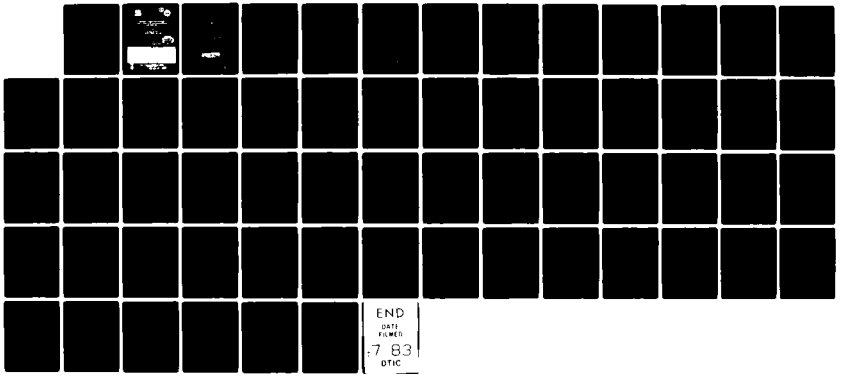
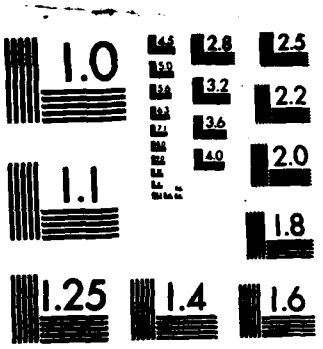
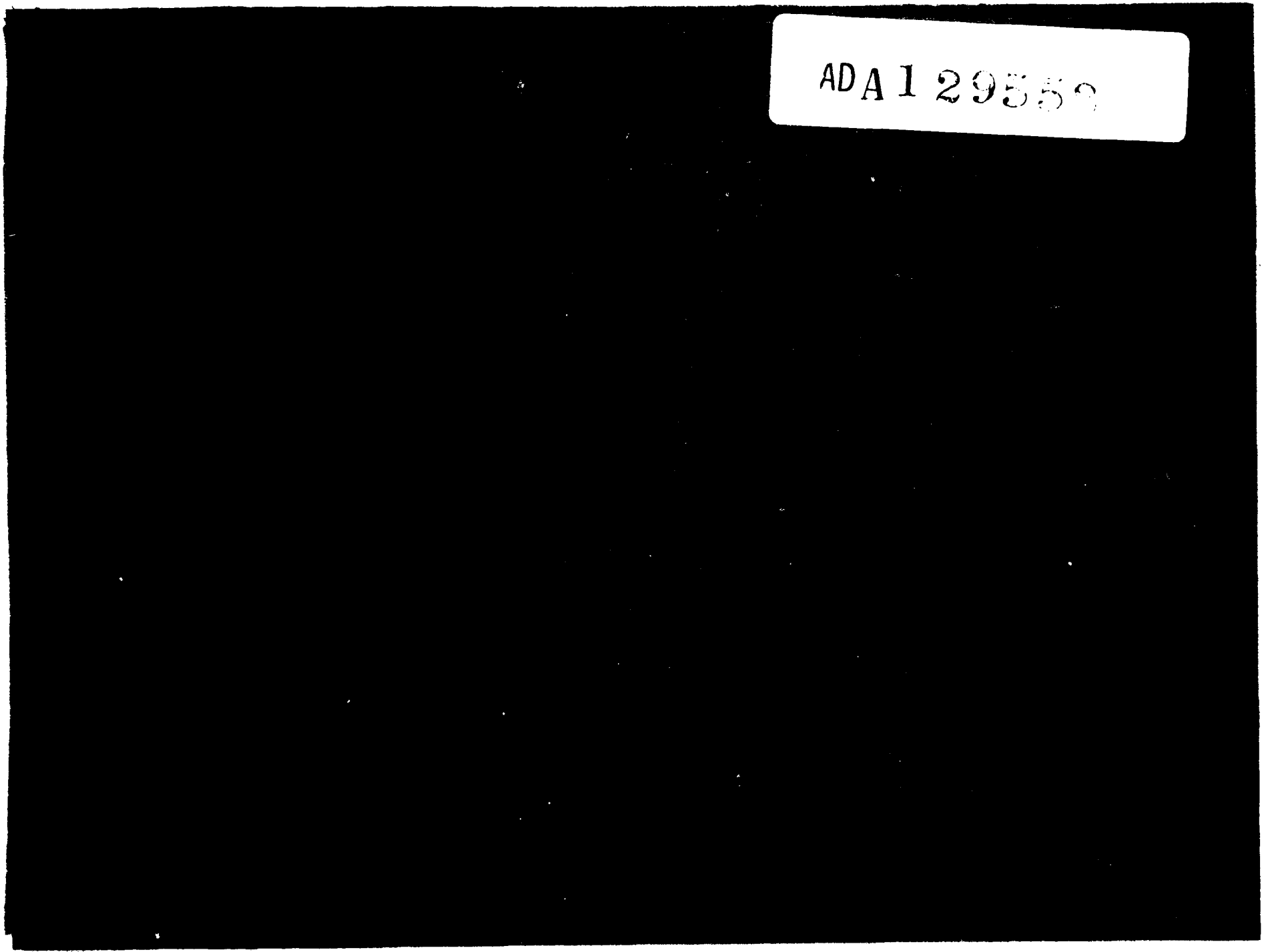
AD-A129 558 MISSISSIPPI SOUND AND ADJACENT AREAS STUDY PLAN FOR DATA-HANDLING AND ANALYSIS(U) ARMY ENGINEER WATERWAYS EXPERIMENT STATION VICKSBURG MS ENVIR.. V LAGARDE 1/1

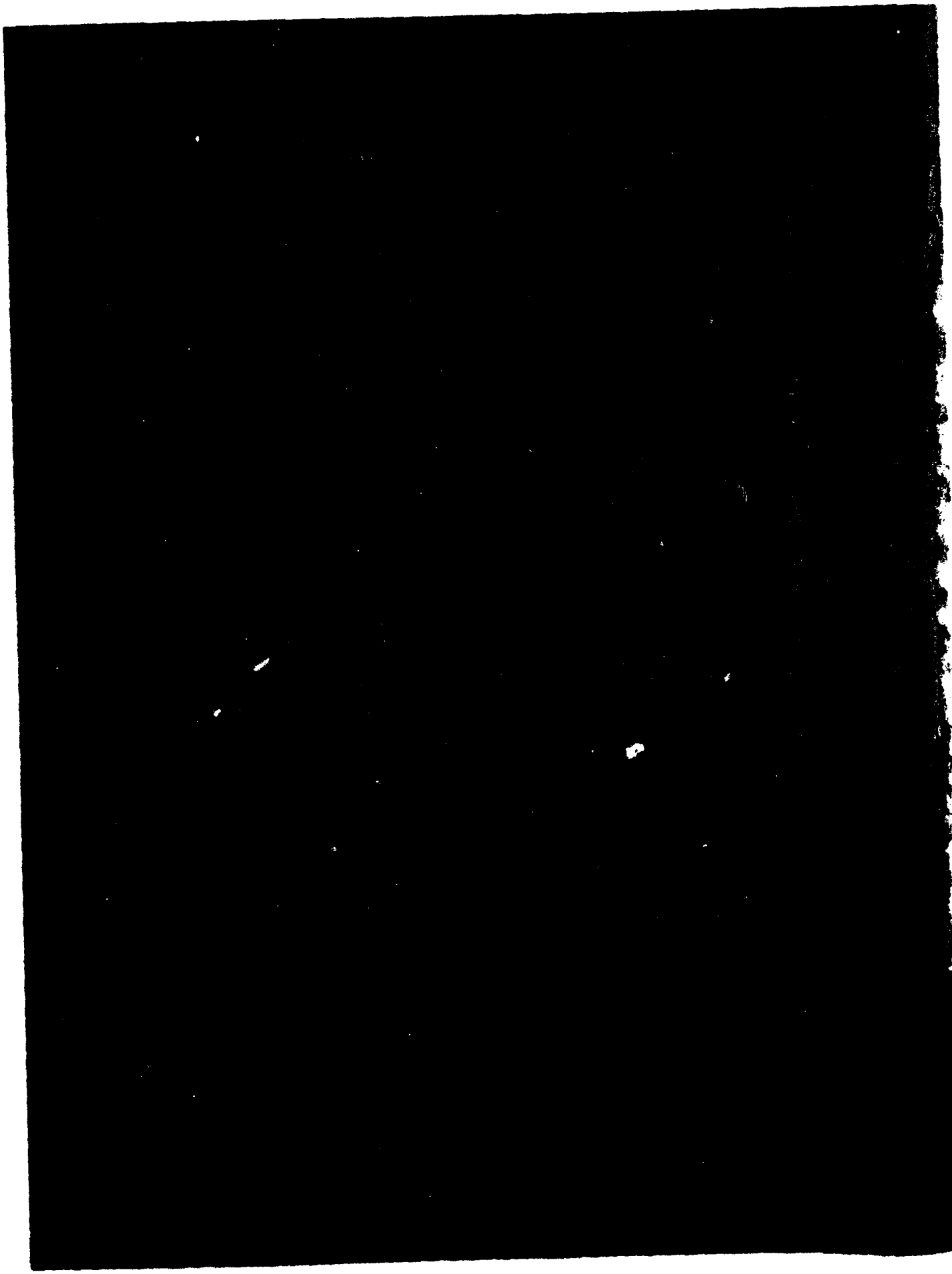
UNCLASSIFIED MAR 83 WES/TR/EL-83-1 F/G 13/2 NL



END
DATE FILMED
7 83
DTIC







Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE	READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER: Technical Report EL-83-1	2. GOVT ACCESSION NO.: AD-A129558
3. RECIPIENT'S CATALOG NUMBER	
4. TITLE (and Subtitle): MISSISSIPPI SOUND AND ADJACENT AREAS STUDY PLAN FOR DATA-HANDLING AND ANALYSIS	5. TYPE OF REPORT & PERIOD COVERED: Final report
	6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s): Victor LaGarde	8. CONTRACT OR GRANT NUMBER(s)
9. PERFORMING ORGANIZATION NAME AND ADDRESS: U. S. Army Engineer Waterways Experiment Station, Environmental Laboratory, P. O. Box 631, Vicksburg, Miss. 39180	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS: U. S. Army Engineer District, Mobile, Mobile, Ala. 36628	12. REPORT DATE: March 1983
	13. NUMBER OF PAGES: 55
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)	15. SECURITY CLASS. (of this report): Unclassified
	15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

Available from National Technical Information Service, 5285 Port Royal Road, Springfield, Va. 22151

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Computers
Computer applications
Data collection systems
Data processing
Mississippi Sound

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

A plan was prepared for coordinating and managing the data required for the Mississippi Sound and Adjacent Area Study, a complex U. S. Army Engineer District, Mobile, planning study. The plan provides a logical decision framework for how data essential to the core of the planning study should be handled. The framework starts with the definition of study problems and needs and traces the information requirements to identify specific data that

(Continued)

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

20. ABSTRACT (Continued).

should be computerized and the hardware and software required. The report provides a series of tabular forms to be filled in by the study manager and staff. The plan is applicable to most complex U. S. Army Corps of Engineers planning studies, i.e., those with multiple objectives covering extensive geographical areas.

