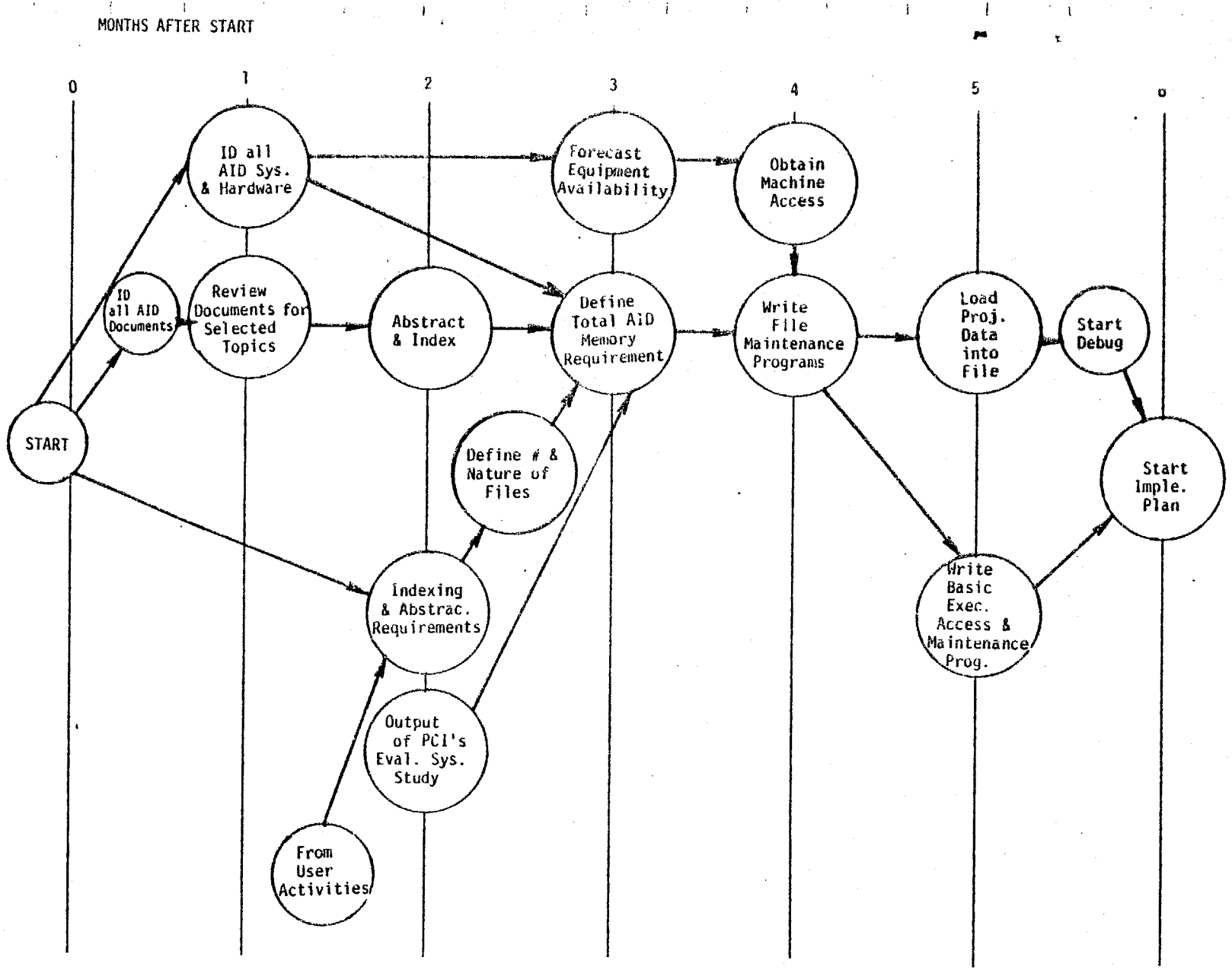


FIGURE 3-4: BASIC SCHEDULE FOR DEVELOPMENT OF THE AID MEMORY

Development of the AID memory will include writing the file maintenance and key executive programs. Two optional activities -- heavily dependent on the availability of quality data as well as on the degree of difficulty in developing the programs -- are the actual loading of project data into file, and the debugging of the basic programs. In any event, the implementation plan for the AID memory can be developed and approved independent of the overall DIS plan.

3. Development of DIS Approaches to Subject and Problem Searches

This activity will consist both of the simulation activities (described in Task 1 as work directly with users) and a separate activity, in cooperation with TAB personnel, to assess the ability and advisability of developing technical subject files independent of special requests -- should or should not DIS encapsulate technical data in advance of active user demand. This will depend heavily upon the number and type of technical fields deemed of interest (which derives naturally from indexing work done to develop the AID memory), the nature of the TAB technical data base, the desirability of active and extensive DIS/TAB cooperative activities (which should be illuminated by the user service activities), and the detection of knowledge gaps (based on the ability of the evaluation system to generate evaluative material of relevant information). A network for activities to develop subject and problem searches is shown in Figure 3-5.

4. Assessment of Agency Personnel and Other Resources

Implementation of the DIS will require both human and physical resources. This activity will inventory the resources available to the Agency in both of these dimensions to determine training, hiring, and equipment needs beyond those currently scheduled. The basic tasks will include:

1. Identification of the skills needed to implement and operate DIS;

2. Work with the Agency's personnel management to determine which individuals have requisite skills and those who could be expected to acquire the requisite skills;

3. Assessment of the hardware and software investments to which the Agency is already committed;

4. Analysis of the above, in conjunction with the system design activity, to develop a plan for staffing DIS development and operations.

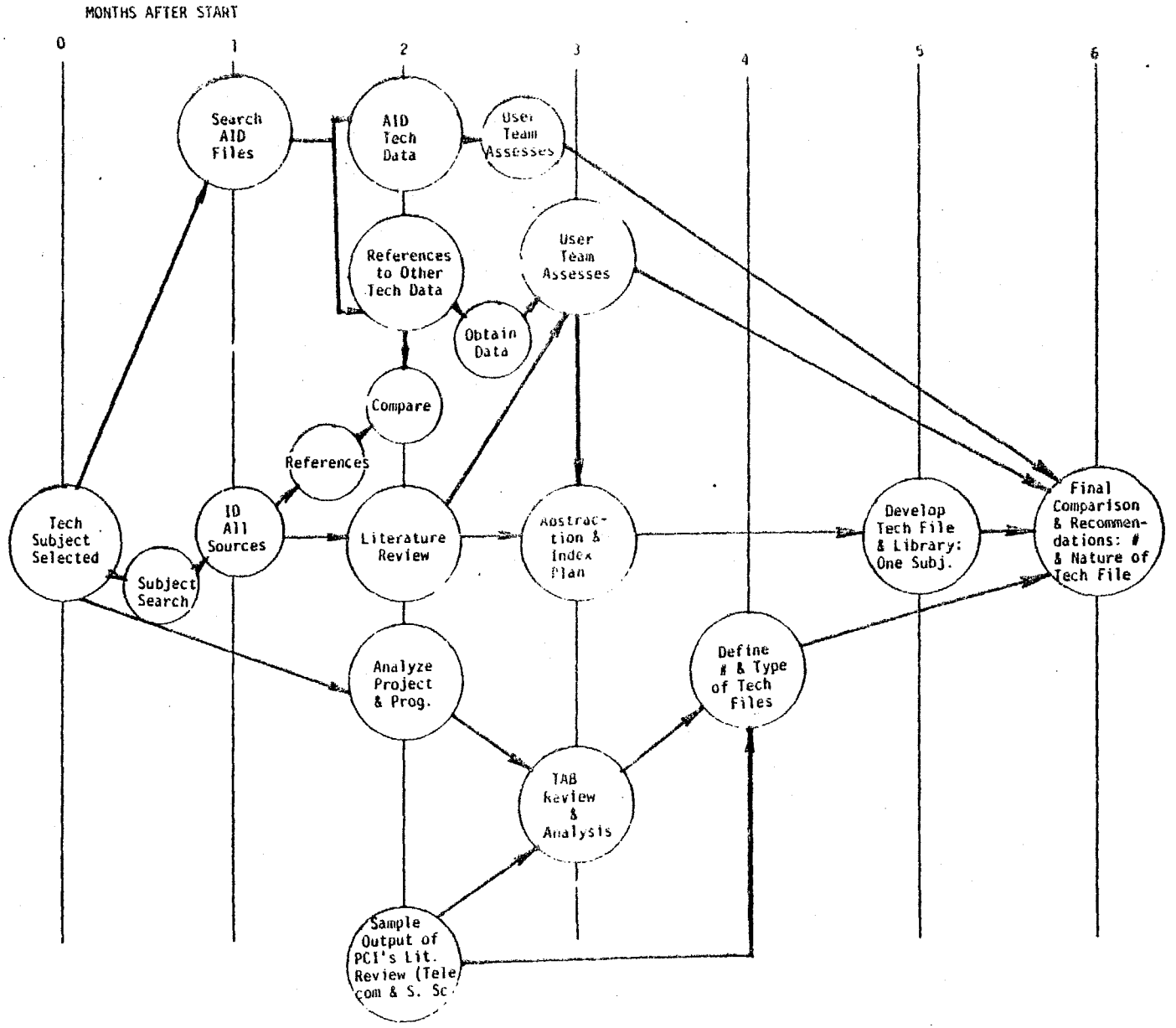


FIGURE 3-5: BASIC SCHEDULE FOR ACTIVITIES IN SUPPORT OF SUBJECT & PROBLEM SEARCHES

Value added by relatively more extensive technical outreach will be assessed by comparing: (1) AID technical data; (2) data obtained by exhausting the technical references within the AID technical data file; (3) a general subject search and literature review. Extensive coordination with TAB will be effected to ensure that DIS obtains full value from TAB expertise and to allow DIS and TAB to mutually assess: (1) the degree of overlap desired between DIS and TAB holdings, and (2) the way that DIS technical files can best serve TAB needs.

5. Analysis and Implementation Planning

This activity synthesizes results from all of the above to develop the items noted on the study Logical Framework (Figure 3-1) and a clear plan for achieving the implementation objectives.

Some of the basic alternatives to be answered by this analysis include:

> What files should DIS maintain?
>
> What files should DIS automate?
>
> Given automation, what are the trade-offs among (a) batch processing only, (b) interactive soft copy only, (c) interactive hard copy, (d) non-real-time (e.g., teletype) input/output?

Other questions to be answered by the study are included in Table 3-2.

D. MANPOWER REQUIREMENTS FOR THE SYSTEM DESIGN PHASE (PHASE IIB)

The overall manpower requirements for the design phase are summarized in Table 3-3. These manpower requirements are understated in that they exclude part-time consultation and assistance of AID technical and program staff.

The limited resources allocated to the key task -- analysis, system design and implementation -- is based on our recognition that this will be a continuing task in which all team members will share.

The resource allocations for each of the tasks are outlined briefly in the following paragraphs.

1. Work with Project Designers

The manpower required for this effort is highly variable -- depending upon the degree of participation of the information team in developing

TABLE 3-2

SOME OF THE BASIC QUESTIONS TO BE ANSWERED BY THE PHASE IB STUDY

1. What are the costs of alternative responses to DIS queries?

2. What is the value of those responses? Alternative responses to be considered in both the semi-automatic and the automatic and special response modes would include:

   - No response;
   - References to DIS holdings only;
   - Abstracts and references of DIS holdings;
   - Analysis and special summarization of these holdings to meet the specialized user needs;
   - Each of the above for non-DIS as well as DIS holdings.

3. What types of special requests is it appropriate for DIS to meet?

4. How can we assign priorities as to which special responses will be met and which won't?

5. What is the best timing and content for semi-automatic responses -- what types of information shall be provided in response to what types of stimuli?

6. How much of what kind of retrospective analysis -- analysis of prior Agency projects -- is justified?

70

TABLE 3-3

SUMMARY OF MANPOWER RESOURCES REQUIRED FOR SYSTEM DESIGN AND PROTOTYPE OPERATIONS

CATEGORY	MAN MONTHS TASK OR ACTIVITY					TOTALS
	WORK DIRECTLY WITH USERS	DEVELOPMENT OF AID MEMORY	DEVELOPMENT OF TECH APPROACHES	ASSESSMENT OF RESOURCE CAPAB.	ANALYSIS AND IMPLEMENTATION PLANNING	
Senior System Analyst	4	5			3	12
Systems Analyst	4	20	3.5	2.5	6	36
Sr. Library Specialist			2			2
Library Specialist	4		3			7
Research Assistant			3	5	4	12
Technical Subject Specialist			2			2
TOTALS	12	25	12.5	7.5	13	70

the project per se. For purposes of this analysis we assume a three-man team. Such a recommended team would include individuals with the following perspectives:

A. A Senior Systems Analyst, with knowledge both of the long-term DIS objectives, as well as the utilization functions as we have defined it;

B. A Library or Information Specialist who can translate information needs into practical query requests of existing sources;

C. A Specialist in AID's planning and evaluation processes, including the Logical Framework.

We assume a four-month period of full-time participation for each of the above.

These resources would of course be in addition to such technical and subject specialists as the USAID felt required for the design task per se.

This team would review pre-program documentation including but not limited to the DAP, and interrogate personnel available in AID/W, to define sources potentially germane to the subject at hand. They would also formulate queries based upon that pre-project documentation, to be addressed both to the AID memory and the other information activities within DIS. (These pre-departure queries will be important both to bet the related information activities started and to provide a baseline against which to judge modified queries occurring subsequently during the effort.)

Having equipped itself with such raw data, data summaries, and bibliographic records as it deems appropriate to the topic at hand, the team would then change its focus of operations to the USAID where it would

directly support project design activities on-site. The team would actively assist and provide information support for preparation of subsequent planning documents -- the PP, the PRP, and PID. During the course of each planning cycle, the team would generate information queries that would in later iterations of the system be expected to generate appropriate information. During these pilot operations it is probable that the other DIS activities would not be able to respond in a timely fashion to such queries. However, these will, as pointed out before, be available as prototype queries to which DIS functions should subsequently be able to respond.