PREFACE

This study was conducted from May to June 1981 at the U. S. Army Engineer Waterways Experiment Station (WES) by personnel of the Environmental Resources Division (ERD), Environmental Laboratory (EL).

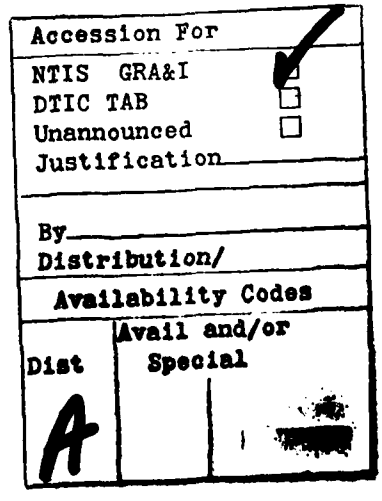
The investigation was authorized by COL R. H. Ryan, District Engineer, U. S. Army Engineer District, Mobile, Ala., on 11 May 1981, in support of the Navigation Branch of the Mobile District.

The study was conducted under the general supervision of Dr. Conrad J. Kirby, Chief, ERD, and Dr. John Harrison, Chief, EL. Dr. Victor LaGarde performed the study and was responsible for preparing this report.

Commanders and Directors of WES during this study and the preparation of this report were COL Nelson P. Conover, CE, and COL Tilford C. Creel, CE. Mr. Fred R. Brown was Technical Director.

This report should be cited as follows:

LaGarde, V. 1983. "Mississippi Sound and Adjacent Areas Study Plan for Data-Handling and Analysis," Technical Report EL-83-1, U. S. Army Engineer Waterways Experiment Station, Vicksburg, Miss.



CONTENTS

MISSISSIPPI SOUND ADJACENT AREAS STUDY PLAN FOR DATA-HANDLING AND ANALYSIS

PART I: INTRODUCTION

Background

1. The Mississippi Sound and Adjacent Areas (MSAA) study, conducted by the U. S. Army Engineer District, Mobile , originated in 1977 when the U. S. Congress authorized the U. S. Army Corps of Engineers to conduct studies to determine whether dredging activities in the gulf coast area could be modified to increase economic efficiency and promote environmental quality.

2. The purpose of the ongoing MSAA study is to (a) provide an overview of the resources and economy of the area, (b) investigate existing dredging and dredged material disposal practices, (c) analyze the effects of these practices on the resources and economy, and (d) determine if these practices should be modified. Modifications to dredging practices are justified if advantages can be gained in environmental quality and increased operations economy, as, for example, in the use of new dredging equipment and the coordination of a regional dredging program.

3. The MSAA study consists of three stages. The first stage, involving a reconnaissance-level investigation, has been completed.* The second stage, Development of Intermediate Plans, is in progress and due for completion in 1983. The object of the second stage is to examine potential solutions to the study problems and identify the alternative plans for modification that will be refined, developed, and analyzed in detail during the third stage of the study.

4. The second stage of the MSAA study consists of three major

---

* U. S. Army Engineer District, Mobile. 1979. "Mississippi Sound and Adjacent Areas Dredge Material Disposal Study, Reconnaissance Study," Mobile, Ala.

efforts: (a) identifying and classifying current dredging problems and regional needs for modification, (b) identifying the modification objectives to be examined, and (c) gathering data to provide the bases for examining alternative means of achieving the objectives.

5. The MSAA study is extraordinarily complex, encompassing the study of water, biological, social, and economic resources in a large, diverse regional area. There is a need for a formal plan for coordinating and managing MSAA study data, data-handling procedures, and analysis procedures which takes into account the study objectives. That is, data management and analysis procedures must be chosen and developed to handle and process the required data so that MSAA study participants can answer the questions dictated by the multiple study objectives.

Objective and Scope

6. The objective of this study was to prepare a plan for coordinating and managing the MSAA study data and data-handling and analysis procedures required to meet the MSAA study objectives.

Overview

7. The procedure for constructing a data management and analysis system for a specific study such as the MSAA study involves a series of steps similar to those in the different stages of a Corps water resources study. Many of these steps summarize, abbreviate, and formalize the information that results from the stage 1 and portions of the stage 2 project work. It is essential to formalize the work and approach to the management system as much as possible.

8. The construction of a data management system is similar to the construction of machinery: a formal planning and design phase yields the maximum return in work efficiency and minimizes the time and funding required for data management system design (or purchase) and use. The need for quality control and completeness is greatest during the

initial phases of the data management planning work and the management system construction work. Proper planning involves starting the planning process at the goals of the management system and working backward to specify what data are required and how the data must be managed to achieve the goals. Figure 1 is an overview of the work steps (phases) that should be followed for the MSAA study. Each step in Figure 1 is described in Part II of this report.

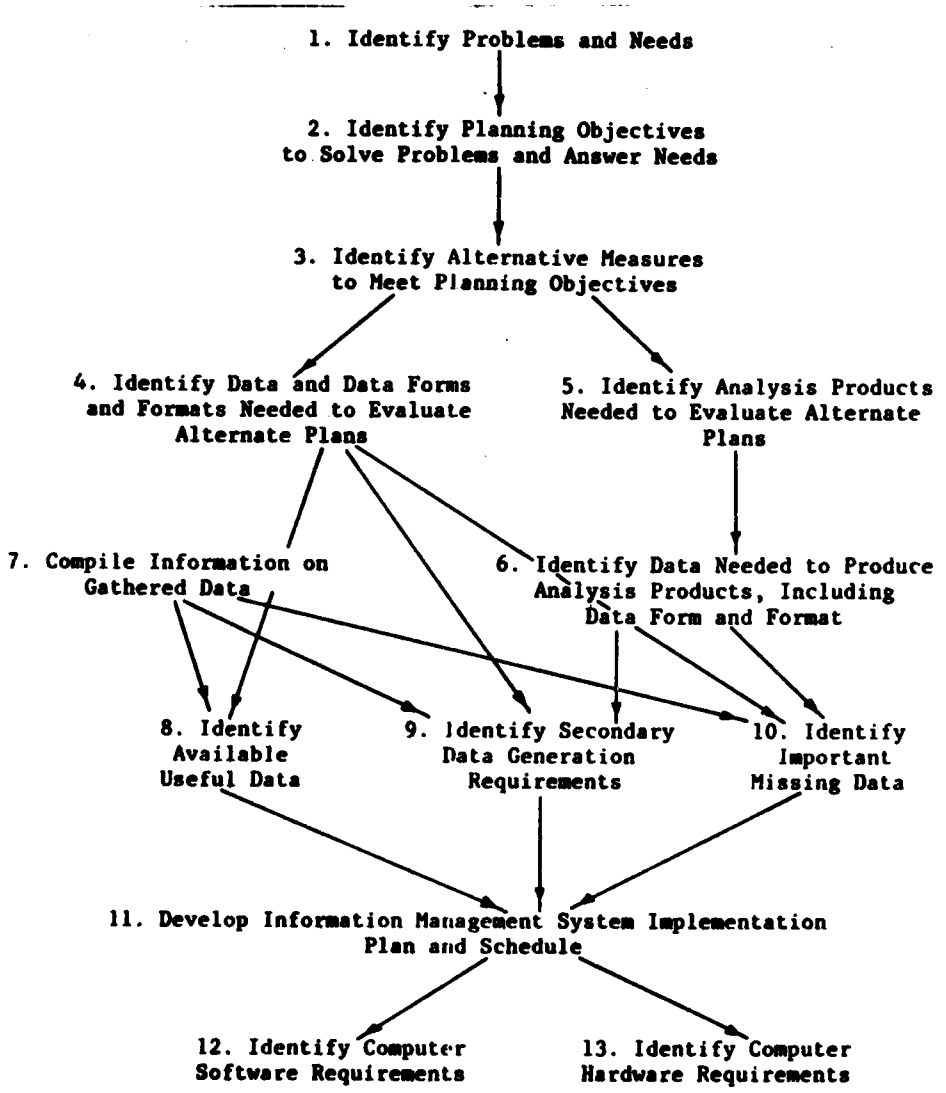


Figure 1. Overview of MSAA data management system plan phases

PART II: DATA MANAGEMENT AND ANALYSIS SYSTEM PLAN

9. Each step in the data management and analysis system planning process outlined in Figure 1 is described in this part of the report.

Work Scheduling and Resources Allocation

10. The following considerations are particularly important in scheduling of work and allocating resources during the system planning process.

a. One person should be responsible for overseeing the total planning process, and one or two persons at most should be responsible for developing each phase shown in Figure 1; the total number of persons involved in the planning process should be kept as small as possible. Five persons is the maximum number required, and not all of these will be involved in the work at the same time.

b. The culmination of work on each phase is the preparation of a table or specific recommendations for actions such as the use of a specific contractor or system. The work on a phase should proceed as shown in Figure 2.