Upon completion of each document in the planning process, the DIS team would return to AID/W and work with both DIS central activities and AID/W reviewers to:

1. Meet next step information requirements of the USAID;
2. Support information needs of reviewers.

During the course of their work, the "on-site" team will have developed four information requests each of which will be cabled to DIS in Washington:

1. Pre-departure, PID;
2. During PP preparation;
3. During PRP preparation;
4. During PIO preparation.

The DIS Washington team will actively respond to these requests and those responses will be assessed by the project design team in terms of relevance, quality and utility. Analysis of these queries and responses should allow us to improve both the query approach (offered the user) and the DIS search and response operations.

2. AID Memory Development

Development of AID memory will consist of:

a. Direct support to the project oriented information activities, and;

b. Integration of the diverse elements of AID memory (e.g., the ARC, the PBAR country program data bank, the evaluation products) into a single functional entity. (Note that integration into a functional entity does not necessarily imply integration within an organizational entity.)

a. Direct Support of Project Specific Information Activities

The preferred information from the AID memory will be the evaluation reports and information, and related data from the country program data bank. Thus, the two subactivities undertaken here will involve both and interrogation of individual's project files, etc., on an ad hoc basis as required to identify the types of evaluation data, final reports, etc., germane to the project needs. It must not be assumed that the evaluative information desired by the project designers will in fact be readily available. Thus, after examining AID's available memory for supporting data, and testing this in terms of the needs of project designers, additional outreach for information will be required. The judgement of the project design teams must of necessity be on one of two alternatives:

1. Information provided is adequate;
2. Information is not adequate.

In the event that the information is not adequate the AID memory component of DIS must both formulate additional queries for the subject searches and more important from the point of view of AID memory, postulate ad hoc evaluations of existing or prior experience as required to generate the necessary information.

In every event where the information is deemed less than fully satisfactory by the project designers, the AID memory component should postulate such evaluations, etc., as needed to obtain the necessary information. However, the decision to go ahead with such evaluations will be a separate management decision based upon:

(1) The increased probability of project or program success that would be associated with such evaluation;

(2) The cost of the evaluation activity.

b. Integration of AID Memory Components within a Single Functional Entity

This activity will both inventory existing and planned AID information sources, develop indexing and search techniques appropriate to accessing any or all of these as required, and develop a plan for automating basic searches and integrating that automation plan with the related automation activities in support of PBAR's PPT, country program data bank, etc.

Note, a key assumption made by the DIS implementation activity is that the project matching activities being undertaken in support of PBAR initiatives will be available at approximately mid-term in DIS first phase implementation activities.

The approximate level of effort of the developement of the AID memory would be almost full-time availability, over six months of:

One Senior Systems Analyst;
Four Systems Analysts (one for each file);
Logistics support (including clerical, keypunch, and machine time).

We define "almost full-time availability" as five man-months during the six months.

3. Technical and Subject Search

The approximate (professional) level of effort to complete this task is estimated as follows:

Literature Review & Analysis	4 man-months
Review of AID Data	2.5 man-months
Library Organization & Development	3 man-months
Indexing	3 man-months
TOTAL	12.5 man-months

NOTE: Exclusive of non-DIS staff. Program, TAB, etc., part-time consultation may total as much as twelve (12) man-months. However, major functions of this part-time assistance overlap the consulation in direct support of project design.

Categories of labor among which this effort divides might be as follows:

Technical Subject Specialist	1 man-month
Senior Library Specialist	2 man-months
Library Specialist	3 man-months
Systems Analyst	3.5 man-months
Research Assistant	3 man-months
TOTAL	12.5 man-months

4. Review of Agency Personnel and Equipment Resources

Manpower requirements for this effort are estimated as:

Systems Analyst	2.5 man-months
Research Assistant	5 man-months

Extensive cooperation of both personnel and management planning staff is assumed.

76

5. Analysis, System Design and Implementation Planning

The basic resources for this task will be provided through the performance of the other 4 tasks. However, it is appropriate to allocate, specifically to the analysis and system design, an additional 3 man-months of Senior Systems Analyst, 6 man-months of Systems Analyst, and 4 man-months of Research Assistant.

SECTION FOUR

MANAGEMENT PLAN

This section of the report sets forth they key management issues for DIS implementation, discussing costs, organizational approach, and staffing. The specific topics covered are:

A. Assumptions
B. AID's Current Expenditures for Information Services
C. Costs of Developing and Operating DIS
D. Organization and Staffing

A. ASSUMPTIONS

The management plan for DIS development assumes that:

1. It is desirable to keep open all DIS options -- ranging from DIS being only AID memory to DIS being the definitive source of information regarding development.

2. A low risk strategy is desired -- premature investments in staff and capital resources are to be avoided.

3. Based on PBAR initiatives, major portions of AID's own memory will be automated and will include a capability for remote real-time interrogation of the centralized data base.

4. There are currently many sources of information being run or subsidized by AID, and the integration, coordination, or replacement of at least some of these activities would be a cost effective DIS role.

The implication of the above assumptions are direct:

1. To keep open the various alternative outcomes of DIS implies a modular design approach, in which additional data bases and capability can be added on as needs become clearer and resources become available. By adopting such an approach, it should be possible to keep major options open not just through design and implementation, but during the operational phase as well.

2. To avoid premature commitment to staff and capital resources, major portions of the prototype and possibly implementation activities should be performed by contractors -- allowing AID staff to phase in and take over activities only after the need for and value of such activities has been proven.

3. Based on the PBAR initiatives, by far the bulk of the automation costs necessary to support DIS will be borne by other activities and agents. Therefore, automation of DIS becomes virtually a zero cost option.

4. DIS analysts will routinely use all available data sources. Through interaction with DIS, based on real value of such information sources to AID project and program designers, it is expected that the better information sources will be strengthened and those of marginal value will atrophy.

B. AID'S CURRENT EXPENDITURES FOR INFORMATION SERVICES

Before discussing recommended investments in the DIS, it is useful to estimate AID's current level of expenditure for related information services. To this end, PCI performed a brief survey of the information activities within AID. The results of this survey are presented in

Table 4-1. The most remarkable finding of our survey was the diversity of information activities that are being undertaken. The information services noted in Table 4-1 are those explicitly identified in AID Handbook No. 18* and by no means exhaust the set of AID information expenditures. Some information activities that are notable for their absence from this table are:

- Grants for information services in population planning (for example, to the University of North Carolina and the East-West Institute);
- Allocations for information service expenditures among the 211d grants;
- The vita system (a nutrition data base);
- The life system (an agricultural data base).
- PARIS (TAB system);
- Information investments not appearing in the AID budget, such as expenditures to enable LDCs to tie in directly to Medlars.

Further, PCI did not interrogate the regional bureaus to determine the expenditures they were making in either formal or informal information services. In addition, note that "zero cost" has been assigned to information activities for offices noted in Table 4-1b. (Based either on that office's estimating their true costs as nil, or that PCI did not contact the responsible officer.) Clearly then, the table tends to understate AID's total expenditures, and we can safely assume that Table 4-1a represents a minimum level of expenditure.

---

* Handbook 18 Information Services: Part IV, Information Sources in AID, June 16, 1975, Agency for International Development, Department of State, Washington, D. C.

TABLE 4-1a

AID OFFICES SURVEYED HAVING ASSOCIATED LIBRARY OR INFORMATION RETRIEVAL COSTS

OFFICE/DIVISION	(1) DIRECT HIRE MAN YEARS	SOURCE OF ESTIMATE	TOTAL COST OF AID MAN YEAR	OTHER COSTS	TOTAL COST
OFFICE OF COMMODITY MANAGEMENT/PROCUREMENT SUPPORT DIVISION	1/10	S. Barondes	$ 4,400	$ -	$ 4,400
OFFICE OF HEALTH	-	D. Newman		? 2/	-
OFFICE OF URBAN DEVELOPMENT	-	W. Miner		? 3/	-
OFFICE OF PRIVATE AND VOLUNTARY COOPERATION	2 1/2	R. McClusky	110,000	? 4/	110,000
SPECIAL ASSISTANT FOR UNIVERSITY REGULATION	1	C. Barker	44,000	1,000 5/	45,000
OFFICE OF CONTRACT MANAGEMENT/ SMALL BUSINESS OFFICE	4	N. Rowe	176,000	-	176,000
OFFICE OF DISASTER RELIEF COORDINATION	1 1/2	M. Heyl	66,000	6/	66,000
OFFICE OF AID RESEARCH AND INSTITUTIONAL GRANTS	-	L. Rhodes	-	160,000+ 7/	160,000+
OFFICE OF AGRICULTURE	1 1/2	M. Galli	66,000	? 8/	66,000
OFFICE OF DATA MANAGEMENT/INFORMATION SYSTEM DIVISION	-	L. Rhodes	-	80,000 9/	80,000
OFFICE OF DATA MANAGEMENT/ AID REFERENCE CENTER	-	L. Faibo	95,216 10/	12,192 11/	107,408
OFFICE OF FINANCIAL MANAGEMENT/STATISTICS AND REPORT DIVISION 11/	1/6	B. Huntington	7,333	-	7,333
OFFICE OF POLICY DEVELOPMENT AND ANALYSIS	2 12/	J. Silverstone	88,000	-	88,000
OFFICE OF DEVELOPMENT PROGRAM REVIEW AND EVALUATION	1 1/2	R. Berg	66,000	-	66,000
OFFICE OF INTERNATIONAL ASSISTANCE COORDINATION/ INTERNATIONAL FINANCIAL DIVISION	1 1/2	C. Gulick	66,000	-	66,000
OFFICE OF MULTILATERAL COORDINATION AND REGIONAL SOCIAL DEVELOPMENT PROGRAM	1/2	F. Lansdale	22,000	-	22,000
OFFICE OF FOOD FOR PEACE	1/2 13/	C. Weden	22,000	-	22,000
OFFICE OF POPULATION	-	J. Speidel	-	750,000 14/	750,000
TOTAL			$832,349	$1,003,192	$1,836,141

1/ "Report on Manpower and Operating Expenses 1974-76"; AID-Office of Financial Management

DIRECT LABOR COST PER YEAR

$$\frac{\text{Total Operating Costs AID/W}^{*} - \text{Overhead (Administrative)}}{\text{\# of AID/W Employees}} = \$25,000$$

* fringe included

OVERHEAD COSTS PER YEAR

$$\frac{\text{Overhead (Administrative}}{\text{\# of AID/W Employees}} = \$19,000$$

Direct Cost of AID Man Year = $44,000

2/ a) AID cost to NLM to allow CDC direct access to MEDLAR is not included.

b) VITA or LIFE program development and implementation is not included.

3/ "Urban development information network" with PADCO, Inc. is not included.

4/ Amortized cost of establishing Technical Assistance Information Clearing House, TAICH, and yearly subscription plus costs of ADS and ARTA are not included.

5/ Amortized cost of AID contribution of $30,000/year for the past five years ($150,000) to the Academy for Education Development, NYC, is not included.

NOTE: AID now subscribes to the service for $1,000/year.

6/ OM is now [illegible]

7/ AID Research and Development Abstract (ARDA)

Labor for Abstracting	$ 60,000
Contract for Publication, etc.	100,000
	$160,000

Amortized costs of the following are not included

RESEARCH GRANTS

From 1/1/62 - 12/3/74 there were 225 contracts = $150,000,000 in 24 problem areas in

1) TAB,
2) PPC, and
3) PHD

but not in bureaus

211-D GRANTS

45 grants from 1968 - 1974

1968	$ 7,000,000
1969	6,000,000
1970	5,500,000
1971	5,000,000
1972	5,500,000
1973	1,000,000
1974	2,500,000
	$34,500,000

8/ Cost of RACA program is not included.

9/ Approximate cost charged AID by Department of State Library for services rendered; cost per year is not a constant.

TABLE 4-7b

ADDITIONAL AID OFFICES LISTED HAVING
LIBRARY OR INFORMATION RETRIEVAL CAPABILITIES,
BUT NO STATED COSTS

OFFICE/DIVISION
OFFICE OF FINANCIAL MANAGEMENT, LOAN DIVISION 1/
OFFICE OF GENERAL COUNCIL 2/
OFFICE OF COMMODITY MANAGEMENT TRANSPORTATION SUPPORT DIVISION 3/
OFFICE OF NUTRITION
BUREAUS 4/
MISSIONS 5/

FOOTNOTES

1/ Information is produced as a function of its operation.

2/ GC is considering a financial arrangement with Department of State Library to purchase services.

3/ Supports information request of Food for Peace (FFP) and Disaster Relief Operation (FDRC)

4/ None of the bureaus were investigated though it is known that the Latin American bureau has a program to code and store on computer selected standard indicators of project progress by functional sectoral categories:

1,500 indicators
30 categories

5/ Each mission has its own library.

10/ COSTS OF ARC SERVICES

The major cost elements of ARC in 1974 are the following: (not including supplies)

A.	Salaries*	80,692.00
B.	18% Fringe** (retirement annual and sick leave)	14,534.00
		95,216.00
C.	Travel*	250.00
D.	Equipment*	560.00
E.	Building Space** ($7.25/Ft^2 x 1570)	11,382.00
		$107,408.00

* Figures provided by ARC

**Figures provided by Office of Financial Management

The cost - effectiveness of ARC services in 1974 may be obtained from the following:

	Total	Cost per Unit
Documents Circulated	5,401	$19.90
Documents Loaned	2,296	$46.81
Total ARC Visitors (AID Govt non-AID and public)	1,783	$60.28
AID ARC Visitors	1,297	$82.86

The cost per visitor figures are probably overstated by a factor of two; the numbers are obtained from the visitors log which not all guests sign.

11/ Described itself as the "manual memory".

12/ Data for only Title IX and Civic Participation section of the office - two other sections not reached.

13/ Data for only one of three sections of FFP.

14/ $750,000 contract with The George Washington University for PHA library services. The Population Program Data System (PPDS) with the National Institute for Commodity Development (NICD) and the Philippine Population Program contract with the University of Chicago and the Philippine National Computer Center costs are not included.