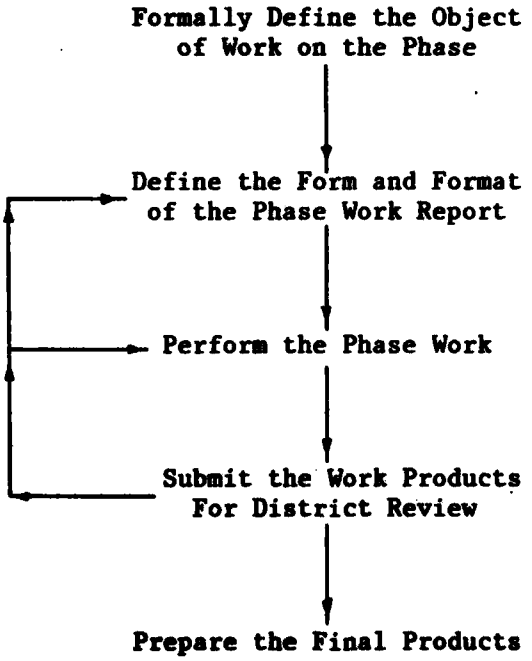


Figure 2. Work sequence on a work phase

c. The payoff for effort expended is greatest during the initial stage of the work and decreases thereafter. Time schedules for work phases are recommended in this report. The time schedules should be firm and changes made only in cases of extreme necessity. The major decisions that must be made, such as the type of computer software capabilities required, can be made if necessary without undue risk after the work on the appropriate phase is partially finished.

d. The work on a phase must be formally reviewed. There should be a review conducted by South Atlantic Division, Mobile District (SAM), personnel at the completion of work on each phase and adjustments made to the work products on the basis of the review. The review effort required of SAM personnel is greatest for the first phases shown in Figure 1 and decreases thereafter. Other reviews of work completed on work phases should take place after any MSAA study review, consultants' meeting, or any action resulting in a discussion or critique of the objectives and status of the MSAA study. The combination of a formal planning procedure as described in this document and study reviews normally results in the discovery of additional or changed data requirements.

e. As the work proceeds to successive phases as shown in Figure 1, the restrictions implied by decisions made in prior phases focus the work effort. The work becomes more technical and restricted in scope, and work phase products become more concrete.

f. To a great extent, the work on phases is intended to take advantage of the work already accomplished in the MSAA reconnaissance study. Products resulting from the phases are intended to provide a systematic framework for design and development of a data management system. The products also (a) provide many materials needed to explain and describe the stage 2 and stage 3 project work, (b) provide a framework for how a planner progresses through the system once it is developed, and (c) provide materials for use in work review meetings.

g. A primary goal of working through the phases is to determine the essential information system capabilities; i.e., those that cannot be compromised and must be available to the user. Another primary goal is to determine those capabilities that can be made indirectly available to the user, with some compromise in system response time or accuracy in order to provide some secondary advantage such as decreased system cost or development time.

Phase 1: Identify Problems and Needs

11. To fulfill the objective for phase 1, a table should be prepared listing all the MSAA problems and needs (Figure 3), starting with those formally identified in the stage 1 study and listed in the Reconnaissance Report.* Additional problems and needs identified after the Reconnaissance Report publication should be added to the list. Each problem or need should be given a concise name and described or defined in a concise manner. The object is to ultimately achieve a list that contains only the problems and needs that the data management system can aid in addressing, with definitions that are acceptable to the SAM staff.

Phase 2: Identify Planning Objectives to Solve Problems and Answer Needs

12. The work performed in this phase is similar to the work performed in Phase 1. A table should be prepared listing all MSAA planning objectives (Figure 4), starting with those formally identified in the stage 1 study and listed in the Reconnaissance Report. Additional planning objectives identified after the report publication should be added to the list. Each objective should be given a concise name and defined or described in a concise manner. The object is to achieve a list that contains only the objectives that the data management system can aid in addressing, with definitions that are acceptable to the SAM staff.

13. The list should be arranged in descending order of priority with minimal discussion and reevaluation. Work results from successive phases will determine whether there is a need to formally reorder the list.

14. The connections between the planning objectives and the problems and needs should be formally displayed as shown in Figure 5. This formal display performs the following functions:

---

* U. S. Army Engineer District, Mobile. 1979. "Mississippi Sound and Adjacent Areas Dredge Material Disposal Study, Reconnaissance Study," Mobile, Ala.

Mississippi Sound and Adjacent Areas Study:

Problems and Needs

Problem or Need	Description

Figure 3. Form for recording study problems and needs

Mississippi Sound and Adjacent Areas Study:

Planning Objectives

Planning Objective	Description

Figure 4. Form for recording planning objectives

Mississippi Sound and Adjacent Areas Study:

Connections Between Problems and Needs and Planning Objectives

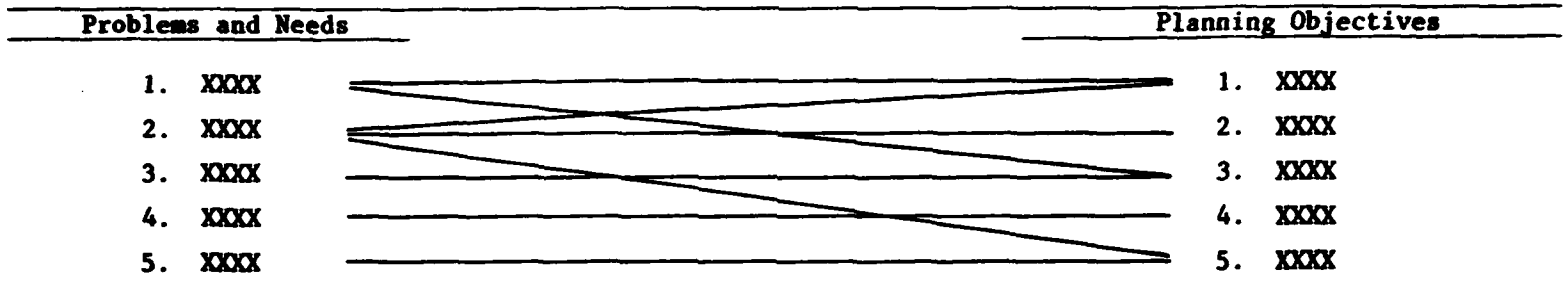


Figure 5. Form for recording connections between problems and needs and planning objectives

a. It provides a vehicle whereby priorities can be better reviewed and possibly reassigned. The priorities set at this point will be one of the study products that have a direct influence on the data processing work scheduling.

b. It provides a concise overview to anyone working on MSAA information management system development.

c. It provides a final check as to whether all problems, needs, and objectives are being addressed.

d. It provides, to persons not familiar with the final management system after development, part of the overview demonstrating why certain data items and analysis products are required for the MSAA study.

Phase 3: Identify Alternative Measures to Meet Planning Objectives

15. The work performed in this phase is similar to the work performed in work phases 1 and 2; a form should be completed (Figure 6), just as for those phases, listing the alternative measured which have been proposed to meet the planning objectives. The entries on this form should be arranged so that the alternatives satisfying the larger number of planning objectives, or else satisfying the more important planning objectives, are placed at the top of the list. An alternative measure should be described specifically in terms of the alternate plan or plans proposed to accomplish the measure; if an alternate plan will satisfy more than one alternative measure, it should be specifically addressed in the description of each measure it will satisfy.

16. This is the most dynamic of all lists and should ultimately contain all alternate plans. The alternate plans considered in the stage 3 study will come from this list.

17. The connection between the planning objectives and the alternative measures should be formally displayed (Figure 7). In addition to providing a concise overview to participants involved in developing the information management system and a vehicle for reviewing the alternate plan priorities, this display summarizes all alternate plans for which data are required.

Mississippi Sound and Adjacent Areas Study:

Alternative Measures to Meet Planning Objectives

Alternative Measure	Description

Figure 6. Form for recording alternative measures to meet planning objectives

Mississippi Sound and Adjacent Areas Study:

Connections Between Planning Objectives and Alternative Measures

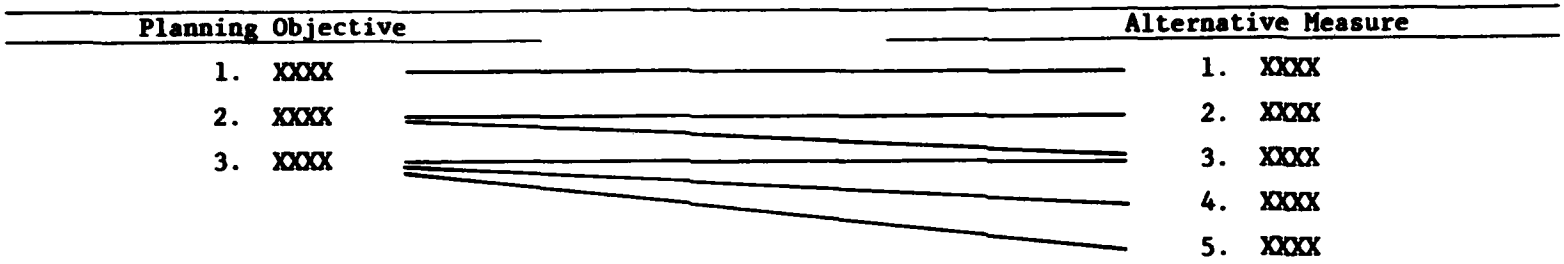


Figure 7. Form for recording connections between planning objectives and alternative measures

Phase 4: Identify Data and Data Forms and Formats Needed to Evaluate Alternate Plans

18. The specification of alternate plans leads directly to a requirement for certain data. The required data (i.e., those that the MSAA study participants require for evaluation and analysis) can consist of the data as it is gathered without further substantive processing, or of secondary data derived from analyzing the gathered data.

19. The work in this phase is concentrated on the data that does not require further substantive processing; i.e., radical change in form and format or combination with other data and submission to non-trivial analysis before being usable in the system. Most biological data require substantive processing since the data must be analyzed and correlated with other data to produce (for example) maps showing distributions of various parameters. Data do not require substantive processing if they can be placed in the information system and retrieved by the user either as a single item or in combination with other data items. A data type can be used both directly and after substantive processing, so that it would be considered in both this and the next work phase.

20. One purpose in separating the data into "directly usable" and "secondary analysis" results is to help determine which data items must be immediately available to the user, for use either as they are or in calculations, and which data items can be placed in the background because they will be either not used or used only for secondary data generation.

21. The table in Figure 8 should be used to compile the required data and the form and format in this work phase. The entries in the Source Form and Format column should be prepared in conjunction with the work on phase 7, Compile Information on Available Data, to make maximum use of available data.

22. All alternate plans should be considered. The same data types required to evaluate one alternate plan will be used for several others. It is highly probable that the total number of different data types will be restricted to a small number.

Mississippi Sound and Adjacent Areas Study:

Data Needed to Evaluate Alternate Plans

Alternate Plan	Required Data		
	Source Form and Format	Data Type	System Needs Form and Format

Figure 8. Form for recording data needed to evaluate alternative plans

23. The entries made under the System Needs Form and Format column should consist of the products the analyst will use. Those products are limited to combinations of map, table, graph, and narrative displays. There may be a requirement to provide the same data type in more than one form and format, or a requirement to provide more than one form and format to complete the display of one data type. An example of this latter situation is a requirement to produce a map showing data locations, and narratives for each location. Each entry under the System Needs Form and Format column should specify map, table, graph, or narrative and be followed by a short descriptive note. When appropriate, the form and format should be displayed in stylized examples and referenced in the description.

24. The entries under the Data Type column should consist of the data types required to meet the user needs.

25. Once the facts for Figure 8 are available, they should be rearranged (except for the Alternate Plan column, which is deleted) into three other forms (Figures 9-11) to emphasize different aspects of the information. This alternate display of the information is needed to design the information system and will also be of interest to system users in defining capabilities available to them.

26. First, the System Needs Form and Format column of Figure 8 should be examined. The entries in this column should be compressed so that every possible entry appears only once. Data in the other columns should be regrouped, ignoring the Alternate Plan column, as shown in Figure 9. The entries in the Data Type column should be regrouped as much as possible; for example, many alternate plans may require the same type of data to produce the same System Need, so that all of these occurrences can be compressed into one Data Type and one System Need. The entries in the Source Form and Format column should be compressed in the same way after regrouping. This formally defines the system needs that the information management system must service.

27. Second, the Data Type entries of Figure 8 should be examined and compressed so that every possible entry appears only once (see Figure 10). The Source Form and Format and the System Needs column entries

Mississippi Sound Adjacent Areas Study:

Data Needed to Evaluate Alternate Plans Arranged by System Needs

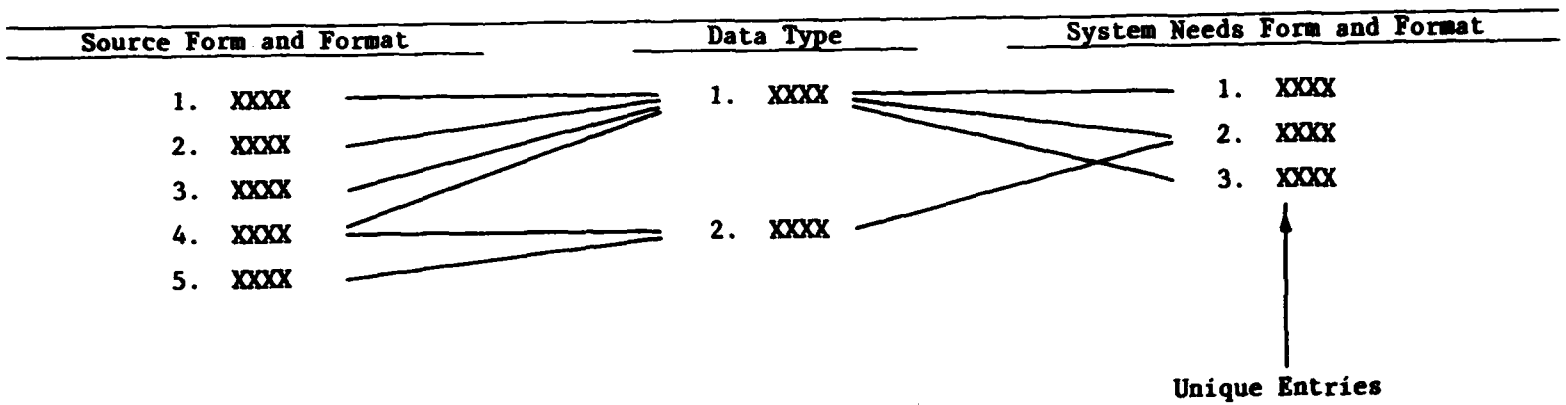


Figure 9. Form for compressing required data according to the system needs

Mississippi Sound and Adjacent Areas Study:

Data Needed to Evaluate Alternate Plans Arranged by Data Type

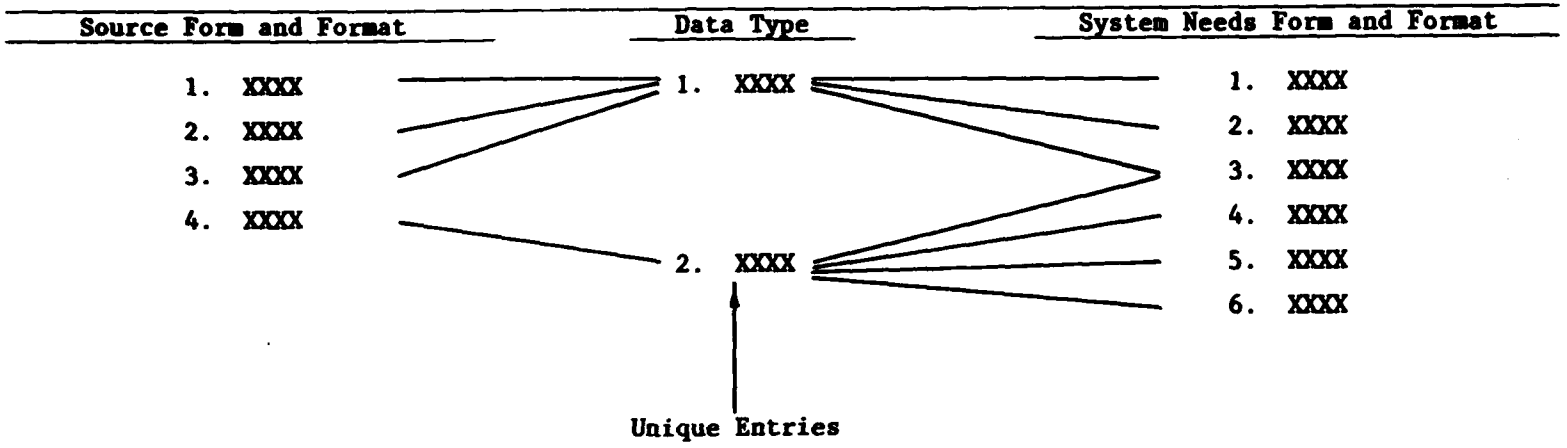


Figure 10. Form for compressing required data according to data type

Mississippi Sound and Adjacent Areas Study:

Data Needed to Evaluate Alternate Plans Arranged by Source

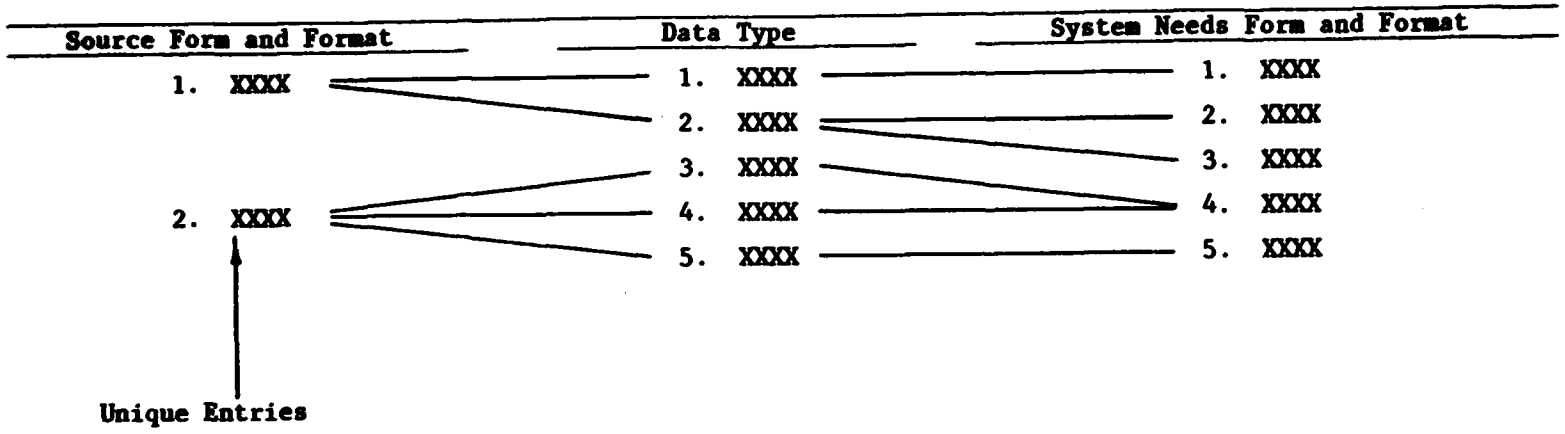


Figure 11. Form for compressing required data according to data source

should then be compressed after regrouping. This exercise provides a formal list of the data types that must be handled by the system.

28. Third, the Source Form and Format entries of Figure 8 should be examined and the form shown in Figure 11 prepared with the other two columns arranged as for Figures 9 and 10. This provides a formal list of the data source forms and formats for which computer software are required.

29. The volume of data contained in the forms given in Figures 9-11 will be greatly reduced from the volume of data in the form given in Figure 8.

Phase 5: Identify Analysis Products Needed to Evaluate Alternate Plans

30. Work in this phase is directed toward identifying data in the form of secondary analysis products required by the MSAA study participants to analyze alternate plans. The table shown in Figure 12 should be completed except for the column Source Form and Format, which is reserved for the next work phase.

31. Descriptions of each column to be completed in Figure 12 are as follows. The Systems Needs Form and Format column contains exactly the same type of information as was described in paragraph 23.