82

Totalling the expenditures in Table 4-1, we see that the Agency is currently spending at least 20 man years of direct hire staff time plus $1,000,000 each year. Based on discussions with AID financial management, we assume a cost per man year of $44,000 per man year,* yielding a total annual expenditure of $1,800,000.

Recognizing the approximate nature of the above estimates, and the clear recognition that we have not identified the total Agency investment in information services, it seems safe to use a figure of $2,000,000 per year as the minimum expenditure AID is currently making in information services.

The fact that AID is spending well over $2,000,000 per year on information services does not mean that it is desirable for DIS to take over these information activities. It does reinforce the Committee's perception that information services should be given a higher priority and more top management visability, and that AID would be well served by having DIS serve as an integrating agent. It is conceivable that DIS's role to integrate existing information activities may allow DIS to "pay for itself" based on the economies it effects. We also know that if DIS could do all of the things that are being done in Table 4-1 and also perform the mission that no one is performing well today -- providing project designers with up-to-date, realistic information that helps them formulate project and evaluation activities -- an expenditure of up to $2,000,000 per year could be justified.

---

* FM's analysis was as follow: Dividing total payroll exclusive of consultants and overtime by the number of employees yields $25,000 per year average salary including fringe. Dividing overhead expenditures by number of employees yields $18,700 per employee. Thus our approximation of $44,000 per man year.

C. COSTS OF DEVELOPING AND OPERATING DIS

1. General

To get a true picture of cost sensitivities, we must consider total cost -- development and operations. On a life-cycle cost basis, higher development costs may result in lower total cost, and vice versa. For example, the decision to automate indexes will result in higher development costs that must be justified based on large savings in operating costs. (A key word in context search of project or report summaries can accomplish in minutes a job otherwise taking days.) Low development costs will be required to establish a "system" that uses human intermediaries to respond uniquely to each request -- yet the operating cost of such a system would be prohibitive if it were to respond to many queries.

To assess cost implications it is best, then, to consider costs of alternative operating systems and then seek minimum cost paths to developing the optimum (most cost-effective) operating system.

Viewing this in the context of a three-phase effort -- development, implementation, and operations. The change in operating costs could vary from negative (less than current) to millions annually. Phase 1 defines the practical options and probable cost effectiveness of each. Phase 2 validates as well as implements, keeping some Phase 3 options open for Phase 2 decision makers. Given DIS plans and objectives as of this moment, the budgetary estimate of costs to develop, implement and operate DIS are as follows:

Phase I (Development):	$350,000
Phase II (Implementation:	$440,000
Phase III (Operation):	$342,000 per year.

84

Phase II and III estimates are, of course, subject to revision during the Phase I activities.

2. Approach to Cost Analysis

As noted earlier, there is to some extent a trade-off between investments in design, implementation, and operations. The largest investments in design and implementation -- e.g., heavy investment in computer assistance -- may yield lowest total (life-cycle) costs. (Based on lowered annual costs of operations such as might be available through a direct-access natural-language system.) Similarly, inadequate investment in design and implementation will result in high operating costs and inefficient service. To clarify the design/implementation/operational cost trade-off, PCI proposes to set a realistic minimum cost of annual operations, compare this to current expenditures, and then set forth a design and implementation strategy that attempts to realize maximum capability within this minimum cost figure. To the extent that additional expenditures are required, they must be justified in terms of value added or cost savings realized from other endeavors.

PCI believes that, given the on-going initiatives in the evaluation system, and the development of various PBAR data banks, the automated DIS capability can be achieved at a negative cost of development. That is, it is our hypothesis that by integrating the various data banks from the high volume user's point of view (the project and program designers), the DIS can effect economies in development of these various systems that will more than offset the management and integrating costs required to achieve that effect.

3. Annual Operating Costs

PCI's estimate of annual costs of an operational DIS is $342,000 per year. This is based on the estimates shown in Table 4-2, which was prepared in cooperation with Vladimir Slamecka, Director of the School of Information and Computer Science at Georgia Institute of Technology. (Professor Slamecka served as one of PCI's Associates for this engagement, and would continue to serve on the DIS design and implementation team.)

Note that the costs shown are "direct costs" -- no provision, other than the space allocation, is made for the government's overhead and fringe costs. This is appropriate when comparing DIS expenditures to current AID budgets, which are similarly based on direct costs. However, it may be inappropriate to use such "unloaded" costs when considering design and implementation phases, where contractor or university support may be required.

The $342,000 of budgeted DIS annual cost is almost undoubtedly less than AID currently spends. ARC expenses alone are approximately $190,000 per year ($107,000 in ARC costs per se, plus $80,000 provided to the Department of State Library). It seems safe to assume that an additional $200,000 per year is being spent in other information activities that would be taken over by DIS.

Given PBAR's plans for development of automated data bases of which DIS would be a primary user, there is no incremental cost for DIS automation. It is already planned to have interactive consoles accessing these various data bases. Those consoles would presumably, in any event, be available to DIS analysts as one of the multiple users. However, it would be appropriate to specify as a DIS unique cost, the cost of perhaps three such "smart" interactive consoles. Given the technology at the moment, the capital cost of the consoles might be around $30,000. Depreciating those consoles over 5 years, we can consider an annualized equipment cost of perhaps $6,000 per year.

86

TABLE 4-2

ESTIMATED ANNUAL OPERATING COSTS FOR THE DIS

Function	Personnel No.	Personnel Salary Level
ACQUISITIONS (5,000 documents/year)		
Acquisitions	1.0	$16,000
Collection Maintenance	1.0	12,000
Circulation	1.0	12,000
INPUT PREPARATION		
Abstracting and Indexing	3.0	48,000
KP/V	1.0	9,000
REFERENCE		
Information Analysts	5.0	90,000
Vocabulary Maintenance	0.5	10,000
Systems Development	0.5	10,000
Programmer	1.0	18,000
MANAGEMENT AND ADMINISTRATION		
Director	1.0	36,000
Clerical Support	4.0	32,000
TOTALS	19.0	293,000
Space: 7,000 ft.^{2} x $7/ft.^{2} =		49,000
TOTAL COSTS		$342,000

(Excludes costs of computer equipment, consoles, and related communications links.)

There is, further, no reason to expect that the frequency, method, or nature of file updating procedures should be substantially different given DIS as the coordinating body than it would be were these systems to be operated independently. Therefore, we assume no additional cost for file updating.

4. Implementation Costs

It is in the nature of the cost trade-offs noted earlier that both operating and design costs are easier to estimate than are implementation costs. Implementation costs will be a function of both implementation approach, design, and operational configuration. A simple example of the "big swing" in implementation costs is the indexing and abstracting of existing documents. If the entire ARC collection is to be so abstracted, then we estimate the associated direct costs at $786,000, based on the calculations shown in Table 4-3. On the other hand, one of the design functions will be to provide criteria for selecting only those documents that are worth abstracting -- potentially reducing the cost of "backlog input" to zero. (Assuming it is not cost effective to abstract the ARC collection.) In that case, implementation costs of abstracting ARC documents would be nil -- the only costs would be design costs (performed during Phase I to test desirability) or operating costs (performed as a by-product of a special query response).