32. The Analysis Procedures column is intended to contain a description of the steps in the analysis process of each procedure necessary to produce the required data. This description could become lengthy if several data types must be handled to prepare the final required data product; it is likely, however, that the same analysis procedures will be used for several different required data products. One object of this work phase is to determine the minimum number of elements needed to build the maximum number of analysis capabilities.

33. The Form and Format for Analysis column is used to record the specific form and format of the data input to the analysis procedures.

34. The Data Types column should be used to record the types of data required for input to the analysis procedures. Many alternate

Mississippi Sound and Adjacent Areas Study:

Analysis Products Needed to Evaluate Alternate Plans

Alternate Plan	Required Data				
	Source Form and Format	Data Types	Form and Format for Analysis	Analysis Procedures	System Needs Form and Format

Figure 12. Form for recording analysis products needed to evaluate alternate plans

plans will require the same data types and data forms and formats.

35. Once the data in Figure 12 are available, they should be rearranged (except for the Alternate Plan column, which is deleted) onto four other forms as shown in Figures 13-16 to emphasize different aspects of the information. The data in Figure 12 should be rearranged as was done for the data in Phase 4 (see Figures 8-11).

36. The rearrangement of data according to System Needs Form and Format (Figure 13) defines the system needs or products that the information system must provide.

37. The rearrangement according to Analysis Procedures (Figure 14) provides a list of the data analysis capabilities required and what products those capabilities can provide. This enables the system designer to place priorities on the required capabilities and provides one means for differentiating among the capabilities offered by different commercially available hardware and software systems.

38. The rearrangement according to Data Type (Figure 15) provides a possible extension of the list of data types prepared in Figure 10 (see paragraph 27) that the information management system must handle.

39. The rearrangement according to Source Form and Format (Figure 16) provides an extension of the list of data source forms and formats prepared in Figure 11 (see paragraph 28) that the system must handle.

Phase 6: Identify Data Needed to Produce Analysis Products, Including Data Form and Format

40. The column headed Source Form and Format in Figure 12 should be completed in this work phase. This work should be performed in conjunction with the work on Phase 7, Compile Information on Gathered Data, and Phase 5, Identify Analysis Products Needed to Evaluate Alternate Plans. This coordination is needed to identify usable available information and to aid in identifying alternate data sources. Work on these three phases is anticipated as an iterative affair in which the work phase products are made consistent with each other and particularly

Mississippi Sound and Adjacent Areas Study:

Analysis Products Needed to Evaluate Alternate Plans Arranged by System Needs

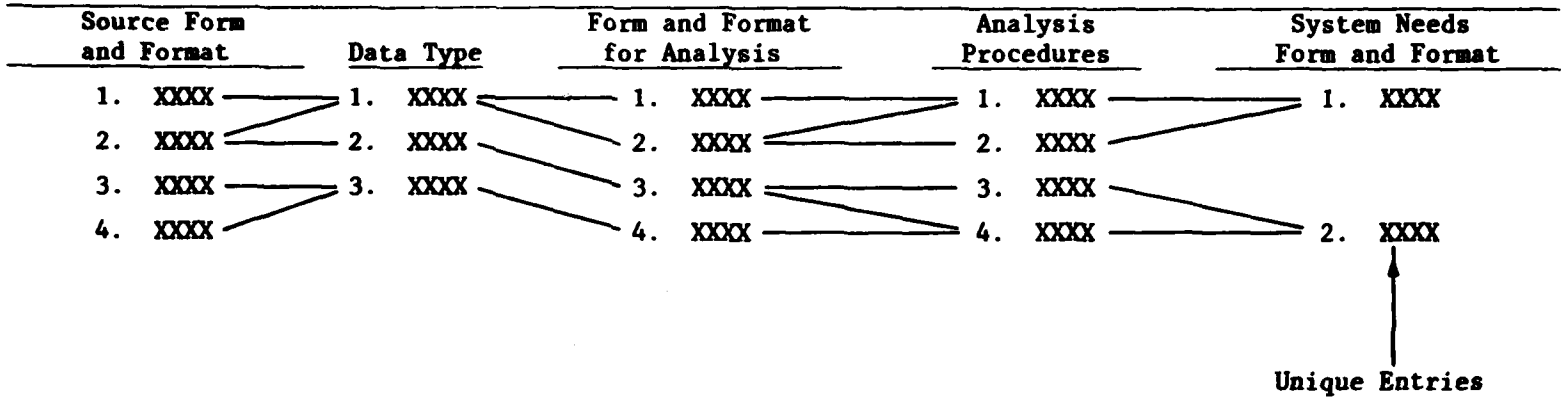


Figure 13. Form for compressing analysis products information by system needs

Mississippi Sound and Alternate Areas Study:

Analysis Products Needed to Evaluate Alternate Plans Arranged by Analysis Procedures

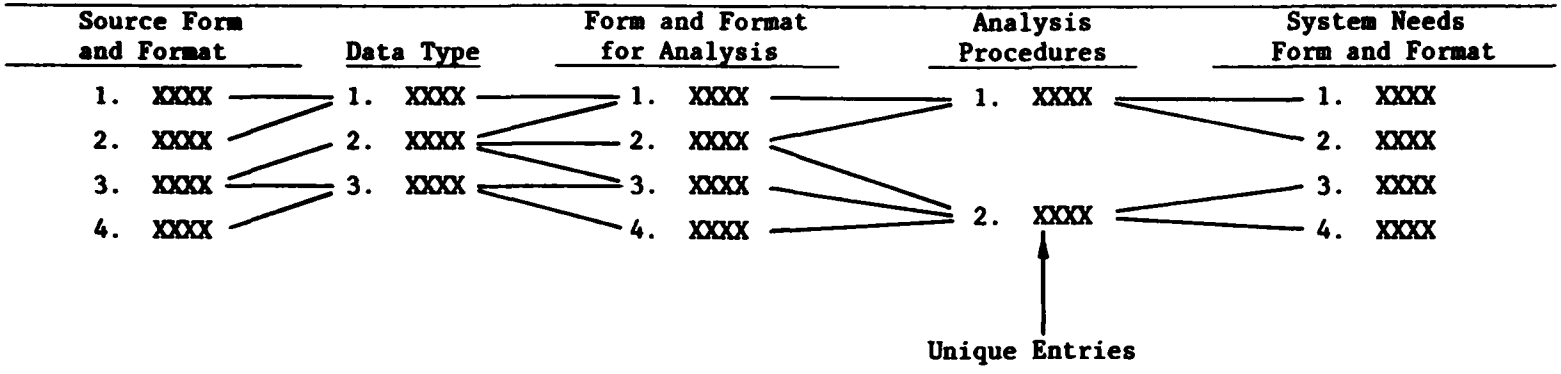


Figure 14. Form for compressing analysis products information by analysis procedures

Mississippi Sound and Adjacent Areas Study:

Analysis Products Needed to Evaluate Alternate Plans Arranged by Data Type

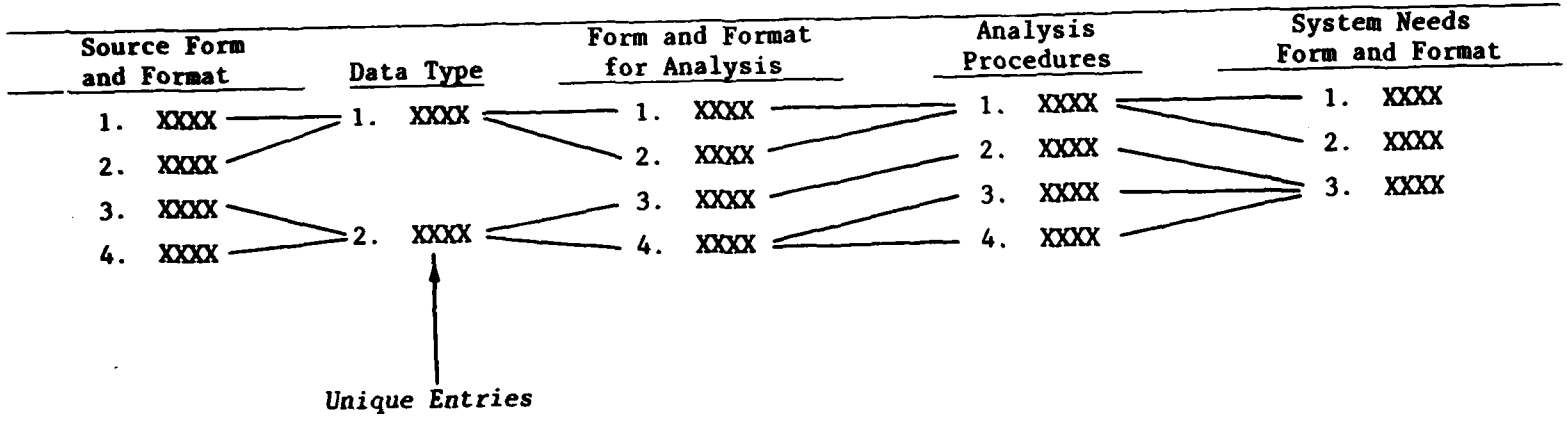


Figure 15. Form for compressing analysis products information by data type

Mississippi Sound and Adjacent Areas Study:

Analysis Products Needed to Evaluate Alternate Plans Arranged by Data Source

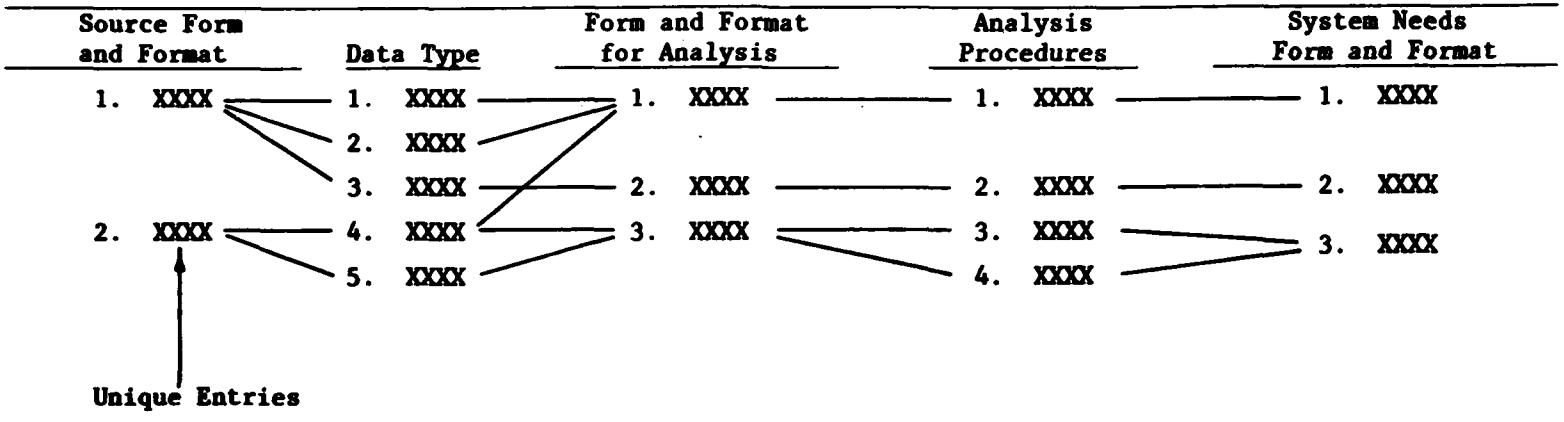


Figure 16. Form for compressing analysis products information by data source form and format

with the available data gathered during the previous work on the MSAA study.

Phase 7: Compile Information on Gathered Data

41. The compilation of available data and the gathering of required data have been major efforts thus far in the MSAA study; this work phase is dedicated to compiling concise descriptions of the data. The descriptions will be needed later to determine (a) the data and (b) the data-handling capabilities required for the MSAA Study and who should provide both. This work effort is separated into the identification of various data sources (see paragraphs 42-45) and the identification and description of data (see paragraph 46).

Data sources

42. All contractors' efforts should be summarized in a table as shown in Figure 17. The Contact Point column in Figure 17 should contain a numbered list of names of persons who have detailed knowledge of the data prepared or gathered for the MSAA study. It will be necessary to telephone each contact point during this work phase to ensure that there is adequate information regarding the data gathered by the contractor.

43. A bibliography of all useful publications, including internal documents and maps, should be compiled and numbered in a table during this work phase (see Figure 18), listing only those documents that can be obtained from the source. Copies should be obtained from the sources and reviewed. The potential usefulness of the document should be rated from one (definitely useful) to ten (definitely not useful) based on the information content of the document and on the appropriateness of the information to the management system development of the MSAA study. Data will later be abstracted from the documents according to their relative priorities.

44. A numbered list of data sources should also be prepared containing information on persons or organizations with special data resources that are not available in documents. This list (Figure 19)

Mississippi Sound and Adjacent Areas Study:

List of Contractors

Contractor: Name, Address	Contact Point: Name, Telephone	Description of Data

Figure 17. Form for listing all contractors

Mississippi Sound and Adjacent Areas Study:

Published Sources of Information

Reference	Potential Usefulness	Available Data Abstract

Figure 18. Form for recording published sources of data information

Mississippi Sound and Adjacent Areas Study:

Special Sources of Undocumented Data

Source	Description of Data and Form and Format

Figure 19. Form for recording special sources of undocumented data

should contain names of persons with special technical knowledge, organizations with data available in computers not readily accessible to SAM personnel, etc. Some entries on this list could be contractors. The list should provide a concise reference for sources of important missing data (if there are any), data additional to that available, explanations of trends that should be exploited in data analysis, and additional documented data.

45. A form (Figure 20) should be prepared containing a list of data available on computers or computer mass storage devices such as magnetic tapes. (Entries on this list may duplicate entries on lists in Figures 17-19.) This list will aid the management system development or implementation process as follows:

a. Published data available through a computer, however indirectly, will be reviewed and a decision made as to whether time and funds could be saved by gathering the data in digital form from the source rather than from the published documents.

b. The MSAA study system will probably require that some data be provided in a messaged or analyzed form (i.e., as secondary data products) rather than in its original form. Any potentially useful computerized data available should be examined carefully and additional funds provided if a contractor (or some other party) can perform the required data messaging or analyses and provide the secondary data products actually required by the MSAA study.

Data

46. After a thorough review, information on the available data should be compiled as shown in Figure 21; each entry on this list should correspond to a specific set of data. The Data Type column corresponds to the column of the same name shown in all previous figures; it is possible for many data entries to have the same Data Type. The Source Code column contains codes relating the data to the sources listed on the forms shown in Figures 17-20. The codes are intended as unique identifiers of data sources and should be numbered according to table and source; i.e., if the List of Contractors table is called Table 1, sources taken from it should be coded 1-1, 1-2, 1-3, etc. A data type entry in Figure 21 should have as many data source codes as there are

Mississippi Sound and Adjacent Areas Study:

Sources of Computerized Data

Source	Data Description and Form and Formats	Required Procedure for Gaining Access to or Using Data

Figure 20. Form for recording sources of computerized data

Mississippi Sound and Adjacent Areas Study:

Available Data Survey

Data Type	Source Code	Description	Form and Format

Figure 21. Form for summarizing the available data

references to the specific data set that have been gathered. The data set should be described briefly in the Description column, and the Form and Format column used to record information on the data form and format.

Phase 8: Identify Available Useful Data

47. The identification of useful data starts with an evaluation of the information resulting from work phases 4, 6, and 7. The data identified as required in the Available Data Survey in Figure 21 should be listed according to data type and source code and described as before (see Figure 22), this time differentiating between the form and format in which the data are available and the form and format required by the MSAA system. The same data can appear more than once in Figure 22 since a given data item may be used directly and as input to analysis procedures used to generate secondary data products. For example, shellfish bed location data may be required both directly and as input to correlation calculations made with several types of physical parameter data. Entry numbers should be assigned each different output form and format required for the same data set and each different form and format available for the same data set; these numbers will be used to indicate which data sets are to be processed.

48. The form shown in Figure 23 should be completed for each entry made on the form shown in Figure 22. On this form will be set up the specific plan for how the entries listed in Figure 22 will be processed to achieve a product that is directly usable by the MSAA analysts, or usable in calculations the results of which will be used by the MSAA analysts.

49. The Data-Gathering, Preparation, and Processing Steps section of Figure 23 should contain a detailed plan for all required operations to be followed in the MSAA for the specific data involved.

50. The data-gathering operation includes any work necessary to gain access to the data source. A data source could be a map, a table in a report, or data in digital form in a computer. While many data sources are available at SAM, data in the preferred form and format may not be. For example, data may have been preferred recorded on magnetic tape rather than summarized in a final report. A procedure should be selected for gathering the data from the source; this should be noted on the Plan along with any required contractual arrangement. The

Mississippi Sound and Adjacent Areas Study:

Available Required Data

Entry	Data Type	Source Code	Description	Form and Format Available	Form and Format Required

Figure 22. Form for summarizing required available data

Mississippi Sound and Adjacent Areas Study:

Plan for Data Handling

Entry Data Type

Source Code and Source

Available Form and Format

Required Form and Format

Data-Gathering, Preparation, and Processing Steps

Cost and Time Estimates

Figure 23. Form for planning data processing steps

number of pieces of data to be gathered should be noted on the Plan, as well as any special physical characteristics, such as scale, size, and quality of materials, that have any bearing on data-gathering.

51. Data preparation includes any work necessary to retrieve the data from the source and place it in a form required for data processing. Noted on the Plan should be (a) the coordinate system used in the data source; (b) any work steps such as superimposing another coordinate system on a map, transforming recorded data in one coordinate system into another, typing tabular data into a computer file, etc.; (c) any equipment or computer software required for any part of the data preparation work; (d) any desirable contractual arrangement for data preparation; and (e) a procedure for quality control of the data preparation process and the data. Quality controlling the data preparation process ensures that the data after preparation are error free or have an acceptable degree of accuracy as required by their intended use. Quality control involves the selection of some reasonable process, however rudimentary, for reviewing the data for possible errors.

52. Data processing transforms the prepared data into a final computerized form in a computer data base. Some data will not require insertion into a computer data base and will forgo this step; all data input to the data processing step should be reordered in a computer file at the completion of the data preparation step. The specific geographic data base computer software capabilities required to process the data, as well as any special equipment, should be listed in the Plan.

53. It is convenient to prepare a block diagram showing the data-gathering, preparation, and processing steps to accompany the information given in Figure 23 for any given data set.

54. A cost and time estimate for accomplishing the tasks of data-gathering, preparation, and processing should be prepared for each data set.

55. In many situations there will be little leeway in choosing between different work steps that must be performed. In some situations several alternate data-handling plans will have to be considered and one plan selected. For example, the required data may be available on a

computer but buried within a large volume of undesirable data, and the cost of separating out the desired data may be much higher than the cost of preparing the data manually from tables or maps. As a second example, the data volume for a given data type may be so small that the cost and time of having a contractor provide the data on magnetic tape may be excessive compared to handling the data at SAM.

Phase 9: Identify Secondary Data Generation Requirements

56. Work on phase 8 and phase 9 should be done simultaneously since phase 8 work includes the consideration of data required for secondary data generation; i.e., the input to phase 9 work.

57. Secondary data generation can be (a) the expression of the information content of a data set in a different form and format, (b) the inference of new information from the old, or (c) the combination of various data sets to produce new data sets containing either the old data in a new form and format or new data inferred from the old data. The expression of the information content of a data set in a different form and format is, for example, the production of maps containing the data set or subsets at different scales; the inference of new information is, for example, the use of salinity samples to produce contour maps showing the mean salinity distribution; the combination of data sets is the production of maps displaying all areas, including shellfish beds, spawning grounds, endangered species habitat, etc., that should be avoided in dredged material disposal; the inference of new from old data sets is the correlation of biological samples with water quality, offshore current, and dredged material disposal data to infer interactions.