Given the uncertainties in estimating implementation costs, and the ambiguities in distinguishing such costs from operating costs, we still must estimate the level of investment that might be required. To develop such an estimate, we will assume:

1. One-fourth of the ARC collection will be abstracted and indexed as an "implementation cost";

2. Implementation per se will require a 12 month elapsed time, during which time two systems analysts and one programmer will be required to mesh DIS needs with PBAR automation efforts.

88

TABLE 4-3

ESTIMATES FOR COMPREHENSIVE ABSTRACTION & INDEXING FOR THE ARC COLLECTION

TABLE 4-3A

ESTIMATED SIZE OF ARC DOCUMENT COLLECTION

DOCUMENTS		NO.
Cataloged (July 1975)		43,600
Uncataloged		
Foreign-language	18,000	
Non-AID (US)	10,000	
AID	6,000	
Subtotal-Uncataloged		34,000
	TOTAL	77,600

TABLE 4-3B

ESTIMATED BACKLOG INPUT COST (DIRECT COSTS OF INPUTTING THE ENTIRE ARC COLLECTION)

Assume:

Selection, 0.05 manhour/document, @ $10.00/hour.
Abstracting, 0.7 manhour/document, @ $8.00/hour.
Indexing, 0.3 manhour/document, @ $8.00/hour.
KP/V, 0.3 manhour/document, @ $4.00/hour.

Document Surrogate = 1,000 characters

Assume that all cataloged and uncataloged documents to be processed for input; then on basis of 80,000 documents, the costs are:

Activity	$ Rate/Document	Cost ($)
Selection	$0.50	$ 40,000
Abstracting		
Indexing	8.00	640,000
KP/V	1.33	106,400
	TOTAL	$786,400

or about $10.00/document. This corresponds to typical costs; higher costs should be anticipated if complex indexing requirements exist.

3. Implementation will include a brief (five man month) definition of telecommunications alternatives for linking USAIDs with DIS;*

4. Five thousand documents and indexes will be microfilmed;

5. The "operational" team will phase in during implementation, so that one-half of the annual operating resources are "implementation phase" expenditures (although they might alternatively be considered "early operations").

Based on the above, direct costs of implementation are established as follows:

1. Abstract and Index ARC: $200,000

2. One man year of systems analyst and two man years of programmer: $76,000.**

3. Five man months to establish telecommunications alternatives: $15,000

4. Microfilming: $150,000.***

5. One half of operating costs: $171,000

Estimated Total Direct Costs: $612,000

This estimate includes items that are discretionary or might be classified differently. Substantial microfilming and abstracting costs may not be required. Further, one might argue that the systems analysis and programming

---

* PCI is currently helping DISC explore methods of improving information transmission to USAIDs and LDCs generally. One potentially effective means is through the use of satellites, which eliminate distance, geological features, and weather as obstacles. Although a purely satellite-based system may not be cost-effective, a combination of existing facilities plus satellite could provide a viable network. PCI's assessment of telecommunications alternatives for social services (under contract to TAB), plus the telecommunications experience of PCI staff, provide a good basis for setting forth feasibility of various communications alternatives.

** Assumes systems analyst at $36,000/year, programmers at $20,000/year.

*** "Retail" costs of purchasing microfiche for one million pages would be $256,000 ($51.00 per 200-page document per Eastman Kodak Labs in Baltimore). Important economies can be effected and the desirability and type of microfilming must be considered as a separate cost-benefit issue.

costs are properly assignable to AID MIS, and that the definition of telecommunications alternatives is not a necessary part of implementation. Excluding these items, "implementation" costs would be simply the partial operating costs incurred during start-up -- or less than $200,000. For cost summarization purposes, we exclude from implementation costs the $171,000 identifiable as operating costs, thus yielding a $440,000 implementation cost. Which is only to restate the problematic nature of estimating implementation costs.

5. Costs of Design and Prototype Operations

Based on the analysis presented in Section III (see Table 3-3), the design and prototype operations phase should require approximately a 70 man-month level of effort.

On the assumption that 50% of this effort will require the services of top-level staff ($36,000 per year) and 50% mid level ($25,000 per year), the estimated direct cost would be:

35 ÷ 12 x $36,000 =	$105,000
35 ÷ 12 x $25,000 =	$87,500
TOTAL:	$192,500

Adding overhead of 200%, the total cost could be considered as high as $576,000. On the other hand, assuming contractor involvement at approximately $5,000/man month, the total cost could be on the order of $350,000. In terms of budgetary (as opposed to real) costs, and assuming 50% contractor effort and 50% direct hire AID staff, costs are approximated as:

Contract (35 man months x $5,000/man month) =	$175,000
Direct Hire Staff (35 man months) =	96,000
TOTAL BUDGETARY COSTS:	$271,000

On this basis, the costs of design are less than the annual operating costs, and probably less than the Agency now spends annually on activities that would be directly replaced by DIS. For budgetary purposes, we assume a design expenditure of $350,000, which would allow the entire effort to be performed by a contractor.

D. ORGANIZATION AND STAFFING

PCI recommends that the bulk of the design and development activities be performed by a contractor. This will minimize start-up time and avoid premature commitment to permanent-hire positions.

The basic study organization is shown in Figure 4-1. The entire design activity could be staffed by contractor personnel, but it is recommended that at some time during the design phase (at least 60 days before implementation), the Agency appoint a DIS Director who would serve in this capacity throughout the implementation phase. Ideally such a Director would be appointed immediately so that he could participate throughout the anticipated 18 months of design and implementation. We do not recommend, however, that the DIS have the same Director during its operational period as it does during its design and implementation phase. As noted earlier, different talents are required for design and implementation than are required to sustain operations. Thus, it is possible that the DIS Director be on short-term assignment to the Agency. However, the Director should have the same ability to influence Agency personnel and resources as does permanent-hire AID management.