58. The form shown in Figure 24 should be completed for each data product when the Available Required Data and Plan for Handling Data forms (Figures 22 and 23) are being put together, using the same numbers in the Entry column to reference data to the information given in those forms. The Calculation Procedure column in Figure 24 should describe the operations necessary to produce the data product from the input data. In most cases, the operation will consist of a computer program and any manual labor steps necessary to control the software operations. In a

Mississippi Sound and Adjacent Areas Study:

Secondary Data Generation Requirements

Available Input Data				Calculation Procedure	Data Product	
Entry	Type	Source Code	Form and Format		Use Description	Form and Format

Figure 24. Form for recording secondary data generation requirements

few cases, the computer will not be required, and the described Calculation Procedure will involve only manual labor.

59. The Data Product Use Description should concentrate on the intended foreseeable uses of the product. Particular attention should be given to whether a one-time product will be required, or whether the product will be repetitively produced with variations; e.g., data correlation studies to demonstrate a relationship between dredging and water quality.

Phase 10: Identify Important Missing Data

60. A major effort in the SAM phase 1 study was directed toward the identification and collection of required data. It is unlikely that any important required data are missing; however, a formal review of the required and available data should be performed for completeness.

61. Any identified missing data should be recorded on the form given in Figure 25. The need for and intended use of the data should be described in the Requirement column and the reason the data are not available in the Reason for Nonavailability column. Various reasons the data are not available include the following.

a. The data need was recently identified by SAM personnel at a public or consultants' meeting or by a contractor.

b. The data were thought to be available; but a review has shown that the data were lost, did not contain the required information, or were too error laden to be of use.

c. The data were scheduled for collection by a contractor, but the contractor was unsuccessful.

d. Some data are available but are limited in such a manner, due to incomplete collection or the lack of collection over as extensive a geographical area as required, that the available data are effectively useless for this study.

62. If important missing data are identified, alternate procedures and data for meeting the study requirement should be reviewed, the impact on the study evaluated, and any steps possible to gather the data within the study schedule examined.

Mississippi Sound and Adjacent Areas Study:

Important Missing Data

Data Type	Requirement	Reason for Nonavailability

Figure 25. Form for recording important missing data

Phase 11: Develop Information Management System Layout

63. A proposed information management system layout is presented in this part of the plan. The work to be performed in phase 11 is to decide what parts of the system are actually required, where they are to be located, and who is to manage and implement the construction of the management system and the insertion of data into the system. An overview of the system is shown in Figure 26.

64. The Developer (see Figure 26) consists of the party responsible for control of the data flow into the system and the parties responsible for performing the data-gathering, data processing, etc. Many of the required data have been gathered and some have been prepared, but few have been processed by the MSAA study team. It is assumed that the MSAA study team will concentrate in the future on its role as user rather than developer.

65. The User (see Figure 26) consists of all parties needing to use the data, principally the MSAA study team members and the development team. The development team must use the data analysis and retrieval capabilities to produce some of the secondary data products (see phase 9 work description) needed for insertion into the master data bases.

Data development

66. The data-gathering, preparation, and processing work has been described in previous parts of this plan: most, if not all data-gathering has been completed by a combination of the MSAA study team and contractors; much data preparation and practically all data processing remains to be performed. It is convenient to visualize the data preparation work as preparing data in a few standard forms and formats to be acceptable to a very few data processing procedures. Similarly, the data processing work can be visualized as processing the data from a few forms and formats into that very restricted set of forms and formats required to be input to the master data bases. The data preparation and processing procedures and the input and output formats can be formally defined, primarily through the work done in phases 8 and 9.

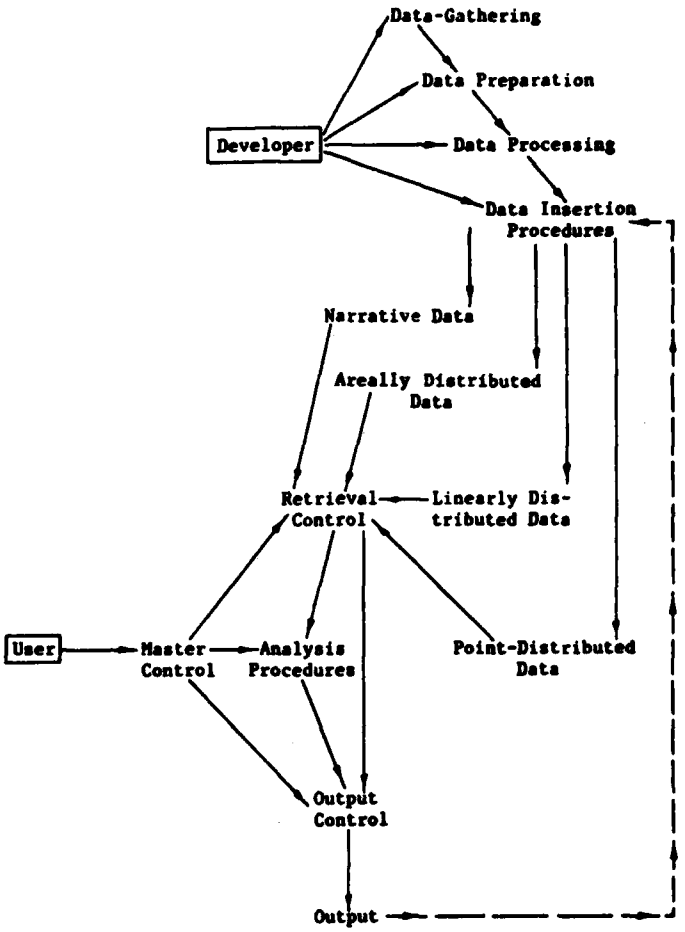


Figure 26. Information management system overview

While the developed data must be readily available to the user through the user-accessed portion of the management system and is therefore best placed in one convenient location, the data development work can be performed in many locations. Part of the work in this phase will involve selecting contractors and Corps personnel to perform data development.

Data storage

67. It is anticipated that much data development work will result in the computer storage of data, but that a significant amount of required data will not be stored in computer files of the management system.

68. For example, it is essential that the MSAA study use a standard set of base maps; for instance, a set of maps of different scale covering the same region. While much data in computer files will be referenced to the geographic coordinate system used on the base maps, the base maps themselves will not be stored in computer files.

69. Other data, such as endangered species and species habitat descriptions, will remain relatively unchanged in content after their preparation. This type of narrative data could be stored in computer files with a minimum effort required to define forms and formats and input the data. The primary power of the computer-based portion of the management system, however, is in the manipulation and analysis of data--operations not required with endangered species narratives. Therefore, the combination of (a) the priority placed on developing necessary data that must be analyzed before it is truly useful and (b) the MSAA study resources may relegate endangered species narrative data to paper files, with a reference to the paper files placed in the computerized portion of the management system.

70. The biological and water quality data will present a more serious decision problem. Considering the MSAA study schedule and resources, the fact that the data are already available on contractor computers, and the need to maintain as simple a management system as possible to allow the analysts to focus on product review and decisions made on the basis of those products, it may be desirable to have present contractors supply secondary data products generated from the original data for inclusion in the management system.

71. In sum, part of the work in this phase will be to decide what data must be placed in computer and paper files and what secondary data products can be provided to computer or paper files after analysis work external to the management system. Any data that are not essential for inclusion in the data files should be given a priority based on the convenience that including them in the data file would provide. If resources permit, the "convenient but unnecessary" data could be placed in the computer files on a priority basis.

Master control

72. The user's activities would be directed through a master control computer program as shown in Figure 26. The master control program provides the user with instructions and operation options and controls the retrieval, analysis, and output computer operations necessary to meet the user's requests; it need contain only those capabilities required by the MSAA study data use. Such a master control program is not elaborate and could be made operational in pieces as different types of data, analysis programs, and output capabilities become available. Part of the work in this phase will be to decide whether an available master control program should be adopted, or whether a new one should be purchased or developed by the MSAA study team. It is most probable that the decision will be to develop this master controller, considering the diverse requirements of the MSAA study and the resources required to adapt available similar programs. The total set of required control capabilities are indicated in the forms completed in previous work phases.

Retrieval control

73. The retrieval control section of the management system would consist of a series of modules embedded within a computer program used for retrieving the various types of data from the different data bases. Each module would be used to locate the data requested by the user and pass that data to the analysis procedures or output control section. The retrieval procedures would differ according to the way the data are formatted in the different master data bases. Some modules in the retrieval control section, such as those used for narrative data retrieval, would be particularly simple. Other modules should provide (a) the capability for retrieving combinations of data and (b) the locations where user-defined combinations of specific conditions exist according to the data. All potentially required retrieval capabilities presently exist in available systems, although no one available system contains all capabilities. Part of the work in this phase will consist of determining which of the potential capabilities are required and how those capabilities should be combined into a single system.

Analysis procedures

74. The analysis procedures should consist of all procedures required for (a) producing secondary data products from available data during the development work and (b) analyzing the master data base data according to the user's requests. In some cases, the same analysis procedure will be used for both. For example, a map containing average sediment properties could be calculated once and inserted into a data base for the purpose of frequent, ready reference; the same procedure would be used to study sediment distribution patterns on the basis of selected sediment data.

75. It is convenient for operational purposes to categorize the data and the analysis procedures by means of statistical or nonstatistical analysis. A well-developed series of powerful stand-alone systems such as the Statistical Analysis System have been used extensively to perform essential statistical analyses on time- and space-repetitive sample data. A decision must be made during this work phase as to whether such a system should be required as part of the capability directly available to the MSAA study planners, or whether the results from the use of such a system should be supplied under contract. If a decision is made to provide this capability directly to the MSAA study planners, a second decision must be made as to whether the statistical analysis system will actually become part of the management system shown in Figure 26 or merely be set up in parallel to feed the management system data. This series of decisions will strongly affect the total cost of the management system.

76. Most nonstatistical analysis procedures are available, although no single system appears to contain all the potentially required procedures. Part of the work in this phase will consist of identifying the procedures to be adapted for or inserted as is into an available system.

Output control

77. All system output would pass through the output control section of the system shown in Figure 26. The output section should contain the standard procedures for the production of tables, graphs,

plots, and listings. All potential requirements can probably be met from available systems. For example, practically all cathode ray tube display devices are supported by program-controlled subsystems providing graph and table format outputs; the same is true of practically all hard-copy plotter display devices. Work on this phase of the study will include the selection of output forms and formats and a review of available procedures. Plans must be prepared for integrating the procedures into the management system.

Data bases

78. It is convenient to place each type of data with uniform form and format characteristics into separate files or data bases.

79. Narrative data. The computer-based narrative-type data shown in Figure 26 could be located in any of the following five types of data files.

a. Bibliography and data source information. This file would contain information such as publication or contractor report references and names of experts on specific types of data. The data would be available to the user through a preselected keyword list made available through the master control and retrieval control procedures.

b. Available data in the computer. Information on the types and formats of computerized data in the management system would be stored in a file. The data would be accessed and searched in a manner similar to that for the bibliographic data.

c. Available data not in the computer. Information on paper-filed data necessary to the MSAA study and on nonessential data available on other computers or at contractors' locations would be contained in a file. The nonessential data would consist of useful but peripheral data, or of data used to generate secondary data products stored in the management system.

d. Narrative accounts not geographically referenced. Some narrative data, such as descriptions of endangered species, are necessary to the MSAA study; but the narratives themselves may apply to many locations or the narratives may not be specifically geographically referenced. These data should be placed in a separate file.

e. Geographically referenced narrative accounts. Data of a narrative nature containing crucial information and referenced to specific locations, such as historical sites

and shipwrecks, should be kept in a separate file. The location data could be stored in a separate point location data file and the narratives retrieved by the user in combination with the tabulation or mapping of the locations.

80. Areally distributed geographically referenced data. Areally distributed data, such as land use, can be categorized according to whether the data are maintained in a grid or line-and-vertex format. All grid data with the same resolution can be maintained in a single file; it may be necessary, however, to have a better resolution for some geographic areas than for others for more accurate data-portrayal purposes. For example, different grid-resolution data files could be used for onshore, riverine, estuarine, and shallow coastal areas, while a coarser grid was used for other offshore areas. The advantages of the grid format are that it requires input data which are relatively inexpensive to prepare and process and it lends itself to efficient analysis operations.

81. The principal advantage of the line-and-vertex format procedure is the ability to computer-plot more smoothly map products that contain lines so that they are more acceptable aesthetically to users in certain situations. A second major advantage of the line-and-vertex format systems is their ability to more easily perform cosmetic surgery on the data (e.g., to shift boundaries--such as land-use boundaries--to alternate locations); this type of operation is desirable when data of questionable quality yield combinations of questionable or impossible conditions when different types of data are overlaid. The line-and-vertex format theoretically can be used to maintain all data in a single file; its use becomes prohibitively expensive, however, as the number of data types per file increases so that multiple files may be necessary. The line-and-vertex format data preparation and processing procedures are much more complex and expensive than those for a grid format system and use many more computer programs and manual labor steps. If a line-and-vertex data file format is used, it is very likely that a grid format system must also be used in parallel since many operations require the transformation of the line-and-vertex format into the grid format

as part of the line-and-vertex format data operating procedure.

82. Systems handling both grid and line-and-vertex data formats are available. A decision must be made during this work phase whether both systems are necessary and, if not, which is to be used.

83. Linearly distributed geographically referenced data. Linearly distributed geographically referenced data, such as channel locations, can be maintained in a single system. A procedure coupling the linearly and areally distributed data must be prepared. The volume of this type of data will probably be small, and the required development will be straightforward. The work in this phase will include the planning and implementation of procedures for handling the linear data.

84. Point-distributed geographically referenced data. Point-location data for the MSAA study can be grouped into two categories. There is a moderate volume of nonrepetitive data, such as historical sites, that can probably be maintained easily in a single data file, making any required system development work straightforward. There is also a large volume of repetitively measured, principally biological data, such as salinity concentration, which consist of many repetitive measurements recorded with a short time interval between measurements.

85. The repetitive point-location data appear to be principally useful for statistical analyses and correlation studies, and it is unclear at this time how much of the collected data should be made available to the user. Part of the work in this phase will be to determine this and whether results of analyses on the repetitive data and some selected "raw" data should be made available, or whether a large volume of the nonrepetitive data must be made immediately available to the user. If a decision is made to have large volumes of the data available and to provide the management system users with the capability to perform standard biological statistical analyses, plans must be made for incorporating one of the standard statistical analysis systems into the management system.

System development plan

86. A development plan and schedule should be prepared. The priorities placed on system development should be directly related to

the priority in which data should be handled to be of maximum aid to the ongoing MSAA study. Work should be phased so that the system will be developed and implemented in parts which are immediately usable. Implementation in parts involves the following steps.

a. Purchase, transfer, or develop an initial software capability to handle a specific class of data.

b. Prepare (1) a set of data for processing into the system, or (2) an analysis of data in the system.

c. Operate the system to process or analyze the data.

d. Examine and critique operational procedures and their cost, time for use, and products.

e. Further develop software capability, if required.

f. If the procedure requires little or no further development, plan and effect the process for producing that class of data handled by the new system feature.

g. Make the new system features and data available to the system users.

Phase 12: Identify Computer Software Requirements

87. The computer software capabilities required to handle the data are identified in work phases 8 and 9 and particularly in work phase 11. The objective of this work phase will be to identify available software that can be used to meet the study requirements, and prepare and implement a plan for gathering the software and using it to develop the required management system. Primary attention should be given to available software.

88. It should be remembered that the management system is a system involving the use of both automated and manual procedures. The software must not only provide the required calculation capability, but must interface efficiently with the form and formats of the input data. This point is of particular importance. An examination of practically any system handling MSAA study-type data will demonstrate that the cost of operating the software is negligible compared to the cost (and time) of manual labor required to prepare the data for processing. The following items should be considered when choosing among available systems or

among software to be used in the MSAA system, or in further developing chosen software.

a. Data editing and system quality control. The greatest time-and-funds consumer is data-error correction. A high degree of quality control, including in depth automated data checking by the software, must be available. Inattention to input data quality control can easily increase the total cost of data processing by a factor of ten. Errors must be recognized and corrected as early as possible in the string of data processing operations; checks in the software can aid the data quality control immeasurably.

b. Number of software steps. The number of sequential software operations required from start to finish in data processing and the number of computer programs used in analyzing, retrieving, outputting, etc., the data from the system should be minimized.

c. Software user control input. The amount of information that must be provided by the software user should be kept to a minimum. When the user must operate a series of computer programs, the input information should be provided once and as early as possible in the string of operations.

d. Technical competence of workers and users. The system, particularly the manual labor work operations, should be set up so that technically simple instructions can be followed in data processing and system use.

89. Choices between prospective available software should be made on the basis of the above considerations as well as the cost and time required to convert, transfer, purchase, and develop software.

Phase 13: Identify Computer Hardware Requirements

90. Computer hardware requirements should be defined in this study phase. The basic capabilities, developed in phases 8, 9, 11, and 12, to be provided by the hardware will probably define a standard set of hardware available at any moderate-size computer facility. The recommendations regarding the actual hardware to be used will be based on a consideration of the following items.

Specialized system software

91. Several specialized data-information systems will be

considered during the proposed study. Each of these specialized systems will provide a portion of the total capability required by the MSAA study. At least one of the systems* was developed in such a manner that the software cannot be transferred effectively from the hardware on which it resides to any other hardware. A decision to use this system as a subsystem within the MSAA system will require the purchase of equipment, allocation of District space and automatic data processing personnel, training of personnel, and further study of how data other than that handled by this system will be handled. Preliminary indications are that there will be a major problem (a) transferring other parts of the MSAA study system to the purchased hardware or (b) efficiently coupling data from the purchased system with data from any other system.

92. Other systems, available through contract,** provide more, but not all, of the required MSAA system capabilities and could be purchased. While these systems can be transferred to most computers, they are effectively restricted to use only on certain machines on which prior transfer work has already been performed. These available systems should be reviewed and considered for use in the MSAA study.

Nonspecialized system software

93. Initial indications are that most of the required software are available and will operate on the SAM in-house computer or on other computers available to the District personnel through present Government contracts.

Location

94. It is not apparent at this time whether the SAM location is the best location for the principal computer hardware. It is possible that the computer and mass memory devices could be located anywhere in the United States and that data input and output peripheral devices are the only hardware items required at the District facilities.

---

* WAAMS, available through Autometrics, Inc., and the U. S. Department of the Interior Fish and Wildlife Service.

** Available through Environmental Science and Research, Inc., Redlands, Calif.

PART III: CONCLUSIONS

95. This document provides a plan for coordinating the MSAA study data, data-handling procedures, and analysis procedures required to meet the MSAA study objectives. The developed plan, if implemented, is expected to provide the following benefits.

a. A formal identification of specific products that are required to meet the study analysts' need for data and for study products.

b. A formal identification of the interrelationships between required data, the data processing work required, and the final products.

c. A means whereby financial and time schedule priorities can be set for the data-handling process and for the selection and development of computer software needed to process the data for analysts' needs.

d. A framework for determining the need for different types of computer information systems.

e. A framework for listing the specific technical capabilities that a required information system must have.

f. An identification of which data are essential to the study goals and which should have a low priority.

g. An identification of data that should be made immediately available to study participants.

h. A formal, condensed, operational overview of the study for data-handling operations to show study participants, particularly participants not familiar with Corps study procedures, the data being studied, the study products, and the study capability to analyze alternatives.

In accordance with letter from DAEN-RDC, DAEN-ASI dated 22 July 1977, Subject: Facsimile Catalog Cards for Laboratory Technical Publications, a facsimile catalog card in Library of Congress MARC format is reproduced below.

LaGarde, Victor
Mississippi Sound and Adjacent Areas Study Plan for Data-handling and Analysis / by Victor LaGarde (Environmental Laboratory, U.S. Army Engineer Waterways Experiment Station). -- Vicksburg, Miss. : The Station, Springfield, Va. ; available from NTIS, 1983.
55 p. : ill. ; 27 cm. -- (Technical report ; EL-83-1)
Cover title.
"March 1983."
Final report.
"Prepared for U.S. Army Engineer District, Mobile."

1. Computers. 2. Electronic data processing. 3. Mississippi Sound. I. United States. Army. Corps of Engineers. Mobile District. II. U.S. Army Engineer Waterways Experiment Station. Environmental Laboratory. III. Title. IV. Series: Miscellaneous paper (U.S. Army Engineer Waterways Experiment Station) ; EL-83-1.
TA7.W34 no.EL-83-1