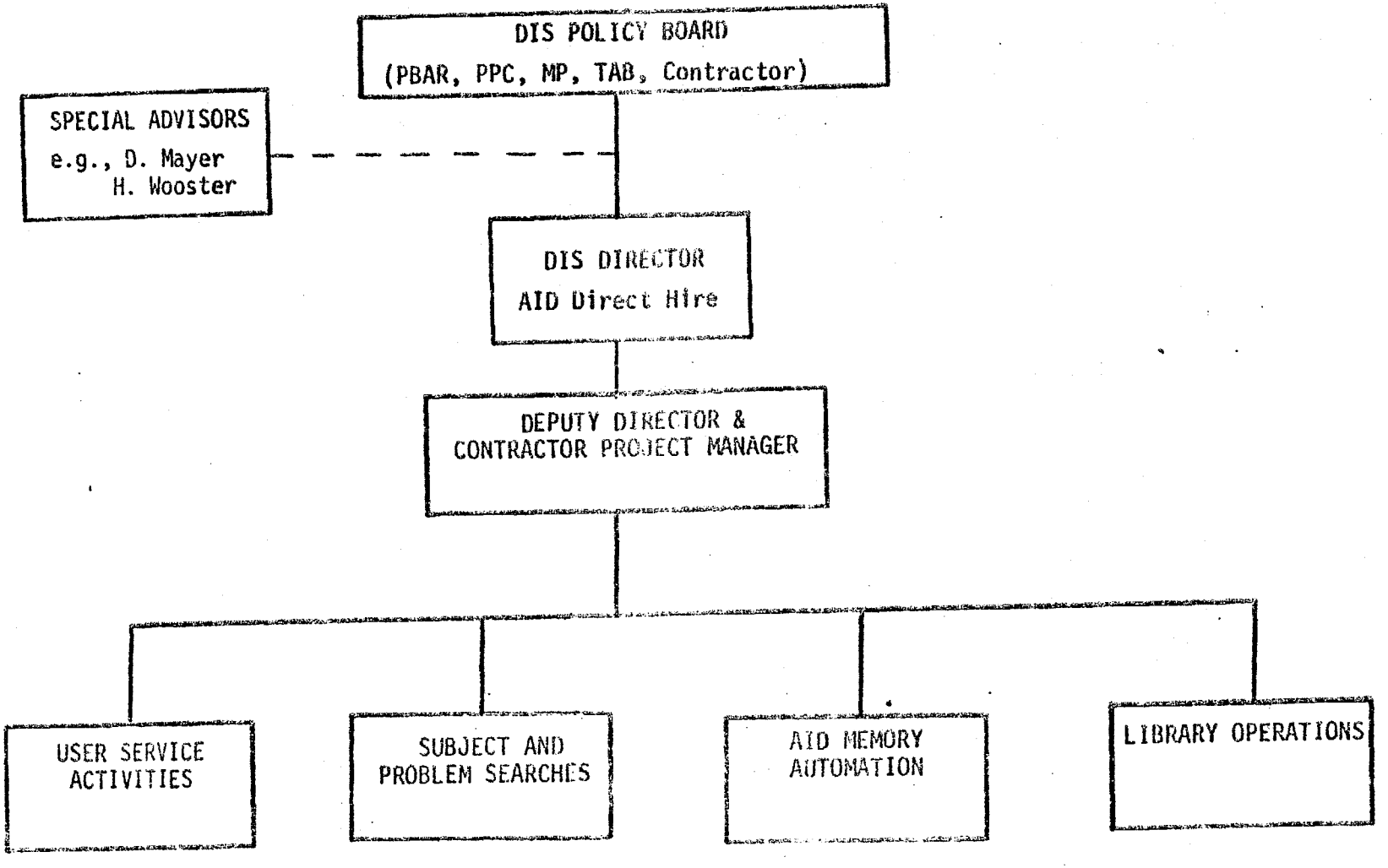
We recommend that DIS, to the extent possible, be an autonomous unit supervised by a Board of Directors consisting of representatives from PBAR, PPC, MP, and TAB. The contractor's Chief Executive should also sit on this Board of Directors.

The Board may also want special advisors reporting directly to it rather than through its contractor or the DIS Director. Two individuals who might serve in this capacity are Mr. David Mayer, who led prior Agency efforts to develop a project-matching system, and Mr. Harold Wooster, who has supervised large-scale government information and analysis processing systems, and is currently employed at the National Library of Medicine.

During the design phase, the head of the contractor team should serve in the dual capacity as head of the contractor team and de facto Deputy Director of DIS. The individual holding this position should have some development-related information systems experience, and experience with the design and implementation of both private sector and government data processing systems.

Subordinate responsibilities within the DIS design organization, as outlined in Figure 4-2, correspond to activities outlined in the scope of work. These staff should be chosen with a view to their ability to perform the basic design tasks, their insight into AID operations, and their ability to establish "job descriptions" for the operational DIS staff positions. In addition to fulfilling the responsibilities as noted in Figure 4-2, each member of the design team will also consider himself as "filling in" and establishing roles for the DIS permanent staff that will ultimately take over his position. Thus, in addition to the functional organization shown in Figure 4-2, the design team will be defining the responsibilities associated with the operational DIS, as noted in Figure 4-3.

Figure 4-1: DIS ORGANIZATION FOR THE DESIGN PHASE



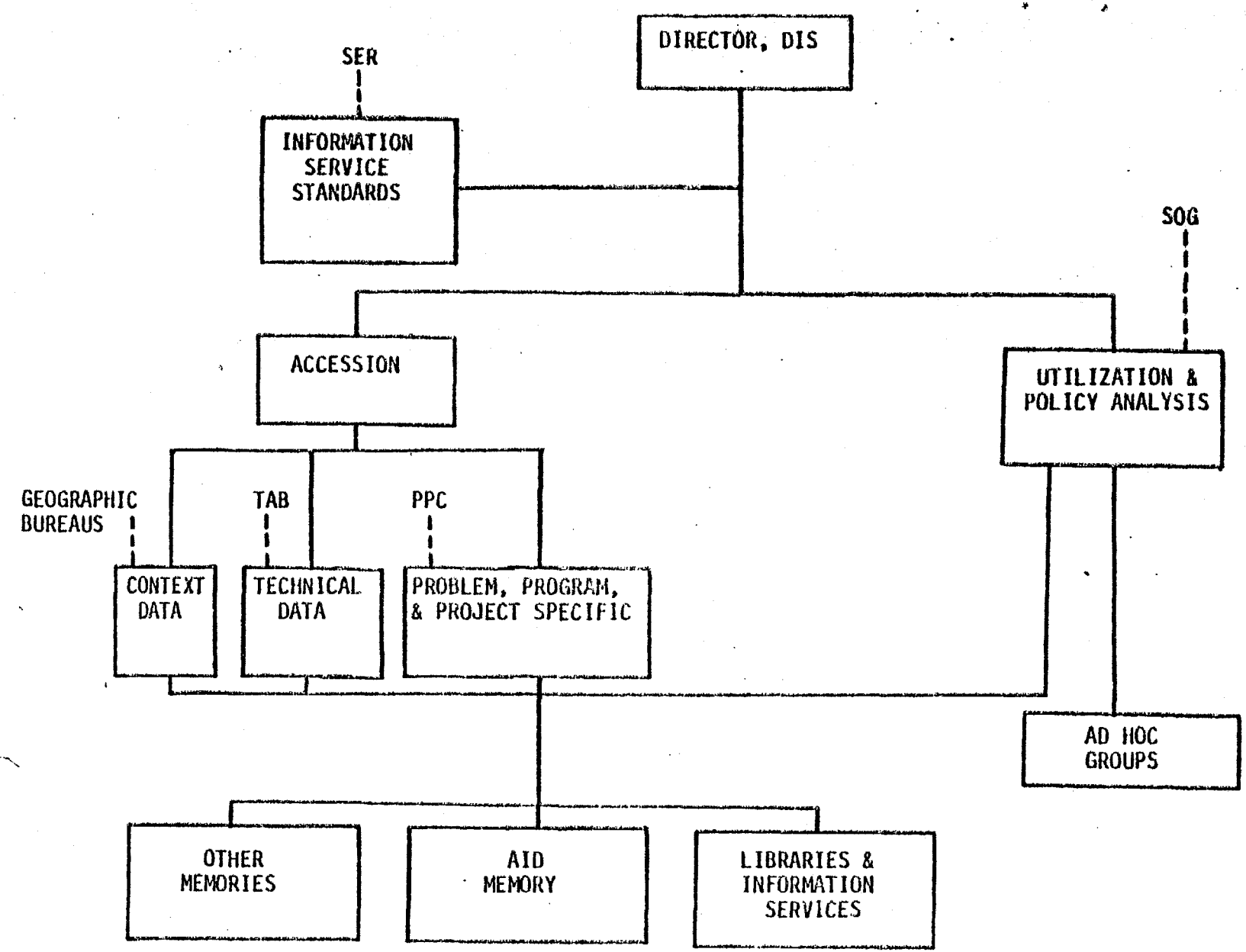


Figure 4-2: ORGANIZATIONAL COMPONENTS AND RESOURCES OF DIS

Attachment: B

Functional Requirements
Development Information Service System

I. Introduction

A. Background

Over the past 18 months, the Agency for International Development has developed a Planning, Budgeting, Accounting, and Reporting System (PBAR). This new operating system will substantially change the way in which A.I.D. conducts its business, and it has generated the need for the development and installation of a new, automated management information system with five major sub-systems.

One of these sub-systems is to be the Development Information Service System, which was recommended by the report of the Task Force on Library and Information Retrieval. The Report, Establishing a Development Information Service, dated August 8, 1975, summarizes the work of the Task Force in:

1. Collecting and analyzing data on the Agency's information activities and processes.

2. Identifying issues, problems gaps and opportunities in the Agency's information system.

B. Purpose

The Development Information Service (DIS) is to provide information on development matters primarily for the use of A.I.D. project designers. DIS will require an information storage, processing and retrieval system; it will utilize information on A.I.D.'s own experience in development as well as "the state of the art" in technical areas. Some elements of the DIS system will lend themselves to automation and may be integrated with the automated aspects of the several PBAR component systems.

Information to be covered by the DIS system will include:

(1) Information regarding experience with development projects, problems and programs, initially covering A.I.D.'s own projects.

(2) Technical data, specifying the state-of-the-art in project-related technologies;

(3) Context data describing social, anthropological, economic, and political variables bearing upon LDC problems/programs.

The DIS system will emphasize analytic not descriptive, information. Rather than giving a user the full body of information related to his projects, DIS will perform the analytic task of culling the literature and synthesizing an information package explicitly tailored to his needs.

The DIS will to the extent possible attain access to, but not necessarily permanent physical possession of, the data needed. The Agency need not and should not commit itself to collecting and maintaining a comprehensive body of data, but should instead obtain the maximum possible data from sources external to AID in response to specific needs.

II. Scope of Work

Building upon the work reflected in the report cited in I.A. above hereafter cited as DIS report, particularly Section II.C. of the report, contractor will:

(1) Identify more precisely the data needed to support project design;

(2) Recommend which of those data should be placed in an automated data base;

(3) Define the type of automated outputs that will be required; and

(4) Indicate the anticipated source of the data to be put into the automated data base.

It is anticipated that the data proposed for the first four files cited on page II-16 of the DIS Report (Project File: Country and Program File; Evaluation and Special Studies File; and the Context File) will largely be included in the PBAR automated systems now in the process of development. Therefore, the contractor will give highest priority to establishing requirements for these four files, submitting the information requested in the previous paragraph to the chairman of the PBAR Task Force by October 15.

The contractor will also determine the need for one or more separate Files to be established just for DIS use. Contractor will provide general recommendations to PBAR Task Force Chairman by November 1, 1975, regarding the content of these latter files and identify that which should be automated. Within one week, the PBAR Task Force Chairman will advise the contractor in writing the extent to which AID accepts the contractor's recommendations. The contractor will then prepare a

draft requirements package for each of the latter approved files in accordance with the attached outline and submit it to AID by December 1, 1975. AID approval and/or comment will be provided by December 12 and a final requirements package will be submitted to AID by December 31, 1975.

The PBAR Task Force Chairman will chair an Advisory Group to work with the contractor. The contractor will be prepared to meet with the group weekly to report on progress and to discuss ideas, concepts, problems, etc. The contractor may request special meetings of the advisory group if he believes it would facilitate his work.

In developing the requirements specified above, the contractor will focus particularly on an agricultural sector assessment, a recently approved health sector loan and a research or R&D project in the Technical Assistance Bureau portfolio. The contractor's selection of an assessment and projects for detailed analysis will be reviewed by the advisory group and subject to the approval of the Chairman of this PBAR Task Force.

ADP FUNCTIONAL REQUIREMENTS

Documentation Outline

I. INTRODUCTION

A. Background
Describe briefly the events which have brought about the need for the proposed system.

B. Purpose
Define the purpose of the proposed system and the objectives to be achieved.

C. Scope
Describe the range or extent of the proposed system, i.e., what are the beginning and ending points of the system.

II. SYSTEM USERS AND USES

Identify potential users and uses of the proposed system.

III. OUTPUT PRODUCTS

A. Purpose
Describe the purpose of each output.

B. Content
Identify data elements to appear on each output, including summary totals.

98

C. Frequency
Define the required production frequency for each recurring output (e.g., first of month, end of FY); define response time requirements for ad hoc or non-recurring outputs.

D. Distribution
Identify recipients of recurring outputs; specify the number of copies for each recipient if more than one copy is required.

IV. DATA ELEMENTS

Describe all data elements to be maintained in the automated data bank; provide, as a minimum, the following information for each data element:

1. Definition.

2. Maximum number of positions; specify the number of decimal positions if applicable.

3. Class (alphabetic, numeric, alphanumeric).

4. Source(s); identify documents or computer files from which each data element is extracted for entry into the system.

5. Validation criteria.

V. INPUTS

A. Source
Identify originating office for each manual input, and computer system for each automated input.

B. Content
Identify data elements to appear on each input; earmark those elements which are not to be entered into the automated data bank.

C. Frequency
State when data must be entered into the data bank, e.g., upon receipt, once each day, first of month, end of FY quarter; identify peak periods of activity.

D. Volume
Estimate the volume of activity for each input period, e.g., daily, weekly, monthly.

VI. GENERAL PROCESSING REQUIREMENTS

A. Special Handling Needs
Define system inputs and/or outputs which require special handling and protection; specify security procedures if applicable.

B. Operational Processes
Describe any formulae, mathematical computations, or operational processes unique to the proposed system.

C. Controls
Define any controls or checks which must be incorporated into the proposed system.

D. Coded Data
Identify all codes which are acceptable in the proposed system; provide code listings and definitions (including structural breakdown) for codes which are not included in current handbooks, manuals, etc.

VII. DATA BANK ACCESS LIMITATIONS

Specify any limitations on accessing current and historical data; e.g., who can update or change data and who can retrieve data.

VIII. INTERFACE REQUIREMENTS

Define the times and points of interface with other organizations, systems, and functions (both existing and proposed).

IX. DATA RETENTION CRITERIA

Define data retention periods for both current and historical data banks.

X. CONVERSION REQUIREMENTS

Identify existing data files which must be entered into the proposed system; describe the types of data on these files and estimate the volume of records.

100 z